FACULDADE DE CIÊNCIAS SOCIAIS E HUMANAS
UNIVERSIDADE NOVA DE LISBOA

Statistical and Geospatial Data Integration for Policy-making
The Role of Data Integration in Public Policies

Rossano Manuel da Costa das Neves Figueiredo

Tese de Doutoramento em Geografia e Planeamento Territorial
Especialidade em Deteção Remota e Sistemas de Informação Geográfica

Versão corrigida e melhorada da tese após provas de Doutoramento

Fevereiro, 2025

Tese apresentada para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Geografia e Planeamento Territorial, Especialidade em Detecção Remota e Sistemas de Informação Geográfica, realizada sob a orientação científica do Professor Doutor Rui Pedro de Sousa Pereira Monteiro Julião da Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa (NOVA FCSH).

Thesis presented to fulfil the requirements to obtain the degree of Doctor in Geography and Territorial Planning, Expertise Area in Remote Sensing and Geographic Information Systems, conducted under the scientific supervision of Doctor Professor Rui Pedro Julião at the School of Social Sciences and Humanities - NOVA University of Lisbon (NOVA FCSH).

DECLARAÇÕES

Declaro que esta tese é o resultado da minha investigação pessoal e independente. O seu conteúdo é original e todas as fontes consultadas estão devidamente mencionadas no texto, nas notas e na bibliografia.

Declaro ainda que tomei conhecimento do Código de Ética da Universidade NOVA de disponível na intranet e que foram respeitados os seus termos no decorrer do presente trabalho de investigação.

O candidato,

Assinado por: **ROSSANO MANUEL DA COSTA DAS NEVES FIGUEIREDO**
Num. de Identificação: 13904800
Data: 2025.02.19 16:07:24+00'00'

____

Lisboa, ... de .... de ....

Declaro que esta tese se encontra em condições de ser apreciado pelo júri a designar.

Declaro ainda que tomei conhecimento do Código de Ética da Universidade NOVA de Lisboa disponível na intranet.

O(A) orientador(a),

Assinado por: **RUI PEDRO DE SOUSA PEREIRA MONTEIRO JULIÃO**
Num. de Identificação: 06868846
Data: 2025.02.19 16:55:17+00'00'

____

Lisboa, ... de .... de ....

ACKNOWLEDGMENTS

I would like to express my gratitude to Prof. Dr. Rui Pedro Julião for supervising and guiding me throughout my PhD journey, highlighting his pragmatic feedback, sense of direction and valuable insights.

I am also grateful to Ana Santos for introducing me to the GEOSTAT projects and the EFGS family, which have inspired this PhD and shaped my professional career in recent years. Without her, I probably would not have had a research topic that motivated and engaged me as much as this one, as well as the time and energy to carry out this research by facilitating the balance between my professional responsibilities and academic work.

To my colleagues at Statistics Portugal for their availability, valuable time, knowledge sharing and content review, especially those directly involved in the CE-SIG project, namely Carla Cardoso, Maria Aurindo and Sara Soares. I would also like to thank Statistics Portugal for allowing me to access all the internal documentation related to the case study and for all technical and non-technical support during the process.

Also thankful to David Sousa for reviewing the content related to the statistical data, in particular the section on statistical quality, and for exchanging ideas when I started to draft the structure of the thesis. His contributions and our discussions were very useful at the beginning of the writing process.

A special and heartfelt thank you to João Oliveira for the love, companionship, unconditional support and patience throughout this journey, from the beginning to the end. Thank you for your trust and motivation in times of uncertainty and discouragement.

A big thank you to my friends who have supported me over the last few years with their friendship and straightforward questions that allowed me to put the ideas together.

Finally, I am deeply indebted to my parents, who have provided me with all the opportunities of this life, and without whom I would not be here achieving this personal milestone.

STATISTICAL AND GEOSPATIAL DATA INTEGRATION FOR POLICY-MAKING

THE ROLE OF DATA INTEGRATION IN PUBLIC POLICIES

ROSSANO MANUEL DA COSTA DAS NEVES FIGUEIREDO

ABSTRACT

The data and methodological gaps that undermine the implementation of the 2030 Agenda for Sustainable Development have pointed to statistical-geospatial data integration as a promising path to provide new geospatial statistics for measuring and monitoring the Sustainable Development Goals and supporting policy-making at multiple geographical levels. The modernisation of official statistics has recognised the potential of new data sources, such as geospatial data, to enrich statistical data and produce high-quality, timely, comparable and detailed statistical outputs, as well as the increased awareness of the importance of location as a matching key to link different data domains, address policy demands and meet data requirements and user needs. The benefits of integrating statistical and geospatial data have been acknowledged by key stakeholders from both statistical and geospatial communities towards the implementation roadmap of the high-level global statistical-geospatial framework - and its European version - as a key governance driver to produce harmonised geospatial statistics in a consistent and systematic manner.

Statistical organisations have made joint efforts to follow general and methodological guidelines, adhere to common standards and best practices, strengthen institutional collaboration, carry out data integration activities and enhance their statistical-geospatial capacity to align and mainstream geospatial production components (data, processes, services and other capabilities) into their statistical production processes from technical and non-technical perspectives. The experiences and practical use cases over the years have confirmed the advantages of geospatial data and technology in modernising official statistics and production systems through authoritative location data, georeferenced statistical/administrative data, common geographies and new analysis and visualisation opportunities that ended up improving the harmonisation and quality of business processes and outputs. Statistical-geospatial data integration was also demonstrated to support policy-making processes by enabling more territorially targeted policy responses and interventions, and contributing to more data-driven and evidence-based decisions that underpin global policy frameworks, regional development agendas and national priorities.

Based on the internationally endorsed frameworks and reference models from the statistical-geospatial operating environment, a Production Model and Assessment Matrix were developed to assist statistical organisations in producing geospatial statistics and evaluate their statistical-geospatial capacity and maturity levels at the national institutional environments, as a complement to existing methodological guidance. The two methodological parts were applied to a project that promotes territorial cohesion, underlining a statistical operation to map accessibility geographies from georeferenced data of facilities and to produce territorial indicators to be displayed in the web dissemination platform embodying search and spatial analysis tools. The results enabled to identify gaps and development areas in governance, data/information, technology, institutional and capacity building issues and provided a set of improvement actions and recommendations to enhance statistical-geospatial integration capabilities. The methodological application in the case study also showcased the added value and operational benefits of statistical-geospatial integration in statistical production in order to produce relevant insights associated with location for better public policy design and implementation, especially related to the provision and access to facilities and services of general interest at the national, regional and local levels.

KEYWORDS: Statistical Data, Geospatial Data, Data Integration, Geospatial Statistics, Policy-making.

INTEGRAÇÃO DE DADOS ESTATÍSTICOS E GEOESPACIAIS PARA ELABORAÇÃO DE POLÍTICAS

O PAPEL DA INTEGRAÇÃO DE DADOS NAS POLÍTICAS PÚBLICAS

ROSSANO MANUEL DA COSTA DAS NEVES FIGUEIREDO

RESUMO

As lacunas nos dados e metodologias que comprometem a execução da Agenda 2030 para o Desenvolvimento Sustentável conduziram à integração de dados estatísticos e geoespaciais como uma via promissora para fornecer novas estatísticas geoespaciais para medir e monitorizar os Objetivos de Desenvolvimento Sustentável e apoiar a elaboração de políticas a vários níveis geográficos. A modernização das estatísticas oficiais reconheceu a potencialidade das novas fontes de dados, como os dados geoespaciais, para enriquecer os dados estatísticos e produzir resultados estatísticos de qualidade, relevantes, comparáveis e detalhados, bem como a maior sensibilização para a importância da localização como chave para ligar diferentes domínios de dados, responder às exigências de políticas e satisfazer os requisitos dos dados e as necessidades dos utilizadores. As vantagens da integração de dados estatísticos e geoespaciais foram reconhecidas pelos principais atores das comunidades estatística e geoespacial, tendo em vista a implementação do modelo estatístico-geoespacial global de alto nível - e da sua versão europeia - como um motor de governação fundamental para a produção de estatísticas geoespaciais harmonizadas de forma consistente e sistemática.

As organizações estatísticas estabeleceram esforços conjuntos para seguir orientações gerais e metodológicas, aderir a normas comuns e boas práticas, reforçar a colaboração institucional, realizar atividades de integração de dados e melhorar a sua capacidade estatístico-geoespacial para alinhar e integrar as componentes de produção geoespacial (dados, processos, serviços e outras capacidades) nos seus processos de produção estatística, de um ponto de vista técnico e não técnico. As experiências e os casos práticos ao longo dos anos confirmaram as vantagens dos dados e da tecnologia geoespaciais na modernização das estatísticas oficiais e dos respetivos sistemas de produção através de dados de localização fidedignos, dados estatísticos/administrativos georreferenciados, geografias comuns e novas oportunidades de análise e visualização de forma a melhorar a harmonização e a qualidade dos processos e dos resultados. Foi também demonstrado que a integração de dados estatísticos-geoespaciais suporta os processos de elaboração de políticas, permitindo respostas e intervenções políticas mais orientadas para o território e contribuindo para decisões mais baseadas em dados e evidências que sustentam os quadros políticos globais, as agendas de desenvolvimento regional e as prioridades nacionais.

Com base nos modelos de referência internacionalmente aprovados no ambiente estatístico-geoespacial operacional, foram desenvolvidos um Modelo de Produção e uma Matriz de Avaliação para apoiar as organizações estatísticas na produção de estatísticas geoespaciais e avaliar a sua capacidade e níveis de maturidade estatístico-geoespaciais nos ambientes institucional e nacional, como complemento das diretrizes metodológicas existentes. As duas componentes metodológicas foram aplicadas a um projeto de promoção da coesão territorial, suportado por uma operação estatística de mapeamento de geografias de acessibilidade a partir de dados georreferenciados de equipamentos e de produção de indicadores territoriais para efeitos de visualização na plataforma online de divulgação que integra ferramentas de pesquisa e análise espacial. Os resultados permitiram identificar lacunas e áreas de desenvolvimento em aspetos de governação, dados/informação, tecnologia, institucionais e de reforço de capacidades, e forneceram um conjunto de acções de melhoria e recomendações para melhorar as capacidades de integração estatístico-geoespacial. A aplicação metodológica no caso estudo também demonstrou o valor acrescentado e os benefícios operacionais da integração estatístico-geoespacial na produção estatística, a fim de produzir conhecimentos relevantes associados à localização para uma melhor elaboração e implementação de políticas públicas, especialmente relacionadas com a oferta e o acesso a equipamentos e serviços de interesse geral a nível nacional, regional e local.

PALAVRAS-CHAVE: Dados Estatísticos, Dados Geoespaciais, Integração de Dados, Estatísticas Geoespaciais, Elaboração de políticas.

TABLE OF CONTENTS

ACRONYMS AND LIST OF ABBREVIATIONS

ABS	Australian Bureau of Statistics
AI	Artificial Intelligence
API	Application Programming Interface
CAPI	Computer-assisted personal interviewing
CATI	Computer-assisted telephone interviewing
CAWI	Computer-assisted web interviewing
CES	Conference of European Statisticians
CoP	European Statistics Code of Practice
CoR	European Committee of the Regions
CSDA	Common Statistical Data Architecture
CSPA	Common Statistical Production Architecture
CSV	Comma separated values
DDI	Data Documentation Initiative
DW	Data Warehouse
DCAT	Data Catalogue Vocabulary
DGGS	Discrete Global Grid System
DG REGIO	Directorate-General for Directorate-General for Regional and Urban Policy
DGT	Directorate-General of Territory
EC	European Commission
EEA	European Environment Agency
EFGS	European Forum for Geography and Statistics
EFTA	European Free Trade Association
EIF	European Interoperability Framework
ELISE	European Location Interoperability Solutions for E-Government
ESS	European Statistical System
EU	European Union
Eurostat	European Statistical Office

FAIR	Findable, Accessible, Interoperable, and Reusable
GAMSO	Generic Activity Model for Statistical Organizations
GDPR	General Data Protection Regulation
GeoGSBPM	Geospatial view of the Generic Statistical Business Process Model
GIS	Geographic Information Systems
GKI	Geospatial Knowledge Infrastructure
GPS	Global Positioning System
GISCO	Geographic Information System of the Commission
GSBPM	Generic Statistical Business Process Model
GFGS	Global Forum for Geography and Statistics
GFM	General Feature Model
GSGF	Global Statistical Geospatial Framework
GSIM	Generic Statistical Information Model
HLG-MOS	High-Level Group for the Modernisation of Official Statistics
HVD	High-Value Datasets
IGIF	Integrated Geospatial Information Framework
IMF	International Monetary Fund
INSPIRE	Infrastructure for Spatial Information in Europe
IoT	Internet of Things
ISO	International Organisation for Standardisation
IT	Information Technology
JRC	Joint Research Centre
LOD	Linked Open Data
ML	Machine Learning
MoU	Memorandum of Understanding
MS	Member State
NGIA	National Geospatial Information Agency
(N)SDI	(National) Spatial Data Infrastructure
NSO	National Statistical Office
NSS	National Statistical System

NUTS	Nomenclature of Territorial Units for Statistics
OGC	Open Geospatial Consortium
OECD	Organisation for Economic Co-operation and Development
OSM	OpenStreetMap
PAPI	Paper-assisted personal interviewing
PARIS21	Partnership in Statistics for Development in the 21st Century
PID	Persistent Unique Identifiers
QAF	Quality Assurance Framework
RDF	Resource Description Framework
SDG	Sustainable Development Goals
SDMX	Statistical Data and Metadata eXchange
SNIG	National Geographic Information System
SP	Statistics Portugal (Instituto Nacional de Estatística, IP)
TJS	Table Joining Service
UN	United Nations
UNECE	United Nations Economic Commission for Europe
UN EG-ISGI	United Nations Expert Group on the Integration of Statistical and Geospatial Information
UNFPOS	United Nations Fundamental Principles of Official Statistics
UN-GGIM	United Nations Committee of Experts on Global Geospatial Information Management
UN-GGIM: Europe	United Nations Committee of Experts on Global Geospatial Information Management Europe
UNSC	United Nations Statistical Commission
UNSD	United Nations Statistics Division
VGI	Volunteered Geographic Information
WFS	Web Feature Service
WG	Working Group
WMS	Web Map Service
W3C	World Wide Web Consortium
XML	eXtensible Mark-up Language

INTRODUCTION

In the last two decades, international organisations and high-level stakeholders increasingly acknowledged statistical-geospatial data integration as a key asset to support global policy frameworks, strategic agendas and development priorities, highlighting the Sustainable Development Goals (SDG) under the 2030 Agenda (Van Halderen et al., 2016). The roadmaps on overarching policy drivers, modernisation of official statistics and geospatial data and technology progress have raised a growing consensus that statistical-geospatial data integration is critical in supporting policy-making at multiple levels, from local to global. Integrating statistical and geospatial data enables maximising the value of data and addressing data and methodological gaps, hindering the implementation of policy instruments for evidence-based decision-making (UN-GGIM: Europe, 2019). It assigns a territorial-based data-driven approach underpinning the policy lifecycle.

The integration of statistical and geospatial data is recognised by the international statistical community as one of the most promising paths for creating new data analysis and dissemination opportunities and providing more timely, reliable, relevant and detailed information to gain deeper and meaningful insights (Eurostat, 2019a; UNECE, 2024a). Advancements related to geospatial data and technology have been addressing the needs and challenges of official statistics on new data sources and new metrics to improve the quality of the statistical outputs. In this context, geospatial data have gained a central role in strategic and technical discussions as a reliable data source to be streamlined in statistical production and to align both statistical and geospatial business processes, models and concepts towards an interoperable environment and standardised geospatial statistics.

A noteworthy milestone was reached in 2019 and 2020 as a global key statistical-geospatial framework was adopted and endorsed by both statistical and geospatial communities to facilitate the integration of statistical and geospatial data. A new maturity stage in this field is emerging towards institutional commitment and practical implementation. However, technical and non-technical challenges still hamper the potential of statistical-geospatial integration to fully add value to data and effectively support policy development and territorially targeted and informed decisions.

1. RELEVANCE AND CONTRIBUTIONS

The relevance of the thesis concerns the need to move forward in the field of statistical-geospatial (data) integration by providing new insights that evidence its added value to policy-making through a comprehensive review of both theoretical and practical issues and a methodology design and application in a case study. This extensive review addresses the underlying governance and production frameworks, guidelines, key elements, best practices and experiences to compile and consolidate the body of knowledge in the field and apply it to the national institutional environment, in particular to the statistical organisation in the context of official statistics. The methodological contributions aim at enriching the existing methodological guidance for the integration of statistical and geospatial data, taking into account the specifications of the case study, as well as complementing the available tools for assessing the capabilities and measuring the statistical-geospatial capacity and maturity of organisations. The results of the methodological application aspire to derive actionable insights that enable identifying improvement areas requiring development work and providing specific recommendations for targeted implementation by the statistical organisation under analysis and surrounding national key stakeholders, in addition to the existing general guidelines outlined through the literature review.

For the purpose of this thesis, it is important to mention that the concept of 'statistical-geospatial integration' encompasses a cross-cutting meaning in which both technical and non-technical issues are reviewed and go beyond data integration itself, underscoring the multi-dimensional perspective of the topic under study.

The thesis follows a global and European roadmap in fostering statistical-geospatial integration over the last years from both statistical and geospatial communities, particularly the work carried out by Statistics Portugal (Instituto Nacional de Estatística, IP) in several related data integration projects and activities. It addresses the discussion and recognition by high-level stakeholders that statistical-geospatial integration can tackle data gaps and requirements concerning global policy frameworks, namely to measure and monitor sustainable development progress (2030 Agenda). In this regard, the thesis endorses this acknowledgement by demonstrating the contribution of geospatial production components (e.g., data, processes and services) in statistical production to develop and deliver more detailed and timely geospatial statistics. These statistical outputs, enriched by location and other geospatial capabilities, provide more comprehensive and accurate insights to support policy-making at the national, regional and local levels. This increasingly accepted idea is what this thesis aims to showcase, verify and validate, namely through the case study.

From the statistical perspective, it tackles the challenges underpinning the modernisation of official statistics, focusing on incorporating new data sources to enhance comparability, timeliness, geographical detail, relevance and quality of the statistical outputs, while addressing the need for innovative information systems, businesses processes and methods in statistical production. It also aims to demonstrate a proof-of-concept for a policy-driven project wherein the use of geospatial data is a fundamental asset in statistical production enabling more informed, effective and targeted policy-making. This proof-of-concept highlights the potential of geospatial data and capabilities to enhance official statistics and underscores its cornerstone role in policy lifecycle and evidence-based decision-making.

The thesis adopts a more practical focus providing actionable insights that address concrete challenges related to statistical-geospatial integration and aligned with institutional environments and organisational needs following policy development agendas. Thus, the methodology was designed for implementation within an organisational context for statistical production and respective national institutional environment while considering outputs that address policy-making requirements over a theoretical exploratory approach.

The outcomes of the thesis are more pragmatic aiming to enhance the technical knowledge and capability development of statistical organisations for building, aligning and improving their statistical-geospatial business processes based on their generic production models and organisational structures. The outcomes include specific improvements and recommendations from different dimensions and action levels towards consistent implementation to facilitate the production of standardised and high-quality geospatial statistics that will provide more meaningful insights for policy-making. The outcomes also provide concrete guidance to enhance the national statistical-geospatial capacity and maturity levels involving improvement and development areas outside the statistical organisation.

It is important to note that the methodology, results and findings are intended to enrich the established reference frameworks and standards from both statistical and geospatial data domains. Instead, they serve as complementary resources to existing implementation guides and methodological materials related to statistical-geospatial integration as they come from a benchmarking exercise and fit-for-purpose applicability with more detailed descriptions.

This thesis expects to leverage statistical-geospatial integration's relevance and critical role in addressing national and international policy agendas and development strategies as well as to raise awareness among national key stakeholders, highlighting political and institutional leaders and data providers. The thesis also aspires to position data integration at the forefront of governance and technical discussions on public policies advocating long-term committees, shaping strategies, defining priority actions and building capacity to formally implement overarching statistical-geospatial frameworks in the national context for better policy-making.

2. OVERALL RESEARCH OBJECTIVES

The primary research goal of this PhD thesis is to demonstrate the added value of integrating statistical and geospatial data for policy-making by showcasing how statistical-geospatial integration capabilities enable more informed and targeted policy responses, support evidence-based decision-making and enhance effective policy development. The thesis seeks to contribute to the body of knowledge and technical development of statistical-geospatial integration, advance its theoretical and methodological foundations and move forward in the institutional implementation of reference frameworks, guidelines and best practices to support better policies from data-driven and improved decisions. The main research goals break down into the following specific objectives:

Objective 1: Conduct a comprehensive review of data, production processes, key stakeholders, frameworks and standards in statistical and geospatial data domains to provide an introductory overview of the field of statistical-geospatial data integration.

Objective 2: Present a detailed overview of the state-of-the-art in statistical-geospatial (data) integration and describe the current development stage, focusing on its application in official statistics and addressing both technical and non-technical key issues, challenges and opportunities within the statistical-geospatial operating environment.

Objective 3: Develop a methodology to describe and evaluate data integration and geospatial-related activities and tasks in the statistical business production model and apply it to a case study (statistical operation) to identify capabilities and gaps, assess the statistical-geospatial capacity and maturity levels and demonstrate its potential and relevance for policy-making.

Objective 4: Propose improvements and recommendations to strengthen statistical-geospatial capacity and capabilities, enabling effective integration of geospatial production components into statistical production in order to produce standardised and high-quality geospatial statistics, tackling multi-level policy-making needs and requirements.

3. THESIS ORGANISATION AND OUTLINE

This thesis is divided into five chapters to structurally accommodate the overall research objectives from a generic perspective to a concrete methodological application and results. The theoretical-conceptual framework underpins the first two chapters and provides an introductory description of statistical and geospatial data domains and the field of statistical-geospatial data integration from the literature review and a state-of-the-art overview.

Chapter I covers both statistical data and geospatial data domains, representing the broad disciplinary areas that support the conceptualisation of the field of statistical-geospatial data integration and underlines the theoretical basis for the design and application of the methodology. Chapter II focuses on statistical-geospatial data integration, breaking down into sub-chapters on technical infrastructures, key elements and a comprehensive analysis of strengths, weaknesses, opportunities and threats from the literature review. This analysis provides a literature summary concerning statistical-geospatial integration and a description of its cornerstones.

Both Chapters I and II objectively provide a systematic and parallel structure in terms of content organisation and narrative while addressing data, production, key stakeholders, frameworks and standards as common topics allowing for a summary of the current state of the research subjects and a comparative review of related literature from a multi-thematic analysis outlook.

Chapter III addresses the methodology by introducing and describing the case study according to the statistical production stages and process components, outlining and systematising the two operational parts of the designed methodology (Production Model and Assessment Matrix) and exhaustively presenting the application of the methodological approach in the case study.

Chapters II and III are the research core of the thesis, where the first one provides an introductory overview and both conceptual and methodological body of knowledge to support the validity and reliability of the designed methodology and its application in the case study. In this regard, the literature review provided an extensive benchmarking exercise by identifying and evaluating several guidelines, best practices, standards and frameworks that support the design and application processes of the methodology and ensure the technical robustness and credibility of the methodological approach.

Chapter IV summarises the results for each operational part of the methodology to facilitate their analysis and understanding and to support a more targeted implementation of the improvements and recommendations presented in the respective chapter. Chapter IV also includes the research findings and discussion points based on the interpretation of the results and aligned with the theoretical-methodological framework from the literature review.

Chapter V presents the main conclusions of the study, highlighting the key findings, and pointing out potential future research efforts and lines of work based on the identified gaps and areas for development and improvement to enhance the statistical-geospatial integration capacity and capabilities of statistical organisations. The contents of this chapter are supported by the results of the methodology, along with the analysis from the literature review. Chapter V is followed by the references list.

I. STATISTICAL DATA AND GEOSPATIAL DATA

This chapter presents a comprehensive overview of statistical and geospatial data, and a review of the statistical and geospatial communities and their operating environment, including the description of the key stakeholders, governance and production models and frameworks in the scope of official statistics and institutional context.

I.1. STATISTICAL DATA

This chapter introduces key elements, structuring components and terminology of official statistics for a comprehensive understanding of the statistical domain at the international level while briefly reporting the implications of using Big Data for statistical production in exploiting the potential of new data sources to increase the quality of statistical outputs. The statistical production process is summarised, the key stakeholders and their roles within the statistical community are identified and described, the existing statistical frameworks are systematised and aspects related to statistical quality and domain-specific standards are covered.

This chapter does not delve into an in-depth literature review and insights related to non-official statistics or other statistics produced by non-certified producers. Thus, the term 'statistical organisation' or NSO (National Statistical Office) is employed when addressing any organisation in which the primary role is the production of official statistics regardless of the statistical activities carried out and the organisation's responsibilities within its statistical system.

I.1.1. STATISTICAL DATA, PRODUCTION AND KEY STAKEHOLDERS

I.1.1.1. STATISTICAL DATA

Statistical data has long been a traditional source of quantitative or numerical information for the analysis of economic, social and environmental phenomena and supporting decision-making in human development. Official statistics are based on statistical data which traditionally "refers to data from a survey or administrative source used to produce statistics" (OECD et al., 2002: 740) describing the demographic, economic, environmental and social development of a country. It can also be any "data collected, processed or disseminated by a statistical organisation for statistical purposes"

(UNSD, 2021: 890) whereas the term “statistics” addresses the statistical output as aggregated data on units or observations (datasets and tables). Hence, the term ‘statistics’ in official statistics is commonly used as the output of the statistical production process and the term ‘data’ when referring to the input for that same process.

According to the Organisation for Economic Co-operation and Development (OECD) (OCDE, 2002: 546), “official statistics are statistics disseminated by the national statistical systems, except those that are formally stated not to be official”, i.e. official statistics are produced by certified producers of statistics. Thus, official statistics are statistics produced by NSO or another producer of official statistics mandated by the national government or certified by the national statistical system to compile statistics for its specific domain (UNSD, 2021). The statistical production process under official statistics is conducted according to fundamental principles and aligned with international statistical frameworks and standards to ensure standardisation and improve their quality across statistical organisations.

At the international scope, official statistics are ruled by cornerstone principles, the United Nations (UN) Fundamental Principles of Official Statistics (UNFPOS), originally developed and adopted at the Conference of European Statisticians (CES) in 1991. In 2014, they were adopted by the UN General Assembly through a resolution (UN, 2014) that emphasised the fundamental importance of official statistics in democratic societies and for national and global policy agendas, particularly concerning sustainable development. Therefore, at the strategic level, they have been playing a key role in the 2030 Agenda and measuring and monitoring the achievement of the SDG indicators and targets. This overarching global framework oversights other policy and domain-specific frameworks and recognises the importance of statistics and their integration with other data sources to support sustainable development.

The 10 UNFPOS underline the modern version of official statistics, as a public good, with common ground rules supported by international statistical cooperation and guide statistical systems and respective statistical organisations in their strategies, programmes and activities supported by compliant political, institutional and legal frameworks. The statistical systems are established to achieve standardisation and harmonisation of official statistics fostering international collaboration, ensuring national coordination and developing common robust approaches and methodologies for statistical production. In addition, they are orientation values for the code of practice and professional ethics to all responsible for producing official statistics. They include principles such as transparency, accountability, relevance, confidence, professional independence, statistical confidentiality, harmonisation, equal access, national coordination, international standards and cooperation.

Official statistics provide authoritative qualitative and quantitative data and information about a wide range of domains (statistical themes or domains), including finance, industry, trade and services, agriculture and fisheries, transport, energy and science. Thus, they are a major source of primary and secondary data from different sources (or providers) through various data collection and acquisition modes.

Primary data is "directly collected by a producer of official statistics exclusively for statistical purposes" (UNSD, 2021: 878) and provided by a primary data source, i.e. an organisation producing primary data. Primary data are typically collected by traditional data collection and acquisition modes, such as surveys (probabilistic and non-probabilistic), usually by applying a questionnaire to a probability sample of the target population, and censuses (population and housing). Historically, primary data has been associated with the census as a survey conducted on the full set of observations of a given population, i.e. a complete enumeration of a country's population (OECD et al., 2002). The main advantage of primary data is feasibility since the data collection and acquisition modes are developed in advance for statistical purposes. However, the associated traditional sources are time-consuming, labour-intensive and expensive, and in most cases substantially increase the burden on respondents posing challenges to official statistics related to data sources, collection and processing.

Secondary data are not directly collected for statistical purposes but instead initially collected by a public or private organisation for administrative or commercial purposes and further acquired and reused by a producer of official statistics. It includes mixed sources when there is a mixed mode of data collection (multisource process). Secondary data include administrative data and Big Data (e.g., sensors, satellite data, etc.). The first one is usually collected by government departments or other public agencies primarily for administrative purposes, i.e. registration, transaction and/or record-keeping, and the second one is by private organisations mostly for commercial purposes. The collection and acquisition of administrative data for statistical purposes have been carried out by statistical organisations in the last few decades, particularly in Northern European countries, due to budget cuts and demands from users for more regular and timely statistics (UNSD, 2021).

Administrative data have been increasingly used in official statistics, particularly related to the temporal and spatial dimensions of data, i.e. periodicity and geographic location of statistics. These data dimensions are of increasing interest to statistical organisations for policy purposes, especially for the SDG (e.g., temporal and spatial resolution and comparability) as well as to modernise official statistics considering emerging challenges on response burden and budgets and new data requirements from user needs. This major transition has witnessed a shift from survey data sources associated with classic approaches to administrative data and other non-traditional data sources by assessing their quality and suitability for statistical purposes (UNECE, 2017).

Administrative data include multi-size administrative registers, i.e. register data, derived from an organisation responsible for implementing an administrative regulation (or group of regulations) in which registers and transactions are viewed as a source of statistical data (OECD et al., 2002). It covers both administrative registers (e.g., list of names and addresses in a certain population) and data resulting from administrative transactions (e.g., bank payments). Moreover, administrative registers are classified as trusted data sources under the control of an appointed government ministry, department or agency to maintain registers and to attend to its administrative function, usually containing master data on entities such as persons, companies, vehicles, licences, buildings, locations or roads.

Despite different national data collection specifications and different strategies of statistical organisations regarding administrative data sources, there are a few commonly recognised advantages in collecting administrative data (UNSD, 2021): i) cost-effectiveness (particularly in data collection); ii) less response burden (and at the same time growing non-response rates and increasing the quality of input data and statistics); iii) timeliness and frequency (increase the frequency of compiling and publishing statistics from faster data availability and preparation); iv) coverage and completeness (reduce or eliminate errors from non-response and other typical errors from sample surveys and more accuracy in statistics from better coverage of target populations and detailed data at the local and regional levels); and v) relevance (better respond to new data needs and user demands and increase the relevance of statistics through filling data gaps for policy monitoring (e.g., SDG calculation). However, there are some quality problems in collecting and using administrative data in statistics, such as differences in units of analysis, concepts, definitions of variables and classifications, incomplete or inaccurate data quality, lack of statistical knowledge and capacities, and need for additional validation checks.

Secondary data sources are traditionally associated with experimental statistics as innovation enablers in statistical organisations. Experimental statistics are statistics that use new data sources and methods to better respond to data requirements and user needs in a more timely manner. Since developing new statistical outputs that comply with established quality principles and requirements and internationally agreed statistical standards usually takes too long, these statistics must fulfil maturity criteria regarding harmonisation, coverage, sound methodology or quality. According to international statistical recommendations, experimental statistics should be marked or labelled as such to enable them to be distinguished from official statistics (UNSD, 2021). The statistical community is increasingly moving forward in integrating experimental and official statistics under the same production process due to resource constraints, good quality (sometimes almost comparable to official statistics) and user demands.

The potential and implications of non-traditional and emerging data sources of large data volumes, commonly referred to as 'Big Data' for producing official statistics have been widely recognised (Hassani et al., 2014; Daas et al., 2015; Reimsbach-Kounatze, 2015). The key properties of Big Data are related to the 3Vs: i) volume (size of the dataset); ii) velocity (data-provisioning rate and time in which it is necessary to act on them, or the speed at which data is created, processed and analysed); iii) variety (heterogeneity of data types, including semi-structured and unstructured data from diverse sources and related to the capability to link these diverse datasets). Due to these properties, Big Data has the potential to tackle data and analysis gaps related to the burden on respondents, coverage, timeliness and improvement of estimates. Filling primary and secondary data gaps by partially replacing traditional data sources will allow reducing data collection costs and respondent burden in the long term.

Big Data should be seen as a complement at this point and not a full replacer of the existing traditional data sources to produce statistics as the most credible scenario is that most statistical activities will continue to be survey-based and according to the traditional approach (ESS, 2014). Moreover, its competitive advantage cannot be underestimated in the trade-off between quality, timeliness and accuracy and in the ability to produce new short-term statistics with higher representativeness, geographic coverage and acceptable levels of quality (UNECE, 2014).

The exploitation of Big Data for statistical purposes has been included in the agenda of the international statistical community as an ongoing strategic debate under continuous review, particularly related to its advantages, disadvantages, risks and opportunities. Digital transformation, the proliferation of readily accessible digital data, the development of Artificial Intelligence (AI) and Machine Learning (ML) techniques and the rise of the Big Data industry are some of the drivers that triggered this discussion. This novel paradigm has changed how data is collected, acquired and processed decreasing the time gap between collection and dissemination, and how operations within an organisation are carried out to handle the size and complexity of large flows of data. Big Data has also opened possibilities for acquiring data in various fields and extended the analysis to new social and economic phenomena with more timely and geographically detailed data (UNSD, 2021).

As outlined by UNSD (2021), statistical organisations have been accessing Big Data sources in the last years, including websites or sensor data managed by the public sector (e.g., road sensors) for experimental statistics or even official statistics. Data classification and coding, data preparation, editing and imputing and imagery analysis are working areas that use ML to produce official statistics in a more efficient manner, i.e. automation, without overlooking lessons learned and quality considerations (UNECE, 2021a). According to Yun et al., (Yun et al., 2018), although ML may be a useful and relevant tool in official statistics, quality issues should be carefully assessed with a new quality approach and different measures than the traditional ones since statistical organisations have no control over data generation from Big Data sources. Also, most of the available Big Data sources are not adjusted to the traditional statistical production or conceptual/methodological frameworks in official statistics that ensure confidence related to statistical principles and standards, particularly related to impartiality, confidentiality and methodology (UNECE, 2014).

In an overview, OECD (2015) outlined three main limitations and associated errors in using Big Data for official statistics: i) poor data quality; ii) inappropriate analytics; and iii) changing the environment. The first one underlines that the quality of data is intrinsically dependent on its context and how it fits the user needs. The second one highlights the need for rigorous scientific methods and high skills in data analysis to prevent the risk of inappropriate use of data and analytics. It also highlights the need for internationally agreed standards to ensure statistical validity, quality and trust in official statistics. The third one briefly addresses the risk of data analytics' robustness in the ever-changing data environment and includes non-technical aspects related to capacity building, data governance, digital policy and legal/technical frameworks.

The recognised limitations and challenges of Big Data in official statistics lead to a different statistical paradigm based on a less business-oriented view of the statistical production process and quality assurance, namely through exploratory methodological approaches and *ad hoc* case studies (UNSD, 2019). These types of methods and studies aim to find tailored solutions on how to extract, interpret, transform and store the input data and make them structured and manageable for analysis and processing according to statistical needs.

In the European context, two memoranda on Big Data were adopted: the Scheveningen in 2013 (ESSC, 2013) and the Bucharest in 2018 (ESSC, 2018). These memoranda had the main objective of drawing strategic lines on Big Data and were followed by a roadmap and action plan covering several strategic dimensions such as governance, policy, legislation, ethics/communication, methods and pilot projects for the European Statistical System (ESS).

The first memorandum acknowledged some of the requirements for the use of Big Data in official statistics and highlighted the need for the ESS to support related developments, including in methodology, quality assessment and Information Technology (IT). It also addressed privacy and data protection issues, recommended specialised training programs and emphasised the importance of partnerships between different stakeholders. Later, the second memorandum encouraged the ESS to implement practical cases of using Big Data in statistical production, develop experimental statistics and consider the achievements and strategic orientations for Big Data and smart statistics.

Table 1 provides a comprehensive comparative analysis of primary, secondary, and Big Data sources, offering a systematic overview of their key features in the context of official statistics.

Table 1.Primary sources, secondary sources and Big Data sources (European Commission, 2021).

	Primary sources	**Secondary sources**	**Big Data sources**
Data are designed to be used in statistical production	yes	no	no
Concepts, definitions and classifications are stated and known	yes	often	rarely
Target (sub-)population is defined	yes	often	no
Metadata available	yes	often	no
Data are structured	yes	yes	rarely
Data refer to units of the population of interest	yes	usually	no
Data need “heavy” preprocessing to be used in statistical production	no	no	yes
Interest variables are directly available	yes	yes	no
Auxiliary variables are directly available	yes	often	no
Data cover target (sub-)population	yes (census) no (surveys)	often	not yet
Data are representative (or lack of representativeness is intentional and/or can be adjusted for in analyses)	yes	often	no
Data values are “clean”	no	sometimes	rarely

Mixed data collection and acquisition modes can also be used (e.g., through joint web-based transfer or extraction solutions) and reflect the emerging shift from the dominance of data collection to multimode inputs. In this case, using paradata - i.e. data about the data collection process - is fundamental in official statistics. This topic addresses an ongoing discussion about complementing or replacing traditional data collection modes linked to primary data in which statistical organisations have already recognised the need to change their data collection processes (Yung et al., 2018).

Regardless of the type of data sources, statistical data is intrinsically related to two key concepts in official statistics: metadata and statistical confidentiality. Metadata are the data that define and describe other data and are a key technical element in statistical quality (UNSD, 2019). This concept is divided into structural metadata and reference metadata. Structural metadata enable the users to understand the data and inform them about the specifications of the statistical outputs through identifiers and descriptors that support the user in processing, organising and discovering statistical data. Structural metadata can include column names, dimension names, variable names and attribute descriptions and are usually published on the web. Reference metadata provide an additional layer of information on the dataset as a whole, for instance, a general description of the dataset, classifications used and evaluation of process quality. It comprises conceptual, methodological and quality metadata, usually supported by technical documentation describing statistical processes and operations and is applied for quality reporting (quality description). Whereas structural metadata must go together with statistical data, reference metadata can be separated from the dataset itself having a broader nature.

Statistical confidentiality has been also a critical value and cornerstone element in official statistics to ensure that individual data is confidential and collected or acquired for exclusively statistical purposes, for instance, through anonymisation techniques and dissemination of aggregated data (UNSD, 2021). This technical requirement comes from the fact that both types of data at an individual level, i.e. microdata, can provide personal or sensitive data enabling the identification and tracking either directly or indirectly of natural and legal persons through attributes such as name, address, location, identifier, etc. Thus, the international statistical community and statistical organisations have been ensuring that microdata is not released or accessed to keep transparency and trust from the data providers and users, and at the same time continue to develop new statistical methods and techniques on statistical confidentiality.

I.1.1.2. STATISTICAL PRODUCTION

In official statistics, the statistical production process is a conceptual model describing phases, sub-processes and activities of a statistical operation carried out within a statistical information system to produce statistical outputs according to the respective statistical regulatory frameworks, guidelines of statistical systems and statistical programmes. This model breaks down into successive and interconnected phases and sub-processes that are composed of a series of activities which are a combination of actions that result in a certain set of products (OECD et al., 2002). Statistical production was traditionally viewed as a 'value chain' - a range of activities carried out by an organisation to deliver a product or service in a specific industry or field - wherein each statistical domain had a different line of production at the operational level, although structurally they shared the same value chain. Thus, several lines of production were designed and implemented in isolation according to the statistical domain which was later recognised by the international statistical community as highly inefficient, namely due to the lack of flexibility (UNECE, 2014).

The modernisation approach of statistical production underlying a more business-oriented perspective changed its traditional view and established a move-forward trend adopted by many statistical organisations over the years. The statistical production cycle was thereafter increasingly recognised as non-linear whereas some sub-processes and/or activities are interdependent from each other and their arrangement and application can be flexible, interactive and diversified. The nonlinearity of the model covers activities of identifying concepts, classifications and variables, designing data collection and building processing and analysing components that can be carried out whenever possible in the development, production and dissemination of official statistics.

The statistical production briefly described below focuses on the production process carried out by the NSO and other producers of official statistics. The operational process of producing official statistics can be outlined and described in four main logical phases (UNSD, 2021):

i) Identifying user needs: this phase aims to support decision-making by systematising information allowing to determine the performance of a new statistical operation that follows previously identified and grounded information needs. It includes sub-processes and activities related to coordination between the organisational units and the development of feasibility studies that should summarise the main technical components, resources and other non-technical requirements of the projects. Although sub-processes in this phase are not usually applied to ongoing statistical operations, some of the activities are reevaluated at the end of the production cycle which may lead to methodological changes that initially supported the definition of the operation. Before the next collecting data phase, activities related to their design should be executed by specifying the modes and methods of data collection and acquisition based on methodological options.

ii) Collecting data: this phase encompasses several modes and methods of data collection and acquisition, including traditional collection, administrative data sources and other emerging ones through extraction and semi-(automatic) transfer (see Table 2). They can be applied in a simultaneous way or stepwise with a well-defined transition strategy between them within the production cycle of the statistical operation.

Table 2. Methods and respective modes of data collection and acquisition (source: author).

Methods of data collection and acquisition	Modes of data collection and acquisition
Collection by interview	- In-person interview: Computer-assisted personal interviewing (CAPI) and/or Paper-assisted personal interviewing (PAPI) - Telephonic interview: Computer-assisted telephone interviewing (CATI)
Collection by auto-fill	- Web-based auto-fill: Computer-assisted web interviewing (CAWI), file upload and/or automatic data transfer - Paper-based auto-fill
Automatic collection	- Web scraping - Application Programming Interface (API)

	- Internet of Things (IoT) - App (e.g., social media) - Sensors - Scanner data
Data collection	- Secure transmission (e.g., web services, SFTP, cloud)

These methods and respective modes are under three main ways of collecting and acquiring data (UNSD, 2021): i) by carrying out direct enquires (i.e. surveys) among persons, households, businesses and institutions - called statistical units; ii) by acquiring administrative data from government and other administrative sources; and iii) by using other data sources, such as Big Data, commercial data and geospatial data**.** For the aim of this thesis, it is important to mention that the CAPI mode enables the automatic collection of additional data and metadata, including geospatial data regarding the location (i.e. location data or geolocalisation) and time of the interview, respectively.

The phase of collecting data is the first operational phase and aims to collect the microdata required for statistical production. In the case of ongoing statistical operations, this phase is the first one in the production process, whereas, for new statistical operations, this phase works as a subsequent phase after activities related to designing, building and testing the statistical production components. This phase is the most intensive and resource-demanding requiring concerns related to data safety and institutional collaboration. This phase covers activities related to sample creation and selection (i.e. applied for sample surveys), preparation work and all actions related to the moment when data is being collected, including preliminary data validation and quality assessment via metadata and paradata.

iii) Processing: this phase is the first operational one and aims to prepare, review and validate the previously collected microdata and to calculate data aggregation as input for statistical analysis. Therefore, the sub-processes and activities under this phase are intrinsically related to the analysis phase since they are usually conducted simultaneously and their respective sub-processes are carried out interactively. It includes activities related to microdata integration, additional validation of the microdata, and error identification and solving for further analysis.

iv) Analysis of the statistics, and reporting and disseminating the statistical outputs: this phase can be broken down into three phases or sub-phases. It encompasses sub-processes and activities related to preparing, validating and interpreting results, and making the statistical outputs accessible to users through different platforms and dissemination channels. The statistical outputs can be micro datasets, aggregated statistics, statistical analysis, statistical services, metadata and other statistical products, such as statistical studies regarding a specific theme or statistical yearbooks as a compendium of statistical information about a wide range of statistical themes and development-related topics (UNSD, 2021). The sub-process of applying disclosure control is particularly relevant to statistical quality to ensure the confidentiality of statistical outputs according to sound methodologies and national/international data protection regulatory legislation. In the functioning of international statistical systems, this phase also comprises data and metadata transmission to the centralised statistical organisation responsible for compiling and disseminating official statistics. Lastly, sub-processes related to managing the release and promotion of the statistical outputs and user support are also carried out during this phase.

These four phases can also be conceptually assigned to three main stages - conceptualisation, operation and (quality) evaluation - resulting in a statistical output as an outcome to the users, generic public and civil society. Also, they can be broken down into more phases, sub-processes and activities, varying according to the statistical operation and/or product. A more detailed description of the statistical business production process will be provided in the subchapter on statistical frameworks.

I.1.1.3. KEY STAKEHOLDERS AND ROLES

Key stakeholders in the international statistical community are: i) Global and Regional institutions responsible for governance and coordination of the Statistical Systems and Statistical Production; ii) National Statistical Authorities, including NSO and other public bodies responsible for the production of official statistics; and iii) Other bodies, projects and initiatives for statistical collaboration, cooperation and modernisation of official statistics.

Although they are not reviewed in this subchapter, it is important to highlight that the users of official statistics are also relevant stakeholders within the statistical community since official statistics are produced for them and designed according to their needs, feedback and demands. The users of official statistics can be grouped into government, international and regional organisations, businesses, media, the academic/research community, non-government organisations and the general public encompassing the largest share of users (UNSD, 2021).

In general, international statistical organisations and statistical systems are the high-level designers of the visions, roadmaps and strategies of the statistical community that will reflect the statistical programmes and activities carried out by the statistical organisations at the national level, ensuring the harmonisation and quality of official statistics. At the technical level, they develop common statistical methods, standards, classifications, concepts and procedures under statistical frameworks and support the transposing of national statistics at the international level for comparability purposes. At the strategic level, they establish guidelines and recommendations that should be translated into actions and promote statistical coordination and cooperation at multiple levels within and outside the statistical community for statistical development and capacity building.

The **UN Statistical Commission (UNSC)** is the highest body of the global statistical system leading the strategic way on the international dimension of official statistics as the highest governance/decision-making body for international statistical programmes and activities. It was established in 1947 by the UN Economic and Social Council (ECOSOC) to develop national statistics for improved comparability, to coordinate the statistical activities of statistical organisations, develop a central statistical service, advise in technical matters of collection, analysis and dissemination and promote the modernisation of statistics. It overlooks the work carried out by the UN Statistics Division (UNSD), a functional commission, in facilitating international statistical coordination with other UN family organisations, regional commissions, committee experts and groups of experts. The UNSD as the secretariat of the UNSC coordinates many statistical cooperation aspects to ensure harmonisation between countries, adequate generation for both national and international use and statistical capacity in developing countries and regions (UNSD, 2021). To achieve these objectives the five UN Regional Commissions (Africa, Asia and Pacific, Europe, Latin America and the Caribbean and Western Asia) support the regional implementation of the global statistical programmes, for instance, addressing data gaps in terms of coverage, timeliness and disaggregation level. Each ECOSOC regional commission has its statistical division enhancing National Statistical Systems (NSS) through methodological guidance, modernisation activities and capacity development.

The overarching agenda of UNSD is committed to enhancing global statistical systems, compiling and disseminating global statistical data, developing international statistical standards, and supporting countries in strengthening their NSS through methodological and strategic guidance. Over the years, the UNSC has produced several key resolutions[1] where the starting point was the idea of an integrated system for the collection, processing and dissemination of international statistics aligned with the UN policy framework. This goal aimed to support the monitoring progress from the firstly established indicators of the 2015 Millennium Development Goals (MDG) and the following SDG under the 2030 Agenda, adopted in 2017 by the UN General Assembly. The UNSC also manages statistical operations at the international scope, such as the 2020 World Population and Housing Census Programmes that recognised census data as one of the primary data sources for the global indicator framework for the SDG.

Other international organisations and specific institutions and agencies within and outside the UN family also work on official statistics at the global level. The World Health Organisation (WHO), the International Labour Office, the UN Educational, Scientific and Cultural Organisation (UNESCO), the UN Food and Agriculture Organisation (FAO), the UN Population Fund (UNFPA) and other UN institutions collect, develop and harmonise statistics and standards in their respective domains. The **International Monetary Fund (IMF)** and the **World Bank** are leading international statistical cooperation partners in which the first one focuses on economic, financial and monetary statistics and standards and handbooks on these topics, and the second one on capacity

[1] To highlight key UNSC resolutions: Resolution 1566 (L) of 3 May 1971, Resolution 6 of 24 July 2006, Resolution E/RES/2013/21 of 24 July 2013, Resolution A/RES/69/282 of June 2015, Resolution E/RES/2015/10 of June 2015, Resolution A/RES/71/313 of 6 July 2017, Resolution 5 of 18 June 2020, Resolution 3 of 8 June 2022.

building programmes in developing countries (UNSD, 2021). The **Organisation for Economic Cooperation and Development (OECD)** has also been committed to developing and harmonising economic statistics. The **Partnership in Statistics for Development in the 21st Century (PARIS21)** has been claimed as a key stakeholder in statistical capacity building in developing countries, assessing and strengthening statistical systems and promoting cooperation in official statistics, particularly in topics related to SDG monitoring and reporting.

Statistical systems are composed of statistical organisations and have the main objective of cooperation and coordination on statistical activities at either national or international scope. Statistical systems enable a common framework for agreed and comparable concepts, classifications, methods and standards to ensure the harmonisation and quality of official statistics within and between countries and over time. International Statistical Systems encompass high-level international statistical organisations and other entities primarily or exclusively in developing, producing and disseminating official statistics at global, regional and sub-regional levels.

At the national level, according to OECD (OECD et al., 2002), the NSS is "the ensemble of statistical organisations and units within a country that jointly collect, process and disseminate official statistics on behalf of the national government" (OECD et al., 2002: 220). They are entrusted with the professional leadership and strategic vision of official statistics at the national level and are responsible for coordinating the main statistical activities and programmes in the country. The NSS comprise National Statistical Authorities - further described in more detail - and follow the national statistical agenda and cooperation agreements between these authorities and other stakeholders around the statistical activities. Similar to international statistical systems, the NSS are binding to the legal framework and in compliance with the technical framework at the national level. For instance, the legal concept of statistical confidentiality differs from country to country establishing different rules to apply data confidentiality and statistical metadata management.

NSS are adjustable to the government structure and political system which can be centralised, decentralised or federal (UNSD, 2021). In the first case, official statistics are predominantly produced by the NSO, and in the second case, the responsibility to produce and compile official statistics relies not only on the NSO but also on several national or regional government ministries, departments or agencies. Decentralised national statistical systems are common in countries where national governments delegate autonomous status to their sub-national/regional entities whereas these entities are subject to central administration (e.g., China, Spain or France). The third case includes two independent layers of producers of official statistics, the federal/central and the sub-national levels, and are usually associated with federal systems (e.g., United States of America, Brazil or Germany). The last two types of organisational structure regarding statistical systems are more focused on developing, producing and disseminating regionalised official statistics.

The **ESS** was built from the early acknowledgement by the European Community of the importance of reliable and comparable statistics for the planning and implementation of European policies. The ESS comprises the Statistical Authority of the European Commission (EC), Eurostat, the NSO of the 27 European Union (EU) Member States (MS) and the European Free Trade Association (EFTA) countries, such as Norway. In the ESS context, NSO are designated as 'National Statistics Institutes and the other producers of official statistics as 'Other National Authorities'. Their mission is to provide independent high-quality statistical information at European, national and regional levels and to make this information available to everyone for decision-making, research and debate. They are coordinated by Eurostat, which has the leadership role to ensure the harmonisation of statistics and EU policies in all statistical fields and promote close cooperation and support with national statistical authorities and NSS.

ESS has other partners and coordinates its work and responsibilities with other EC services and agencies, such as the European Central Bank, and other partnerships including candidate countries (e.g., Albania, Montenegro, Serbia or North Macedonia, etc.) and international organisations such as UN Economic Commission for Europe (UNECE), OECD, IMF and World Bank. The EC, European Parliament and Council are the political drivers of the ESS working across the political level among the governance bodies.

National Statistical Authorities include **NSO and other producers of official statistics** that have been mandated by the national government or certified by the national statistical system to compile statistics for its specific domain. NSO can have different names as 'Central Statistical Office' or 'National Statistical Institute' according to the national or international statistical systems, whereas the second term is formally used in the ESS. At the national level, the NSO are the most specialised statistical organisation and is responsible for compiling, producing and disseminating official statistics from all domains and statistical themes. Other producers of official statistics address any organisational entity within a government ministry, department or agency and have to be professionally independent and exclusively or primarily focused on statistical production. Thus, specialised statistical departments or units of government ministries produce official statistics alongside the NSO in areas of responsibility of the respective departments and ministries (e.g., education, health, labour and justice, etc.).

The **International Organisation for Standardisation (ISO)** is the most globally recognised organisation related to standards taking the leadership role in international standardisation. This organisation is composed of national member bodies and technical committees responsible for carrying out the development of international standards through collaboration with other international organisations, both public and private, governmental and non-governmental agencies. The work of preparing international standards is conducted by the member bodies who have expressed an interest in the specific subject for which a technical committee has been established, and which counts on their participation and technical support. ISO can be considered as a general key stakeholder applied to statistical, geospatial and other domains, producing both general and domain-specific standards.

The **CES** gathers leading statisticians and statistical experts from around 60 countries to guide the statistical work by providing guidelines and recommendations, setting standards for statistical production, assessing statistical systems and compiling a series of in-depth reviews related to emerging trends in the statistical environment. The responsibility of the Secretariat of CES is UNECE as these two organisations join efforts to promote statistical innovation to tackle emerging challenges and opportunities in official statistics and exchange experiences between statistical organisations.

In the action scope of UNECE, the **High-level Group for the Modernisation of Official Statistics (HLG-MOS)** was established by the CES to actively foster the development of statistical organisations while working together in a collaborative and voluntary way towards the modernisation of official statistics. It embodies Chief Statisticians of several national and international organisations that define the vision, mission and priorities in the international statistical community. This group focuses on modernising statistical concepts, standards and business processes through the implementation of new technologies, methods and capabilities in statistical organisations while attending to the emerging trends, challenges and opportunities in the ever-changing digital environment and the implications to official statistics. HLG-MOS also works very closely in developing and supporting statistical frameworks/models and standards through several task teams assigned to their maintenance to ensure stability over time and discussion forums to follow up and gather feedback from colleagues and experts through revision rounds and global consultations.

The Supporting Standards Group is the operational responsible for the development, promotion and maintenance/revision of such standards and frameworks/models and has been particularly important for statistical organisations in the last years in improving the efficiency of statistical production processes and better meeting user needs. Some of those standards and frameworks/models will be described in more detail in the next subchapter.

HLG-MOS also works in projects and other modernisation initiatives through several work programmes in areas such as Big Data, Data Integration, Data Governance for Interoperability and ML helping the implementation of new technologies, methods and other capabilities in statistical organisations, for instance, thorough training material.

I.1.2. STATISTICAL FRAMEWORKS

The HLG-MOS's Supporting Standards Group has developed several common frameworks and standards - 'ModernStats models' - to be widely used by the international statistical community and to improve and streamline statistical production carried out by statistical organisations. The **Generic Statistical Business Process Model (GSBPM), Generic Statistical Information Model (GSIM), Generic Activity Model for Statistical Organizations (GAMSO)** and **Common Statistical Production Architecture**

(CSPA) are the main ones developed and peer-reviewed by the international statistical community (GEOSTAT 4, 2021a).

These high-level global frameworks comprise reference models and give a generic overview of the statistical framework environment in the international official statistics community. Moreover, these international frameworks - considered 'industry assets' - are custodians of (or administrated by) the HLG-MOS's Supporting Standards Group which also supports and facilitates the implementation of such frameworks and related standards needed for statistical modernisation.

It is important to mention that other statistical frameworks and respective models have been developed to supplement or support the ModernStats activities within the international statistical community. The Common Statistical Data Architecture (CSDA) (UNECE, 2018) is a good example of guiding statistical organisations on how to organise and structure their processes and systems for efficient and effective data and metadata management, including dealing with new types of data sources, such as Big Data to be embodied in the statistical production process.

The **GSBPM** is one of the cornerstones of the standards-based modernisation strategy of HLG-MOS and has been widely adopted as a standard business process[2] model by the international statistical community since its development in 2008. The model can be seen as a background template to design, plan and implement processes related to digital transformation within NSO (PARIS21, 2022). GSBPM is in its 5.1 version released in 2019 (UNECE, 2019a) and aligned with both current versions of GSIM (2.0) and GAMSO (1.2).

From a more general perspective, this framework can be defined as an "ordered collection of related and structural logical activities and tasks performed by statistical producers to convert data input into statistical information" (Ariza-López et al., 2021:5). GSBPM describes and guides the overall process of statistical business production and

[2] In general terms, a business process model is defined as how organisations create products, services and policies. It is a succession of interconnected activities that, starting from an identifiable input, results in a defined output as a product or service. In addition, a business model is a business design structure guiding an organisation in creating and retaining value.

the sub-processes needed to produce official statistics in a general and process-oriented way, i.e. adaptable to any organisation (UNSD, 2021).

This reference model is flexible since it can be directly implemented or used as a baseline for a customised version to fit the organisational context and national specifications. GSBPM is designed to be applicable regardless of the data source (e.g., survey, administrative registers, etc.) being used for data integration and metadata standards, however it mainly focuses on primary data sources or traditional statistical sources (e.g., survey data) and primary statistics. Thus, it acknowledges the growing importance of integrating statistical data with geospatial data in which some sub-processes include activities using geospatial data. In addition, it can also be used as a template for process documentation, for harmonising information and technological infrastructures for statistical production and providing a framework for process quality assessment and improvement (e.g., quality indicators for survey and administrative data).

The framework breaks down into three levels: i) Level 0, also known as the 'management level' comprises the statistical business process together with the overarching processes (i.e. management sub-systems); ii) Level 1, the model comprising eight phases of the statistical business process, divided into 44 sub-processes; and iii) Level 2, the sub-processes within each phase (Figure 1). The Overarching Processes are: Quality Management, Metadata Management, Data Management, Process Data Management, Knowledge Management and Provider Management. The sub-processes within each phase identify possible steps in the statistical business process and the inter-dependencies between them following a logical sequence but not strict nor mandatory.

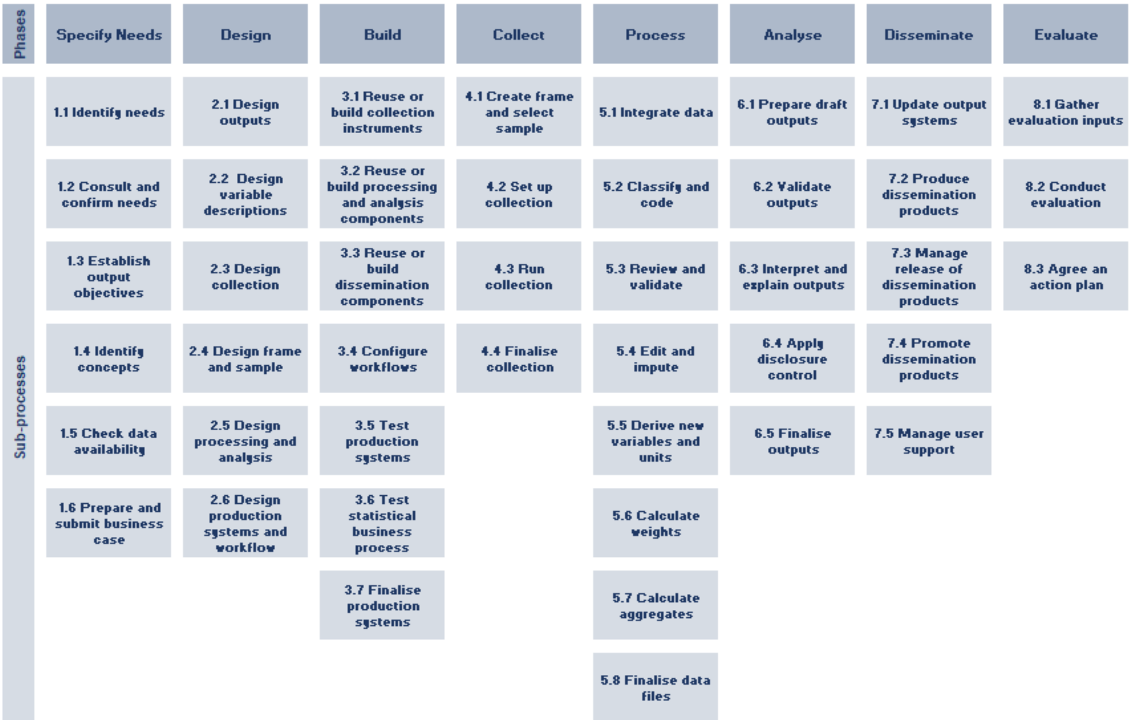

Figure 1. GSBPM structure with the 8 Phases (Level 1) and respective 44-sub-processes (Level 2) (source: author, from UNECE, 2019a).

Level 1 is made up of the following eight phases of the statistical business process: i) Specify Needs; ii) Design; iii) Build; iv) Collect; v) Process; vi) Analyse; vii) Disseminate; and viii) Evaluate. In terms of the descriptions of each phase and the number of sub-processes, the following contents are briefly presented:

i. Specify Needs phase is prompt when a need for new statistical products is identified or feedback about current statistics is applied including activities related to stakeholder engagement, current and future statistical needs and preparation of business cases to meet those needs.

ii. Design phase describes the development and design activities and elements needed to define or refine the statistical products and services, and any related work needed to define the statistical outputs, concepts, methodologies, collections methods and operational processes.

iii. Build phase addresses the development and improvement of systems used in the collection, processing, analysis and dissemination of data and focuses on the end-to-end process.

iv. Collect phase involves the different collection methods (e.g., acquisition, collection, extraction, transfer) and all necessary information (e.g., data, metadata and paradata) for further processing.

v. Process phase describes the processing of input data and their preparation for analysis, including carrying out the necessary tabulation, data checking cleaning, transformation and preparation for analysis.

vi. Analyse phase involves activities where statistical content is prepared and statistical outputs are produced and examined to ensure they "fit for purpose" before dissemination to users (e.g., disclosure control).

vii. Disseminate phase manages the release of the statistical products to users, including all activities associated with assembling and releasing a range of static and dynamic products throughout different channels.

viii. Evaluate phase manages the evaluation of a specific instance of a statistical business process and can take place either at the end of a specific sub-process or in a continuous manner during the statistical production process.

Despite NSO and other producers of official statistics, GSBPM is also used by statistical departments of international organisations, such as IMF, to modernise data management functions to increase the efficiency of the set of business processes and sub-processes and improve data quality. In a broader context, GSBPM is also being used as a framework to strengthen data governance through building a new data management platform, defining roles and responsibilities, creating directives and procedures and allocating resources for data projects.

The **GSIM** is an internationally endorsed reference framework representing the core pieces of information needed to produce statistical outputs (Ariza-López et al., 2021). It is an overarching conceptual framework of internationally agreed definitions, attributes and relationships describing the pieces of information (information objects) within the statistical production process. It establishes a conceptual model mapping terms, definitions and their relationships and generic descriptions, management and use of data and metadata supporting the production of official statistics. An illustrative example: a statistical classification is based on a concept which is a type of population and measured by a variable.

GSIM helps statistical organisations deal with the data/information flows within the production process by providing standardised information classes to facilitate the reuse and sharing of methods, tools and processes and improve communication within and between statistical organisations (UNECE, 2024b). The consistently described information classes can constitute a common language between the different roles in the statistical organisation (business and IT experts) and statistical domain subjects as well as be used as inputs and outputs to automate processes, streamline the production of official statistics and align standards at the national and international levels.

GSIM was first developed by and for the international statistical community in 2012 and has been reviewed several times to incorporate new developments and user needs. The last version (version 1.2) came out in 2019 August (UNECE, 2019b) and the current version (version 2.0) was released in February 2024 to improve how the information flow is harmonised inside the processes within the statistical production across different thematic domains. Compared to the previous version the main changes involved adding, removing and renaming information classes to simply the model and clarify meanings, and updating their relationships in some focus areas, such as metadata and design of business processes.

The GSIM is made up of five top-level information object groups: i) Base Group provides features that are reusable by other objects to support functionality and changes in the information objects in the whole framework; ii) Business Group handles the designs and plans of the statistical programmes and the processes required to deliver them; iii) Concept Group addresses the meaning of information to provide an understanding of what the data is measuring throughout the statistical business process. This group includes statistical content such as definitions of concepts, variables, populations, unit types, statistical classifications and code lists; iv) Exchange Group describes the information that comes in and goes out of a statistical organisation covering exchange channels and provision agreements related to data provision and data delivery; and v) Structure Group is used to structure information throughout the statistical process to understand how data is structured, the format and storage location and jointly considers both data and metadata.

The implementation of GSIM has created a ready environment for the reuse and sharing of methods, components and processes enabling statistical organisations to rethink and redesign their business processes by defining common objects in the statistical production. Moreover, this model has also strongly contributed to semantic interoperability enabling statistical organisations and stakeholders to improve their communication and exchange and comparability of data and information through a common language, terminology and conceptual baseline. In the context of digital transformation, this model can be used to shape digital data flows as a template to design and develop data structures and concepts (PARIS21, 2022).

In terms of standards, this framework does not provide any type of standard developed by its structure, however it is expected to be implemented using external standards and models that support technical implementation within the organisation. Therefore, GSIM works as an add-in conceptual model to implement standards, including both statistical and non-statistical standards, or other relevant standards for statistical information. The Statistical Data and Metadata eXchange (SDMX) is an example of a statistical standard enabling mapping to implementation models and establishing a link between GSIM and its technical implementation.

The GSIM and GSBPM are complementary for the creation and management of statistical information and the production of official statistics providing added value when both frameworks are applied together. GSIM helps to describe GSBPM sub-processes by identifying and defining the information objects as inputs (e.g., dataset or variable) and outputs (e.g., transformed or new information objects) flowing through the GSBPM phases and sub-processes. Also, GSIM facilitates the metadata management overarching the production process (i.e. GSBPM) and supports a consistent approach regarding metadata.

The **GAMSO** describes overarching activities and management processes taking place within statistical organisations and defines capabilities needed to support statistical production (UNECE, 2019b). GAMSO is in version 1.2 (January 2019) following the review of the current version of GSBPM (version 5.1) to ensure consistency and clarity between both reference models in which there were no structural changes, but only minor improvements concerning the descriptions. This is a requirement since GAMSO is an extension and complementary framework to the GSBPM by identifying additional management activities of the statistical production (GEOSTAT 4, 2021a).

GAMSO focuses on broader aspects of organisational management for modernising statistical production rather than the specific production process. It includes four activity areas covering high-level management and strategic aspects: i) Strategy and Leadership; ii) Capability Development; iii) Corporate Support; and iv) Production. It also provides a common vocabulary to support international collaboration activities and a conceptual ground for resource planning within the statistical organisation (Ariza-López et al., 2021). Moreover, depending on the specificities of some statistical organisations, such as the level of centralisation of systems and degrees of responsibility for coordination, additional activities might be applied.

The framework is applicable across all kinds of government and organisational levels of statistical organisations, without overlooking the administrative and political context. For instance, this framework can be implemented for cost measurement to produce official statistics or as a tool to measure and communicate the value of modernising statistical activities within an organisation.

The **CSPA** is a reference industry/enterprise architecture for official statistics developed to face common challenges and emerging information needs in the statistical environment, including rigid processes and methods, ageing technological environments, less-holistic architectures and IT solutions and insufficient resources (UNECE, 2021b). CSPA provides a means for the international statistical community to develop, share and reuse shareable statistical components in statistical production within and across statistical organisations through a more collaborative development approach.

CSPA aims to enhance interoperability and reduce the cost of developing and maintaining processes and systems through a set of agreed common principles and definitions (GEOSTAT 4, 2021a). The current version (version 2.0) is a living document in order to stay update and relevant in constantly changing statistical and business environments, making it a non-static reference architecture with a long-term implementation perspective. It has also a descriptive feature since is a standard non-normative framework focusing on supporting the facilitation, sharing and reuse of solutions and services across and within statistical organisations, and guiding their transformation and modernisation processes (UNECE, 2021b). CSPA will allow statistical organisations to make their business processes and systems more flexible and align the components of statistical production, namely by promoting international collaboration.

The scope of the CSPA recognises four architecture areas: i) Business Architecture (defines what the industry does and how it is done); ii) Information Architecture (describes the information, its flows and uses across the industry, and information management); iii) Application Architecture (describes the set of practices used to select, define and design software components and their relationships); and iv) Technology Architecture (describes the infrastructure technology supporting the other architecture perspective). The Application Architecture area is important from a geospatial perspective as it includes services on classification (management and use of geographical classifications), registry (register of location data), geography (geospatial data) and statistical metadata (statistical metadata aligned with geospatial metadata).

CSPA builds on and uses other HLG-MOS models (GSBPM, GAMSO and GSIM) to ensure alignment for the development, sharing and reuse of the statistical components and a common understanding of the different statistical production elements. Hence, the common vocabulary provided by CSPA is built on a vision of standardised architecture and shareable services to streamline statistical pipelines and promote interoperability and harmonisation across statistical production.

I.1.3. STATISTICAL QUALITY AND STANDARDS

I.1.3.1. STATISTICAL QUALITY

One key dimension of official statistics is quality. The aim of NSO and respective NSS is to develop, produce and disseminate high-quality statistics according to quality frameworks supported by statistical quality assurance requirements (Eurostat, 2014). The extent to which these quality assurance requirements have been fulfilled addresses the quality assessment of statistical activities and the evaluation/review of the processes and resulting statistical outputs in the final stage of the statistical production process (UNSD, 2021). In addition, quality management covers a higher level of coordination regarding all activities related to quality (e.g., planning, control and improvement) at each stage and in the overall statistical production process to detect quality errors, correct them and ultimately improve the quality of statistical outputs. Statistical quality seeks to meet established quality confidence ensuring that statistical outputs are fit-for-purpose according to user needs and that are produced effectively and efficiently through the evaluation of such achievement under a statistical quality framework.

Quality management operates under statistical quality frameworks with associated standards, tools and methods for quality assessment and risk management mechanisms in a multidimensional perspective (UNSD, 2019). Statistical quality frameworks also encompass quality assurance aspects, including recommendations and codes of practice with a normative nature. Some of them were inspired by general quality frameworks applied to any organisation, products and services, for instance, the ones developed by ISO focusing on certification and standardisation (e.g., the ISO 9000 family standards focusing on quality management systems). Over the last decades, high-level international organisations and statistical systems such as the OECD, the UNSD and the ESS, in collaboration with national statistical authorities, have developed a comprehensive set of international statistical frameworks, standards[3] and quality recommendations covering almost all statistical domains (OECD, 2007). Some examples are reviewed in this subchapter.

Some of the above-mentioned international statistical information and process-oriented standards and frameworks/models, such as GSBPM and GSIM, also cover quality issues. These statistical frameworks and standards aim to ensure a common agreed ground on the quality of statistical terminology, processes and outputs. Furthermore, statistical quality management and metadata management are also intrinsically linked in the overall statistical production process including the quality criteria of official statistics and user confidence, i.e. reference metadata.

Lastly, it is important to outline that current and emerging trends and resulting challenges in official statistics - some already mentioned above - may require updating

---

[3] As defined in European legislation (Article 1, paragraph 6 of Directive 98/34/EC#), a standard is a technical specification, approved by a recognised standardisation body for repeated or continuous application, with which compliance is not compulsory and which is adopted as an international standard, a regional standard or a national standard.

traditional quality frameworks as well as revising the concept of statistical quality and its dimensions to accommodate the implications of modernising official statistics (UNECE, 2021a).

I.1.3.2. STATISTICAL QUALITY FRAMEWORKS AND STANDARDS

International organisations in the statistical domain and statistical systems developed and adopted statistical quality frameworks containing recommendations, implementation guidelines and quality assurance tools and methods. These can be either quality assurance frameworks or quality management frameworks encompassing systems and procedures supporting quality assurance or quality management within statistical organisations, respectively (UNSD, 2019; UNSD, 2021). These frameworks aim to guide NSS and respective statistical organisations in the production of high-quality official statistics and have been updated considering new data sources, new data providers and the SDG indicators (UNSD, 2019). These statistical quality frameworks are also aligned with the legal and technical frameworks of the respective statistical systems and have a firm legal basis and robust mechanisms for quality assurance in official statistics.

The UNSC adopted the UN National Quality Assurance Frameworks Manual for Official Statistics in 2019 which includes the frameworks, implementation guidance and recommendations to manage quality in official statistics at the international level (UNSD, 2019). The Manual aims to respond to the challenges and opportunities in official statistics under the 2030 Agenda for Sustainable Development by providing guidance to support countries in developing and implementing an NQAF. Moreover, it also identifies several benefits of implementing a quality framework, including systematic monitoring and ongoing identification of risks and quality issues. It includes standard guidance and training materials in quality assurance within the NSS through the coordinating role of the NSO.

The Manual identifies core overarching recommendations and addresses quality assurance in official statistics based on a set of principles and associated requirements on four levels - statistical system, institutional environment, statistical processes and statistical outputs - and linked to the 10 UNFPOS. Quality terminology is also introduced and related to other international statistical community frameworks (e.g., GSBPM, GAMSO and GSIM). In this regard, it provides a more general perspective of the GSBPM overarching process regarding quality assurance in production processes (Ariza-López et al., 2021). At the operational level, specific development and implementation actions are provided, including quality considerations for different data sources (e.g., administrative data and other data sources such as geospatial data).

Several NSO have developed and implemented their individual quality assurance frameworks inspired by this manual and others have their quality principles, policies and guidelines from their respective statistical systems (UNSD, 2021). The statistical offices of Canada, Finland, Australia, the United Kingdom and South Africa are examples of designing and developing their quality frameworks and guidelines.

The current (and third) edition of the European Statistics Code of Practice (ES CoP) was adopted in 2017 (Eurostat, 2017) following the second revision process, alongside the ESS Quality Declaration that endorses the commitment to independent and high-quality official statistics at European, national and regional levels. As the cornerstone of ESS quality management, the ES CoP sets the standard for quality European statistics according to a set of principles connected to the institutional environment, statistical processes and outputs. It is a self-regulatory instrument built on 16 statistical quality principles grouped under these three thematic areas in which each principle includes a set of indicators of best practices and standards providing implementation guidance and ensuring transparency and comparability within the ESS. The following table (Table 3) presents the ES CoP 16 Principles, their description and the number of indicators, grouped by thematic area.

Table 3. European statistics Code of Practice (ES CoP) (ESS, 2017).

Area	Principle	Name	Description	Indicators
Institutional environment	Principe 1	Professional Independence	Professional independence of statistical authorities from other policy, regulatory or administrative departments and bodies, as well as from the private sector, ensures the credibility of European statistics	8
	Principle 1bis	Coordination and cooperation	Statistical authorities actively cooperate within the partnership of the ESS, so as to ensure the development, production and	3

Area	Principle	Name	Description	Indicators
			dissemination of European statistics	
	Principle 2	Mandate for Data Collection and Access to Data	Statistical authorities have a clear legal mandate to collect and access information from multiple data sources for European statistical purposes. Other stakeholders may be compelled by law to allow access to or deliver data at the request of statistical authorities	4
	Principle 3	Adequacy of Resources	The resources available to statistical authorities are sufficient to meet European statistics requirements	4
	Principle 4	Commitment to Quality	Statistical authorities are committed to quality by systematically and regularly identifying strengths and weaknesses to continuously improve process and output quality	4
	Principle 5	Statistical Confidentiality and Data Protection	The privacy of data providers, the confidentiality of the information they provide, its use only for statistical purposes and the data security are guaranteed	6
	Principle 6	Impartiality and Objectivity	Statistical authorities develop, produce and disseminate European statistics respecting scientific independence and in an objective, professional and transparent manner in which all users are treated equitably	8
Statistical Processes	Principle 7	Sound Methodology	Sound methodology underpins quality statistics requiring adequate tools, procedures and expertise	7
	Principle 8	Appropriate Statistical Procedures	Appropriate statistical procedures, implemented throughout the statistical processes, underpin quality statistics	7
	Principle 9	Non-excessive Burden on Respondents	The response burden is proportionate to the needs of the users and is not excessive for respondents. The statistical authorities monitor the response burden and set targets for its reduction over time	6
	Principle 10	Cost Effectiveness	Resources are used in an effective manner and statistical processes are optimised	4

Area	Principle	Name	Description	Indicators
Statistical Output	Principle 11	Relevance	European statistics meet the user needs as well as consider and anticipate their emerging needs and priorities	3
	Principle 12	Accuracy and Reliability	European statistics represent reality in an accurate and reliable way	3
	Principle 13	Timeliness and Punctuality	European statistics are released in a timely and punctual manner	5
	Principle 14	Coherence and Comparability	European statistics are internally consistent, comparable between regions and countries, and over time enabling to join related data from different data sources	5
	Principle 15	Accessibility and Clarity	European statistics are presented in a clear and understandable form, released in a suitable and convenient manner, available and accessible on an impartial basis with supporting metadata and guidance	7

This common quality framework complements and strengthens the European legal framework relating to quality, which is based on the Amended Regulation (EC) No 223/2009 on European statistics (2015).

One of the elements of the ESS quality framework is the Quality Assurance Framework (ESS QAF). It is a self-regulatory statistical quality framework complementing the ES CoP and the general quality management principles at process/output and institutional levels. Thus, the ESS QAF assists the implementation of the ES CoP by the ESS statistical authorities in terms of planning, monitoring and assessment, and transposes ES CoP principles and indicators into concrete actions. The ESS QAF is currently in its second version (2.0 version, 2019), following the 2017 revision of the ES CoP.

Whereas the ES CoP establishes high-level strategic and conceptual guidance, the ESS QAF provides more operational-oriented guidance with examples based on more detailed quality assessments and reporting activities considering national circumstances. In this sense, the ESS QAF represents a collection of non-binding methods, tools and good practices on statistical quality for further use and/or already in use. Each principle and indicator have a set of suggested institutional methods and/or process/output methods with a more detailed description. At this level, the framework covers sector-specific quality assurance methods and tools where the quality of statistical processes and outputs are assessed and reported based on standardised rules and reporting structures.

Furthermore, the ESS QAF together with the ES CoP and statistical sectors formed the quality policy of Eurostat which centrally carries out the quality assurance work through regular quality reviews and evaluation procedures. In summary, the ESS quality policy has four levels of quality assurance: i) the ES CoP provides the two highest levels concerning the 16 Principles and 84 Indicators and addressing quality of the institutional environment; ii) the third level encompasses the ESS QAF and provides good practices, methods and tools for practical implementation of the ES CoP; and iii) the fourth level covers sector-specific quality assurance methods and tools where the quality of statistical processes and outputs are assessed and reported based on standardised rules and reporting structures.

Standardisation is a fundamental element of quality enabling the modernisation of official statistics and making statistical production more efficient (Eurostat, 2020). Statistical quality frameworks are usually built on both general and statistical quality standards, i.e. domain-specific standards for statistical production (UNSD, 2019). In official statistics, standards are generally a set of documented requirements to be fulfilled based on user needs concerning either a statistical output or product, a process, the entire statistical production or even a quality management system.

In general, statistical standards aim to promote best practices and the use of common concepts in official statistics compiled by national statistical authorities and international organisations within the statistical community (OECD, 2007). The role of standards in statistical quality management, with an emphasis on metadata management, has been increasingly recognised due to a more dynamic data ecosystem where multiple stakeholders are producing and exchanging statistical data in different formats and statistical outputs using various methods and tools.

At the global scope, the SDMX is an international standard that facilitates and improves the exchange of statistical data and metadata via IT technology and is embodied in technical specifications documentation with several versions released since 2004. The information model is its conceptual cornerstone by describing the standardised object model for modelling the structure of the statistical datasets, the coding schemas for classification and the rules for data exchange, including data quality checking. Although SDMX was initially designed for official statistics, particularly for aggregated statistical data, it has been evolving and becoming flexible to support microdata and several unstructured data formats from many different domains. More recently, SDMX version 3.0 (2021) introduced new features, improvements and changes by expanding technical specifications for geospatial metadata and improving the management of associated data. This update included new geospatial-related concepts, new specialised code lists on geospatial features and a new component representation type that allowed the use of SDMX datasets as geospatial input data for Geographic Information Systems (GIS) applications for spatial representation, i.e. maps. Moreover, it enabled connecting statistical data to geographical characteristics and embodying detailed geospatial structural and reference metadata in the exchange of statistical data (UNECE, 2021c). This new version relied on the collaboration of the geospatial community under a global roadmap on statistical-geospatial data integration, i.e. strengthening the bridge between statistical and geospatial data and related communities. In addition, the Data Documentation Initiative (DDI) is another example of a modern standard used by the statistical community and it is recommended by CSPA to increase the sharing of data and information between systems and architectures and consequently modernise statistical production. Both DDI and SMDX should provide a standardised foundation to represent statistical data and metadata and support the statistical community to operate their own general data models and metadata capabilities.

In the context of the ESS, quality standards are defined in the ESS Handbook on Quality and Metadata Reports (Eurostat, 2020) and the quality reports are compiled for the individual statistical datasets providing relevant information for users on methods for data collection and validation, both at national and European levels. The EHQMR is included in the Catalogue of ESS Standards, therefore being recognised as an ESS standard. The Single Integrated Metadata Structure (SIMS) 2.0, incorporated in the EHQMR, is the most well-known and used dynamic inventory and the conceptual framework for European statistical quality and reference metadata reporting providing definitions and guidelines for producers and users. Hence, SIMS enables streamlining and harmonisation of quality reporting and metadata across the statistical domains and countries of the ESS (Eurostat, 2020).

The EHQMR also includes the ESS version of the SDMX (Euro-SDMX Metadata Structure) with descriptions of the concepts and sub-concepts applied in the European statistical environment. In this regard, Eurostat has been developing publicly available SDMX tools for data and metadata providers, data receivers and developers, highlighting the SDMX Converter as a tool that converts statistical datasets between eleven different formats.

In the end, changes occurring in the statistical operating environment, such as new or revised standards for data and metadata management, data protection and statistical confidentiality concerns related to microdata and new techniques and methods pose several quality challenges to statistical organisations. This new paradigm will require modernisation of the statistical production in terms of overarching quality management and innovative quality assessment and risk management mechanisms, balancing between driving innovation and standardisation benefits with a more flexible approach to statistical quality activities.

I.2. GEOSPATIAL DATA

This chapter primarily focuses on the public usage and institutional application of geospatial data for statistical organisations considering global and regional policy frameworks and overarching trends in the geospatial domain, especially in geospatial data production and management. The key stakeholders and their roles within the geospatial community are identified and described, and the existing geospatial infrastructures and frameworks are systematised while aspects related to quality and domain-specific standards are introduced. Given that the geospatial domain is more open, heterogeneous and business-oriented than the statistical domain, particular attention will be devoted to topics related to geospatial data for statistical purposes.

In addition, this chapter is not an in-depth review of geospatial data nor provides insights from the extensive technical work and academic research on the topic, particularly concerning Geographic Information Science (GIScience) and surrounding issues.

I.2.1. GEOSPATIAL DATA, PRODUCTION AND KEY STAKEHOLDERS

I.2.1.1. GEOSPATIAL DATA

Geospatial data - traditionally called ‘spatial data’ - are the “sum of our interpretation of geographic phenomena” (Guptil, 2001:14775) in the digital age, the primary information used by GIS and other software tools for spatial analysis. As pointed out by Guptil (Guptil, 2001), spatial data describes the phenomena on or near the Earth’s surface addressing any piece of information related to location. GIS are computer-based systems capable of storing, manipulating and visualising geospatial data, i.e. any geospatially referenced in the digital environment (Kadmon, 2001). In addition, geospatial data can be related to the concept of ‘geographic information’ or ‘geoinformation’, encompassing all cartographically represented and georeferenceable data, going beyond traditional cartographic information (Julião, 2001). For the aim of this thesis, the term ‘geospatial data’ will be used throughout the entire work since it is more comprehensive and extensively used in many related research and application contexts that do not necessarily involve geographic location and its graphic representation in the form of a map.

The UN Committee of Experts on Global Geospatial Information Management (UN-GGIM), a leading global organisation in the geospatial domain, defines geospatial data as a description of the physical location of features on, above and below the surface of the Earth, and the relationships between these geographical features (UN-GGIM, 2022). From an institutional perspective, geospatial data is also referred to as data with direct or indirect reference to a specific location or geographical area, according to the legal definition in the European INSPIRE Directive (EU, 2007). The term ‘spatial data’ can also be interchanged with ‘geodata’ or ‘location data’ by establishing a digital connection between a place, associated individuals and businesses, and the surrounding natural and built environment.

The conceptual link between geospatial data and location data is the most comprehensive and acceptable within the statistical community, where its three elementary components involve location (where), attribute(s) (what or who) and time (when). Moreover, the UNSD addresses geospatial data as an alternative data source for official statistics, referring to data with implicit location information that could be attached to administrative records (UNSD, 2021). Therefore, the statistical community recognises the importance of geospatial data since it provides the content and context for understanding natural and human systems which has always analysed data about people, built and natural environments (Hadley, 2018).

In the GIS environment, there are two basic types/formats of geospatial data, vector and raster. They establish the conceptual division from the 1990s between discrete objects and continuous surfaces with different models, structures and implications in processing, analysis and digital representation (Goodchild, 2010). Vector data is built on points, lines and polygons to spatially represent well-defined objects with discrete boundaries, such as buildings (points), roads (lines) and cities (polygons). These basic elements of the vectorial model are topologically connected, from polygons to points stored as geographic coordinates (x,y) for accurate positioning. Raster data is formed by a grid, often composed of linear sequences of cells or collections of adjacent squared cells along two perpendicular axes, to spatially represent continuous objects. Thus, a grid is a structured arrangement of cells that creates a framework to organise data allowing a certain absolute value to be associated with a discrete part. Raster models store data that varies continuously (e.g., elevation surface) and are related to the concept of spatial resolution connected to the size of grid cells or pixels (smaller sizes of grid cells provide higher/finer spatial resolution and therefore, more disaggregated and spatially accurate data).

Similar to statistics, geospatial data can be either official or non-official. Official geospatial data is commonly referred to as 'authoritative geospatial data' addressing data collected, maintained and published by official data sources and competent and reputable authorities, for instance, a governmental agency or national geospatial authority, who assigns an authenticity certificate. The national geospatial authorities, including mapping, cadastral and land registration authorities - responsible for the production and provision of authoritative geospatial data - play a key role in providing technical and methodological guidance in geospatial data management for building national geospatial infrastructures and enhancing geospatial capabilities. They provide reliable and trusted geospatial data complying with agreed legal and technical frameworks at the national and international levels and are more oriented to public usage and regulatory purposes.

Authoritative geospatial data also supports sound and robust policy and legal frameworks and strengthens the geospatial ecosystem by providing reliable and trusted data as a public good for achieving societal, environmental and economic benefits and sustainable development challenges (UN-GGIM, 2024). However, drivers and trends in geospatial data creation and management and the changing technological environment over the last years (e.g., new data sources and user requirements) have made the differences between authoritative and non-authoritative data less clear (UN-GGIM, 2020a).

The Global Fundamental Geospatial Data Themes is an overarching example of an international framework promoting core authoritative geospatial data to tackle data gaps and needs in the SDG, especially in developing countries that face problems in data quality, timeliness and interoperability (UNECA, 2007; Scott & Rajabifard, 2017). The list of the 14 themes was created as core reference datasets in global geospatial information management to strengthen geospatial data infrastructures and capabilities and to be implemented within the scope of SDG (UN-GGIM, 2018; UN-GGIM, 2019). These minimum primary datasets cover traditional examples of authoritative geospatial data, such as geology and soils, population distribution, orthoimagery, land cover/land use and land parcels, i.e. cadastral data, in which these last two are datasets traditionally provided by national geospatial authorities (See et al., 2016). Authoritative location data, such as addresses and buildings, are also included to provide global authoritative location references alongside indicative relevant data standards to support government administrations at all levels. The minimum list was extended to national fundamental data themes to support the goals and targets of the global indicator framework and strengthen the geospatial capabilities of countries (Scott and Rajabifard, 2017).

Non-authoritative geospatial data include any geospatial data centrally provided by the private sector, i.e. businesses from the geospatial industry and major web companies and crowdsourced geospatial data, including Volunteered Geographic Information (VGI), and related to user-generated content. Crowdsourced geospatial data and VGI encompass data collection from voluntary and non-expert action representing a paradigm shift in the way geospatial data are created and shared (Elwood et al., 2011; Zhang & Zhu, 2018). Thus, users around the world from different backgrounds and expertise can actively contribute to a geospatial database or infrastructure by creating new data content, adding missing data, correcting/altering and/or updating existing one. Crowdsourcing is being used as a source of geospatial data when there is no authoritative data available - or even as auxiliary data when such data exists - as well as an alternative to traditional geospatial infrastructures (e.g., open data) turning out to be a very convenient data source when the quality of crowdsourced data is better (UNECE, 2016).

The Internet, the proliferation of mobile devices, location-based services and IoT devices broadcasting location data allowed more citizens to have more direct contact with geospatial data and technologies (e.g., Google Maps). Some of these technological trends, particularly in GIS technology, changed the statistical landscape from the dominance of data collection to multimode inputs built on microdata. Thus, citizen engagement has become a major focus of geospatial technology through independent location-based technologies (i.e. location recording) and intuitive user interfaces (e.g., Global Positioning System - GPS - embedded in smartphones) enabling citizens to provide data and other kinds of input using web services according to their specific needs (See et al., 2016; Dangermond & Goodchild, 2020). Due to requiring lower financial resources, crowdsourcing practices focusing on volunteered data collection and citizen-centred collaborative platforms have become important sources of active crowdsourced geospatial data, such as OpenStreetMap (OSM). OSM creates, updates and maintains the most complete and free base mapping source via an open volunteered community with more or less expertise that reached the milestone of 100 million edits (Dangermond and Goodchild, 2020; Coetzee et al., 2021). In contrast, passive crowdsourced geospatial data are collected via an out-of-knowledge agreement, i.e. not voluntarily requested or provided by the user, such as social media data and any mobile technology recording location.

These forms of non-authoritative geospatial data collection raise quality problems, ethical concerns and issues related to copyright, ownership, data privacy and licensing (See et al., 2016). Unstructured data, lack of data collection control, transparency in data management, interoperability and quality assurance measures assessing bias effects for official statistics are just a few examples (Zhang & Zhu, 2018). Assessing and correcting the quality implications of crowdsourced geospatial data and VGI should be a priority in future research (e.g., data standards) towards more reliability and usability with authoritative data. The value and potential of non-authoritative geospatial data cannot be overlooked despite some of these technical shortcomings, particularly when these types of data collection are becoming more established in a way to tackle data gaps in terms of spatial/temporal coverage, thematic attributes and general data completeness aspects. Even recognising that they cannot fill all data needs, a reasonable trade-off between quality, flexibility and standardisation should be assured while at the same time addressing important questions about data availability and frequency of data collection and dissemination (e.g., metadata and cost-effectiveness, etc.).

Geospatial data are increasingly being used to address key global and regional challenges and support multi-level policy frameworks through high-level international coordination on geospatial data management topics, providing strategic recommendations, connecting to policies and actions for sustainable development and responding to user needs (Haldorson et al., 2016). Geospatial data has been actively contributing to and aligning with global development agendas by promoting efforts on their maintenance, management and quality, particularly in authoritative data, to better inform governments for policy formulation and monitor progress at national and global levels (UN-GGIM, 2020b). The UN 2030 Agenda generally recognises the value of geospatial data as a fundamental asset to achieve global goals and national targets as being implemented and integrated into sustainable development processes, especially at local and national levels (Scott & Rajabifard, 2017). This acknowledgement comes from the UN resolution (UN A/RES/70/1, 2015) that engaged the geospatial community to exploit a wide range of geospatial data sources while promoting public-private partnerships.

The importance given to geospatial data also comes from the need to have geographic location embedded in data inputs and outputs to consistently support, track and report the SDG indicators and targets. Moreover, geospatial data has been increasingly playing a key role as complementary to the traditional data and administrative sources under the SDG roadmap tackling data disaggregation needs (e.g., smaller or non-administrative, functional geographies) and making indicators more accurate and comparable across countries and regions. For instance, the integration and aggregation of geospatial data with statistical data is essential to produce some indicators, such as the 11.3.1 (Ratio of land consumption rate to population growth rate). This idea was also endorsed at the first UN World Geospatial Information Congress in 2018 as an overarching theme called “The Geospatial Way to a Better World” opening the discussion about the geospatial dimension in the SDG and recognise that data for the SDG must be geospatial to support the measuring and monitoring their progress over time and across countries and regions (Kraak et al., 2020).

Considering geospatial data as official data for the SDG, the SDG Geospatial Roadmap was designed to provide action-oriented guidance on using geospatial data and technologies for measuring, monitoring and reporting the geospatially enabled SDG indicators (UN-GGIM, 2021). At the global level, 46 of the 169 SDG targets (27%) require geospatial data to monitor the progress of each of the 17 UN SDG in which user requirements have been investigated and identified through showcasing common and specific examples from several countries (UN-GGIM: Europe, 2016). Thus, geospatial data and services are increasingly required for policy formulation, implementation and monitoring at the EU level, especially in policies with territorial impact (Bernard et al., 2005). One of the most promising geospatial data sources for policy-based applications is Earth Observation (EO) data. According to UN-GGIM (UN-GGIM, 2020a), EO data is one of the major future technological trends in the geospatial data community, shaping the direction of geospatial data creation, management and analysis for policy purposes.

EO developments have created opportunities for the modernisation of official statistics and challenges for the NSO in the context of the overarching policy agendas, including the global indicator framework for the SDG that requires more timely data at a more disaggregated level (CEOS, 2018; UNECE, 2021). Technological advances in remote sensing have provided more data availability about the Earth's surface, i.e. digital trails, with a higher spatial resolution and temporal frequency for lower data collection and processing costs enabling timelier outputs at a more disaggregated level (UNSD, 2017). The expansion of EO data and technology also resulted from innovations in the space industry regarding data collection techniques, the rise of AI and Big Data (e.g., advanced ML techniques for image analysis and data extraction) and improvements in computational capabilities (e.g., cloud computing) have helped to overcome processing and time constraints. These developments improved EO cost-effectiveness - monitoring solutions are less expensive and challenging - alongside enhancements in imagery quality, remote sensing software and open cloud processing tools (e.g., Google Earth Engine) (UNECE, 2019c). Therefore, EO data, particularly satellite imagery data and related technology, hold great potential for NSO in sustainable development monitoring and reporting, namely to track environmental and land use changes or climate metrics, and provide more quality, accuracy, and timeliness to official statistics. In this regard, the UN SDG have been a decisive driver in the use of EO data and technology in official statistics to support the production and reporting of specific indicators (30 out of 232), mapping the progress across different geographic areas and over time, and fostering international collaboration for standardised data collection and processing approaches (UNECE, 2019c).

Over the past years, NSO have been using EO data and satellite-derived data, assessing the viability and exploiting the practical applications to address current and emerging needs and to improve official statistics (UNSD, 2017; UNECE, 2019c) (e.g., the UN-promoted System of Environmental-Economic Accounting). Statistical programmes are being expanded to incorporate EO data and technology into official statistics (e.g., environmental statistics) while research initiatives and experimental work promoted by the statistical systems are being undertaken to turn case studies into business cases (proof-of-concept) and provide guidance and recommendations. In the ESS context, the Copernicus Programme, through its Sentinel satellites and supporting stakeholders' collaboration network, has been helping NSO capacity in dealing with EO data and integrating such data with traditional statistical methods by providing standardised methods for data collection and processing and ready-to-use products and services (e.g., Land Monitoring Service). During the development of such activities, it has been clear that NSO lack the EO expertise to tackle their statistical needs and quality requirements and that building the capacity to produce official statistics from EO data and technology is a key issue ahead (UNECE, 2019c).

The recognition of geospatial data as a strategic asset for policy lifecycle and decision-making has been limited in supporting sustainable development. According to Scott and Rajabifard (Scott and Rajabifard, 2017) connecting both political and technical levels has been a constant challenge over the years, with different implementation and maturity levels that have been particularly struggling for least developed and developing countries. Thus, there is also a general lack of awareness and understanding of the value of geospatial data in policy-making which emphasises the need for capacity development, education programs and engagement and communication initiatives (UN-GGIM, 2020b). The role of governments in developing a modern geospatial infrastructure is unclear and much work remains to be done to raise awareness of the value and benefits of geospatial data at the policy level (Coetzee et al., 2021). Moreover, the lack of investment, shortages of human resources and restrictions in data sharing, associated with political unwillingness and absence of commitment between data providers and users, are also barriers undermining the work undertaken so far (Eurostat, 2019a; UN-GGIM, 2020b).

I.2.1.2. SPATIAL DATA INFRASTRUCTURES

Spatial Data Infrastructures (SDI) have been one of the main research topics and cross-cutting practical applications of geospatial data, particularly by the GIS community, being shaped over the years by geospatial technology trends and in response to changing political, institutional and socio-economic circumstances (Coetzee et al., 2021). SDI have been generally defined as a framework of policies, institutional arrangements, technologies, data and people that enable the effective sharing and use of geographic information” primarily focusing on geospatial data and its use, namely for sustainable national development (Bernard et al., 2005). Thus, an SDI embodies cross-cutting issues from technical and non-technical components built on geospatial data from a variety of themes and domains. SDI aim to reduce duplication of efforts among governments in terms of data collection and financial resources and make geospatial data more centrally available, accessible, shareable and usable through harmonisation and standards.

During the 1990s, several countries developed SDI at the governmental level, i.e. National SDI (NSDI) to collect and use data for national administrative and policy purposes. Due to existing technical and institutional arrangements, these NSDI started to promote the adoption of standards focusing more on organisational aspects (Rajabifard et al., 2003; Craglia, 2015). Whereas global and regional institutions were mainly focused on SDI development for publishing data, local SDI were primarily addressing operational needs for daily management and decision-making by local authorities. Most of SDI projects were rather isolated and exploratory focusing on collaboration issues and standards (Masser, 2005).

In the early 2000s, GIS technology, like any other type of software technology, shifted from the desktop domain with physical restrictions to a distributed GIS on the web and to distributed computing, enabling geospatial-related systems to be linked and accessed as a single virtual system (Tait, 2005). This shift was a key driver for the second generation that focused more on data users, data sharing and decentralised structures, i.e. service-driven infrastructure (Masser, 2005). It also opened the way to a new wave of web geospatial applications, modernising the concept of geoportal (Maguire & Longley, 2005). IT development also provided new opportunities to geoportals on metadata services, capabilities (e.g., mapping, geocoding, routing, etc.) and geospatial web services (e.g., cataloguing and mapping) supporting users to better find, explore and use the available geospatial data and services in the SDI (Bernard et al., 2005).

During the 2000s, many geospatial experts within governments questioned the SDI conceptual model for national development as a data infrastructure providing authoritative geospatial data and connecting to other information systems with different data domains (Coetzee et al., 2021). As outlined by the authors (Coetzee et al., 2021), the geospatial environment was changing, gaps in geospatial data management and operationalisation of SDI were widely recognised and no overarching guiding and decision mechanism was promoting common approaches when bringing these topics to global-level discussion. The urgent need for action towards this global policy framework changed the SDI vision, particularly related to data, emerging technologies, producers and users, a more participative geospatial industry and strategic alignment of SDI with development strategies (Rajabifard et al., 2003). However, several organisational, institutional and political obstacles have harnessed SDI effective and sustainable implementation, especially concerning data sharing and access and administrative sectorial boundaries (Masser, 2005).

Furthermore, Coetzee et al. (2021) identified and described new and emerging conditions and pointed out the impacts and significant changes to the geospatial ecosystem that go beyond the traditional scope of SDI. It included more location in decision-making at all levels, new geospatial data sources and services (e.g., commercial sources), technological advances (e.g., IoT and IT developments) and more automation, analytics and intelligence (e.g., machine-to-machine, geo-analytics and real-time analysis). In terms of non-technical aspects, changes in user expectations were also outlined, referring to more demanding users, including non-geospatial experts. Organisations and their structures and operating procedures will need to be more flexible and collaborative. It also highlighted geospatial data processing and sophisticated geospatial analysis (e.g., open algorithms and processing protocols) from dynamic data with a higher spatio-temporal resolution to support decision-making (Coetzee et al., 2021).

In the last years, SDI have witnessed a paradigm shift to a Geospatial Knowledge Infrastructure (GKI) used for data management, integration, analysis, modelling, aggregation and dissemination across domains and organisations despite many technical, policy and legal challenges (UN-GGIM, 2020a). It comes from a future vision of a dynamic and mainstream geospatial ecosystem - a subset of the wider digital ecosystem - supported by sustainability principles, interconnectivity and multi-stakeholders (Coetzee et al., 2021). It is supported by a progressive knowledge management paradigm that surpasses the traditional dataflow and supply chains and information management to improve the decision quality and value. Meanwhile, countries have been in various stages in the development of their NSDI with the mission to improve the availability and access to several geospatial data themes to address national policies and international agendas (UNECE, 2016). Open data and e-government initiatives were developed to better attend data requirements and user needs.

Both authoritative geospatial data and VGI have a crucial role in the SDI's future development through a series of data collection operations carried out on a specific thematic domain or data initiatives from the public sector according to legal mandates and administrative duties (Masser, 2020). It is expected that national governments and mapping agencies will continue their traditional mission and share the production of authoritative geospatial data enabling to leverage the quality and reliability between public and private data providers (Coetzee et al., 2021). In this regard, the way these key stakeholders respond to the changes happening in the wider digital environment and how they will exploit emerging societal trends and policy agenda opportunities will have impacts on the future geospatial ecosystem. As mentioned by Masser (Masser, 2005) the implementation of SDI by governments is a long-term task that often reflects the need for reinvention in response to political, institutional and technological changes.

I.2.1.3. GEOSPATIAL DATA PRODUCTION

There is no business geospatial production model under a regulatory framework as it is defined in official statistics. However, there are underlying steps in the production of geospatial data based on dispersed best practices and many standards within the geospatial community. According to Sun et al., (Sun et al., 2019), the traditional lifecycle of geospatial data can be divided into six phases: i) data acquisition; ii) data processing; iii) data storage; iv) data management; v) data sharing; and vi) data application. These steps are generic and may vary depending on the specific data application or the organisation responsible for the geospatial production.

Data acquisition involves data collection, measurement and capture processes. Raw data is collected from various sources in many forms, such as digital data and EO systems (ground-based, aerial and space-based sensors) which is related to the mode and method of data collection and related technologies. It can include remote sensing (i.e. satellite imagery), GPS surveys, field surveys (e.g., geological surveys), location data assigned to administrative data, data purchased from third-party providers, location-based services (e.g., social media) or crowdsourced data from citizens, i.e. user-generated content. Although the emerging trend promotes the use of existing data to eliminate duplicated work and data overlaps, new data acquisition operations can occur when a production/business need is identified and duly justified, namely related to policy demands, new data requirements and/or user needs. A new relevant data theme to be included in the NSDI that contributes to a specific national policy, improvements in data quality, compliance with a new legal framework (e.g., confidentiality and privacy), new technologies or IT solutions and the need for more timely and/or spatially accurate data to produce better outputs may be reasons to undertake a new geospatial data collection activity.

Data processing can include preprocessing procedures to correct errors and other techniques to prepare the data for further analysis, such as data cleaning (often in routines), data transformation (e.g., to alter projections or convert between coordinate systems), data mining and data integration. Therefore, this step covers quality control and management procedures attending quality dimensions, such as accuracy, precision, lineage, consistency and completeness. This step can also include georeferencing.

Data storage is the process of storing and organising the collected geospatial data using specialised software, i.e. oriented towards object-relational database systems that can handle geospatial data (e.g., PostGIS or Oracle Spatial). Collected geospatial data are often stored in an in-house geospatial database (or Data Warehouse - DW) located on the internal network and designed to store, index, manage and query according to technical requirements and the IT environment/strategy. Rules and norms about data formats and structures, vocabularies, ontologies and identifiers need to be defined in this phase to make data operational within the organisation, for instance in terms of data discovery and access and to avoid data duplication. Thus, the data management phase is closely linked to the previous phases to ensure effective maintenance and updating of the geospatial data. Data security, especially containing sensitive information, should be guaranteed in this phase.

Data sharing involves the procedures allowing the end-users to access and (re)use the geospatial data covering data distribution, transfer and sharing mechanisms that should meet the data requirements and demands of the user community. This phase includes data dissemination and publishing, for instance through spatial visualisation instruments, such as static or interactive maps and web geospatial-based services (e.g., story maps).

Data application is a broader phase as it depends on the use of geospatial data by the end-users and stakeholders for public/collective or individual/private purposes, including data interpretation and spatial analysis to extract meaningful information and generate knowledge and insights supporting various human-based applications.

I.2.1.4. KEY STAKEHOLDERS AND ROLES

Key stakeholders in the international geospatial community are: i) Global and regional Geospatial Agencies responsible for governance and coordination of geospatial-related activities; ii) National Geospatial Authorities, including the National Geospatial Information Agencies (NGIA) and other public institutions responsible for the production and provision of authoritative geospatial data and services; iii) Global and regional institutions involved in geospatial data and services regulation, implementation and monitoring; and iv) Other organisations, projects and initiatives for geospatial data collaboration, cooperation and innovation.

The global and regional geospatial organisations and agencies are responsible for governance, innovation, strategic leadership and high-level coordination and guidance of geospatial programmes, activities, actions and deliverables.

At the global level, the **UN-GGIM** has played the leading role in global geospatial information management over the years. It was established in 2011 to set the agenda and direction for geospatial data management within national and global policy frameworks, and to strengthen international cooperation in this field. The UN-GGIM provides high-level global coordination of activities related to geospatial data management for sustainable development, with particular focus on the integration of geospatial data with statistical and socio-economic data. It develops strategies and actions to enhance geospatial capabilities and capacity building and disseminates best practices, recommendations and experiences on legal and institutional instruments, data and information management models and effective governance.

The UN-GGIM ensures that national geospatial authorities communicate and work jointly with each other and with NSO to contribute to a more effective data management environment addressing global strategic drivers and user needs and requirements (UNSD, 2021). The lines of work cover international geospatial standards, user needs in geospatial data management, policy and technical frameworks, data interoperability and fundamental geospatial datasets. UN-GGIM has been internationally recognised by wider geospatial community, including academia, industry, the private sector, as a UN effective and productive intergovernmental mechanism that has had a global reach and impact on location-based policy and related development demands (Coetzee et al., 2021).

In the European context, the regional committee of the UN-GGIM, **UN-GGIM: Europe**, was formally established in 2014. The regional committee carries out a series of activities to demonstrate the benefits of authoritative geospatial data in achieving and monitoring the SDG, focusing on data integration and core geospatial data that underpin harmonisation and interoperability issues. Other UN-GGIM regional committees operate under their territory of influence and equally follow the UN-GGIM overarching vision and strategy considering the regional specifications and national circumstances of their respective countries. They support the implementation of geospatial frameworks and standards, explore ways to promote geospatial-statistical data integration, showcase the added value of geospatial data and contribute to the development of relevant policy and data strategies regarding geospatial data management.

The **NGIA** are the stakeholders responsible for collecting, maintaining and publishing authoritative geospatial data and services, designing a national geospatial governance model and building geospatial infrastructure and capabilities for national development. These regulatory organisations have national leadership in geospatial data management and are assigned to implement the global geospatial roadmap according to the circumstances of their country and the needs of the national government. In the European context, these authorities are also formally addressed as the National Mapping and Cadastre Agencies.

Many countries have more than one NGIA, with different agencies dealing with different geospatial data themes, organised according to the country's administrative system. Some NGIA are responsible for managing address data registers, land administration through the maintenance of cadastral data or the publication of administrative geographies. In this regard, custodianship mandates are crucial mechanisms to avoid cross-government duplication in the acquisition and management of authoritative geospatial data (UN-GGIM, 2020b). In addition, other national producers of authoritative geospatial data and services may be in charge of building and maintain the NSDI in accordance with the established national geospatial data framework and agreed custodianship guidelines.

In the group of institutional stakeholders involved in geospatial data and services regulation, implementation and monitoring, there is the **Geographic Information System of the Commission (GISCO)** within Eurostat. At the European level, GISCO is responsible for meeting the EC's geospatial data needs and requirements at the EU, national and regional levels. GISCO provides geospatial reference data as a set of core geospatial datasets (e.g., administrative units, statistical units, buildings, addresses, cadastral parcels, transport network, etc.) that are regularly used by a range of EC services and authorities, such as the **Joint Research Centre (JRC)** and the **European Environment Agency (EEA)**. In addition, GISCO provides geospatial services and GIS-based software, supports cartographic and spatial analysis activities, coordinates activities across the ESS related to geospatial data and fosters the use of GIS to support EC activities. It also promotes the benefits of statistical-geospatial data integration by chairing the Working Group (WG) on the integration of statistical and geospatial information that involves representatives from NSO and NGIA and discuss different topics related to this field.

The organisations and agencies responsible for designing geospatial standards are key international players in promoting open geospatial standards and interoperability services that refer to recommended practices to facilitate the development, sharing and use of geospatial data, services and technologies. The national and international cooperation and involvement of these types of organisations and agencies is becoming increasingly important in developing, publishing and maintaining a set of largely recognised common standards and terminology (UN-GGIM, 2020a).

Geospatial standards can be developed by any of various types of organisations, including both public and private, such as a consortium (group of corporations), a national government (including its bodies and institutes), a professional association or a purpose-made standards organisation. There are some national and international organisations responsible for the development of geospatial standards: i) the **ISO Technical Committee 211 Geographic information/Geomatics (ISO/TC211)**; ii) the **Open Geospatial Consortium (OGC)**; and iii) the **Open Source Geospatial Foundation (OSGeo)**. These international standards organisations aim to develop, maintain and make publicly available standards facilitating the access, sharing and use of geospatial data and services in a consensual process with representative members from government, industry and academia (UN-GGIM, 2015). They facilitate the implementation and adoption of both general and specific-domain standards for the global geospatial community. Their activities are relevant within and beyond the geospatial community, following geospatial technology and industry trends. In addition, the International Hydrographic Organisation in partnership with technology standards organisations also plays a crucial role in developing geospatial standards ensuring technical interoperability across information systems from different domains and future applications of geospatial technology.

ISO/TC211 is a technical committee established within ISO focusing on standards and technical specifications related to digital geospatial data, including reference models, metadata, location-based services, classification systems and web services.

OGC is an international formal organisation on standards built on partnerships with commercial, governmental, non-profit and research organisations to develop, implement and disseminate open standards for geospatial content. As a worldwide community committed to improving the use of and access to geospatial data, OGC creates free and publicly available geospatial standards that enable new technologies. Over the last decades, OGC has addressed the problems created by the lack of interoperability from the various data models and structures by building a set of technical specifications within the various fields of the GIS community (Goodchild, 2010). OGC has also been planning and developing work programmes by evaluating the impacts of emerging technology trends in the geospatial industry, identifying gaps and addressing standardisation issues according to the definition of strategic priorities (UN-GGIM, 2020b). Thus, OGC manages a flexible collaborative research and development process attending the geospatial environment and anticipating and solving real-world geospatial challenges experienced by OGC members (GEOSTAT 4, 2021a). OGC is one of the stakeholders within the geospatial community that more closely works with the geospatial industry and various domains and business sectors that use geospatial data, emphasising its key commercial role (e.g., open standards in the private sector).

ISO/TC211 and OGC have had a cooperative agreement and joint efforts to develop standards and implement specifications that meet the needs of the geospatial industry, governments, academia and user communities. The most important and leading standards in the geospatial domain are provided by ISO/TC211 and by the specifications developed by the OGC, particularly concerning SDI and spatial visualisation services (Bernard et al., 2005). Web Mapping Service (WMS) and Web Feature Service (WFS) are examples of geospatial web services used for direct visualisation and access of geospatial on the Internet that have emerged from the joint efforts between ISO/TC211 and OGC over the years.

OSGeo is a non-profit non-governmental organisation aiming to support and foster open-source geospatial data and related technologies. This organisation also provides financial, organisational and legal support to the global open-source geospatial community in developing open standards through partnerships with OGC and ISO/TC 211 while actively supporting interoperability with open formats and the international standards community.

Other relevant players can be outlined concerning the broader standards community and the wider digital environment. The World Wide Web Consortium (W3C) is an international community developing Web standards - or Internet standards - for Web technology, including protocols, design principles, technical specifications, guidelines and recommendations. Concerning geospatial data, W3C has developed standards for authentication, authorisation, and security that could be implemented in developing an SDI (UN-GGIM, 2015).

The **Geospatial World Forum** is a collaborative platform of geospatial experts and leaders that aims to share and implement a common vision across the global geospatial community by building bridges between public policymakers, national mapping agencies, the geospatial industry, international organisations, scientific/academic communities and general users. This interactive platform engages on topics related to worldwide geospatial trends and drivers such as the digital transition in the geospatial industry and its role in the world economy, GKI, geospatial intelligence and innovative geospatial technologies.

EuroGeographics is a membership association based on voluntary collaboration of experts from the European NGIA with the main goal of supporting the development of the European SDI (GEOSTAT 3, 2019a). EuroGeographics aims to facilitate access to and use of authoritative pan-European geospatial data and services based on harmonisation principles and common data requirements and technical specifications. The members of this association are building the operational European Location Services (ELS), which aim to change the way users access and use location-based products based on authoritative geospatial data. At the global level, EuroGeographics acts as a non-governmental organisation in consultative status with the UN Economic and Social Council or as an invited participant in certain UN-GGIM sessions.

Finally, the geospatial industry and the **private sector** are key stakeholders towards a more global, integrated and sustainable geospatial ecosystem. The geospatial industry and private companies are growing in an increasingly competitive economic market by providing geospatial data, services and solutions, driven by the increasing importance of location in human activities. It includes providers of geospatial services (e.g., GIS technology) and private consortiums bringing together companies in the field.

I.2.2. GEOSPATIAL FRAMEWORKS

Both geospatial frameworks and infrastructures are made up of geospatial data, information systems, architectures, technologies, business models, and technical, legal and policy frameworks to deliver geospatial datasets and services. Since the geospatial domain does not have a set of regulatory frameworks established by a high-level overarching organisation like in official statistics, some geospatial frameworks will be outlined in this subchapter. This list of examples aims to be illustrative of the current geospatial frameworks and infrastructures with national, regional or global implementation and in different organisational/institutional contexts that are leading the way forward.

The **Integrated Geospatial Information Framework (IGIF)**, as the overarching geospatial framework of UN-GGIM, provides a starting point and guidance for countries to develop, integrate and strengthen geospatial data management and resources (UN-GGIM, 2020b). The vision of the framework relies on the efficient use of geospatial data by all countries to effectively measure, monitor and achieve sustainable social, economic and environmental development.

IGIF aims to provide strategic guidance at the national, regional and local levels to prepare, support and implement specific action plans and initiatives enhancing national geospatial data management and geospatial capabilities. IGIF also focuses on needs and gaps in geospatial practices to facilitate implementation while building capacity, supporting innovation and providing the leadership, coordination and standards required to deliver integrated and harmonised geospatial data. IGIF was built on past efforts in implementing SDI and NSDI while recognising their shortcomings, outlining data collection and technological aspects and helping countries to successfully implement related capabilities through good practices (Scott, 2020). In this regard, IGIF can replace the traditional NSDI concept by grouping together technology, policies, standards, good practices and human resources necessary to acquire, process, store, analyse, disseminate and use geospatial data (UN-GGIM, 2021).

IGIF comprises three parts conceptually and methodologically intertwined: i) Part 1: Overarching Strategic Framework; ii) Part 2: Implementation Guide; and iii) Country-level Action Plans. The Overarching Strategic Framework is the strategic policy guide providing the overarching strategy supported by 7 Underpinning Principles, 8 Goals and 9 Strategic Pathways (Figure 2). These elements demonstrate the holistic and cross-cutting approach of the framework that looks for long-term and sustainable implementation of its vision.

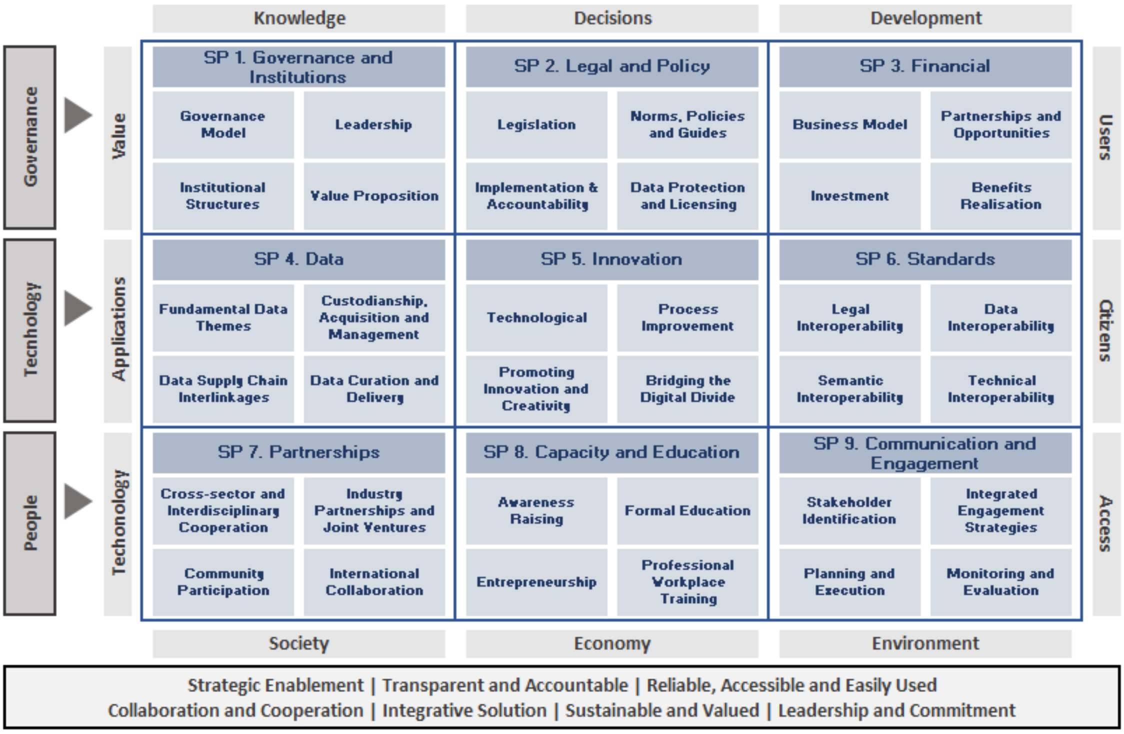

Figure 2. IGIF structure, anchored by seven underpinning principles, nine strategic pathways and the three main areas of influence (source: author, from UN-GGIM, 2020b).

The 7 Underpinning Principles represent the key guiding values when implementing the framework. The principles are: i) Strategic Enablement (political and financial support); ii) Transparent and Accountable (key accountability and transparency guidelines); iii) Reliable, Accessible and Easily Used (data requirements for research, development and innovation); iv) Collaboration and Cooperation (between government, business, academia, and civil society to promote data sharing and reduce duplication of effort across the government sector); v) Integrative Solution (joint work between citizens, organisations, systems, and legal and policy structures); vi) Sustainable and Valued (national efficiency and productivity enhancement in the long term); and vii) Leadership and Commitment (often at the highest level to enhance the long-term value of investments).

These principles are applied via the 8 Goals: i) Effective Geospatial Information Management; ii) Increased Capacity, Capability and Knowledge Transfer; iii) Integrated Geospatial Information Systems and Services; iv) Economic Return on Investment; v) Sustainable Education and Training Programs; vi) International Cooperation and

Partnerships Leveraged; vii) Enhanced National Engagement and Communication; and viii) Enriched Societal Value and Benefits.

The Strategic Pathways are built under three areas of influence: governance, technology and people. Each strategic pathway has guiding principles and is operationalised by specific objectives to assist countries in a hierarchical and interdependent perspective. The 9 Strategic Pathways are: i) Governance and Institutions; ii) Legal and Policy; iii) Financial; iv) Data; v) Innovation; vi) Standards; vii) Partnerships; viii) Capacity and Education; and ix) Communication and Engagement.

The Implementation Guide provides operational steps and actions that need to be taken by governments and stakeholders to implement the IGIF and comprises reference guides, good practices and expected deliverables and outcomes after actions have been carried out. It provides flexible implementation enabling countries to develop their action plans in order to meet their national priorities and needs while considering their geospatial maturity level. Lastly, the Country-level Action Plans are not binding programmes addressing a more technical and fit-to-purpose approach, providing templates to operationalise the framework.

The **Infrastructure for Spatial Information in the European Community (INSPIRE)** was established as the European SDI to ensure compatibility and usability of geospatial data across the MS to support Community environmental policy decisions and environmental-related activities. The INSPIRE Directive 2007/2/EC is the core legal framework of the European geospatial data environment aiming for harmonised, easily findable, accessible and (re)usable geospatial datasets and services provided by the Community institutions and MS through a Geoportal as a central access point.

The INSPIRE Directive aims to create metadata, establish network services, ensure Pan-European cross-border and cross-thematic technical interoperability of geospatial datasets and services, facilitate data exchange and enable access, sharing and use by government authorities, businesses and citizens through common requirements. The infrastructure builds upon existing or newly collected geospatial datasets managed by the MS covering 34 data themes that should conform to harmonisation and interoperability conditions, data specifications and technical arrangements, i.e. Implementing Rules. These Implementing Rules are legally binding in their entirety and cover the following areas: metadata, data specifications, network service, data and service sharing, spatial data services and monitoring and reporting. Hence, technical rules on common data models, code lists, metadata and services are established. Technical recommendations are also provided, but these are not legally binding.

Although the cornerstone element addresses harmonisation, the technical guidelines for implementing the Directive remain non-binding (e.g., data collection and calculation methods) and only refer to geospatial standards, when appropriate. The few mandatory components (e.g., Implementing Rules) were behind schedule or did not achieve the quality requirements in many MS which compromises the goals of comparability and interoperability (CoR, 2021). Thus, INSPIRE lacks robust legislation and binding mandates at the technical level, i.e. less indicative/suggestive nature and open to misunderstanding, to make MS more compliant with the technical guidelines. Gaps in INSPIRE implementation are also caused by institutional and legal obstacles related to national legal restrictions on data regulation, political contexts and administrative/organisational environments contributing to different national maturity levels. Lastly, there is no clear leadership due to the absence of an official specialised EC geospatial agency working as a high-level governance and coordination body in the geospatial domain at the European level. Regardless of the identified gaps, a consensus may be established that INSPIRE is a role model in terms of SDI development, and formulation of public policy at the European level (Masser, 2020).

The **European Interoperability Framework (EIF)** was adopted in 2017 (Communication 2017/134) by the EC, given its primary need to digitalise the public sector, address interoperability challenges and move forward with the current European interoperability strategy with new focus areas, more detailed guidance and up-to-date recommendations. The interoperability action plan was established to guide the MS throughout the EIF implementation covering national actions and recommendations related to governance, engagement and interoperability activities, such as interoperable digital public services. It also works as a legal framework built on harmonised legislation towards the establishment of the digital single market and a coherent European interoperable environment across borders, organisations and policy sectors. EIF acts as a common core supported by two interoperability elements - National Interoperability

Framework and Domain Interoperability Framework - that should be developed in a uniform and coordinated manner. These two elements are either transposed directly or tailored considering national specifications to meet the specific needs of the country while allowing a degree of flexibility.

The EIF is structured by one conceptual model, 4 interoperability levels/layers supporting the interoperability model, 12 general and underlying principles on interoperability and 47 specific recommendations spread out by the elements of the framework. The principles are grouped into 4 categories: i) one principle setting the context for EU actions on interoperability; ii) 4 core interoperability principles; iii) 4 principles related to generic user needs and expectations; and iv) 3 foundation principles for cooperation among public administrations. The conceptual model is aligned with the interoperability principles proposing a standard approach ('interoperability by design') to designing and implementing European public services in which each principle has one or more associated recommendations providing objective practical guidance.

In addition, two additional cross-cutting and background layers organise the model: integrated public service governance (organisational structures, roles and responsibilities and formal interoperability agreements) and interoperability governance (holistic approach for full interoperability via standards and specifications). The latter can be illustrated by the INSPIRE Directive as an interoperability framework supported by legal interoperability, coordination structures and technical arrangements.

The model consists of basic components, such as open data, integrated services delivery and service governance, digital by default, reuse of data and services, catalogues for describing, finding and using services, and security and privacy. These components cover high-quality authoritative data sources focusing on machine-readable formats towards innovation and a data-driven economy and are linked to the EU Directive on the reuse of public sector information (Directive 2013/37/EU). This Directive encourages MS to make public information available for access and reuse as open data and in a format that ensures interoperability with implications for the future development of the geospatial field in Europe.

Under the geospatial domain, the **EU Location Framework (EULF) Blueprint** provides guidance for EIF implementation and is closely connected to the interoperability principles and scope of the general framework. The EULF Blueprint aims to establish a coherent location framework to facilitate the integration, exchange and sharing of location data and related services towards European location-enabled e-government, i.e. location interoperability[4] based on geospatial data and digital standards. In this regard, the EULF Blueprint is intrinsically related to the INSPIRE since it will be built using the Directive to explore the use of location data for optimising digital services and as a key element to address the goals of the Digital Single Market strategy. Moreover, when evaluating the deliverables for monitoring, assessing and reporting on the status of location interoperability in policy and digital public services, the degree of INSPIRE implementation is included as a focus area.

From 2017 to 2021, the EULF Blueprint has expanded to add more actionable guidance, new data quality recommendations and technical content, additional best practices, legislative revisions, new business models and technologies, and policy content updates. It also extended links to other frameworks (e.g., EIF and IGIF) enabling a more cross-domain implementation.

Lastly, the **European Location Interoperability Solutions for E-Government (ELISE)** action has been working over the years as a package of legal/policy, organisational, semantic and technical interoperability solutions to enable digital governments through geospatial data and Location Intelligence[5]. It helps to support digital government transformation by making the best use of location data and technologies in an interoperable manner for all citizens, businesses and public administrations. The ELISE builds on top of the EULF providing location-related solutions for all levels of the EIF through a set of outputs such as studies, a framework for guidance and monitoring, applications and a Geo Knowledge Base Service to enhance knowledge exchange and capacity building (e.g., webinars and workshops).

---

[4] At a general level, location interoperability is the ability of organisations, systems and devices to exchange and make use of location data with a coherent and consistent approach. In the context of digital government, this definition can be expanded on using location or geographical area for government policy and digital public services, involving coherent interactions between public administrations, businesses and citizens (ELISE Glossary).

[5] The process of deriving meaningful insight from geospatial data relationships - people, places or things, to solve particular challenges such as demographic or environmental analysis, resources management and traffic planning. Location intelligence tools consist of a combination of GIS software, web mapping solutions, and position technologies such as GPS and location-based data and services.

Over the years, besides complementing EIF concerning location interoperability, ELISE had several achievements. It established a cross-reference between the EULF and IGIF, helped to put INSPIRE into practice with tools for data providers and assessed the role of SDI in future business models and data ecosystems towards a more user-driven approach. ELISE raised awareness of new approaches and applications regarding the potential of location-enabled digital government and assessed new policies (e.g., General Data Protection Regulation - GDPR (EU, 2016) - and the European Data Strategy) and technologies (e.g., AI and API).

I.2.3. GEOSPATIAL QUALITY AND STANDARDS

I.2.3.1. GEOSPATIAL QUALITY

In contrast to the statistical domain, the geospatial domain lacks a comprehensive global regulatory framework to ensure geospatial data quality through harmonised assurance tools and consistent approaches. It comprises a collection of international business-oriented geospatial standards - often in the form of open standards - established by international standards organisations. These standards aim to ensure consistent requirements and harmonised approaches in the creation, management and publication of geospatial data and services and to achieve well-defined and comparable quality in evaluation and reporting through standardised quality measures (ISO, 2013). In addition, there are standalone best practices, guidelines and recommendations for the management of geospatial data and services provided by some of the key stakeholders in the geospatial community according to specific applications and operational contexts.

I.2.3.2. GEOSPATIAL STANDARDS

Similar to statistical quality, the quality of geospatial data and services is intrinsically related to the needs and requirements of applications and users whereas geospatial production should be supported by specific-domain quality management. To secure those needs and requirements, standards are required as documented and consensual agreements between providers and consumers, setting up rules, guidelines or technical specifications (UN-GGIM, 2015). Geospatial standardisation is crucial in delivering geospatial data, services and products to all users with different levels of geospatial expertise and making geospatial content readable and understandable to everyone.

Geospatial standards provide rules, guidelines, characteristics and recommended practices to facilitate developing, sharing and using geospatial data in the digital environment, including GIS software and services. They enable connecting providers, regulators and consumers within the geospatial community and establishing agreements and common understandings between the user communities and geospatial software, hardware and procedures so people can implement these types of technologies and capabilities with as little effort, time and cost as possible. Geospatial standards are also important to achieve location interoperability allowing geospatial data and services to be exchanged and used between organisations, systems and devices within a cooperative and sharing environment in a consistent and coherent manner.

The geospatial community uses standards for geospatial data management developed by leading international standards organisations, such as the ISO and OGC. Standards in the geospatial domain are much more open to the market, particularly the industrial and technological fields, more likely to be aligned to general standards and more business and user-oriented (Ariza-López et al., 2021). According to the authors (Ariza-López et al., 2021), a more consolidated international leadership in this field has dedicated in the last decades more efforts to the development of standards for geospatial data and interoperability, highlighting ISO standards on formats, metadata and web services.

ISO standards are generally considered by the geospatial community to be guidelines and reference models for geospatial production (data/information and technology), as they usually translate the formalisation of OGC standards, which are community-driven and more oriented towards tailored development for practical and up-to-date applications. Thus, ISO standards are more formal and less flexible compared to OGC standards as they are supported by rigorous standardisation processes with less frequent revision cycles and are more oriented towards organisations requiring ISO certification. Nevertheless, it is important to note that both ISO and OGC standards are complementary, with minor technical differences, and that OGC standards often provide the technical specifications that influence the next versions of ISO standards. However, the adoption of both types of geospatial standards is voluntary and not legally binding due to the less regulated geospatial environment, i.e. no regulatory frameworks.

Most of the geospatial information infrastructures from the public and private sectors that have demonstrated economic, productivity and efficiency benefits are built on open geospatial standards, highlighting their interoperability capabilities (UN-GGIM, 2015). Open standards mean that they are publicly available, that anyone can access and use them, are detached from any intellectual property and are based on neutrality principles, i.e. they should come from a consensus decision, focus on user needs and no organisation shall own or manage them. Open standards ensure that users can access them on reasonable terms, which is particularly important for developing countries and countries with low geospatial maturity levels. Voluntary standards can also become mandatory within a legal or organisational framework if adopted by a national governments or businesses.

Open geospatial standards aim to achieve higher levels of interoperability and quality by providing internationally agreed standards on data models (i.e. common data formats and information structures) for online access, exchange and download, metadata and distributed functionalities handling geospatial data and services. They also increase the ability of a wider range of users from different data/information communities to access, use and exchange geospatial data more easily and consistently. Thus, open geospatial standards are cornerstones for higher maturity levels in geospatial capacity and should be supported by robust policy, technical and legal frameworks for data access and sharing. However, many non-technical constraints have hampered their full implementation in governments and other public institutions, including a lack of consensus, institutional arrangements, leadership commitment and effective governance and coordination. In addition, budget cuts, legal barriers and the lack of a long-term strategic agenda and experience in this area have undermined the potential of open geospatial standards.

According to UN-GGIM (UN-GGIM, 2015), there are two key types of geospatial standards. The first focuses on data/information (e.g., data specifications, schema, syntax, encodings, metadata, etc.) and the second focuses on technology and applications (e.g., interfaces, API, services, etc.). The data/information standards define content (data and metadata) models (e.g., concepts and topics) and legal and administrative requirements such as confidentiality, licensing, security, data sharing and quality. The technology standards define technical requirements to develop interfaces (e.g., API) enabling communication between different systems and infrastructures addressing technological advances and emerging digital trends (e.g., Linked Open Data - LOD - and cloud computing). These types of geospatial standards establish file formats so that data can be exchanged and understood by everyone, from the source to the end user throughout the data lifecycle.

Geospatial standards can also be classified into three categories: i) data; ii) services; and iii) metadata (UNECE, 2024c). The first ones ensure geospatial data are stored and transferred across different systems in common and easily readable formats (e.g., GeoPackage). The second ones involve web-based services for visualising, sharing and downloading geospatial data (e.g., WMS and WFS). The third ones address consistent storage, description, composition (cataloguing and tagging) and display of the metadata associated with the geospatial data to facilitate discovery and usability (e.g., OGC Catalogue Service for the Web and ISO Metadata Specifications). They are useful for creating geoportals and catalogues of geospatial datasets for discovery purposes.

In recent years, geospatial technology standards regarding data and metadata models, services, discovery and access have seen greater growth due to the expansion of the Internet, taking advantage of modern web development practices. OGC has developed standardised web interfaces, including API and well-known web services standards oriented towards the exchange of geospatial content (e.g. WMS and WFS), to facilitate the access and use of geospatial data and its integration with other types of data.

Technological, legal and policy trends and user requirements impacting the collection, use, management and visualisation of geospatial data have also been drivers for enhancing existing geospatial standards and developing a generation of new ones and related best practices (UN-GGIM, 2015). For instance, W3C has been addressing web semantic issues concerning geospatial data and supporting the development of geospatially-enabled web standards (e.g., GeoSPARQL) throughout vocabulary that recognise geospatial features and rules for handling geospatial data in web technologies.

Also, the General Feature Model (GFM) has traditionally been used by the geospatial community, especially by institutional stakeholders, and is the most illustrative example of INSPIRE adopting this model for spatial object types and their properties. The GFM has also been extensively used in the geospatial community to describe geospatial data and metadata in a consistent manner.

ISO has published several geospatial standards under the direct responsibility of ISO/TC 211, including general and domain-specific geospatial data and technology standards for geospatial data sharing (ISO 19100 series). They encompass technical specifications for geospatial data and services for the management, collection, processing, analysis, access, visualisation and transfer of data in digital format between different users, infrastructures, applications and systems. Some of the overarching and most relevant **ISO/TC 211 geospatial standards** to date are summarised and briefly described below:

ISO 19157-1:2023 Geographic information - Data quality (Part 1: General requirements): defines the principles for describing the quality of geospatial data by establishing a set of components and the process for developing additional ones, specifying the content structure of data quality measures, describing general procedures for data quality evaluation and principles for data quality reporting (ISO, 2023).

ISO 19115-1:2014, Geographic information - Metadata: describe metadata fundamentals and conceptual and application schemas for both geospatial data and services (e.g., acquisition and processing and XML schema, etc.) (ISO, 2014). The ISO 19115-1:2014 is broken down into key parts, Part 1 (Fundamentals) and Part 2 (Extensions for acquisition and processing). The first defines the schema required for the description of digital geospatial data and services (e.g., the minimum set of metadata, providing information about the identification, extent, quality, spatial and temporal aspects, content, spatial reference and other properties). The second defines the schema required for the enhanced description of the acquisition and processing of geospatial data, including imagery.

ISO 19107:2019, Geographic information - Spatial schema: it provides specifications on conceptual schemas for describing the spatial characteristics of geographic entities, and a set of consistent spatial operations with these schemas (e.g., data query, management and exchange, etc.) (ISO, 2019a).

ISO 19108:2002, Geographic information - Temporal schema: it defines concepts for describing temporal features of geospatial data and provides a basis for defining the temporal aspects of the respective metadata (ISO, 2002).

ISO 19119:2016, Geographic information - Services: it defines requirements for the creation of services enabling one service to be specified independently of one or more via distributed computing platforms, i.e. interoperable service implementations (ISO, 2016).

ISO 19111:2019, Geographic information - Referencing by coordinates: it defines the conceptual schema for the description of referencing by coordinates and describes the minimum data required to define spatial, parametric[6], temporal and mixed coordinate reference systems (ISO, 2019b).

ISO 19112:2019, Geographic information - Spatial referencing by geographic identifiers: it defines the conceptual schema for spatial references based on geographic identifiers establishing a general model for spatial referencing using geographic identifiers and defining the components of a spatial reference system. It also specifies a conceptual scheme and structure for gazetteers to be constructed in a consistent manner (ISO, 2019c).

International guidelines and joint efforts have been developed towards a more mature geospatial standardisation, including future roadmaps, institutional partnerships, geospatial community strategies and SDI initiatives focusing on data integration and interoperability. The Global Fundamental Geospatial Data Themes and IGIF are key drivers in fostering the adoption of standards in the geospatial domain, addressing international and national needs on core geospatial datasets and providing guidance, best practices and compliance mechanisms, namely related to geospatial data and metadata requirements (e.g., data content, conceptual model, delivery issues, etc.). In the European context, INSPIRE has foster the use of active standards, particularly in

[6] Type of coordinate reference systems that use a non-spatial parameter that varies monotonically with height or depth.

the technical details embodied in the technical recommendations on metadata, interoperability, network services, sharing and coordination (e.g., ISO 19115, ISO 19107, ISO 19112 and ISO 19137, etc.).

The High-Value Datasets (HVD) (Implementing Regulation 2023/138) are another example of promoting open geospatial data standards and contributing to a higher level of open data maturity in Europe. They are core reference datasets made available in open, machine-readable and reusable formats through API. The HVD are part of the European strategy for data (EC, 2020) which aims to create a European single market for data to address issues, such as availability, interoperability and quality of data, and foresees the creation of common European data space and sectoral data spaces. The HVD are one of the key data actions in the European horizontal data sharing legislation - the Implementing Act on HVD was adopted by the EC in December 2022 and published in January 2023 -, on high-quality open government data held by the public sector for businesses, new information products and innovation. The Implementing Act on HVD defines the list of datasets for each of the six thematic categories (geospatial, earth observation and environment, meteorological, statistics, companies and company ownership, and mobility) and the requirements for their provision, such as key attributes, granularity, formats, license, among others. The Implementing Act - under the Open Data Directive (EU, 2019) on open data and the reuse of public sector information - also aims to make the data available for reuse free of charge, preferably under open access licenses, acknowledging socio-economic benefits, especially for the public sector.

Innovative solutions that take advantage of Internet technologies and are fully connected to IT standards, as well as organisational changes, are needed to ensure the full implementation of geospatial standards, preferably open and web-oriented, considering the ever-changing technological environment and user requirements (UN-GGIM, 2015). The development of non-domain-specific standards (non-specific proprietary formats and encodings) should be promoted to enable the connectivity between geospatial and non-geospatial systems and user interfaces, and ultimately make geospatial data and services understandable and usable for every user community.

II. STATISTICAL-GEOSPATIAL DATA INTEGRATION

This chapter provides a comprehensive overview of statistical-geospatial data integration (also referred to as data integration) in the context of official statistics. It encompasses a review of the technical and non-technical aspects of statistical-geospatial data integration outlining the main data integration methods and processes to produce geospatial statistics (or geospatially enabled statistics). The key stakeholders and their roles within the statistical-geospatial operating environment are identified and described alongside relevant projects and initiatives fostering statistical-geospatial data integration.

Existing statistical-geospatial frameworks and infrastructures with national, regional or global implementation and in different organisational contexts are presented as overarching strategies and reference guidance driving harmonisation and building capacity between the statistical and geospatial communities. Key elements supporting statistical-geospatial data integration and enabling the production of geospatial statistics are outlined and described as cross-cutting themes and interconnected action areas. Lastly, a holistic analysis of the current state-of-the-play of statistical-geospatial data integration is carried out to gain general insights and a comprehensive understanding of both internal and external factors and identify key issues for future developments.

II.1. STATISTICAL-GEOSPATIAL DATA, PRODUCTION AND KEY STAKEHOLDERS

II.1.1. STATISTICAL-GEOSPATIAL DATA

In general, data integration is the process of combining data from different sources to obtain a unified view (Magnani & Montesi, 2010) or linking different data sources into one to increase analytical power and get comprehensive analysis and new insights (Eurostat, 2019a). In another sense, it is the practice of combining data from different data sources into a single dataset, within the overall data management, to meet both business and application processes and/or information and user needs.

Since the late 19th and early 20th, geography has been understood as a fundamental component of statistical organisations, mainly through geographic classifications supporting statistical production, primarily for aggregating and mapping statistics, i.e. statistical maps (UNECE, 2021c). This was one of the first connections between official statistics and geospatial data in which collected statistical data were linked to geographical areas, usually administrative boundaries. In census operations, statistical units started to be defined as census tracts (in North American countries) and enumeration areas or building blocks (in European countries) to collect, process, analyse, classify and display census data (e.g., choropleth maps). This practice gained relevance in the early 20th century when governments recognised the value of visualising census data to understand population distribution, socioeconomic patterns and regional variations, and it has been extensively used in census rounds ever since.

The development of GIS technology has also extensively contributed to providing both conceptual and technological breakthroughs that enabled integrating, analysing and spatially visualising statistical data. GIS development introduced new ways of mapping statistical outputs and better understanding them. The increasing importance of GIS for official statistics, particularly from the early 2000s, resulted in a handbook as a reference technical guide to support countries in carrying out population and housing censuses by focusing on digital mapping issues (UNSD, 2000). Alongside the emerging rise of digital technologies, GIS tools allowed NSO to digitalise the census data, describe spatial patterns and interpolate values for the location to tackle sample gaps, contributing to innovative ways of statistical-geospatial data integration in statistical production (PARIS21 and Statistics Sweden, 2021). This overall strategic path crossed over the last two decades has increased the spatial context of official statistics and the geospatial awareness by the users to have more disaggregated data, perform spatial analysis and spatially visualise statistical outputs.

The development of spatial statistics (the use of geographical principles in computing statistics or statistical modelling of spatial processes) in the 1980s and 1990s, provided more accurate statistical analysis involving spatial interpolation techniques and contributed to the development of geostatistics and its application to geospatial data integration. Spatial statistics have been built around the first law of geography (also known as Tobler’s law) - which states that near things are more related than distant things - and associated concepts related to spatial relationships to help quantify patterns in data and have a deeper understanding about them. The book written by Noel A. C. Cressie on statistics for geospatial data (Cressie, 1993) provides one of the first comprehensive overviews of spatial statistics and related methods and applications. The book also continues the discussion on the integration of geospatial data with statistical methods supported by some examples (e.g., non-spatial statistics for geospatial data). Daniel A. Griffith has also extensively contributed to the advancements in spatial statistics by developing statistical methods and techniques for analysing geospatial data and incorporating spatial dependencies, i.e. spatial autocorrelation, into traditional statistical models (Griffith, 2003).

GIScience gained more academic attention in the late 20th and early 21st centuries and contributed significantly to the development of spatial analysis and spatial statistics as key research topics, including data integration techniques and the conceptual framework for integrating statistical and geospatial data, including quality aspects (Goodchild, 1992; Goodchild et al., 2003). The 'Handbook of Spatial Statistics' (Gelfand et al., 2010) was a milestone in such topics as a key publication concerning spatial statistics and its applications in various fields, including official statistics, encompassing both theoretical and applicational aspects. The field of spatial statistics has been expanding over the years with new methods and applications, including spatial-based ML tools and advanced capabilities embedded in GIS environments, to improve analytical workflows and turn data into richer information. In addition, developments in GIS software and technology also corrected inconsistencies and errors that often arise from integrating geospatial data with different sources (e.g., statistical datasets), such as spatial overlaps and inaccuracies in geographic extent and location (UNECE, 2016).

In recent years, the importance of integrating geospatial data with statistics is increasingly being recognised by international organisations and key stakeholders, especially concerning social, economic and environmental policies at the global, national and regional levels (Van Halderen et al., 2016). The benefits of data integration are widely acknowledged by both statistical and geospatial data providers. It enhances the information systems, supports the policy lifecycle and forecasting in providing information with a suitable spatial resolution for more effective policy interventions and enriches data from both authoritative and emerging data sources while avoiding collecting the same data many times (Eurostat, 2019a). Data integration also enhances the value of statistical and geospatial data itself in terms of quality and interoperability and offers new possibilities for data analysis and presentation as well as deeper insights for decision-making (UNECE, 2024a).

The 2030 Agenda has brought attention to the need for harmonised data of increasing quality, reliability, accuracy, comparability, timeliness and granularity in supporting the measurement and monitoring of the SDG. Thus, integrating statistical and geospatial data is being acknowledged as one of the most promising ways to maximise the data value for evidence-based decision-making strengthening the territorial dimension of the SDG indicators and providing meaningful contributions to address data and methodological gaps (UN-GGIM: Europe, 2019). Geospatial data and its integration with statistical data have supported the operationalisation and computing of some SDG indicators with higher geographical disaggregation and delivering them at the sub-national and sub-regional levels. In this regard, global efforts to move forward on data integration have been focusing on agreed data requirements for harmonisation and interoperability and common methodologies and standards to ensure communication between information systems and the same ground in data production. These recommendations enable delivering comparable outputs in different countries and regions for various cross-border and multi-scale applications. The geospatial community has been a key partner in the SDG framework promoting the use of EO data and other technical and methodological solutions related to geospatial data in SDG calculation for more informed decisions in impact assessment, progress monitoring and future planning (UN-GGIM: Europe, 2015).

The UN and other international and regional organisations inside and outside of the UN family have conducted global and regional efforts in the last decade for greater data integration. Both statistical organisations and national geospatial authorities have played a central role in this mission. Although an increasingly acknowledged need for data integration in the field of geospatial data management to support governance and decision-making towards sustainable development, more focus has been placed on the technical side rather than the communication between involved stakeholders (Man, 2013). Also, despite extensive efforts and a lot of work done by many international and national key players in the last years at multiple levels, there are many technical, institutional, legal and organisational constraints and challenges preventing or hindering full statistical-geospatial data integration (Eurostat, 2019a).

II.1.2. STATISTICAL-GEOSPATIAL INFORMATION PRODUCTION

The international statistical community defines data integration as “any activity in the statistical business process when data from one or more sources are integrated” (UNECE, 2017:5) which could be a mixture of various data sources and the result is an integrated dataset produced to support statistical production. Data integration also implies that at least two different data sources (input), either from internal or external data sources, are combined to create an integrated dataset (output), such as administrative data. Data integration methods are techniques, methodologies and operational approaches used to integrate, combine, link or match different data sources. Data integration processes are sequential or interactive steps, i.e. activities and tasks, to run a long-term data integration lifecycle.

In official statistics, data integration processes are carried out on multisource statistics based on multiple data sources, including the combinations of one or more surveys, administrative registers or Big Data (Yung et al., 2018). Within a statistical organisation, data integration can be considered as a business process under the statistical business process, i.e. a set of process steps to perform one or more functions to deliver a statistical output, in which the sequential flow between activities/tasks is undertaken to integrate different data sources (UNECE, 2021a). Data integration can happen at any stage of the statistical business production whenever data are combined or integrated from at least two different data sources, covering activities in developing, producing and disseminating official statistics (UNECE, 2017).

Some examples of data integration activities are applied in official statistics, such as micro-validation (at unit record level), imputation of missed values, supporting fieldwork during data collection, producing a survey sampling frame, combining datasets from different statistical domains to national accounts, spatial analysis, applying statistical disclosure control methods and in statistical cartography. These activities aim to make statistical production more efficient and flexible, improve the quality of input/processed data, add value to dataflows and enrich statistical outputs. Other common types of data integration in official statistics exist, such as administrative data sources with survey and other traditional data, new data sources (e.g., Big Data) with traditional data sources and data at the micro level with data at the macro level (e.g., for scientific research). Data integration can also be used for validation and imputation purposes, the first to validate official sources with data from other sources or to check the quality and validity of the information produced, and the second to impute missing values in a dataset.

Statistical organisations have carried out these activities to modernise official statistics, align their processes and systems, and build common infrastructures and services. However, they have been developing their information architectures, business/production processes, IT systems and technological/technical solutions which usually delay progress on data integration, standardisation and interoperability within and outside the statistical community.

One of the most common types of data integration in official statistics is statistical-geospatial data integration which commonly generates geospatial statistics (or geospatially enabled statistics) as output. Geospatial data aim to enrich the statistical data throughout statistical production by focusing on the location dimension which is transversal to all statistical phenomena (UNECE, 2017). Geospatially enabling data is the cornerstone of this type of data integration, linking data to a location, spatialising information and making data geographically understandable (UN-GGIM, 2020b). Thus, location works as a digital integrative and cross-cutting joining key (Eurostat, 2019a).

Geocoding is "the process of transforming a description of the location or unreferenced location information (e.g., address or name of a place, etc.) to the location's measurable position on the earth's surface" (GEOSTAT 4, 2022a:14) in the digital environment. Geocoding can be conducted by joining or linking location data with tabular data within a database, a desktop GIS environment (open source or commercial software) or as a service (e.g., web-based or API) by taking descriptive location information (e.g., address) as input and returning the respective geographic coordinates, i.e. point spatial object at the digital format. It is commonly associated with the terms 'geoenabling' or 'linking' depending on the type of application and is increasingly supported by advanced geospatial technologies that include automated data acquisition and spatial intelligence capabilities to improve the matching process (UN-GGIM, 2021).

Georeferencing is an intertwined concept as the "process of referencing data against a known geospatial coordinate system by matching to known points of reference in the coordinate system" (GEOSTAT 4, 2022a:15), assigning data to their digital precise and absolute location and covering a set of overarching processes in which geocoding is included. In this regard, geocoding is considered a subset concept of georeferencing. Whereas in geocoding the coordinate values (geocodes) are commonly unknown, in georeferencing the coordinates are linked to a defined geodetic reference system (framework for defining locations in space), encompassing a set of parameters that unequivocally position a spatial object in the GIS environment and ensures an accurate mapping. Although not so well-established, especially among NSO, geocoding can also be operationalised automatically and interactively by extracting relevant parts of location data to allow matching with other information and coding each statistical/administrative unit record.

In official statistics, geocoding is a method that geospatially enables statistical unit records or other non-spatial data (e.g., administrative records, such as address or dwellings registers) by assigning or linking location information (e.g., coordinates or geographical codes) to each statistical unit record (e.g., housing unit or business). Geocoding statistical data is considered a requirement within statistical production because it facilitates data integration and enables (dis)aggregation by location into smaller or larger geographic units for statistical analysis (UNSC, 2019).

According to the Australian Bureau of Statistics (ABS, 2018), there are three main types of geocoding depending on the location data in the unit record: i) point and point-in-polygon geocoding through geographic coordinates from field data collection or administrative records; ii) address geocoding based on full physical address description which can include a range of address types and models; and iii) locality geocoding based on partial physical addresses, such as postcode, locality, municipality and/or region, which is the less popular geocoding method for unit record data.

Address geocoding is one of the most common geocoding options since address data is considered the most elemental geospatial object linked to the household, enabling linking processes between statistical data and geospatial data. It is more spatially accurate than locality geocoding, especially when high-quality address data (full physical address description) are available, underpinned by internationally or nationally endorsed standards (e.g., standard format and official coding system/index) along with a robust regulatory framework. Addresses have been increasingly important in official statistics at both national and global levels. Address data are basic location data and the most geospatial elemental level in providing the lowest common geographic denominator - or unit - assigned to each statistical unit for data collection and dissemination purposes (UNECE, 2016).

The digital spatial representation of the statistical units linked to direct and physical location data (geographic coordinates) or indirect and descriptive location data (e.g., address, city name, etc.) are the geocodes (GEOSTAT 4, 2022a). The geocode is a single geographic coordinate or a unique code used to determine the location of an object on the Earth's surface, frequently represented by a point or a polygon (ABS, 2018). Geocodes enable the production of geocoded data through a direct link to a set of coordinates using alphanumeric codes of geographical areas or locations (e.g., administrative units). Thus, geocodes can be point-based data (geographic coordinates of an address or centroid of a building) or small area geographies (e.g., census tracts, small statistical areas or building blocks), stored as a geometry data type. Larger geographies can be used as geocodes when smaller geographies are unavailable for statistical purposes, but high-precision geocodes are more recommended for data linking at the unit record level (GEOSTAT 2, 2017). Geocodes are also used in many countries in their unified identity systems for persons, addresses or businesses in which these identifiers are usually anonymised and incorporated into the statistical production taking privacy protection measures (UNECE, 2017). However, the lack of standardisation on the identifiers at both national and international levels, sometimes from poor collaboration and an absence of understanding of technical issues between NSO and NGIA, is widely acknowledged as a data integration challenge.

According to UN-GGIM: Europe (UN-GGIM: Europe, 2021), the main data integration methods for integrating statistical and geospatial data are: i) point-based system; ii) area-based data integration; iii) spatial join operation; iv) Persistent Unique Identifiers (PID); v) ontologies; and vi) Resource Description Framework (RDF) for Linked

Data. In addition, UNECE (2017) divides data integration methods into two main groups: record linkage and statistical matching.

The **point-based system** enables identifying and combining records corresponding to the same statistical units (e.g., building or dwelling) from two or more data sources based on matching or pairing of unique identifiers, i.e. PID. It can also be known as a geographic linkage which occurs when matching the unique identifier to a series of geographic codes (geocodes) that apply to the location of the statistical unit (e.g., address) (UNECE, 2016). It is considered the conceptual cornerstone of statistical-geospatial data integration and is recognised and described in global frameworks, and a fundamental feature in the development of a geocoding infrastructure to support the production of geospatial statistics. The point-based system is a technical and methodological framework supported by a geocoding infrastructure that enables a statistical unit record to be linked to a precise geographical reference, i.e. geographic coordinates (x, y and z) (GEOSTAT 4, 2021b). Location data from addresses, buildings, dwellings and cadastral parcels are eligible to be spatially represented by vector point coordinates. Physical address location data is commonly considered the universal input data for this geocoding approach (GEOSTAT 2, 2017). Hence, the point-based geocoding infrastructure allows for accurate spatial representation of statistical data collected from survey and administrative data sources using location as a matching key.

Due to the high-precision point referencing setting, this production system enables data aggregation procedures (e.g., geocoded population data) into larger geographies or stable analysis units (e.g., grid cells) in a bottom-up approach. It also tackles a major methodological shortcoming related to areal data by adapting to changes occurring in output geographies, particularly in administrative units or other geographies derived from them for statistical purposes (e.g., Nomenclature of Territorial Units for Statistics - NUTS - in the EU level). Thus, it ensures comparable analysis over time and space with more adaptability and flexibility to produce and release geospatial statistics.

The point-based system can be conceptually and topologically connected to location data at the polygon level while ensuring consistency and hierarchical linkage to create an integrated location data framework (e.g., Statistics Sweden's case on the national authoritative location data framework). Point coordinates of buildings and addresses (central coordinate/centroid and/or building door) could be linked to the polygon of the cadastral parcel where spatially intersected or are located, enabling the integration of all these object types into the geocoding infrastructure (GEOSTAT 2, 2017). This type of location data framework will also support geospatial data quality assessment by enabling point-in-polygon operations with high-quality polygon data (location accuracy), checking topological consistency and conducting data comparison in time series to detect errors.

The point-based system is used in many European countries, particularly in most ESS countries, demonstrating a reasonable degree of geographical coverage of point coordinates of buildings, addresses and/or cadastral parcels (GEOSTAT 2, 2016). However, it encompasses several practical challenges concerning its development and implementation. Availability and accessibility of authoritative high-quality geospatial reference data (e.g., addresses and buildings), completeness and geographic coverage of the datasets, and lack of data interoperability and standards are examples hampering the effective implementation of point-based system (GEOSTAT 2, 2017; GEOSTAT 4, 2021b).

The **area-based data integration** involves data integration from different sources with different samples of the same target population to provide insights into the relationship of non-joint variables in the datasets and deals with similar statistical units rather than identical ones, as is the case in record linkage (UNECE, 2017). This method links statistical data in tabular form with geometric areas through area identifiers, representing an indirect spatial reference of statistical data, and is used for spatial analysis and dissemination in statistical production (UN-GGIM: Europe, 2021). It combines polygons associated with any output geography with statistical tables using geocodes stored in both statistical tables and geometries either directly in the statistical table at the statistical unit record or linked in some way to the record, i.e. correspondence table (GEOSTAT 2, 2016). It is traditionally used in surveys and censuses where previously geocoded population data is assigned to fixed output areas (e.g., enumeration district, census blocks or small statistical areas) and frequently aggregated within boundaries to a higher geographical level for dissemination (UNECE, 2015).

A standard data joining method linking statistical data tables with geometries is the Table Joining Service (TJS), an OGC standard that enables the description and exchange of tabular data containing attributes about geographic objects (OGC, 2010). It provides a standard interface to join tabular/attribute data stored in one environment with geospatial data addressing the corresponding geometry stored in another environment (GEOSTAT 3, 2019b; GEOSTAT 4b, 2021). The geographic identifiers that must be included in each table record - location attributes - will enable one to join alphanumeric data to the respective geometry and spatially represent it on a map. OGC TJS was applied with the support of ISO TC/211 19112:2003 (Geographic information - Spatial Referencing by geographic identifiers) to ensure standardised and consistent use of geographic identifiers by defining a conceptual schema and establishing a general model. The OGC specifications of this standard can be used as a technical guideline for statistical organisations to develop their joining services and applications in a more automated and efficient way, according to their needs and adapted to the data and technological components. Alternatively, they can be used to provide open and more user-oriented web-based interfaces that allow users to employ their statistical datasets and geospatial data (via file input, upload or URL links), to visualise a basic map from the join operation, and to download it in different formats (GEOSTAT 4, 2022b).

A **spatial join operation** is a macro-level linkage method that refers to joining data from multiple sources that cannot be individually linked or matched using geocodes but can be linked through spatial aggregation by location. In other words, the spatial join operation joins attributes from spatial objects (usually point-based data) to another one (usually area-based data) based on spatial relationships, i.e. point-to-polygon geospatial operations. This operation is a standard geoprocessing tool in GIS for spatial analysis, matching rows of the join features values to the target values based on their relative spatial location and defined geographic criteria. Spatial join operations are also used to spatially overlay point-based population data with grid cells to create datasets of population grids enabling data aggregation and grid statistics for dissemination, being commonly used for comparative analysis of statistical grid data from censuses (GEOSTAT 1A, 2012). Grid statistics are processed and reported using gridded geographies with a consistent size (usually squared), identified with a unique geocode and independent of the underlying geography (e.g., administrative units or statistical areas), typically with a single explicit value inherent in each cell (UN-GGIM, 2021).

According to UN-GGIM: Europe (UN-GGIM: Europe, 2021), **PID** are geocoding requirements or technical prerequisites for the implementation of the two previous data integration methods, which enable statistical/administrative data to be linked at record unit level with an indirect spatial reference by using common identifiers. PID compute data linkage from different data repositories and information systems using semantics (semantic interoperability) to join (and exchange) data through a common language, terminology and conceptual basis that defines the meaning of data elements and the logical relationships between them (statistical data and corresponding spatial objects). Thus, PID are intertwined with the data integration core term 'linking' as the 'process of connecting structured data sources using a system of unique identifiers' (UN-GGIM, 2021:28) based on standard web technologies.

By using PID or common identifiers in unit record data and location data, geometries and coordinates do not need to be necessarily stored with the unit record data, especially when such data are effectively and safely stored in a data management environment with maintenance and management routines under a version-management policy (e.g., versioning history at the record level and documentation on identifier lifecycle management practices).

PID can be deployed throughout micro-level linkage, i.e. microdata integration, or macro-level linkage (Yung et al., 2018). The first one addresses individual units in multiple datasets that can be associated with each other by the presence of unique identifiers. In the second one, although linkage operations can be performed due to the absence of such identifiers, data can be aggregated at some level. In the European context, PID are highly recommended and implemented as general requirements for EU and pan-European data themes, compliant with INSPIRE, as a mechanism of lifecycle attributes and versioning and aligned with the key recommendations on data integration and specifications of core geospatial datasets for policy applications. Thus, the use of PID for consistent geometries and codes of statistical/administrative geographies - both current and historical - facilitates combinations of geocoded microdata and any statistical/administrative geography, including grid systems, in a more flexible, convenient and effective manner (GEOSTAT 1B, 2013).

Linked Data technologies can support the use of PID alongside harmonised vocabulary collections (unique coding systems and common ontologies), correspondence tables and metadata standards throughout the data lifecycle. These improvements will address issues related to semantic and syntactic heterogeneity from different data sources, types and structures, as the lack of formal and cross-domain ontologies for statistical and geospatial data is widely recognised (UN-GGIM, 2021). Many efforts have focused on developing targeted ontologies, hierarchies of ontological levels, general semantic foundations and consistent vocabularies to annotate, enrich and model semantic data relationships and provide a common understanding and application of heterogeneous datasets and related semantics issues (Ariza-López et al., 2021).

Lastly, statistical and geospatial data can be transformed into **RDF** as both formats and data models to be employed in **Linked Data** technologies (UN-GGIM: Europe, 2021). RDF is a standard model, developed by the W3C for interchanging data on the web that facilitates merging operations from different schemas and allows structured and semi-structured data to be linked, published and shared across different applications (W3C, 2014). It is a data integration method supported by a technical framework that provides a more user-friendly way of representing and describing resources and their relationships on the web in a machine-readable format, enabling the creation of data graphs (RDF graphs) from, for example, Excel and Comma-Separated Values files.

In the case of statistical and geospatial data, both data models can be accessed on the web, represented and queried using SPARQL (the query language for RDF) and GeoSPARQL, respectively, providing a stable and common semantic structure on the web (data vocabulary and models). GeoSPARQL, an extension of SPARQL, is one of the major RDF applications that defines the vocabulary for semantically representing geospatial data and is designed to support geospatial querying and reasoning (Perry & Herring, 2012). GeoSPARQL is a suitable way to publish geospatial data as LOD, while RDF enables statistical and geospatial data from different sources to be linked (using geography as a common element) and delivered in a single application, along with publishing structured metadata from the shared ontologies and mapped data vocabularies (Eurostat, 2019b).

Many statistical organisations represent and share their datasets as RDF within open data web-based catalogues using Data Catalogue Vocabulary (DCAT), its extension, DCAT-AP, or other modelling ontologies (e.g., XKOS) to make their data and metadata compliance with the Findable, Accessible, Interoperable and Reusable (FAIR) data principles (FAIR, 2016). These principles aim to make data more findable through catalogues and metadata, accessible using open and standardised protocols, interoperable via content encoding (e.g., HyperText Markup Language) and reusable by license information and data product specifications. In this regard, the use of open standards and non-proprietary data formats, such as eXtensible Mark-up Language (XML) and GeoJSON, and a range of web-based best practices to represent geospatial data and metadata (e.g., API and machine-readable tools) increases the accessibility and usability of integrated data.

In addition, W3C has suggested the DCAT standard to many statistical organisations as a standard supporting the discovery use case of all data types. Statistical data and structural metadata in RDF format are used for linked data and metadata harmonisation applications to publish and exchange data from various sources making data more easily accessible and reusable to the users. These innovative ways of sharing and visualising statistical content, including the statistical classifications and geographical variables (e.g., NUTS classification), through RDF modelling and LOD-based technologies, are modernising the dissemination of official statistics and fostering interoperability.

The development of DCAT application profiles is also endorsed towards a more interoperable architecture, highlighting the good practice case of Statistics Sweden in developing both specific StatDCAT and GeoDCAT application profiles (Haldorson & Moström, 2018). On one hand, the StatDCAT-AP is used to describe statistical datasets and LOD graph formats for statistical classifications in which statistical data from different sources and time series can be linked to create more comprehensive datasets, enabling data integration and temporal analysis. On the other hand, GeoDCAT-AP is an extension used for describing geospatial datasets, data series and services. rk

Considerable work has been done on structural and semantic metadata, both conceptually and technically (e.g., GSIM, API, etc.), however, the number of statistical organisations sharing their data and metadata via LOD technologies is still limited despite the opportunities to build standardised and on-the-fly data integration (Do et al., 2015). Linked data practices, in particular those related to LOD, have been most applicable to data integration by providing means for representing, linking and querying statistical and geospatial data from different data sources on the web provided in both statistical and geospatial ontologies (UN-GGIM: Europe, 2021). There is also a need for more API - preferably open - and other technological solutions to support different data domains and user communities in data integration (e.g., smart statistical and geospatial data provision and ready-to-use tools). Meanwhile, the data quality and consistency of linked statistical-geospatial data as well as privacy and security concerns are some of the current challenges associated with this last group of data integration methods. These issues are particularly relevant when dealing with sensitive data and semantic interoperability to create standardised vocabularies and ontologies that capture semantics from both data domains.

Furthermore, although data integration methods are generally well documented, especially by the geospatial community, for example through technical papers, methodological guidance and good practices, there are still technical barriers and limitations to their effective use (UN-GGIM: Europe, 2021). Despite the growing institutional literature available as well as extensive methodological considerations, there are potential challenges related to data integration in official statistics that require new skills and innovative IT solutions. Improvements are needed to harmonise existing standards from both statistical and geospatial communities (e.g., in data formats for exchange and application services) or to create new ones, taking into account technological trends, to make them more compatible with each other. Standardisation will enable the development and implementation of shareable technological solutions and streamline data integration methods across statistical production.

II.1.3. KEY STAKEHOLDERS AND ROLES

In most countries, statistical production and geospatial data management are generally assigned to different public institutions and governmental bodies (e.g., NSO and NGIA), with some exceptions like in Mexico and Brazil that combine both domains in one organisation. Thus, official statistics and activities related to geospatial data are traditionally detached, with each domain establishing its own information systems and production frameworks, and the statistical and geospatial communities not actively communicating or interacting with each other (Ariza-López et. al, 2021). Nevertheless, a few international organisations and high-level stakeholders have been promoting statistical-geospatial data integration and building synergies and institutional partnerships between the two communities, particularly driven by the 2030 Agenda for more accurate, timely, detailed and comparable data, i.e. SDG roadmap.

The UN Expert Group on the Integration of Statistical and Geospatial Information (UN EG-ISGI) is a group under the UN-GGIM intergovernmental body, established in 2013 to raise awareness and promote the importance of integrating statistical and geospatial data within global, regional and national policy frameworks and in supporting decision-making and policy development at multiple levels. It was also established to support the regional and national implementation of the **Global Statistical Geospatial Framework (GSGF)**. This overarching framework was developed by this key player as the first strategic bridge and global standard to facilitate data integration from both the statistical and geospatial communities in order to produce geospatially enabled statistics in a harmonised manner. In this context, the EG-ISGI has also recognised the importance of IGIF as an enabler of the GSGF and for the statistical domain, strengthening its interconnections with the GSGF and enhancing the value that geospatial data can add to official statistics.

The UN EG-ISGI reflects the UN-GGIM's strategic pillars of sharing knowledge, raising awareness and strategic leadership concerning statistical-geospatial data integration providing a high-level coordination of the activities, developing and sharing guidelines and best practices and strengthening capacity and institutional collaboration. One of the main activities of the UN EG-ISGI is to foster interoperability between statistical, geospatial and administrative data communities in supporting the GSGF implementation, namely by examining maturity levels and capabilities of countries.

This expert group also has European enforcement through the UN-GGIM: Europe WG on Data Integration focusing on integrating geospatial data with statistical data (and other data/information) taking into account the European statistical-geospatial operating environment. This WG is composed of experts from the Member States (MS), affiliated with geospatial and geospatial organisations, private companies and observer organisations, such as Eurostat, JRC and EEA. Together with the other four WGs (SDG, IGIF, Data Strategy and Policy and Geodetic Reference Frames), they compose the five lines of work and key priority areas to meet strategic regional alignment of the data needs and requirements with the global programme.

In the agenda of statistical-geospatial data integration, the line of work on SDG - as a subgroup of the WG on Data Integration - aims to showcase the added value of integrating geospatial data with other data to address the SDG indicators by developing technical and methodological solutions, such as with EO data. Moreover, the WG on Core Data focuses on data interoperability and harmonisation through core geospatial data themes and related recommendations that meet user needs and requirements.

In addition to UN-GGIM, UNECE develops capacity in statistical-geospatial data integration by producing and sharing guidelines, methodological materials and best practices, and fostering innovation in line with the modernisation of official statistics for the 2030 Agenda implementation and SDG. UNECE has been actively providing workshops, sessions, interactive activities and training initiatives to promote networking and discussion among key stakeholders and to identify obstacles, solutions, lessons learned and future challenges and opportunities at both national and international levels. UNECE also fosters institutional partnerships and supports greater collaboration between the statistical and geospatial communities and promotes the use of common standards to increase interoperability across the region.

One key line of work of UNECE has been related to standards issues on data integration to identify domains where integration is hampered by a lack of common standards and set actions towards statistical and geospatial standards harmonisation. In this regard, the HLG-MOS has been providing guidance for the international statistical community on how data integration activities fit into the statistical business process by promoting the development of data integration strategies and practices, encouraging joint approaches and researching key topics related to quality, methods and technologies. Throughout its work on standards, interoperability and governance issues, this High-Level Group has been prompting the data integration agenda within international statistical programmes and overarching policy frameworks and strategies.

The Global Forum for Geography and Statistics (GFGS) is a global network focusing on the exchange of experiences, use cases and best practices related to statistical-geospatial data integration to strengthen capacity and improve capabilities. In collaboration with UNECE and EFTA, GSGF has conducted a series of webinars (or 'coffee talks') to present, share and discuss emerging topics, innovative methods and practical examples related to statistical and geospatial data, such as grid statistics, geospatial and statistical standards and functional geographies for statistical dissemination.

At the European level, the **European Forum for Geography and Statistics (EFGS)**, which works closely with Eurostat, has focused on developing activities and providing guidelines and methodological materials on geospatial statistics to advance the integration of statistical and geospatial data across Europe, in particular within the ESS context. The EFGS works as a voluntary professional network with several national contact points from more than 40 countries and regions involving experts from NSO and NGIA and other professionals working in statistical-geospatial data integration in the European context. EFGS aims to improve the capabilities of NSO regarding statistical-geospatial data integration towards a more effective European statistical-geospatial data ecosystem, build bridges and strengthen cooperation between the European statistical and geospatial data communities. A key overarching priority and operational cornerstone has been to provide a common guidance for a harmonised implementation of the GSGF in Europe, namely through the GEOSTAT projects over the years, alongside the annual conferences that promote networking between both communities.

EFGS worked in the **GEOSTAT projects** from 2010 to 2022 in collaboration with and funded by Eurostat, firstly focusing on a population grid dataset to represent census data towards the regional adaptation of the global statistical-geospatial framework (GSGF). They aim to establish a harmonised production infrastructure for standardised geospatial statistics under the ESS long-term strategy considering the potential of new data sources (data revolution), new metrics for statistics with higher geographical detail and new forms of governance (ESS, 2014). Initiatives such as the GEOSTAT projects were fundamental in developing a consistent and systematic approach to integrate statistical and geospatial data in a heterogeneous and complex data integration landscape (UNECE, 2017).

The **PARIS21** is also a key global player and a major contributor to statistical-geospatial data integration by promoting better production and use of statistics throughout the policy lifecycle and for sustainable development, especially in developing countries. PARIS21 is an active stakeholder by enhancing statistical capacity through a global network, developing innovative solutions for statistics (e.g., new data sources) and building and sharing knowledge via discussion and strategy papers. It also compiles and shares guidelines and good practices on the integration of statistics with geospatial data in NSS and their data ecosystems, with the mission to assist and support countries with lower levels of maturity and capacity. In addition, the SDG constitute a key action area in which PARIS21 is fully engaged and can be seen as an operational driver to promote initiatives addressed to NSO and NGIA for the development and implementation of consistent approaches based on the data requirements under the global policy framework.

The **Directorate-General for Regional and Urban Policy (DG REGIO)** is the EC department responsible for the EU policy on regions and cities. Over the years, DG REGIO has done much to promote statistical-geospatial data integration in the European context. It is part of the EC stakeholders that are responsible for defining economic, social and environmental development policies, together with other policy DGs (e.g., in environment and agriculture). As designers of regional and territory development and environmental policies, this group of stakeholders relies on the use of official statistics and (authoritative) geospatial data to produce geospatial statistics fitting to their needs on policy development for regions and types of territories. Hence, DG REGIO developed a series of experiences and outcomes on the combination of comparable statistical indicators with available geospatial pan-European datasets, while assessing the data integration opportunities and challenges for several applications, such as land use/land cover in urban areas, accessibility to services and transport networks. It has also been working closely with UN-GGIM: Europe and EFGS to find synergies and mutual gaps in data integration issues. It also works with the JRC to support territorial-based policies and SDG implementation at the urban and regional levels. In addition, the JRC supports

EU digital policies through research efforts related to digital technologies, data and digital platforms, contributing to the advancement of data integration in Europe and to the European Data Strategy at the governance, legal, technical and technological levels.

The **European Committee of the Regions (CoR)** also plays a central role in fostering statistical-geospatial data integration at the European level, outlining the importance of integrating regionalised statistical data with geospatial data and a harmonised statistical-geospatial framework for better EU policy-making (CoR, 2021). The CoR also highlights the need for disaggregated data (local and regional statistics) for more informed decision-making, as well as for NSO and local/regional authorities to tackle data gaps and support SDG indicators towards a territorial approach for their EU implementation (CoR, 2019). In general, it recognises the strategic and operational value of statistical-geospatial data integration for the design, implementation and monitoring territorial-based EU policies and of geospatial data for carrying out territorial analysis for policy processes.

Lastly, the administrative data community plays a crucial role in fostering greater statistical-geospatial data integration by shaping the production and management of administrative data sources according to agreed requirements and standards and promoting common understanding and institutional agreements (GEOSTAT 4, 2021b). This data community includes global, regional and national public institutions responsible for administrative data collection and maintenance of public administrative data repositories (population and business registries, land and cadastral registries, tax authorities, etc.) and other bodies responsible for legal frameworks involving administrative data (e.g., EU Digital Single Market). The administrative data community also encompasses private data providers from businesses and companies.

II.2. STATISTICAL-GEOSPATIAL FRAMEWORKS

The **GSGF** was jointly developed by the UNSC and UN EG-ISGI as a culmination of their visions and synergies over the years, attending to development priorities and agendas from national to global levels. It follows the work and efforts of the global statistical and geospatial communities to define a common international framework for the acquisition, management and use of statistical and geospatial data, acknowledging data integration as a foundation for multi-level government action and one of the key challenges for evidence-based decision-making (UNECE, 2016). It was conceptually and methodologically based on a national statistical-geospatial framework of Australia (Statistical Spatial Framework), inspired by IGIF and related national experiences (e.g., address data management), and further formally adopted by UN-GGIM in 2018 and endorsed by the UNSC in 2019. The GSGF also comes from the range of national initiatives, practical examples and best practices to support global implementation.

The GSGF is an overcharging framework to integrate statistical and geospatial data and produce harmonised and standardised geospatially enabled statistics, facilitating data-driven and evidence-based decision-making to support policy development at multiple levels (UNSC & UN-GGIM, 2019). The GSGF is a common high-level global statistical-geospatial framework, particularly oriented to the global statistical and geospatial communities, that provides a consistent production process for geospatial statistics and common data integration approaches based on location information from statistical, administrative and other data sources. It addresses a strategic roadmap and a guiding implementation mechanism with some degree of flexibility and adaptability, considering the different circumstances and gaps of countries, regions and organisations.

The GSGF is built on five **Principles** and four **Key Elements** to turn input statistical and geospatial data into integrated, harmonised, standardised, interoperable and comparable outputs used for analysis, diffusion and decision-making (Figure 3).

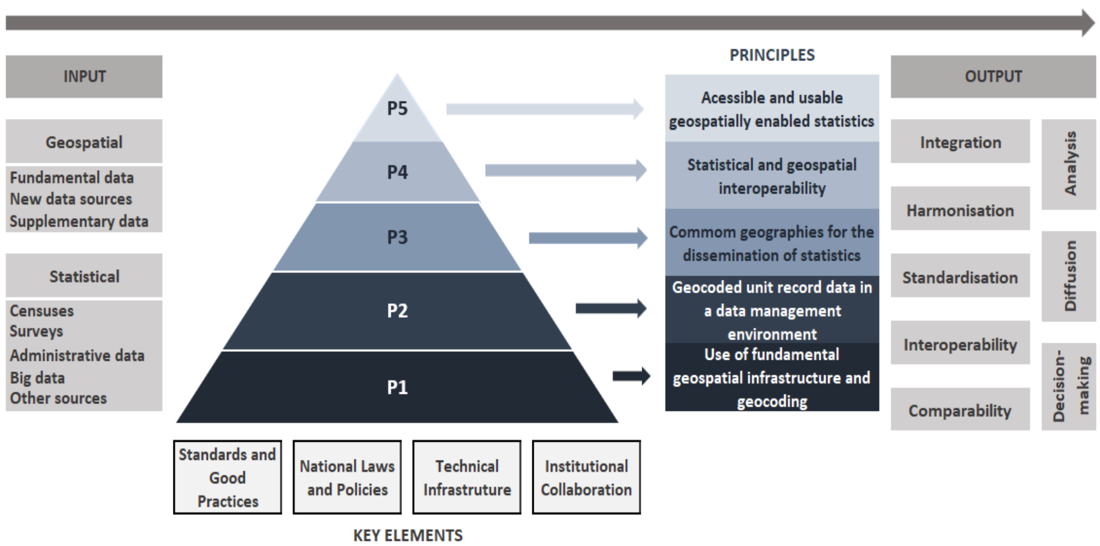

Figure 3. Global Statistical Geospatial Framework (GSGF) from inputs to outputs through the Principles and Key Elements (source: author, from UNSC & UN-GGIM, 2019).

The high-level five Principles are the conceptual cornerstones of GSGF guiding throughout the consistent and systematic production process for geospatial statistics and are defined by a set of goals supported by common standards, tools, methods and best practices. They outline the broad processes to enable data integration and are hierarchically structured and connected at different levels:

Principle 1 (Use of fundamental geospatial infrastructure and geocoding): It recommends a common and consistent methodological approach to place each statistical unit in time and space through a fundamental geospatial infrastructure with geocoding practices, preferably at a point-based level. The goal is to obtain a high-quality, standardised location description to assign accurate and precise coordinates and/or a small geographic area or standard grid reference to each statistical unit.

Principle 2 (Geocoded unit record in a data management environment): It supports the process of linking and storing high-precision geographic references (e.g., geocodes and PID) to each microdata/statistical unit record within a secure data management environment. These technical requirements will enable data aggregation and data linking processes from various data types and sources in a sustainable manner.

Principle 3 (Common geographies for the dissemination of statistics): It applies geography as a tool for data integration by using a common and agreed set of geographies to the display, storage, analysis and reporting of statistical data and enabling comparisons over time.

The Principle 3 aims to address the inconsistent network of dissemination geographies at the national, regional and global levels as well as the absence of small statistical areas in several countries which undermines national and international data comparability. Therefore, it encourages the development of global and regional datasets, namely for monitoring the SDG indicators, and their full integration into the authoritative geospatial infrastructures, including the NSDI (Principle 1).

The GSGF considers two main geographical classifications - administrative units and gridded geographies - from which other geographies (e.g., statistical areas) are derived according to the selected criteria. It advocates the adoption of grid systems as one of the most recognisable agreed frames of common geographic areas to encode, store, analyse, display and report statistical data from different data sources enabling the production and dissemination of geospatial statistics (UN-GGIM, 2021).

Grid systems are increasingly popular due to their flexibility since they provide a set of advantages: i) they can be constructed hierarchically in terms of cell size (the smallest size of the geographic unit) to match the study area; ii) grid cells can be assembled fitting into the geography of interest based on a specific application; iii) all grid cells have the same size allowing for easy comparability; iv) grids are stable over time; and v) grids can integrate easily of other types of data. Using these grid systems to geospatially enable statistical data and produce gridded statistics usually requires georeferenced point datasets with high spatial accuracy (Neves & Moreira, 2017). Both aggregation methods for converting point-based data into grids and disaggregation methods for estimating grid data can be carried out to produce harmonised grid datasets by adopting bottom-up and top-down approaches to produce geospatial statistics (GEOSTAT 1B, 2013).

The Discrete Global Grid System (DGGS) is a type of worldwide comparable grid system as a hierarchy of equal area cells (e.g., rectangular/squared or hexagonal) with sequentially finer spatial resolution in which individual data records can be assigned to a cell, respectively referencing by zonal identifiers with its structured geometry. This reference system works as a globally unifying key and common geometry that easily brings both statistical and geospatial data together, increasing analytical potential. The implementation of this system also constitutes an effective alternative to existing geospatial frameworks due to its scalable ability for data aggregation and integration application and compliance with OGC standards (Bousquin, 2021).

Principle 4 (Statistical and geospatial interoperability): It defines the preconditions and technical requirements for the development of a statistical-geospatial data ecosystem enabling effective linking of datasets across different systems and applications as well as efficient data exchange, discovery, access and use of geospatially enabled data. Hence, it focuses on interoperability between statistical and geospatial data from common standards and consistent good practices and approaches to overcome structural, syntactic and semantic barriers from different communities and data providers.

Principle 5 (Accessible and usable geospatially enabled statistics): It outlines the need for data custodians to make geospatial statistics in an accessible and usable manner according to agreed international standards, good practices and technological solutions, for instance, through standard web services, LOD and machine-readable methods. Operational guidelines and policies related to data confidentiality and privacy protection are other aspects of this principle.

The **Key Elements** are cross-cutting dimensions to the five Principles, allowing data to be obtained from several sources and ensuring that the Principles are applied throughout the production process for geospatial statistics. The four Key Elements are:

Standards and Good Practices: It is applied across the statistical and geospatial communities and extended to IT and other related domains, including formal statistical and geospatial standards, good practices and guidelines.

National Laws and Policy: It includes legal frameworks and professional and social infrastructures enabling or constraining activities. It may cover international and national data protection, privacy and confidentiality legislation and regulations, ethics and social license requirements, open data and data stewardship policies and data access and sharing agreements.

Technical Infrastructure: It covers a broad range of technical capabilities at national and regional levels, such as technical skills, agreed methodologies and business processes, IT infrastructures and systems, and data management environments.

Institutional collaboration: It addresses the crucial role of institutional collaboration across statistical, geospatial and administrative communities, the need for political commitment, formal agreements and a collaborative approach between stakeholders (e.g., knowledge exchange and education initiatives).

UN-GGIM in its ongoing mission to promote the adoption and implementation of the GSGF developed an implementation guide to provide more practical guidance to countries on how to implement and operationalise the framework, showcasing several national examples (UN-GGIM, 2021). This document reviewed key topics, their applicability and connections to the GSGF Principles, provided objective explanations and recommendations, identified resources and further reading materials, and proposed an agreed terminology for both statistical and geospatial communities.

The European version of the GSGF (**GSGF Europe**) results from the work carried out in the series of the GEOSTAT projects (EFGS and Eurostat) towards standardising geospatial statistics and strengthening the statistical-geospatial capacity in the European context. It provides a high-level summary of the conceptualisation and interpretation of the GSGF (the high-level policy document) in the European context, attending the European statistical-geospatial operating environment. Thus, it works as a regional adaptation and implementation of the global framework.

The GSGF Europe is supported by the same five Principles and four Key Elements of the global framework to produce harmonised and standardised geospatial statistics (GEOSTAT 4, 2021b). The GSGF Europe describes these Principles, summarises their focus points and objectives, interprets them in the light of the regional context (e.g., ESS) and assesses them from different cross-cutting aspects that should be considered to achieve an effective statistical-geospatial framework.

Principle 1 aims to develop an organised geospatial infrastructure based on precise and standardised location references and consistent geocoding approaches according to nationally agreed standards and good practices.

Principle 2 aims to enable geospatial consistency in data focusing on the process of placing and storing each statistical unit record in an effective data management environment that can handle changes in geographies and on privacy and confidentiality aspects, for instance, defining data maintenance policies and custodianship roles.

Principle 3 aims to ensure common geographies focusing on metadata to support data aggregation and integration processes, standard mechanisms for conversion between geographies and aggregation and disaggregation methods enhancing quality, consistency and comparability of data. This principle delivers a higher maturity level in the European context due to the ESS's solid conceptual and methodological framework for reporting official statistics to ensure European comparable regional statistics. The NUTS, functional territorial typologies at both local and regional levels (TERCET regulation, 2017[7]), the 1km^2 Population Grid provided by the GEOSTAT project and INSPIRE Data Themes on statistical units and geographical grid systems are examples supporting the application of this principle in the European context. Other wide ranges of statistical classifications as well as cross-cutting classifications from different domains at national, regional and global levels can also foster the implementation of this principle.

Principle 4 aims to enable interoperable data via common data, standards and processes focusing on laws and policies supporting cooperation and good practices for greater standardisation and interoperability throughout the production process of geospatial statistics.

Principle 5 aims to provide easy access and usability to geospatially enabled data focusing on data discovery, access, analysis, visualisation and dissemination capabilities highlighting the need for data custodians to make high-quality geospatial statistics accessible and usable to all users while ensuring privacy and confidentiality mechanisms.

GSGF Europe also outlines the relations between the Principles addressing the foundational role of Principle 1 and shared contribution with Principle 2 to secure conditions for the full implementation of Principle 3, which at the same time established functional links with Principle 5, and Principle 4 has a cross-cutting relation with all other GSGF Principles (GEOSTAT 4, 2021a). The description of the Key Elements is extended to organisational, legal, semantic and other non-technical aspects to produce accessible and usable geospatial statistics.

The main difference between the GSGF Europe and the global framework is the set of Requirements ("What") and Recommendations ("How") on which the European version of the GSGF is operationally built, supporting the practical application of the Principles and Key Elements (Figure 4). They provide detailed and equally non-binding implementation guidance instructions and detailed actions, promoting a more top-down operational approach and turning the framework into small, concrete and manageable pieces.

[7] https://eur-lex.europa.eu/eli/reg/2017/2391/oj/eng.

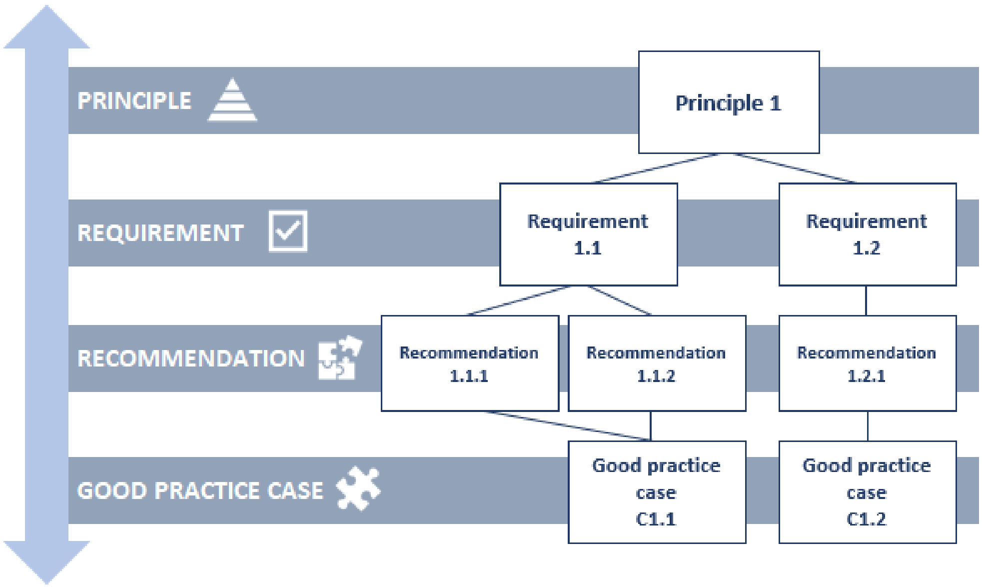

Figure 4. Hierarchical implementation structure of the European version of the Global Statistical Geospatial Framework (GSGF Europe) (source: author, from GEOSTAT 4, 2022c).

The Requirements and Recommendations underline the structure of the framework where each Principle is followed by several Requirements that break down into more specific Recommendations, which in turn connect to some good practice cases. In a nutshell, they work as a methodological roadmap, a self-assessment tool and a checklist to assist organisations in producing geospatial statistics in a systematic and consistent way.

The first version of GSGF Europe was developed in the GEOSTAT 3 project to adapt the global version of GSGF by drafting a guide for harmonised implementation considering the regional statistical-geospatial operating environment and does not intend to replace the global guidance but rather ensure overarching compliance. The framework was further reinforced in the GEOSTAT 4 project addressing geospatial statistics. In this regard, it is important to provide a brief overview of the four editions of the **GEOSTAT projects** that strongly contributed to the development of statistical-geospatial data integration in the European context:

GEOSTAT 1A (2010-2011) was the first edition producing a prototype 1 km^2 grid dataset for 2006 population data and developing methodological foundations for generating European grid statistics to support the creation of European population grid datasets (GEOSTAT 1A, 2012). This project aimed to spatially represent various variables in a 1 km^2 grid dataset using the 2011 Census data to ensure data comparability across European countries. It provided recommendations and guidelines on producing a European population grid dataset from various national sources towards a harmonised approach, including confidentiality and quality considerations related to grid data.

GEOSTAT 1B (2012-2013) continued with the work of GEOSTAT 1A in developing best practices and guidelines to produce grid statistics at the European level and spatially referenced statistics in a hierarchical system of stable and neutral grids. It produced the first version of a grid dataset for the 2011 Population Census and a detailed manual guiding on how to create population grid statistics as part of the statistical business process (GSBPM) (GEOSTAT 1B, 2014). Additional technical guidelines on the methodology to produce a European harmonised population grid from the national data through aggregation or disaggregation methods were also developed. The project also showcased the usability and potential of grid statistics for spatial analysis through a case study on accessibility to emergency hospitals, including statistical confidentiality issues.

GEOSTAT 2 (2015-2017) proposed a generic model for a national point-based geospatial reference framework for statistics based on geocoded location data and promoted the integration of geospatial data into the statistical business process (GSBPM) to enable the consistent production of geospatial statistics. Both conceptual and methodological aspects to develop and maintain a point-based foundation for the production of geospatial statistics were addressed. The guidance document was the main outcome of the project, along with national practices and recommendations from several NSO in building and maintaining this type of geocoding technical infrastructure (GEOSTAT 2, 2017). GEOSTAT 2 also presented the results from a 2015 survey on geocoding practices in the European NSO. The results showed significant progress in most countries regarding the implementation of the point-based production model and in geocoding of statistical data with full or partial coverage and high spatial accuracy (GEOSTAT 2, 2016). Legal restrictions, lack of resources and geospatial capacity, and poor institutional cooperation and collaborative environment were identified as key barriers to effective geocoding practices.

GEOSTAT 3 (2017-2019) focused on the development of the European version of the GSGF by adapting the global framework to the ESS and the wider European context and its statistical-geospatial operating environment. The work resulted in an implementation guide, supported by good practice cases linked to the GSGF Principles, requirements and recommendations, breaking down the framework to a more operational level. This implementation guide was designed to be a flexible roadmap to implement the GSGF in the European context with the ability to be extended and revised due to technical and institutional changes. GEOSTAT 3 also tested and assessed the usefulness of the GSGF in Statistics Sweden, resulting in a national case study report that outlined some ideas for new activities and improvement actions, and identified key statistical-geospatial capability elements (e.g., information, standards and institutional arrangements) (Haldorson & Moström, 2018). This report also aimed to support other European NSO to improve their statistical-geospatial capacity for SDG measuring and monitoring the SDG. It also highlighted the opportunity for the research community to work closely with NSO, the importance of having more in-house geospatial expertise/capacity within the statistical organisation and streamlining geospatial data in statistical business production.

GEOSTAT 4 (2020-2022) aimed to enhance the GSGF Europe by expanding the ESS methodological guidance, improving previous materials, and updating the list of Requirements and Recommendations (GEOSTAT 4, 2022c), which resulted in a new version of the GSGF Europe (GEOSTAT 4, 2021b). GEOSTAT 4 produced an extensive collection of materials and outcomes supporting the implementation of the GSGF Europe, encompassing both technical and non-technical aspects under the framework's scope. It included the first draft of the GSGF Europe Reference Architecture based on a common enterprise approach built on a centralised geospatial data repository supporting standardised geospatial processes in statistical production, including the listing and mapping of stakeholders and roles, geospatial business services and geospatial activities. The GSGF Europe Reference Architecture was designed to be the basic operational structure of the European statistical and geospatial community in order to build shareable solutions and the same view of the operating environment (GEOSTAT 4, 2021c). The GSGF Europe Reference Architecture is a key distinctive component compared to the global framework as provides more concrete implementation guidelines and analytical tools on architectural descriptions and conceptual models.

Quality of geospatial data management for official statistics was a new line of work, establishing guidelines on quality-related components of geospatial data and processes within statistical production, such as geospatial enhancements of the ESS QAF and recommendations for geospatial quality reporting (e.g., quality indicators) (GEOSTAT 4, 2021d). Other outcomes included an extended terminology for the statistical and geospatial communities (GEOSTAT 4, 2022a) and a review of the GSGF and frameworks environment in which links between the GSGF and the surrounding nine frameworks were identified and described (GEOSTAT 4, 2021a). Confidentiality management in geospatial statistics, considerations related to survey data and Big Data in data collection and guidance on concrete business were also provided as core outcomes (GEOSTAT 4, 2021e; GEOSTAT 4, 2022d; GEOSTAT 4, 2022e). Lastly, the 'GSGF Europe: User Guide' was developed to describe all project materials and to assist users in implementing the GSGF Europe by providing a roadmap and methodological overview for the production of geospatial statistics in a consistent, systematic and harmonised way (GEOSTAT 4, 2022d). This document is an implementation tutorial taking into account the activities and needs of users and addressing relevant aspects, such as strategy and leadership, production and corporate support.

Furthermore, much work and effort has been done in the European context, at both technical and organisational levels, to improve the capacity to integrate statistical and geospatial data and to produce standardised geospatial statistics. Notwithstanding the significant results and milestones achieved by the GEOSTAT projects, their breakthroughs and insights are just the starting point in the GSGF Europe implementation journey.

As another key geospatial-statistical framework, the **Geospatial View of the GSBPM (GeoGSBPM)** was developed by the Geospatial Task Team under the Supporting Standards Group of the HLG-MOS and its first and current version was released in 2021.

The reference document (UNECE, 2021c) identifies and describes geospatial-related activities and considerations needed to produce geospatial statistics (or geospatially enabled statistics, as it is referred to in the framework), considering two overarching frameworks, the GSBPM and the GSGF (see Figure 5). On one hand, the GSBPM, as the core standard process model in the statistical community, has a connection point to Principle 4 of GSGF (statistical and geospatial interoperability), allowing a common understanding from both statistical and geospatial communities. On the other hand, the GSGF can be streamlined and integrated into the statistical production process using the GSBPM to produce consistent and harmonised geospatial statistics.

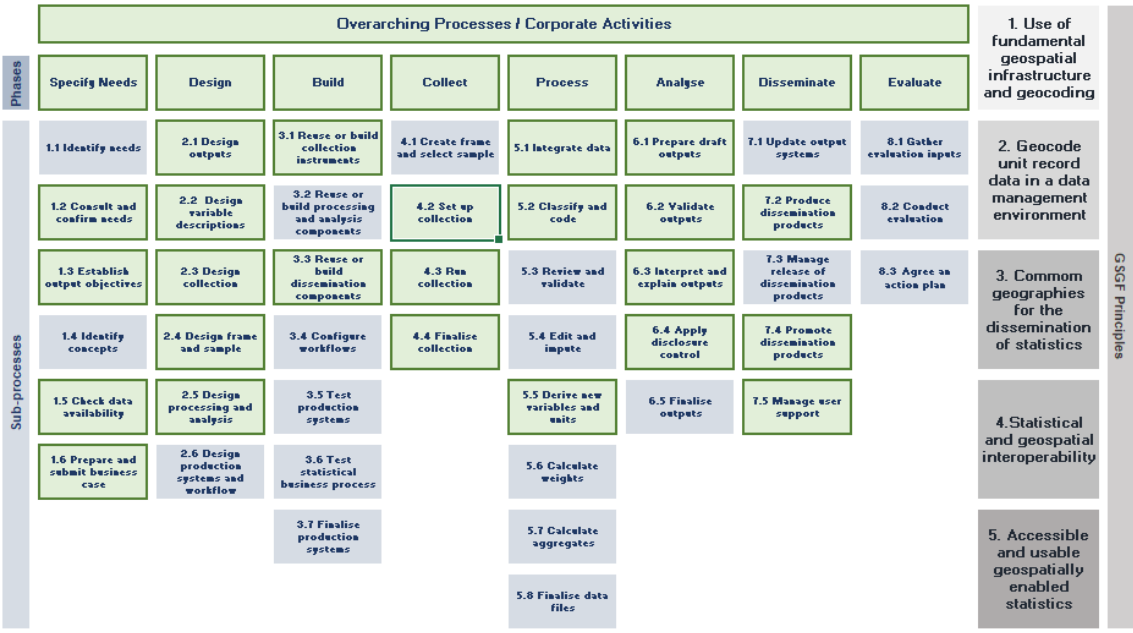

Figure 5. Geospatial View of the Generic Statistical Business Process Model (GeoGSBPM) and production process elements with geospatial-related activities and considerations (source: author, adapted from UNECE, 2021c).

GeoGSBPM enables the operationalisation of the GSGF Principles by statistical organisations as a common framework to help them identify common activities to produce geospatial statistics and to manage the quality and metadata of statistical and geospatial data and services (GEOSTAT 4, 2021a). GeoGSPM can be considered an integrated framework strengthening and formalising the bridge between both statistical and geospatial communities and respective data domains towards more interoperability and common standards. It can also be seen as a preliminary guideline to apply core business processes carried out by statistical organisations to a geospatial product (see

Ariza-López et al., 2021). Moreover, it enables the standardisation of the collection, processing, analysing and dissemination of geospatial data throughout statistical production and improve its management, usability and quality in the business processes.

GeoGSBPM follows the structure of the eight GSBPM phases and overarching processes/corporate-level activities (strategic collaboration and cooperation, metadata management and quality management) describing what geospatial-related activities should be included in each phase and high-level management processes and activities while considering the GSGF Principles. In total, 21 geospatial-related activities are mapped under each sub-process and corresponding phase of GSBPM or overarching processes/corporate activities. The distribution of geospatial-related activities is presented in the table below (Table 4) for each GSGF Principle and the GSBPM phases and overarching processes/corporate activities.

Table 4. Geospatial-related activities by GSGF Principle, GSBPM phase and Overarching processes/Corporate activities (source: author, adapted from UNECE, 2021c).

GSGF Principle	Geospatial-related activities	Specify Needs	Design	Build	Collect	Process	Analyse	Disseminate	Evaluate	Overarching processes / Corporate activities
Principle 1	5	X	X		X	X				X
Principle 2	3		X			X				X
Principle 3	3	X	X		X					
Principle 4	5		X	X			X	X		X
Principle 5	5	X	X	X			X			X

Although it is not a high-level statistical-geospatial framework for global or regional implementation, the PARIS21 partnership together with Statistics Sweden produced a stand-alone guide on geospatial data integration in official statistics (PARIS21 and Statistics Sweden, 2021). This guide provides practical step-by-step guidance inspired by the five GSGF Principles and recommends eight main tactical steps for NSO to establish stronger partnerships with NGIA towards an effective statistical-geospatial data environment, being more oriented to the statistical community. This guidance emphasises the importance of institutional collaboration as a core foundation of data integration processes and endorses the need to embed geospatial capabilities in the strategies and activities of NSS and other producers of official statistics in order to build interoperability within the statistical ecosystem,

This step-by-step strategic and methodological guide outlines specific tasks related to management issues, human resources and capacity building (technical skills development) related to statistical-geospatial data integration. The steps should be implemented through a coherent set of tasks in which those tasks may address one or more steps and generally focus on external aspects related to resources, data, technologies and data interoperability. It also breaks down into specific recommendations, good practices and case studies from the NSO, inspired by the structure of GSGF Europe, to facilitate the sequence implementation of the steps.

II.3. STATISTICAL-GEOSPATIAL INFRASTRUCTURES

Statistical-geospatial infrastructures operate within and are connected to a data management environment as the technical centre streamlining data integration processes (data and workflows) that enable the systematic and consistent production of geospatial statistics. These infrastructures are operational cornerstones of the previous statistical-geospatial frameworks, supporting data and information lifecycle and outlining the interactions between statistical processes using geospatial content and geospatial processes (GEOSTAT 4, 2021a). They involve data and information streams (inputs and outputs), business processes, information systems and data architecture, methods, and all technological, legal and other non-technical components that support statistical-geospatial integration activities throughout statistical production.

The technical infrastructure, based on a point-based geocoding system, is one of the conceptual and methodological cornerstones of statistical-geospatial data integration, which is closely connected to the conceptual development and implementation of the GSGF. The point-based foundation can be seen as a consistent framework built on geospatial reference data needed to produce and disseminate geospatial statistics based on location data, such as geocoded address, building and dwelling registers (GEOSTAT 2, 2017).

This infrastructure plays a key role in mainstreaming the integration of statistical and geospatial data in statistical production, sharing the GSGF Principles 1 and 2 on promoting an organised geospatial infrastructure and geocoding practices at the unit record level in a secure data management environment. It provides the conceptual and technical basis to enhance geospatial data management - preferably in a centralised data reference repository -, to support consistent geocoding approaches in geospatially enabling individual records (microdata) and to enable data aggregation processes using common geographies (GSGF Principle 3). The ability to spatially discriminate and assign individual locations (e.g., dwellings and buildings) to each statistical unit record provides greater flexibility in production and maintenance, increased exploitation of spatial analysis and adaptability to changes over time. Thus, it allows a more straightforward and territorially flexible methodological approach for data aggregation and new geographical variables compared to the traditional area-based (Figure 6) approach with fixed output areas, mostly used in surveys and census operations (GEOSTAT 4, 2021b).

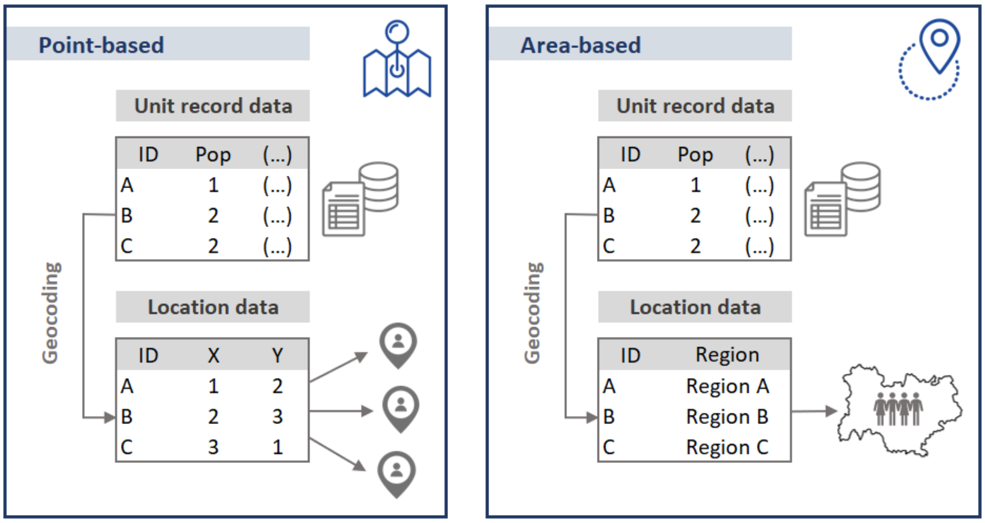

Figure 6. Conceptual layout between point-based and area-based approaches for geocoding infrastructures (source: author, from GEOSTAT 2, 2017).

It is widely acknowledged by the international statistical community and experts in data integration that the point-based foundation is the best option to fully integrate geospatial data into statistical production since it tackles current challenges from NSO and user needs related to relevance, efficiency and timeliness (GEOSTAT 2, 2017). It enables assigning a spatial context to support decision-making and policy design (relevance), expands the range of statistical outputs at geographical and thematic levels at lower cost due to its flexible production setup (efficiency), and provides more timely outputs by linking administrative data to location data (timeliness). The point-based location is recommended by GeoGSBPM due to time stability and territorial flexibility in terms of data aggregation for dissemination purposes (UNECE, 2021).

The main features of the **point-based geocoding infrastructure** in terms of general characteristics, data organisation and implementation approaches are outlined and described as follows. Firstly, the point-based geocoding infrastructure has three **general characteristics/principles**:

1) Use of high-quality point-based location data, regularly updated with time stamps, preferably authoritative location data with high-level maturity to ensure long-term temporal maintenance.

2) Geocoding of statistical unit, and related statistical data, at the unit record level to ensure topological and geometrical accuracy requirements.

3) Use of standardised and consistent identifiers/geocodes to link the statistical unit record with location data (high-precision geocodes) to allow a consistent and hierarchical linking and flexible choice of the location data objects.

The first characteristic/principle addresses the most important precondition for an effective implementation of point-based geocoding infrastructure. It is related to the access to high-quality geospatial reference data (e.g., address or building data) and the requirements for qualifying such data sources as eligible to be used to geocode statistical data at the unit record level, i.e. quality profiles (GEOSTAT 4, 2021b).

Secondly, the **statistical-geospatial data model** comprises three different **data tiers** (Figure 7) making the distinction between geospatial reference data on which the geocoding infrastructure is built and geospatial data needed to create statistical content in the context of statistical production (official statistics).

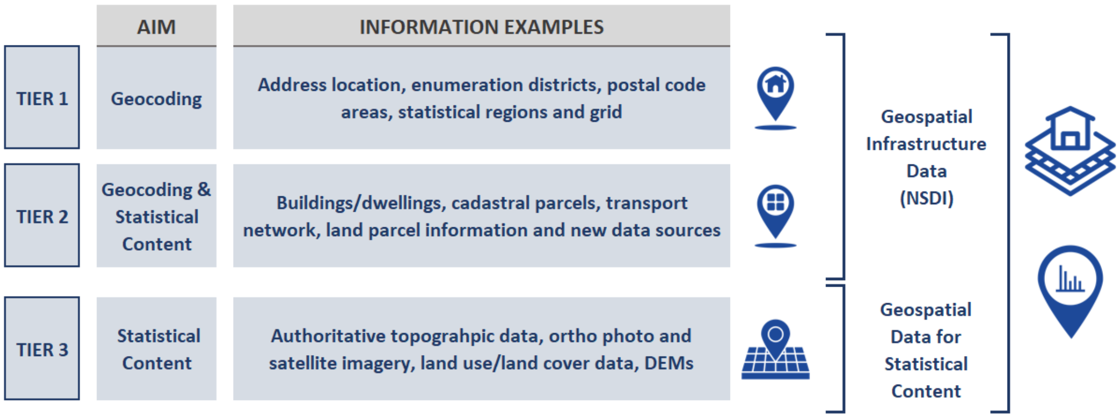

Figure 7. Tiers of information (statistical-geospatial data model) required to support statistical production and produce geospatial statistics (source: author, from GEOSTAT 2, 2017).

The **Tier 1** involves the location data used exclusively to geocode, such as address data, census enumeration districts, postal code areas, statistical and administrative geographies, and grids. The **Tier 2** includes the geospatial data to geocode and to create or enhance statistical content, including building/dwellings data, cadastral parcels, transport network as well as new data sources, such as sensor data. The **Tier 3** comprises the geospatial data for statistical purposes that cannot be used directly to geocode statistical or administrative data, including authoritative datasets on topography and land cover/land use from the NGIA, and other products from EO data. Both Tier 1 and Tier 2 are fundamental to setting up the geospatial infrastructure data and Tier 3 can have a complementary or independent nature regarding Tiers 1 and 2, in which the first one needs to be combined with data from the second one to produce relevant statistical content (e.g., detailed geospatial statistics) throughout statistical production.

Thirdly, the point-based geocoding infrastructure can be implemented through three different **approaches from both organisational and methodological perspectives** (GEOSTAT 2, 2017):

Internal: Both location and statistical data are collected and managed completely within the NSO holding full control of data, processes and quality mechanisms. All the costs of such internal data collection and maintenance routines must be supported by the NSO and identifiers/geocodes may not be consistent and harmonised across different public institutions and data producers.

Mixed: Location data are collected and managed outside the NSO and statistical data are within the statistical organisation. In this approach, NGIA or a consortium encompassing the NGIA is traditionally responsible for the authoritative location data collection and maintenance, sharing costs and promoting institutional cooperation. The challenges may include national geospatial data management policy, NSO commitment and institutional partnerships in the overall data environment.

External: Both location data and statistical data are collected and managed outside the NSO through the NGIA or other providers of geospatial data. Shared costs, flexibility in terms of production, i.e. administrative data with higher frequency release, and data consistency are some of the benefits of this approach. However, there is no or limited direct control in data collection raising quality and confidentiality issues.

Although the geocoding infrastructure is structurally related to the GSGF Principles 1 and 2, it can establish connection points with all statistical production (GSBPM) to ensure its set-up, maintenance and production lifecycle (GEOSTAT 2, 2017). The following GSBPM main phases outline key issues related to the geocoding infrastructure to support the production of standardised geospatial statistics:

The **Specify Needs** phase involves identifying the user needs and defining the requirements for geospatial data (e.g., available datasets in the NSDI and level of geographical granularity of the input data) in order to assess their suitability for statistical purposes. It should also promote the awareness of geospatial statistics to a broader audience (e.g., statistical data linked to location, more detailed statistics or new geographical classifications).

In the **Design** phase, core geospatial data sources should be systematically identified and assessed according to a set of quality criteria (e.g., geographic coverage, spatial accuracy, attribute completeness, etc.), along with the data processing capacity and technical specifications at the organisational or national level. This assessment work will lead to a more appropriate selection of location data based on their application prior geocoding and will provide robust technical conditions to ensure harmonisation for data aggregation processes. Institutional arrangements on data provision and supply chains may also need to be established and the output formats of geospatial statistics need to be described (e.g., specifying if will be released as an open data product).

In the **Build** phase, a methodological description of the geocoding infrastructure should be drafted including production, storage and maintenance practices and quality routines to create a systematic production environment, i.e. processes are repeatable, accountable and streamlined across statistical production. Data architecture and supporting conceptual models as well as information systems and technology specifications should also be extensively described during this phase.

In the **Collect** phase, location data are obtained and managed as well as it is recommended that historical records are maintained and life-stamped in a consistent manner to handle updates and avoid data duplication, i.e. register the same location reference twice, and mismatches between location data objects and statistical unit records. The GeoGSBPM gives preference to geocode the collected data at the point level rather than using area-based geocoding during data collection (UNECE, 2021).

In the **Process** phase, the quality assessment of geospatial data and metadata is carried out - preferably, at the individual record level - through data verification routines to report the errors to and be corrected by the data providers/custodians (hybrid and data broke approaches), i.e. data directly edited by the source. The identifiers (e.g., PID) also need to be checked for consistency and coding system (logical rules) to enable effective data linkage. Alternative approaches to ad hoc data correction can be applied, such as address validation tools, homogenisation of address data or interpolation of address location points (GEOSTAT 2, 2017). The geocoding methods where the statistical unit records are geocoded (secure data management environment) are conducted and the geospatial statistics products begin to be prepared (e.g., data aggregation from dissemination geographies or output geographical variables).

The **Analyse** phase involves managing confidentiality aspects related to a greater disclosure risk (e.g. geographic differencing and clustering) due to the higher geographical detail of the geocoded statistical data during data collection, where geospatial data may allow the identification of individuals by their location. Knowing the detailed geographical location of a person's household significantly reduces the variables required to identify that person individually - directly or indirectly -, and if a population density in a certain area is low, disclosure risk increases (Loonis et al., 2018). This means that the identity and characteristics of a statistical unit, such as a person, household or company, can be found in the released data, including more sensitive variables that can lead to an almost complete identification (e.g., characteristics of an individual in a particular area).

Traditionally, statistical disclosure control applied by the statistical community does not take into account geospatial data and related spatial features, while aiming to reduce the usefulness of the data in exchange for more protection and minimising the risk of disclosure. Thus, the geospatial component adds complexity, uncertainty and requirements to the management of statistical confidentiality and poses new challenges to statistical disclosure control in order to avoid revealing the identity of characteristics of an individual, household or company (GEOSTAT 4, 2021e).

Innovative statistical disclosure control methods (e.g., perturbative or suppressive methods) that can handle geospatial data and embody spatial autocorrelation principles should be applied in this phase to ensure personal or confidential data protection and reduce disclosure risks. New confidentiality methodological approaches dealing with georeferenced (micro)data should also ensure that statistical confidentiality, data protection and privacy of personal data and data providers are preserved, while avoiding identification risks associated with the location of individual data. This is particularly relevant when geospatial data are disseminated at several geographical breakdowns, allowing users to derive information on new and smaller areas from overlapping areas (Costemalle, 2019). This addresses geographic differencing problems where the combination or intersection between two or more geographical areas/variables may lead to confidentiality breaches and potentially increase the risk of disclosure. Along with identity and attribute disclosure, disclosure by (geographic) differencing is one of the most common statistical disclosure risk issues, where the microdata can be accessed from an output table that takes the difference of two tables with geographical variables, such as grids and NUTS (Dékány, 2019).

Furthermore, Statistics Finland is a leading statistical organisation in geocoding practices, streamlining geospatial processes in statistical production and applying the GSGF in its information infrastructure and enterprise architecture. In the past, Statistics Finland acknowledged a core problem related to data integration, particularly in combining statistical and geospatial data to support statistical production, including cases of duplication, overlapping, non-uniform and redundant outputs. The Geospatial Reference Architecture was created as an architecture part of Enterprise Architecture To address these issues. This change in the environment of the data repositories aimed to standardise geospatial processes and technologies throughout statistical production supporting the data processing stage and implementing modular geospatial services. It was endorsed by a strategy map that included a geospatial production centralised model, logical allocation of geospatial data to the DW (IT and business architecture principles), consistent use and delivery of geospatial data according to user needs, and cooperation with other national geospatial data producers.

Statistics Finland's geospatial data are processed and stored only once for all statistical and information processes and centralised in the geospatial data repository, which provides a logical description of how data are organised in the overall architecture. These implementation practices follow some of the recommendations and guidelines of GSGF covered by Principles 1 and 2 related to geocoding and the data management environment in which data are organised according to the GSGF conceptual and information model. In practice, statistical units are automatically linked to location through geospatial PID and all national standards-based location data are centrally stored in the geospatial data repository within a logical data architecture (only once) that considers the geographic relationships between objects (geospatial and non-geospatial).

The architecture model is mainly sustained by the interoperability between data repositories (e.g., organisation data repository, population and social data and energy data, etc.) through shared geospatial services (e.g., address standardisation and editing, geocoding, building service, etc.) and by the information model (buildings, geographical classifications and location links). Generic services have also been implemented to handle repeated processing tasks and to ensure coherent location data across different statistical datasets. It also considers the GSBPM and defines the responsibilities and roles of the stakeholders involved in the operating environment where geospatial experts produce shared geospatial data and statisticians use it to carry out their processes.

The geocoding process plays a fundamental role in the architecture for geospatial data in Statistics Finland. The input data are stored in the data repositories for data collection, including address data from the Finnish Patent and Registration Office, directly collected to complete the address data from the Digital and Population Data Services Agency. The stored data is further allocated to the logical data repositories for data processing, where the address data (alphanumeric) are transferred to the Business Register (Organisation Data Repository) and the associated location data (e.g., georeferenced address point) to the Population and Social Data Repository. Both these data repositories are synchronised. The Business Register uses the geocoding services to link addresses to coordinates and building identifiers (location-building link), and the Geospatial Data Repository retrieves area information for valid addresses using the building service and edits the invalid ones using the address editing service. Thus, the geospatial data repository containing points of buildings and other geographical areas through shared geospatial services transfers geospatial data to the other data domain repositories using the geospatial identifier that is sent by that specific data repository.

This architecture model also aimed to improve interoperability within the NSO before addressing external interoperability issues across the national data ecosystem. In this regard, organisational interoperability issues were firstly prioritised and then in the second stage, technical interoperability issues were addressed through the geospatial data conceptual model scheme towards building a common internal technical and non-technical understanding. The conceptual model places geographical features at the centre of the model enabling both organisational and technical interoperability, including the stakeholders, architecture descriptions, information model and shared services. Thus, the conceptual data model (Geospatial Data Repository) is a key interoperability component that enables geospatial data to be linked with non-geospatial data in a consistent and efficient way, while providing good geocoding quality (e.g., high matching rates).

The Geospatial Reference Architecture of Statistics Finland served as inspiration for the GSGF Europe Reference Architecture as the basic structure and operational architecture describing data, information, services, processes, actors and roles that make up the framework (GEOSTAT 4, 2021c). This reference architecture provides a template solution for building an architecture that aligns technology and IT strategy with business needs and a common vocabulary to design and implement it in a standardised manner. The successful experience from Statistics Finland was a best practice model in describing these aspects to enable the same perspective of the statistical-geospatial environment and common ground on conceptual models (interoperability) by both communities. The architectural description of the statistical and geospatial processes and their relationships in statistical production (e.g., quality feedback and service use) and the mapping of the key national stakeholders and their roles in improving cooperation were also documented based on the experience of Statistics Finland.

Statistics Finland has also been a pioneer statistical organisation regarding statistical-geospatial data integration in the European context, adapting the GSGF into the national operating environment and linking it to the policy framework, namely through a discussion and collaboration network with experts from the public and private sectors. This statistical organisation has been closely working with national stakeholders to draft and implement a roadmap as a strategic plan guiding the progress towards the vision of the Finnish version of the GSGF, including long-term objectives, development activities and governance methods (e.g., organisational interoperability). The roadmap highlights the better use of location data in organisational operations, improved (location) interoperability and development of the statistical-geospatial ecosystem.

Although the concept behind the geocoding infrastructure is relatively easy to understand, its development and implementation is not a straightforward process and presents many challenges and potential constraints (GEOSTAT 4, 2021f). The challenges are related to the main prerequisite to successfully implementing a point-based geocoding infrastructure, namely access to high-quality geospatial reference data. Due to the lack of a nationally consistent framework and a long-term strategy for authoritative geospatial data, which requires investment, standards, regulations and institutional commitment, incomplete geographic coverage of data, heterogeneity of data models, poor data quality and difficult data access from legal or financial reasons are the main constraints. According to a GISCO/GEOSTAT 4 survey (GEOSTAT 4 & Eurostat, 2020), 80% of the 40 surveyed countries have already geocoded population data at the point-based level demonstrating that geocoding practices are becoming mainstream. However, according to the same survey other countries do not have complete coverage of address locations and/or buildings, or do not comply with international standards, for which legal mandates to address such data gaps are recommended to improve data availability at multiple territorial levels.

The GKI, as a location-centric infrastructure that goes beyond SDI and other traditional geodata infrastructures, can be seen as a key conceptual and technological driver to promote statistical geodata integration as it covers advanced data integration capabilities. The GKI focuses on the creation of geospatial knowledge integrating geospatial approaches, data and technologies building on data integration, particularly via automated and real-time processes. The GKI aims to converge the geospatial ecosystem to the broader digital ecosystem to create more valuable and automated knowledge. This future vision will act as a bridge to reduce the geospatial digital divide within the global digital ecosystem, ensuring that geospatial data and capabilities are mainstreamed everywhere across governments, businesses and users.

The GKI is supported by 6 principles and 6 elements with each element breaking into several components considered as expectations or more concrete goals (38 in total) that address both technical and non-technical aspects. The 6 principles are: i) User-centric; ii) Innovation, dynamic and agile; iii) Decentralised; iv) Real-time and predictive; v) Knowledge focus; and vi) Collaborative. The 6 components are: i) Geospatial Dimension of the Data Ecosystem; ii) Foundation Data Infrastructure; iii) Integrated Policy Framework; iv) Industry Ecosystem; v) Partnerships and Collaborations; vi) Applications, Analytics and Modelling. The following table (Table 5) compares typical (N)SDI and GKI through their main differences in terms of capability.

Table 5. Comparison between a typical SDI and a GKI (inspired by Geospatial World Forum & UNSD, 2020).

(N)SDI	GKI
- Data-centric - Data-on-demand - Centralised system - Data environment (data-driven) - Specialised and scattered technology and standards - Manual processes (e.g., data integration and aggregation)	- Analytics-centric (fit for analytics data) - Knowledge-on-demand - Distributed system - Knowledge environment (user-driven) - General and integrated technology and standards

(N)SDI	GKI
- Desktop/web-portal - 2D representation and visualisation - Supply-centric - Static data - Limited data range (i.e. traditional data sources and data types) - Professional users only - Linear and independent - Descriptive analysis	- Automated and sustainable processes (e.g., data update and real-time data integration) - 4D/5D representation and visualisation - User-centric - Dynamic data with multi-source data (e.g., crowdsourced data, mobile data, EO data, IoT, etc.) - Both professional users and non-geospatial experts - Intelligent search - On-the-fly data analysis (real-time processing and sharing) - Predictive analytics, modelling and simulation

GKI will play a central role in the vision of a future geospatial ecosystem with novel digital government processes, modern data licensing, user-centric technology platforms and services from AI/ML methods, advanced API for the exchange of data and services, open data and analytical software and stakeholder collaboration mechanisms (Coetzee et al., 2021). It will also make statistical and geospatial data more interoperable, of good quality and easily accessible by providing more sustainable and automated data integration processes and standardised interfaces (UN-GGIM: Europe, 2021). Thus, GKI developments will contribute to building a shared statistical and geospatial knowledge ecosystem through more integration and standardisation conditions between statistical and geospatial data and processes. According to UN-GGIM: Europe (UN-GGIM: Europe, 2021), both statistical and geospatial organisations need to invest in resources and capacity building in a cross-domain perspective towards common data spaces, focusing on new standards and innovative tools, and establish overarching governance structures.

Some of the IGIF action plans have covered many of the GKI elements and components in their respective countries and organisations. On one hand, the IGIF pathways build the GKI components while considering the emerging trends and drivers in geospatial data management identified by the UN-GGIM. On the other hand, the outcomes of implementing the GKI will support future IGIF development.

II.4. KEY ELEMENTS

Statistical-geospatial data integration can be structured around the following three key elements that are thematically cross-cutting and represent interconnected strategic and actionable areas: i) governance; ii) data, information and technology; and iii) institutional collaboration and capacity. These key elements should be targeted as foundational components when integrating statistical and geospatial data and producing geospatial statistics, covering both technical and non-technical capabilities.

The first key element focuses on high-level guidance and actions related to policy and legal frameworks and how policy agendas, legislation and other overarching (production) frameworks with political and leadership support can contribute to statistical-geospatial data integration. It outlines requirements for effective governance and collaboration activities carried out by the main international, regional and national stakeholders, particularly from the statistical and geospatial communities. Financial issues and governance standards are also covered as their implementation are fundamental non-technical prerequisites for the systematic production of harmonised and standardised geospatial statistics.

The second key element entails the technical capabilities and requirements, best practices and methodological guidance supporting the statistical-geospatial production process. It covers issues related to the technical infrastructure and data/information lifecycle (data collection, management, analysis, visualisation and dissemination), including business processes and services. It also focuses on the development of common technical standards for data harmonisation and interoperability (e.g., data/metadata models) to integrate, manage and use both statistical and geospatial data throughout statistical production.

The third key element covers non-technical issues supporting statistical-geospatial data integration, including institutional collaboration and cooperation, capacity building and capability development. The role of the statistical, geospatial and administrative communities in producing geospatial statistics will be briefly described to demonstrate how collaboration and coordination arrangements between these data communities are important to ensure sustainable data provision and data integration.

While the third key element addresses intra-institutional and inter-institutional issues at the national level (e.g., NSO, NGIA and other national stakeholders), the first key element covers overarching institutional capabilities at the international level, including high-level coordination and synergies in the broader organisational context. In addition, although production frameworks are supported by the technical infrastructure and business processes (second key element) these frameworks are outlined in the first key element, as they require governance action (top-down) to be implemented in the organisations. Technical standards, such as those for data and metadata management, are addressed in the second key element since they support technical capabilities in integrating statistical and geospatial data (e.g., harmonisation and interoperability).

II.4.1. GOVERNANCE

Effective governance in statistical-geospatial data integration can be achieved when global/regional development agendas and key policy and legal frameworks are fully aligned. In addition, both statistical and geospatial communities share the same perspective over an operating environment to produce geospatial statistics, i.e. a common understanding of data and business production processes (GEOSTAT 4, 2021b). This key element is cross-cutting in nature, encompassing the other two key elements, as high-level policy frameworks and governance initiatives (e.g., the 2030 Agenda, the European Data Strategy, etc.) will shape the strategic direction of both technical and non-technical capabilities supporting statistical-geospatial data integration activities.

High-level international and national decisions, strongly supported by political commitment and embedding data integration in policy frameworks and legislation, are the first step in establishing data, information and technology requirements, strengthening institutional collaboration and fostering capacity building and capability development. It covers the strategy and leadership levels, specifically international and national leaders, politicians, policy designers, legislators and high-level stakeholders, that establish future development visions and strategies within and outside the statistical and geospatial communities. These levels are crucial because they are at the forefront of providing financial support and sustainability for information systems, the investments needed to support data management practices, technical infrastructures and maintenance of standards. Political power and decision-makers also ensure that the means of funding are available for data integration activities, such as innovation projects, thematic grants and training initiatives, while assuring that the financial resources are appropriately allocated and effectively spent by the key stakeholders.

Governance coordination is fundamental as it builds synergies and common understanding between international and national institutions (e.g., UN-GGIM, Eurostat and EFGS with NSO and NGIA) to advance statistical-geospatial data integration. High-level stakeholders and key players can develop conceptual and methodological guidelines, establish knowledge exchange platforms, endorse cooperation networks, bring together data producers and users, and identify shared issues and priorities. However, activities to promote statistical-geospatial data integration will only be effective if governance actions between these stakeholders are coordinated and aligned in a transparent and collaborative manner, in particular to avoid misinterpretation of roles and responsibilities, duplication of work and overlapping outcomes.

Policies and laws must formally incorporate data (and metadata) specifications from various domains, not only specifically oriented to the statistical or geospatial data. They need to be more data and technology-driven (e.g., taking advantage of the development of interoperable IT systems) and be designed according to standards, quality and accessibility requirements. The development of standards is usually based on a top-bottom approach, where policy agendas support the creation of standards to address data and technology gaps and user needs as well as legal instruments are defined to promote their general adoption, with a more or less binding character. Actions and recommendations on data governance should also be defined when designing digital strategies under the national and international policy agendas (e.g., HVD) to facilitate integration, harmonisation and standardisation across data domains and business sectors.

Digital governance and policy lifecycle (from design to monitoring and evaluation) must be side by side and strategically aligned with data governance measures, considering the latest developments in digital technologies and user needs. This is particularly relevant for creating the legal conditions for data interoperability (legal interoperability) in geospatial infrastructures (e.g., NSDI) to make data more compatible, usable and shareable from cross-domain and cross-border perspectives, enabling organisations and countries to operate under different policy and legal contexts and data ecosystems (UN-GGIM, 2021). Whether a country has specific laws for one type of data (e.g., statistical or geospatial data) or these laws are part of a larger legal framework with more general implementation guidelines, an intermediate approach can provide appropriate legal solutions. These solutions may combine several laws and regulations that are functionally interconnected and address both data governance and digital/technological issues to ensure a more comprehensive legal framework.

Open data policies and legal frameworks should also embody more inclusive and flexible pieces of laws and regulations, setting common requirements for core data domains and reducing legal restrictions for data access and sharing (Eurostat, 2019a). Legal flexibility is becoming increasingly important as legal acts and regulations need to be more responsive to current and future digital and technological trends (AI is a prime example), rather than being reactive by default, as innovation will not be limited or confined to existing policy and legal frameworks. Therefore, legal frameworks and regulatory arrangements around data need to incorporate not only statistical and geospatial open standards but also internationally agreed, binding and fit-for-purpose open standards from other domains (e.g., IT and web technology) to improve technical interoperability for data access and use. This may require legislative reform at the national and international levels in the future, and an active intervention from politicians, regulators and lawmakers will then be crucial to achieving this legal transition, along with digital transformation.

Furthermore, the increased data geospatialisation and collection of location data combined with statistical data at the unit record level (georeferenced microdata) raises issues of data protection and confidentiality as well as ethics concerns. Policy frameworks and legal instruments need to be carefully assessed in order to establish more-or-less guidelines and binding solutions, which will then be formally put into practice through consistent methodologies and best practices by organisations, given the ever-changing data landscape and technology environment. In addition, balancing the mission to make more data publicly available and accessible, while preserving data privacy and security, and respecting intellectual property rights and ethical values, is a current governance dilemma with many discussions and challenges ahead.

Both international and national policy and legal frameworks need to be designed considering different data types from various domains and thematic applications, while accommodating some degree of flexibility, and at the same time, embodying some mandatory minimum requirements. This trade-off between legal flexibility and enforceable commitment will allow them to accommodate governance and institutional arrangements to foster the necessary changes, while taking into account the specific characteristics of organisations and countries (e.g., institutional environment, etc.).

Strategic alignment at the governance and leadership levels is also required to establish basic data within a national data framework through consistent data/metadata specifications (e.g., dataset profiles) and uniform standards to ensure data quality, integrity and interoperability with less redundancy, inconsistency and duplication. A national data governance model is required because authoritative datasets are typically managed at the intra-institutional level in closed data ecosystems and information silos, without clearly defined data governance roles and responsibilities and effective regulatory mechanisms (e.g., to ensure compliance with standards).

Production frameworks with a strong political and governance endorsement (e.g., GSBPM, IGIF and GSGF) are essential to leverage the different institutional, organisational and legal environments of both statistical and geospatial communities and ensure a consistent operational implementation approach. Internationally adopted governance standards and norms, especially from the statistical and geospatial domains, can be seen as guidance requirements to further apply technical standards in production frameworks at both national and international levels. Thus, governance standards (legal and organisational interoperability) will create the strategic alignment and leadership and corporate conditions to implement technical capabilities at the organisational and national levels (technical and semantic interoperability) across production processes, including in data lifecycle management practices.

II.4.2. DATA, INFORMATION AND TECHNOLOGY

Data, information and technical infrastructure are key technical capabilities to integrate statistical and geospatial data. They encompass data architecture, information infrastructure, IT systems, data and metadata management, business processes (inputs/outputs, workflows and tasks), services, technologies and all other technical capabilities and operational requirements to support statistical-geospatial integration activities and produce geospatial statistics. It also covers technical standards to standardise data and metadata and to harmonise methodologies, services and processes within the same organisation and across organisations (organisational interoperability) to share the same understanding of different statistical-geospatial production issues. All these technical components are designed and operationalised to streamline geospatial data and capabilities during statistical production in a consistent and optimised manner.

Statistical-geospatial production is primarily supported by functional relationships between a national geospatial infrastructure (NSDI), a common geospatial data repository with geocoding capabilities, and a data management environment to geospatially enabling statistical/administrative data and data aggregation using common geographies. In order to ensure that each statistical/administrative unit record is spatially represented accurately and consistently, agreed technical specifications and data and metadata requirements need to be defined together between NSO, NGIA and other data providers. These technical specifications and data and metadata requirements can include an agreed system of geocodes, a standardised structure of unique identifiers, standards, common data descriptions and formats (e.g., time stamps and geographical classifications), GIS software version or coordinate reference system. In addition, if the NGIA already provides good practices on this issue, it is appropriate for statistical organisations to follow them to ensure geospatial data quality (e.g., consistency and completeness) within the national data framework, highlighting the geospatial data repository.

The location data collected - preferably at a point-based level - need to be validated to ensure that they are suitable for geocoding and data integration, i.e. geospatially enabled statistical/administrative microdata, especially those from private producers and non-traditional data sources, i.e. outside the NSDI. This can be achieved through a point-of-entry validation tool embedded in the collection instrument, or through procedures to assess identifiers and check data consistency (logical rules in the data entry procedure), preferably via established routines that allow feedback mechanisms to data providers. Thus, checking the availability and integrity of geospatial data and metadata is fundamental to successfully building an in-house geocoding infrastructure and ensuring that the defined data requirements meet production needs related to the data management environment. These practices on geospatial data validation are already being successfully used by many statistical organisations for population or business registers showing results in terms of quality improvement (GEOSTAT 3, 2019a). Their implementation also demonstrated increased correct data linkage/matching, easier time stamping of data and efficient lifecycle management of records by directly enabling compatibility and harmonisation between datasets and with the NSDI, namely through access to authoritative location data.

A reference framework of common geographies at the national and international levels needs to be established to support data integration from different sources, to improve data aggregation/disaggregation and to facilitate analysis and dissemination of geospatial statistics. It is expected that this framework be continuously adaptable to new data requirements and user needs, while ensuring conditions of accessibility and usability, including geographical versioning for conversion. Besides the traditional administrative and statistical geographies, functional geographies need to be considered, particularly the use of gridded geographies within a global or regional comparable grid system, such as the DGGS that has multiple resolutions. Moreover, a common set of geographies attached to standards contributes to better data quality, consistency, comparability and usability, and simplifies the visualisation, analysis and interpretation of information creating a common geographical denominator for data and metadata integration (GEOSTAT 4, 2021b).

Technical capabilities support both technical and semantic interoperability dimensions. The first one addresses applications and infrastructures/architectures that enable linking different information systems and services, including data integration services, interface specifications, data and metadata standards and communication protocols. The second one covers both semantic and syntactic aspects, including agreed data formats and terminology and means to exchange data and information (e.g., concepts, vocabularies, taxonomies, schema and code lists) in a way everyone understands and manipulates data.

Cross-domain reusable data structures/models and more sophisticated semantic queries coupled with developments in standard web technologies (e.g., RDF, Hypertext Transfer Protocol and Uniform Resource Identifiers) have significantly improved semantic interoperability over the years. In this context, efforts to connect both statistical-geospatial metadata (e.g., SMDX 3.0) and emerging semantic web standards (e.g., RDF) have been technical drivers for continuous improvement data integration, sharing, querying and discovery. These improvements have expanded LOD applications and made data from different sources and types more easily structured, interlinked and published on web pages and more automatically machine-readable.

Standards for documenting metadata for geospatial data and services, especially when received from external sources, need to be used by statistical organisations and aligned with statistical standards. A fully integrated metadata management environment requires conceptual alignment and structural harmonisation between existing statistical and geospatial metadata, along with digital and technology standards. In addition, formal ontologies can improve semantic interoperability through standardised terminology and shared vocabularies between both statistical and geospatial data domains that enabling to be consistently integrated and interpreted by users, machines and application systems, flowing smoothly from the data source to the end user. In a broader scope, multi-level technical interoperability connection points need to be established by aligning standardised technological and services solutions from both statistical and geospatial domains (GSGF) with general interoperability frameworks (e.g., EIF) (Ariza-López et al., 2021).

Technical specifications on data, metadata and technology need to be established when releasing geospatial statistics to facilitate the description, cataloguing, discovery, accessibility and usability of the outputs and increase their reach to the user community. Consideration should be given to open data licensing terms and conditions, and free web services compliance with internationally agreed standards (e.g., OGC) and non-proprietary formats (e.g., API, WMS and WFS) to ensure easy data access and use (open data). These technical solutions will also promote a more service-oriented and user-friendly experience through innovative and customised data exploitation, analysis and visualisation tools, and data integration and extraction services.

Business or application services support some of the statistical and geospatial processes, enabling data and workflows to run smoothly and efficiently at some stages of the statistical production in order to produce geospatial statistics (GEOSTAT 4, 2021b; GEOSTAT 4, 2021c). These modular services can be either geospatial services (dealing with geospatial content) and/or statistical-geospatial services that combine statistical and geospatial services to ensure data standardisation and interoperability, and alignment and monitoring of processes.

The geospatial services include geospatial data loading for acquisition, storage and maintenance in the geospatial data repository, geocoding to converter descriptive location data to direct locations (e.g., addresses to geographic coordinates) and coordinate transformation to convert input geographic coordinates to the reference coordinate system (e.g., ETRS89 and WGS84). Geocoding services play a key role in producing geospatial statistics, preferably in an open and business format. They can automatically be linked to geographic coordinates and a building's unique identifier with the support of address standardisation services. Statistical-geospatial services include data linkage to assign unique codes, usually stored and managed by the organisation's data repository/databases through statistical processes, to the geospatial data object (e.g., a building) by using location as a matching key variable to integrate geometry with statistical data.

OGC's TJS has been a statistical-geospatial service increasingly used as a data joining service and implemented as a proof-of-concept for specific applications since it provides a simple web interface compliant with existing statistical and geospatial standards. It enables statisticians to combinate statistical and geospatial data, and to visualise the outcome in a user-friendly way without any desktop GIS software or geospatial expertise. Lastly, OGC WMS, WMF and API Maps are open service standards used to support GIS-based dissemination solutions that provide easy and flexible visualisation, access and use of geospatial data, while ensuring consistent and interoperable publication and sharing over the web and integration with other different applications (e.g., web, desktop, mobile, etc.).

II.4.3. INSTITUTIONAL COLLABORATION AND CAPACITY

This key element covers issues related to institutional collaboration and cooperation, communication and engagement, organisational interoperability, corporate support, mapping of stakeholders and roles, capacity building and capability development, including skills and training.

Statistical-geospatial operating environments of most countries and regions often run under complex institutional and legal contexts that pose non-technical obstacles to all key stakeholders, especially to statistical and geospatial data providers. Data integration is not only a technical process involving data, information and technology components within statistical production but also requires effective cooperation and collaboration between different stakeholders within and across organisations and data domains (multi-disciplinary and multi-stakeholder approach). Institutional collaboration is one of the cornerstones of statistical-geospatial data integration, as the production of geospatial statistics is strongly supported by the joint work of the statistical and geospatial communities – in particular, the NSO and NGIA - in cooperation with the administrative data community and other data providers. The administrative data community has been playing an increasingly active role in geospatial statistics in ensuring both technical, semantic and organisational interoperability, including aligning their business processes and harmonising data and metadata specifications (UN-GGIM: Europe, 2021).

The three above-mentioned data communities support the production of geospatial statistics at the different stages and action levels by carrying out different roles in the national and international institutional environments and respective data ecosystems (GEOSTAT 4, 2022c). The statistical community supports the geospatial community in the development of geospatial infrastructures (e.g., NSDI), namely by establishing core geospatial datasets for statistical purposes (e.g., authoritative location data) while assessing geospatial data availability and designing geocoding capabilities within those infrastructures. This collaboration at the first stage of the production of geospatial statistics (Specify Needs and Design phases) is extremely important as it avoids the same data being collected several times by different data communities and ensures that data meet the agreed quality requirements for geocoding, analysis and outputs. Additional guidance and consensus-based standards on data and quality requirements are needed when data are provided by the private sector, including Big Data sources, adopting different institutional collaboration strategies.

Both statistical and administrative data communities ensure the best geocoding performance for each statistical/administrative unit record in a timely manner as administrative data sources better meet the ever-increasing user needs and demands (e.g., policy-makers) on timeliness, comparability and relevance of statistical outputs (GEOSTAT 2, 2017). Administrative data custodians typically provide regularly updated administrative data and time-stamped registers and are an emerging alternative data source to traditional surveys and census operations in which close cooperation with the geospatial community is essential to ensure consistent geocoding of administrative records (e.g., standardised geocodes). In this regard, the geospatial community provides methodological guidance and technical support on geocoding practices towards a point-based production model, namely through automated geocoding processes, modern matching/linking procedures and geospatial services. Long-term institutional collaboration between the statistical and geospatial communities within a sound regulatory framework is fundamental to the development and implementation of common geographies for geospatial statistics, ensuring compliance with agreed management prerequisites, comparable data over space and time and easy conversion of historical geographies.

The active involvement of all data communities is also required to ensure and foster multi-dimensional interoperability, involving both technical and non-technical capabilities for standardisation and harmonisation. They need to work together to develop agreed technical guidelines (e.g., data models/formats, metadata specifications, methods, etc.) and standards that meet user needs and align business processes, information models, technologies, services solutions and specific-domain laws (e.g., statistical systems and geospatial data). Therefore, open and regular communication and coordination as well as institutional commitment to data integration activities between NSO and NGIA are needed to overcome silo barriers, reduce duplication of data, work and costs, and reach formal agreements on standards and technology issues. The formal assignment of agreed roles and responsibilities is also required for clear data collection and management obligations, and custodianship mandates, including for provision, access and use of geospatial and administrative data.

Both statistical and geospatial communities share responsibilities at the end of the production process in checking geospatial statistics are released in an accessible and usable format to users. This common goal covers the challenge of managing and preserving data privacy and confidentiality issues, while continuously looking for greater spatial granularity of statistical outputs with higher quality and reliability requirements (e.g., metadata standards for discovery). As geospatial statistics increase disclosure risks and chances of data privacy breaches, it is fundamental to incorporate geospatial confidentiality into statistical confidentiality (e.g., classical statistical non-perturbative/post-tabular methods that consider spatial features and deal with geospatial data) and develop modernised methods to manage confidentiality throughout the statistical production. Thus, both data communities need to make joint efforts and exchange knowledge in developing geospatial-based disclosure control methods and in designing technical solutions to maintain privacy when releasing geospatial statistics at higher disaggregated levels (e.g., grid or small statistical areas). Collaboration should also be established with the administrative data community to safeguard the confidentiality of sensitive (micro)data, while respecting national or international legal obligations on statistical confidentiality and data protection.

Statistical-geospatial dissemination solutions should be explored and tested by both data communities, preferably data-driven, user-oriented and compliant with open standards, following the development of GIS and digital technologies. Statistical-geospatial methodological development and best practices will improve the understanding, analysis and visualisation of geospatial statistics through innovative statistical-geospatial tools and products based on technology trends and user needs. Cooperative capacity building initiatives and knowledge exchange networks with the academia, the research community and the private sector will be decisive in fostering innovation and modernisation, namely through exploratory approaches, common interoperable services and new standards from emerging technologies (GEOSTAT 4, 2021b). An ongoing engagement relationship and close interaction with users - for example, following statistical quality reporting practices - is also important to efficiently identify and respond to their needs and demands in terms of input data requirements (e.g., spatial resolution) and geographies of interest for visualising geospatial statistics.

The statistical and geospatial communities, with the guidance and technical coordination of high-level organisations and key international players, need to invest in and strengthen the capacity development of NSO and national statistical-geospatial capabilities. Statistical organisations need to promote initiatives for building in-house geospatial capacity to ensure that they have the appropriate skills, expertise and knowledge to successfully carry out geospatial-related activities in statistical production (Haldorson & Moström, 2018). Having GIS experts working together with statistical production teams (e.g., from data collection to the dissemination units) facilitates finding work synergies, tackling gaps in information models, IT systems, technologies, methods, standards and frameworks, and building a common understanding to align and monitor both types of business processes and services.

International cooperation and technical assistance between countries and organisations with higher statistical-geospatial maturity levels and other countries and organisations with less experience and body of knowledge are needed to facilitate and leverage the implementation of frameworks, standards, technical guidelines and best practices. The three data communities also need to work together with international standardisation bodies (e.g., ISO) and designers of statistical and geospatial standards (e.g., OGC) to develop and adopt internationally and nationally agreed standards, ensure interoperability of data and services, improve data quality and accessibility (Eurostat, 2019a). Moreover, close cooperation between NSO, administrative data community and international standardisation bodies is increasingly relevant since statistical organisations need to change their business processes by taking advantage of Big Data and developments in semantic web technology to integrate traditional and non-traditional data sources (Harwood & Mayer, 2016). This cooperation at the international level will help to build a common understanding and dialogue around technical and non-technical capabilities of different operating environments, extend the application of a larger range of technologies (e.g., LOD) by multiple types of users, and make statistical-geospatial data integration processes easier and more efficient.

Ultimately, capacity building activities, such as training sessions and professional training programmes on statistical-geospatial data integration will contribute to enhancing the capabilities and skills development of human resources (know-how) in NSO and NGIA, as well as in other key stakeholders from the public sector. In this regard, geospatial literacy initiatives can be developed to be taught to experts from the statistical and administrative communities and geospatial experts can improve skills related to EO data and processing and geospatial programming to apply it in statistical production (PARIS21 and Statistics Sweden, 2021). In addition to raising awareness of the integrative role of geospatial data in official statistics across the institutional environment, strategic alliances and educational partnerships (e.g., internships with the geospatial industry and funded projects) can also improve and endure capability development at the organisational and national levels.

II.5. STRENGTHS, WEAKNESSES, OPPORTUNITIES AND THREATS

The production of geospatial statistics in official statistics is entering a new maturity stage, following an extensive amount of technical and non-technical guidance, methodological resources, supportive materials, case studies and best practices over the years, particularly to support the implementation of GSGF (Europe). In this context, it is necessary to describe and assess the current situation of statistical-geospatial data integration in order to support the design of a future vision and underpinning strategies, to define the priorities and next development steps, and to formulate recommendations and improvement actions.

A comprehensive analysis identified the strengths, weaknesses, opportunities and threats (SWOT) addressing statistical-geospatial data integration, namely outlined by some of the key stakeholders and resulting work from the literature review. On one hand, this SWOT analysis enables reporting the current state and recent developments in the field by assessing internal and external factors and identifying gaps and shortcomings, including overlapping work and outcomes. On the other hand, it helps to raise awareness by identifying benefits and synergies, to define future strategic directions and lines of action, and to develop a better understanding of the constraints and challenges that are hampering the progress of activities related to statistical-geospatial data integration. It provides deeper insights into the value and impact of statistical-geospatial data integration, drawing on the perspectives of both statistical and geospatial communities, addressing technical and non-technical capabilities issues (key elements). It also enriches the body of knowledge and expands the guidance documentation to better assist in the implementation of overarching frameworks and reference standards, and to better inform and support policy-making processes.

Each factor within the SWOT categories (internal and external) will be allocated to the aforementioned key element(s) based on the thematic and actionable scope to have a more comprehensive analysis and extensive evaluation of the progress, issues and obstacles in integrating statistical and geospatial capabilities and align their business processes. The following table (Table 6) summarises the SWOT descriptions of statistical-geospatial data integration and links them to the three Key Elements: (1) Governance; (2) Data, Information and Technology; and (3) Institutional Collaboration and Capacity.

Table 6. SWOT matrix and key elements on statistical-geospatial data integration (source: author).

SWOT		**Key Elements**		
		(1)	**(2)**	**(3)**
S	The existence of a recognised global high-level framework for statistical-geospatial data integration that provides a common and flexible guideline to the international statistical and geospatial communities for producing geospatial statistics in a harmonised and standardised manner.	X		
S	Some countries already have established (point-based) geocoding practices, efficient and sustainable technical infrastructures and well-structured/documented data management environments that could support statistical-geospatial data integration and respective production process, i.e. mainstream the use of geospatial data and services within statistical data/workflows and align business processes.		X	
S	The existence of flexible frameworks for the modernisation of official statistics (e.g., GSBPM and GAMSO) that can be adapted to accommodate geospatial data and capabilities on the existing statistical production, business architectures, conceptual models and organisational structures.	X	X	

SWOT		Key Elements		
		(1)	(2)	(3)
S	Good ensemble of technical guidelines and manuals, best practices for benchmarking and concrete business cases to promote the potential of statistical-geospatial data integration, particularly on new applications and products based on the combination of emerging data sources (e.g., Big Data and EO data) and traditional data sources, including statistical and geospatial.		X	X
S	Several NSO and NGIA have cooperation agreements and collaboration mechanisms (e.g., data sharing agreements and memorandums of understanding) and work closely with statistical/geospatial counterparts suggesting that there is a relatively good cooperation level at the national level despite different forms and types of cooperation between countries.	X		X
S	Several NSO and NGIA have/had participated in regional or international WG and projects related to statistical-geospatial data integration to improve their input data, business processes, outputs and quality management.		X	X
S	Location is an integrative and valuable element for all digital data on society, economy and environment enabling linking and integrating various data sources and types that have location dimension, i.e. assigning people and business to a place or geographical location by combining geospatial data with environmental, socio-economic and other statistical data. Location also enables connecting different systems and data sources and improves the interoperability of datasets across data domains.		X	
S	Statistical-geospatial data integration covers cross-cutting dimensions of sustainable development, supports fundamental data domains (or minimum primary sets of data) at national and international levels and creates meaningful data relationships from multiple themes important for the realisation of economic, social and environmental benefits across local, national, sub-regional, regional and global levels.	X	X	
S	Outputs (i.e. geospatial statistics) with higher geographical granularity and the right spatial/temporal resolution provide a more accurate geographic context of the phenomenon, improve the understanding of social, economic and environmental dynamics and deliver new insights.		X	
S	Increased (spatial) analytical power and research potential (scientific and academic communities) providing more accurate and reliable descriptive, predictive and prescriptive analysis for more targeted		X	

SWOT		Key Elements		
		(1)	(2)	(3)
	responses and actions that need user demands. New possibilities for data representation and spatial visualisation have also arisen from statistical-geospatial data integration (e.g., grid-based output geographies and dynamic mapping).			
S	Many standards and standardised geographies are already being used by some NSO and NGIA, especially by their existing authoritative data.	X	X	
W	Lack of policy and governance robust initiative and clear strategic guidance with critical gaps and absence of connection points or synergies with national statistical/geospatial frameworks needed to achieve sustainable development agenda and other policy frameworks (especially in developing or low-income countries).	X		
W	Disparities and gaps between technological/digital advances around data and the legal and policy frameworks in which new and complex legal and regulatory issues have emerged due to potential risks and opportunities from innovation (data privacy, data sensitivity, ethical concerns, etc.).	X	X	
W	Lack of extensive awareness, engagement and understanding of the value and impact of statistical-geospatial (and cross-domain) data integration, geospatial data and their integrative role in the data ecosystem (e.g., NSDI) and communication initiatives, especially towards geospatial organisations.	X		X
W	Lack or poor organisational interoperability in which most organisations cannot effectively document, integrate and align their business processes and responsibilities to achieve commonly agreed goals on statistical-geospatial data integration, more specifically in terms of requirements related to interoperability and from the user community. Bureaucratic internal systems/processes that are difficult to change lead to a lack of organisational interoperability and modernisation.	X		X
W	Lack or poor understanding and collaboration on both technical and non-technical issues between stakeholders, unclear responsibilities allocated to data custodians alongside poor cooperation between data producers and absence of effective communication with data users to listen to their needs.		X	X
W	Lack or poor semantic and technical interoperability between different data sources and domains due to the lack of standardised unique identifiers, different data collection and storage methods,		X	

SWOT		Key Elements		
		(1)	(2)	(3)
	business production processes incompatible information systems, semantics and data/metadata standards.			
W	Access to data is restricted by legal or financial reasons, particularly concerning geospatial data which turns this type of data too expensive and national laws and regulations restrict free data sharing and impose limitations on public access and use for all users.	X	X	
W	Absence of a reference global framework for, and management of, common geographic areas for analysis and dissemination in an accessible and usable format widely agreed upon and implemented by countries and regions.	X	X	
W	Few countries have their quality frameworks and measures covering both statistical and geospatial aspects within the statistical production process, including quality indicators to assess geospatial data quality as well as confidentiality protection/disclosure control methods which demonstrate limitations in quality issues within statistical-geospatial data integration.	X	X	
W	The absence of a global authority and/or regulatory framework in the geospatial domain underlines no official and consistent legal, technical and organisational system for geospatial data creation and management and quality assurance, as it happens with the statistical domain. For instance, it is common to have a legal mandate regarding official statistics under legal and institutional frameworks built in the different statistical systems whereas no such legal instrument exists for the collection/acquisition of authoritative geospatial data. This type of organisation/regulatory framework could contribute to minimise existing data duplication, manage licensing issues, privacy concerns and legal aspects in the geospatial domain.	X		X
W	Both statistical and geospatial communities have gaps and problems related to legal and regulatory frameworks for global implementation, particularly concerning norms and standards in which a key legal difference between statistical and geospatial data addresses the robust regulation of statistical confidentiality that is not covered by the geospatial domain.	X		
W	Limitations in hardware and software components which compromise performance, the development of an effective technical infrastructure for data and the ability to conduct data integration processes in a sustainable and efficient manner.		X	

SWOT		Key Elements		
		(1)	(2)	(3)
W	Lack of ready-to-use data and insufficient data quality, especially in geospatial data, including constraints on currency, geographical coverage, completeness and spatial accuracy of data (compared to statistical data).		X	
W	Lack of statistical and geospatial capacity and (digital) data literacy at all stages and action levels of policy lifecycle and decision-making.			X
W	Complex and diverse situations and maturity levels regarding statistical-geospatial data integration in countries in which developments in this field are pacing at different speeds at the global and European level.	X	X	X
W	Lack of skills (know-how), human resources and/or training in several themes, including advanced scripting and programming in GIS for increased automation in data production, web mapping tools and services (e.g., API and LOD), interoperability and standards, and use of EO data and technology in official statistics. In addition, the lack of multidisciplinary teams with IT, data science and data quality skills also hamper the effective development of data integration activities.			X
W	The statistical community is more incorporating geospatial data and capabilities in their processes rather than the geospatial community using statistical data, processes and standards within their workflows/processes which demonstrates some disparities in the extent of data integration activities and maturity level.		X	X
O	Greater engagement and communication with policy and decision-makers and to a much broader target audience by promoting the awareness of the benefits and impact of statistical-geospatial data integration and its importance for evidence-based decision-making as a key message to support the 2030 Agenda and other global/regional/national policy frameworks and thematic overarching strategies (e.g., digital and data transition, and rural development, etc.).	X		
O	New policy and regulatory frameworks and strategic drivers related to digital transformation (e.g., digital government initiatives, open data regulations and data governance guidelines in the public sector) can speed up developments in statistical-geospatial data integration, namely in aspects related to data infrastructure and management, data interoperability and sharing, innovation, technical skills and training.	X		

SWOT		Key Elements		
		(1)	(2)	(3)
O	Room for some improvements and modernisation towards a sustainable and shared statistical-geospatial operating environment built under global/regional statistical and geospatial open data infrastructures and common data spaces in which worldwide statistical-based projects requiring geospatial data and capabilities (e.g., SDG and Population and Housing Censuses) can play a transformative role in unifying production, conceptual and legal frameworks.		X	
O	Multi-stakeholder and multi-disciplinary approaches to progress statistical-geospatial data integration well-established by developed countries to be shared with and implemented by governments of developing or low-income countries (e.g., best practices, lessons learned, guidelines and sharing technical solutions and innovation).			X
O	Both statistical and geospatial communities are committed to following the development of technological trends related to interoperability, creating common solutions for data use, reuse and sharing (e.g., linked data) and embodying other data domains and information production frameworks. The establishment of task forces and WG can be an important means for addressing future technical challenges and ensuring progress while recognising the growing need for interoperability to realise efficiencies and cost savings and share it with decision-makers.		X	X
O	New, emerging and non-traditional data sources (e.g., Big Data, VGI, open data, commercial mapping data, cellular data, sensor data, etc.) from technological developments and modern data collection/acquisition and data creation mechanisms (non-human agents and open-source technology). It may lead to increased geospatialisation of information (at lower costs) and to a more heterogeneous data availability and richness which can significantly improve the timeliness, completeness and relevance of future official statistics, more specifically geospatial statistics.		X	
O	Robust legal framework and national mandates to produce and use authoritative data (e.g., location data and fundamental data themes in a standardised form and according to agreed requirements). National leadership for INSPIRE and NSDI development are also important opportunities to foster national geospatial data standardisation and enhance data integration issues.	X	X	X
O	Big Data, EO data innovation, and technological developments from digital transformation and data revolution can be key drivers in		X	

SWOT		Key Elements		
		(1)	(2)	(3)
	fostering/enhancing statistical-geospatial data integration within statistical and geospatial communities and modernising official statistics. The increasing availability of high-resolution satellite imagery can provide more detailed and regular insights to better support evidence-based decision-making.			
O	The statistical community can take advantage of the current and future trends and drivers in geospatial data creation, maintenance and management alongside the digital and technological advancements and innovations in the geospatial industry, such as the GKI, Digital Twins, Geospatial AI, geospatial analytics and user-oriented services based on web technologies.		X	X
O	The statistical community can learn from the experience gained by the geospatial community over the years that has had a perspective more open to the market, business sectors, companies and users focusing mainly on the interoperability of systems and web-oriented and open commercial standards addressing the industrial and technological fields.			X
T	Absence of clear overarching leadership, holistic vision and high-level strategy transcending across different governance and policy frameworks (e.g., 2030 Agenda, Digital Agenda, IGIF) data domains and cross-cutting thematic areas fostering statistical-geospatial data integration.	X		
T	Lack of long-term political willingness and commitment, high-level coordination, binding agreements, institutional mandates and financial support leading to unsustainable funding resources, duplication of work, conflicting overlaps, legal gaps and excessive bureaucratic procedures in statistical-geospatial activities.	X		
T	Failure of institutional actors to address the increasingly complex user needs, data demands and requirements and expertise/human capital exigencies in the multi-level development agendas and overarching policy drivers.	X	X	X
T	No clear guidance and technical developments around data quality assessment, methods and tools for data integration, interoperability issues, standards, web services and tools for data dissemination (e.g., INSPIRE services and proof-of-concepts for tools or services to be tested and evaluated).	X	X	
T	Too much dependency on the development of digital technologies and their use and affordability, particularly by the governments and	X	X	

SWOT		Key Elements		
		(1)	(2)	(3)
	public institutions including NSO and NGIA and the risk of emerging digital disrupts and ethical concerns on data protection and fundamental rights. The fast pace of technological changes will require constant upskilling and investment.			
T	Not meeting the demand for training related to GIS skills, especially in the human resources housed in statistical organisations, and other future skills and capabilities requirements (reskilling and upskilling) considering the acceleration of digital and technological developments (e.g., Data Science, automation, etc.). High staff turnover and loss of geospatial expertise (especially in EO data and processing) within the NSO may be a future risk in producing new geospatial statistics.			X
T	Future challenges related to data collection, access restrictions and both methodological and technological aspects, particularly related to the evolution of data acquisition methods that will add more size and complexity, even with lower costs.		X	
T	Traditional institutional barriers and organisational obstacles for intra and inter-institutional coordination and collaboration remain due to cultural and societal specificities of each country and region (e.g., resistance to change, absence of an experimental/innovation culture or sense of mutual support, etc.). The lack of cross-administrative mechanisms for collaboration also compromises the planning and implementation of data integration activities, namely related to standards.	X		X
T	The value-cost of geospatial statistics decreases based on declining budgets and public funds, increasing demands by users and proliferation of alternative players in the field of official statistics, namely private stakeholders providing data on a more regular and detailed basis, which may compromise the commitment to quality and credibility of the statistical organisations as well as other principles of official statistics. The budgetary difficulties in the public sector and insufficient long-term resourcing to support necessary changes may compromise planning and development of data integration activities and innovation in producing geospatial statistics.	X	X	
T	Lower interaction or unbalanced communication between the statistical and the geospatial communities in which gaps are getting bigger, and the organisational drivers are going towards more	X		X

SWOT		Key Elements		
		(1)	(2)	(3)
	oppositive directions (e.g., no cohesive approach to produce standardised geospatial statistics).			

On the basis of the SWOT matrix analysis, factors related to governance and institutional collaboration and capacity are the main obstacles to greater statistical-geospatial data integration, including issues related to policy and legal alignment and limitations in human and budgetary resources. Factors related to the awareness of the value and benefits of statistical-geospatial, including communication and engagement initiatives between the statistical and geospatial communities and with high-level stakeholders, have a considerable impact on the development of data integration activities. In this regard, it is recognised that a top-bottom action flow is necessary, from long-term political commitment, high-level coordination and clear leadership involving both statistical and geospatial communities, along with governance mechanisms to strengthen the institutional and legal environments across different operating systems and data ecosystems.

Some of these observations have been noted in the results of the UNECE survey on the ‘Integration of Statistical and Geospatial Information’, conducted in 2023 to all NSO and NGIA of the UNECE region (UNECE, 2024a). This survey aimed to assess the current state of ongoing data integration activities, covering both technical and non-technical capabilities, to identify key issues and obstacles and to help set the strategic direction for future projects and actions for developing statistical-geospatial capacity.

Furthermore, interoperability and data harmonisation and standardisation issues are the main technical barriers to the development and use of common standards. Global and national efforts to overcome differences and gaps between statistical, geospatial and other data communities are necessary in terms of production frameworks, technical infrastructures, data formats, IT systems, business processes, services and technology standards. In addition, different governance models, legal frameworks and financial support, as well as a lack of a common understanding due to scarce institutional collaboration and commitment environment hinder the adoption and innovation of standards. Challenges related to data requirements and quality (e.g., confidentiality, integrity and reliability) also have a major impact on data integration activities of NSO and NGIA and on national statistical-geospatial capacity, especially in technical infrastructures and organisational skills, as new data sources emerge and user needs change, driven by evolving digital and technological trends.

III. METHODOLOGY

This chapter describes the case study, considering a set of criteria and considerations, including the main production stages and key processes of the selected statistical operation (CE-SIG project), and depicts the two operational parts that underpin the developed methodology: i) **Production Model**; and ii) **Assessment Matrix**. The methodological application in the case study is presented in a systematic and extensive manner, following the structure of the methodology, in order to facilitate the compilation and interpretation of the results.

III.1. CASE STUDY

The selection process of the case study aimed to demonstrate the contribution of geospatial data, processes, services and capabilities and its integration with statistical data and production pipelines for more standardised and high-quality geospatial statistics. The developed methodology is mainly oriented to statistical organisations and other producers of statistics within the statistical systems since it is based on the statistical business production model, a reference framework for the international statistical community. Nevertheless, it can be applied to other organisations managing statistical and geospatial data and producing geospatial statistics to support policy-making processes, including monitoring and evaluation.

The selection of the case study was based on the following criteria and considerations. The case study should:

- Have a fundamental geospatial dimension in which geospatial data, processes, services and capabilities should be a prevailing part, without overlooking the aspects related to statistical-geospatial integration.
- Encompass, preferably, cross-cutting societal domains, disciplines of application or fields of human activity, such as education, health or environment, to be able to demonstrate how statistical-geospatial integration can contribute to policy-making in multiple areas of action.
- Cover most statistical business production phases and sub-processes (GSBPM) - highlighting design, build, collect, process, analyse and dissemination phases - and the overarching processes/corporate activities by covering geospatial-related and

data integration activities and tasks in the development of the statistical operation. The statistical operation will be the unit of analysis in applying the methodology to support a more pragmatically oriented evaluation exercise and design more objective improvement actions and recommendations for more concrete implementation measures.

- Demonstrate progress and show evidence of improvement and enhancements in statistical-geospatial integration within statistical production, particularly related to the statistical organisation's geospatial capabilities. This requirement addresses efficiency and sustainability in the statistical production lifecycle from a long-term perspective.
- Illustrate how statistical-geospatial integration can provide additional value, enhance the information capacity of statistical data and add new insights into traditional statistical outputs, products and services following the future vision and strategies in modernising official statistics.
- Design, develop and provide innovative solutions addressing data, information and technology aspects either in production or dissemination, including changes in the technical capabilities and components conducted by the statistical organisation to embody statistical-geospatial requirements, guidelines and good practices. Standards for data harmonisation and interoperability will be a relevant criterion, especially when addressing web-based technologies and solutions.
- Encompass institutional cooperation and collaboration mechanisms between the statistical organisation and other public institutions and governmental agencies, including for data provision and access.
- It should not be a stand-alone case but rather partially or fully embodied in the technical infrastructure, reference architecture and statistical production of the statistical organisation to produce and update regular geospatial statistics.
- Illustrate the statistical community and production system of a certain country/region, especially concerning specifications related to the NSS and good practices of statistical-geospatial integration in the respective national operating environment. This criterion aims to assess the state of development and modernisation of the statistical organisation, particularly through adopting both statistical and geospatial frameworks and standards.

- Provide extensive documentation and supporting materials on its design, development, production and outcomes, including detailed and comprehensive information about technical and non-technical aspects related to statistical-geospatial integration.

A large number of activities, projects, studies, use cases, business models and proofs-of-concept on statistical-geospatial data integration have been carried out by different statistical and geospatial organisations from different countries, providing technical and non-technical guidelines, best practices, recommendations and expanding the body of knowledge in the field over the last years. These developments and contributions can be applied in the developed methodology through the key conceptual and methodological aspects outlined in the previous chapter, especially concerning statistical-geospatial frameworks.

III.1.1. SELECTION AND DESCRIPTION

The case study is the Map of Facilities and Services of General Interest[8] (CE-SIG), a new statistical product, developed by SP, under the project on Territorial Cohesion and Social Services of General Interest (2020 Technical Assistance Operation Program[9]). Alongside a set of asymmetry indicators at the local and interregional levels (IASSLOCAL), this new statistical product follows the design and implementation of new statistical products to support the monitoring of the 2021-2027 European Structural and Investment Funds in providing enhanced evidence-based decision-making and promoting territorial cohesion. These two new statistical products will contribute to a more informed policy lifecycle and forecasting by producing more detailed information for accurate and effective policy interventions at the local and sub-regional levels.

The case study selection was driven by advantaged access to internal data and methodological documentation as well as other working materials (e.g., minutes of meetings, testing results, etc.) that supported the design and development of the statistical product. On this subject matter, more direct access to relevant documentation and straightforward communication with the involved staff provided more

[8] 'Carta de Equipamentos e Serviços Sociais de Interesse Geral' translated into Portuguese.
[9] 'Programa Operacional Assistência Técnica' (POAT) translated into Portuguese.

comprehensive and detailed insights regarding statistical production, its processes and related technical and non-technical issues, gaps, challenges and improvements.

The data compilation and (geo)processing carried out in the CE-SIG from a statistical operation perspective will enable the calculation of the territorial asymmetry indicators (e.g., service provision, coverage, accessibility, service areas, etc.) and the production of other statistical outputs on population, business, labour and real state markets and income. Thus, the CE-SIG and IASSLOCAL are operationally interlinked aiming to evaluate the provision level of facilities and services of general interest[10] (e.g., schools, hospitals, museums, etc.) to the population and follow up the progress of asymmetries across the territories which is very useful for territorial-based policy design and implementation. In this regard, the project is aligned at the governance level with the overarching strategy of strengthening public policy planning, monitoring and evaluation.

CE-SIG is a geospatial-based dissemination platform with mapping capabilities (WebGIS) allowing the users to consult sectoral and integrated information about the georeferenced facilities and respective services from different domains (education, health, culture, etc) and both public and private sectors. The dissemination platform provides descriptive variables on the facilities and services (e.g., capacity, number of users, type of management body, institutional nature, etc.) and shares information about their accessibility and demand geographies (service areas and catchment areas, respectively) and associated metrics (e.g., covered surface, dwellings, population). Pop-up information windows for each facility are provided with associated variables related to location and institutional nature as well as other specific variables by facility typology.

The users can search and spatially visualise by two modes addressing two levels of analysis, the facility/service and territory. In the first one, the user can search and view the facilities by sector and type of facility and service, their associated accessibility/demand areas, and the second one by territorial levels (NUTS, municipality

[10] According to the metadata management system of SP, the concept of 'service of general interest' addresses the services aiming 'to meet the essential needs and fundamental rights of citizens, based on the principles of solidarity and equal access, constituting a fundamental element in promoting economic, social, and territorial cohesion, as well as sustainable development' (INE, 2024). It involves services from various sectors such as education, health, culture, civil protection, social services, among others in which accessibility, quality and cost constitute important indicators for its assessment.

and parish) providing an integrated layout of the facilities in the territory and a description on the territorial units related to the provision of facilities/services. Both criteria allow the provision and extraction of descriptive indicators for the selected facility/accessibility geography/territorial unit. The following table (Table 7) summarises the two search and visualisation modes available to users in the CE-SIG dissemination platform.

Table 7. Two search and visualisation modes of the CE-SIG dissemination platform (source: author).

By territory	By facility
● Territorial levels of search: NUTS 1, NUTS 2, NUTS 3, Municipality and Parish ● Integrated visualisation of the facilities in the territory ● Availability and extraction of indicators for the selected territorial level ● Description of the territorial units regarding the facilities/services supply	● Search and visualisation by sector, type of facility and service ● Search and visualisation of the service areas and catchment areas associated with the facility ● Availability, consultation and extraction of the descriptive indicators for the selected facility and for the geography of service areas and catchment areas

The CE-SIG dissemination platform focuses on the spatial visualisation and access to the information resulting from SP's work of collecting, integrating, harmonising and aggregating different datasets from available data sources into a unified and standardised dataset ready for (geo)processing and analysis as well as able to be easily visualised and used by the users. In this regard, institutional collaboration and arrangements with national public institutions from different government sectors and communication between stakeholders were required to guarantee secure and enduring data access and sharing, establish an agreed data model and manage data validation procedures.

The platform aims to design and implement a reference architecture - i.e. a template solution for a particular domain - from the business, data/information, application and technology perspectives that benefit from SP's enterprise architecture and respective technical infrastructure already in place. This consideration ensures that data is regularly updated and new input data from administrative data sources (public institutions from government sectors), statistical operations (NSS programmes and activities) and surveys are integrated into the dataflows and workflows of the in-house statistical production. The data/information architecture of SP is built on four data repositories that constitute the main infrastructure domains: i) Building Geographic Database[11] (BGE); ii) Economic Units Integrated Database[12] (BIUE); iii) National Buildings Database[13] (BNE); and iv) Resident Population Register (BPR), under development from the integration of administrative data sources. Alongside the Agricultural Holdings Database[14] (BEA), the BIUE and BNE constitute the statistical unit management files of SP. This statistical product introduces a new element within the information infrastructure and data management environment, the Facilities and Services Integrated Database[15] (BIES), a master dataset enabling the maintenance and management of the registers on the facilities and services, and the production of periodic reference frames by the first level of facility typology. Alongside the CE-SIG dissemination platform, an external company designed and developed BIES via an outsourcing service.

CE-SIG is aligned with the SP mission and goals on data collection and acquisition mandates, access to and integration of administrative and private data (National Data Infrastructure), institutional collaboration and technical infrastructure underpinned by core datasets supporting statistical production, such as population and buildings/dwellings.

In terms of statistical production, CE-SIG can be broken down into four main stages: i) Data acquisition and validation; ii) Integration into the data management environment and information infrastructure; iii) Spatial analysis and calculation of indicators; and iv) Metadata management and dissemination platform. Some dataflows and workflows between these stages are interactive highlighting the non-linearity of the GSBPM (and GeoGSBPM) for the development of a statistical operation.

[11] 'Base Geográfica de Edifícios' translated into Portuguese.
[12] 'Base Integrada de Unidades Económicas' translated into Portuguese.
[13] 'Base Nacional de Edifícios' translated into Portuguese.
[14] 'Base de Explorações Agrícolas' translated into Portuguese.
[15] 'Base Integrada de Equipamentos e Serviços' translated into Portuguese.

The following Figure (Figure 8) maps the production process (workflow and dataflow) of the CE-SIG project.

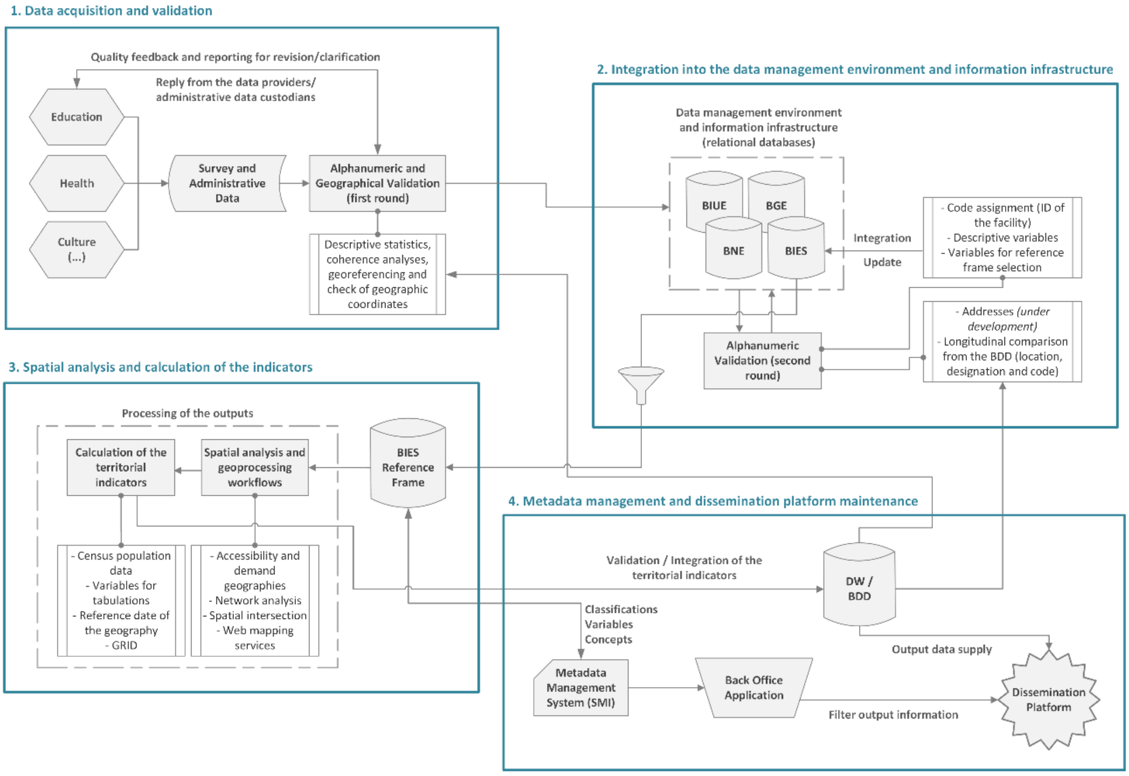

Figure 8. Production process of the CE-SIG project (workflow and dataflow) (source: author).

The four main production stages and key processes supporting the CE-SIG project are described in detail below.

i) Data acquisition and validation: This stage involves institutional collaboration with the data providers/custodians, including the project's presentation and data collection planning when initially checking data availability with the potentially interested public institutions and governmental agencies. Data and metadata features (e.g., data and metadata mode and structures), requirements (e.g., typologies of the facilities and services), technical specifications (e.g., common variables such as the physical address of the facility) and access and licensing conditions are discussed and defined with the data provider/custodian during this stage. Following the agreed information requests, internal and external data acquisition occurs from survey data collected in statistical operations (conducted by SP and under the NSS) and administrative registers/files sent by multi-sectoral public institutions and governmental agencies. After the Information Infrastructure Unit loads the received data, such input data is initially validated concerning codes and formats, i.e. checking whether data or data models meet the required specifications, in which data assessment and validation procedures are carried out in two steps:

- Firstly, the Data Collection and Management Unit identifies potential syntactic errors and inconsistencies regarding alphanumeric content against pre-defined formatting rules applied to the Dissemination Database[16] (BDD) (e.g., consistency analysis and additional data processing) and defines a harmonised dataset of the facilities with a final structure and set of attributes adjustable to other sources and to be viewed and used by the internal users. This processed dataset from the original uploaded data has a consolidated table attached by default with indicators containing the services provided in the facilities. The Data Collection and Management Unit communicates with the Information Infrastructure Unit to create and manage the versioning of the processed datasets to ensure the imputation and update processes for the various periods (e.g., reference years from the survey frequency and administrative data release). The dataflow and workflow will produce the first master output. This output already contains a unique code sequentially assigned to each facility and service by the Data Collection and Management Unit.
- Secondly, the Geo-Information Unit uses the first master output to georeference the statistical unit record (facility) based on the reported x and y coordinates, according to the in-house Georeferencing Protocol, and uses the original address register as auxiliary data to support the validation of the location information (reserve geocoding). This protocol is a guidance document outlining technical specifications and general recommendations for consistent georeferencing, validation conditions and correcting criteria on coordinates and address points to ensure accurate and standardised georeferenced data for each statistical unit record and a common geospatial

[16] 'Banco de dados de difusão (BDD)' translated into Portuguese.

reference frame (of buildings). The Geo-Information also identifies if the facility already exists as a building point in the BGE (if not, a request for creating a new one is made) and geocodes with the territorial units' references (NUTS classification, administrative units and grid) from the multiple historical series through a series of geospatial analysis, including a point-in-polygon operation. BGE represents the location of the building (as a statistical unit) through geographic location attributes (coordinates and addresses) and descriptive attributes (compliant with the INSPIRE Theme 2, Annex III) and is a dataset that supports the collection of georeferenced census data. The dataflow and workflow will produce the second master output with the same number of records as the first one and additional variables from the geospatial analysis addressing georeferenced (point) data editing and changes in terms of location. In this regard, this enriched output also identifies the records addressing a new facility code, an existing facility code without location changes or an existing facility code with location changes.

The Data Collection and Management Unit uses the georeferencing and geographic validation output (second master output) to conduct a comparative analysis regarding the administrative units from the initially provided data and the recent version of the geographies to detect possible changes and differences between the data supplied by the source and the georeferenced data. The results of this analysis are embodied in a standalone report to be used as a feedback tool to the data providers.

ii) Integration into the data management environment and information infrastructure: This stage uses the second master output to certify the address data for introduction in the BNE by assigning the building code to the new facilities (not existing in the BGE and BNE). BNE establishes a 1:1 relationship with the BGE at the first level, the geospatial data repository of the BNE, through the building code (geospatial object ID) in which BGE contains the ID and (points) geometries of buildings and BNE a set of general variables and specific attributes. In addition, BNE and BGE also can have a 1:n relationship since BNE also contains descriptive and extensive information at the dwelling level in which one building can have more than one dwelling. BNE is an evolutionary version of the National Dwellings File[17] (FNA) addressing an integrated file of buildings and dwellings in which the buildings ultimately contain the register of all respective dwellings, regardless of whether they are intended exclusively for housing or other residential purposes. Whereas the FNA - created from the 2011 Census and updated from administrative and survey data and the 2021 Census - only covers residential dwellings (primary or secondary housing, vacant, demolished, classic, non-classic, etc.), BNE includes both residential and non-residential buildings/dwellings, such as companies and (CE-SIG) facilities.

The Information Infrastructure Unit carries out this workflow by assessing the possible entering of data based on the previous identification of new buildings and the facility status in the previous period (active or non-active) and executing all necessary updates in BNE. Consequently, the results from the previous analysis are used to check data entering or updating records in the BIES. The facilities and services are identified by their unique code (ID) and are associated with a set of variables managed in the corresponding application system in which the facility is related to other data repositories that already operate in the information infrastructure of SP (e.g., BNE and BIUE). BIES also provides annual images of the facilities and services that will be created to support the calculation of indicators and the CE-SIG dissemination platform. This database is also connected to BGE through the building code as a matching key in a conceptual and functional relationship of one to many, i.e. one building can have one or more facilities/services, enabling the extraction of the descriptive variables taking into account the buildings associated with the facility/service.

The Data Collection and Management Unit consolidates and enriches the data to be uploaded in BIES and next a more recent version will be created and records added/updated (e.g., new location) by the Information Infrastructure Unit. A distinction between the insertion of a new facility/service and the update of an already registered facility/service is required to be carried out outside the BIES, namely through geographical validation in a GIS environment (geographic coordinates), to avoid duplication of spatial objects and store location only once.

---

[17] 'Ficheiro Nacional de Alojamentos' translated into Portuguese.

The third master output is produced with the same number of records as the second one but with a new updated structure considering the data reference period and original code from the source. Afterwards, the Information Infrastructure Unit ensures that all updates are performed in BIES and BIUE within the data management environment and technical infrastructure. Moreover, the third master output provides the conditions for the next geoprocessing workflows and calculation of indicators.

iii) Spatial analysis and calculation of the indicators: This stage encompasses a series of geoprocessing workflows and spatial analysis techniques, specifically network/routing analysis to generate the service areas, catchment areas and origin-destination matrix and support the calculation of accessibility and territorial asymmetry indicators (IASSLOCAL). The processes and activities are shared between the Units of Geo-Information and the Territorial Statistics and are mostly performed through database-stored procedures and automatic geoprocessing workflows (Python scripts and Arcpy library).

The reference frame of the facilities (facilities and services active in the most recent period) constitutes the input data for the spatial analysis and is extracted/updated from BGE (georeferenced points of buildings) alongside the attached BIES attributes (e.g., information about users of certain services). BIES provides general variables that will enable associating the outputs from the geoprocessing workflows and spatial analysis with the respective reference frame (most recent year) based on the level and designation of the type of facility and/or service. Hence, the relationship between BGE and BIES is important since all georeferenced points of facilities of the continuous reference periods and typologies are stored in one single geospatial dataset and this task retrieves the right facilities (target population) to be spatially analysed.

The various service areas and catchment areas for each facility of the reference frame are generated and the output geographies are stored in the in-house geospatial data repository. These output geographies are: i) service areas by Euclidian distance (cumulative rings computed using physical distance buffering operation); ii) service areas by walking time-distance; iii) service areas by car time-distance; iv) potential catchment areas; v) effective catchment areas; and vi) normative catchment areas. The accessibility geographies to the facilities and respective services spatially address the time range (isochrones) or the physical access distance to those facilities/services. In contrast, the demand geographies spatially illustrate the potential, effective and normative (administrative) demand for the facilities/services. The following figure (Figure 9) outlines the geographies of accessibility that will support the calculation of indicators and constitute spatial objects to be visualised in the CE-SIG dissemination platform.

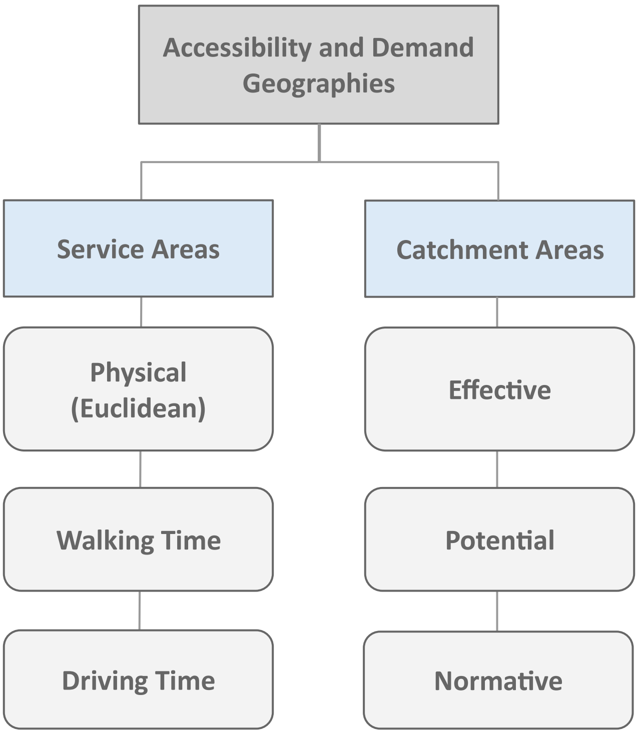

Figure 9. Accessibility and demand geographies in the context of CE-SIG (source: author).

The service areas by walking and car time-distance and the potential and effective catchment areas are computed based on the authoritative street data and navigation/routing network of the ArcGIS Street Map Premium stored at the local server in the Geographical Information Infrastructure (IIG), the SDI of SP.

The geoprocessing workflow produces two main outputs: i) a table with a list of the service/catchment areas with the associated residential building codes (BGE, 2021 Census) that are located within the respective polygons (spatial join operation with the point locations of the buildings); and ii) a table with a list of the service/catchment areas with the surface (Km^2) per distance and time ranges. Each residential building has the parish code associated enabling calculating the indicators by polygon/parish. In addition, each service/catchment area was assigned a unique identifier (made up of the facility code) depending on the type of output geography and respective distance/time range (measured in km and minutes). Data from the origin-destination matrix will be used to calculate indicators, such as average distance and time to the nearest facilities in each parish and municipality.

Afterwards, the Information Infrastructure Unit is responsible for calculating the indicators using the previously mentioned tables and storing the output tables with the results in the DW. The computation of the indicators is a semi-automatic alphanumeric processing workflow operationalised for each reference period and facility and is connected to the geoprocessing workflow by a database wherein the preceding outputs are stored. Although the number of individuals is stored in the data model of BGE, the building code is the only necessary variable to extract the population data (2021 Census) from the DW for calculating some indicators, such as resident population (nº) by service area and population density (nº/Km^2) by catchment area.

In total 336 indicators are calculated scattered by three different thematic types that constitute the unit of analysis: i) area (150 indicators related to the accessibility and demand geographies in terms of the main census variables and surface); ii) facility (48 indicators related to the characterisation of the facilities); and iii) territory (138 indicators related to geographic location, proximity criteria and according to the dissemination geographies, i.e. territorial units such as NUTS and municipalities). All types of indicators are attached to the typologies of the facilities, including non-tertiary education (e.g., primary and secondary schools), tertiary education (e.g., universities), hospitals, fire brigades and museums. The calculated indicators are also attached to the nominal variables of the facilities and respective services, such as the service typology and institutional nature.

Most indicators used geospatial data (georeferenced points of facilities and for statistical-geospatial data integration) and/or spatial analysis (e.g., network analysis for the accessibility and demand geographies) for their calculation within geoprocessing workflows, emphasising their territorial dimension and the production of more geographically detailed statistical outputs. Such geospatial-based indicators that constitute geospatial statistics represent around 93% of the total of calculated indicators. Moreover, a significant part of the indicators used both survey (census data) and administrative data provided by the statistical operations and external sources (i.e. surveyed entities and administrative data custodians), respectively.

The set of indicators with the ‘area’ and ‘territory’ types outline the use of geospatial data and/or spatial analysis to operationalise such indicators. Additionally, the indicators at the facility level underline the use of both survey and administrative - 21 in total -, namely for the indicators that characterise the facility depending on the service typology (descriptive variables). Examples of indicators from the ‘facility’ type that did not use either geospatial data or spatial analysis and only used survey/administrative data are the beds (No.) by hospital, goods of museums (No.) by museum, vacancies at tertiary education (No.) by tertiary education institution and non-teaching staff in non-tertiary education (No.) by non-tertiary education institution. Nevertheless, some indicators at the facility level used geospatial data and/or spatial analysis since they calculate metrics about median time and the resident population living in a certain time range (from the service areas) by foot and car. Parallel indicators are defined in the ‘territory’ type wherein the georeferenced point-based data of the facilities are used as input for their calculation - 138 in total - and georeferencing is considered as a criterion for the geographic location and registered as a conceptual variable in SMI. Internments in hospitals per 1000 inhabitants and surgery rooms in hospitals are examples of these types of indicators. Within this group of indicators, more specifically in the education sector, some of them use census population data to assign the correct school population according to the education level (age groups) of the facility and respective service typology). Hence, the databases containing the census population at the building level (BNE and BGE) were used to calculate these indicators and extract the records based on the defined age categories: i) 3 to 5 years old for pre-primary education; ii) 6 to 14 years old for primary and lower secondary education; iii) 15 to 17 years old for upper secondary education; and iv) 18 to 22 years old for tertiary education.

The following table (Table 8) outlines some examples of indicators, for each type of indicator and service typology, to be displayed in the CE-SIG dissemination platform, including additional examples of the indicators from the 'facility' type that exclusively used survey/administrative data as input, as previously mentioned.

Table 8. Examples of CE-SIG territorial indicators by type of indicator and service typology (source: author).

Type of Indicator	Typology	Designation
Area	Non-tertiary Education	Resident population in the service area of the facility (No.) by non-tertiary education institution and distance range (Euclidian/physical distance); Annual
		Resident population in the effective catchment area of the facility with pre-primary education (No.) by non-tertiary education institution; Annual
	Tertiary Education	Resident population in the service area of the facility (No.) by tertiary education institution and walking time range; Annual
		Population density in the service area of the facility (No./Km^2) by tertiary education institution and time range by car; Annual
	Hospitals	Surface in the service area of the facility (Km^2) by hospital and distance range (Euclidian/physical distance); Annual
		Buildings in the service of the facility (No.) by hospital and time range by car; Annual
	Fire Brigades	Housing units (dwellings) in the service area of the facility (No.) by fire brigade and distance range (Euclidian/physical distance); Annual
		Surface in the potential catchment area of the facility (Km^2) by fire brigade; Annual
	Museums	Resident population in the potential catchment area of the facility (No.) by museum; Annual
		Private households in the potential catchment area of the facility (No.) by museum; Annual
Facility	Non-tertiary Education	Teachers in non-tertiary education (No.) by non-tertiary education institutions; Annual
		Resident population in the service area of the facility within 30 minutes by car (No.) by non-tertiary education institution; Annual

Type of Indicator	Typology	Designation
	Tertiary Education	Students enrolled in tertiary education (No.) by tertiary education institution; Annual
		Graduates of tertiary education (No.) by tertiary education institution; Annual
	Hospitals	Medical doctors (No.) by hospital; Annual
		Median access time by car of the resident population in the potential catchment area of the facility (Minutes) by hospital; Annual
	Fire Brigades	Firemen (No.) by fire brigade and type of labour contract; Annual
		Resident population in the service area of the facility within 60 minutes by car (No.) by fire brigade; Annual
	Museums	Visitors (No.) of museums by museum; Annual
		Resident population in the service area within 15 minutes on foot (No.) by museum; Annual
Territory	Non-tertiary Education	Non-teaching staff in non-tertiary education (No.) by geographic location (georeferencing) (CAOP 2020, NUTS 2024) and institutional nature; Annual
		Proportion of resident population aged between 6 and 14 years old whose proximity to an institution with primary and lower secondary education is less than 30 minutes by car (%) by geographic location (georeferencing) (CAOP 2020 - NUTS 2024); Annual
	Tertiary Education	Median access time by foot of the resident population aged 18 or more to the nearest tertiary education institution (Minutes) by geographic location (georeferencing) (CAOP 2020, NUTS 2013); Annual
		Proportion of the resident population aged between 18 and 22 years old whose proximity to a tertiary education institution is less than 60 minutes by car (%) by geographic location (georeferencing) (CAOP 2020, NUTS 2013); Annual
	Hospitals	Beds (No.) of hospitals by geographic location (georeferencing) (CAOP 2020, NUTS 2013); Annual
		Proportion of the resident population whose proximity to a hospital with emergency service is less than 30 minutes by car (%) by geographic location (georeferencing) (CAOP 2020, NUTS 2013); Annual

Type of Indicator	Typology	Designation
	Fire Brigades	Proportion of the resident population whose proximity to a fire brigade is less than 30 minutes on foot (%) by geographic location (georeferencing) (CAOP 2020, NUTS 2013); Annual
		Median access time by foot of the resident population to the nearest fire brigade (Minutes) by geographic location (georeferencing) (CAOP 2020, NUTS 2013); Annual
	Museums	Goods of museums (No.) by geographic location (georeferencing) (CAOP 2020, NUTS 2024); Annual
		Median access time by foot of the resident population to the nearest museum (Minutes) by geographic location (georeferencing) (CAOP 2020, NUTS 2024); Annual

The Territorial Statistics Unit validates the results of the indicators calculated by the Information Infrastructure Unit through additional validation procedures, some based on geoprocessing workflows (e.g., indicators related to service areas and based on census data) and the remaining in a database management environment (Oracle). The output tables containing the validated results of the indicators are going to the production environment (output systems where data is ready to be used for dissemination) via a production table that will feed the CE-SIG back-office application. The CE-SIG back-office application was also designed and developed externally by a company via an outsourcing service and constitutes a management tool to edit and select the information that will be displayed in the CE-SIG dissemination platform and viewed by the users. This management tool will be described in more detail in the next stage.

iv) Metadata management and dissemination platform maintenance: At this stage, the geospatial data and indicators supporting the CE-SIG dissemination platform are ready to be used as input to feed the output systems and some components related to the dissemination platform are prepared, updated and finalised. During this stage, several processes and activities are simultaneously and interactively conducted by several Units of SP:

- The Methodology Unit applies disclosure control to ensure that output data from the indicators to be disseminated do not breach the appropriate rules on confidentiality, privacy and protection of personal data according to the national and international policy and legal framework in force regarding statistical confidentiality (INE, 2022). Such policy and legal framework covers Law No.22/2008, of 13 May, which establishes the principles, rules and structure of the NSS and in particular the Principle of Statistical Confidentiality (Article 6) regarding Statistical Confidentiality, Law No.22/2008, of 13 May from the Regulation (EC) 223/2009 of 11 March (article 20 and following), amended by Regulation 2015/759 of 29 April, establishing the legal framework for the development, production and dissemination of European Statistics, and Regulation (EU) 557/2013 of 17 June concerning access to confidential data for scientific purposes. The Statistical Confidentiality Policy, established within the NSS context, is formalised as the public commitment by SP, as the central national body responsible for statistical activity coordination and development, to follow the Principle of Statistical Confidentiality in its mission to produce and disseminate independent and impartial official statistics. The national statistical legislation recognises this principle to ensure that statistical confidentiality is guaranteed in law and regulatory measures are established for statistical confidentiality and data protection within the NSS (as happens in the ESS). Moreover, this principle acknowledges all individual data (microdata on natural and legal persons) collected for statistical purposes as confidential and may not be released to users and third parties. In addition, SP should meet the legal requirements regarding the protection of personal data and privacy outlined in the GDPR - Regulation (EU) 2016/679, of 27 April - as well as other legislation and regulatory instruments in which statistical confidentiality is mentioned, such as the Directive (EU) 2019/1024, of 20 June, on open data and the reuse of public sector information. To comply with these legal obligations, SP incorporates a set of technical and organisational practices as well as methods of statistical disclosure control in its statistical production process to avoid and minimise the risks of privacy breaches and identification and ensure the confidentiality, security and integrity of data. According to international best practices, SP applies a series of methods and procedures for protecting and preserving statistical confidentiality in each production cycle of statistical operations. These methods and procedures embody information security requirements (defined in the in-house Information Security Policy and following legal confidentiality and data protection rules) that include both physical security and logical security measures. These measures aim to restrict and prevent direct contact or access to the data and information - especially personal and sensitive data - as well as the supporting architecture and technical infrastructure, including the IT system. Such information security requirements are reviewed, reevaluated and documented when a new statistical operation is carried out, highlighting the importance of the design phase. In general, the preparation of microdata, anonymisation of personal data, aggregation or suppression of data and targeted record swapping are examples of traditional statistical disclosure control methods adopted in SP in the processing and analysis phases to ensure no direct or indirect identification of individual data in the disseminated outputs.

- In the CE-SIG project, statistical disclosure methods are only applied to facilities that have a private institutional nature (for instance, in the health sector, more than 50% of the hospitals are private), although the information is not directly related to the commercial aspects of the data providers and no corporate information for competitiveness reasons will be displayed. This specific confidentiality rule involves that the descriptive information about this type of facilities will not be displayed and public to the users and when a territorial unit (e.g., parish) has less than three private facilities such data (georeferenced points and alphanumeric data) will not be visible in the CE-SIG dissemination platform. Furthermore, the decision to contact the data providers to find out whether it would be possible to disclose the desired information (e.g., staff counts, installed capacity and number of hospitalisations) about these facilities in order to not compromise the main strengths of the project is being implemented.
- The following confidentiality rule was also applied: data may only be released if they refer to three or more statistical units per basic variable

or set of variables (e.g., territorial unit), so that they do not allow any direct or indirect identification of the statistical units, i.e. facility.

  - The outcome is the availability of data files containing the results of the indicators with confidentiality processing completed (based on the policies of confidentiality and protection of personal data in force) to proceed with the preparation of the statistical outputs (geospatial statistics) for dissemination.

- The Geo-Information Unit produces the WMS (OGC) from the geospatial datasets produced in the previous stage (facilities and associated service areas and attachment areas per distance and time ranges) and from the existing geospatial data repository related to common dissemination geographies, such as the administrative units and NUTS classification. These dissemination geographies are associated with a geographical time reference since the boundaries (and respective coding systems) might have changes and updates over time which require a versioning of the historical geographies (integrated into the metadata system). Each time reference has a top-down hierarchical relationship from NUTS1 to the parish in which the Official Administrative Map of Portugal[18] (CAOP) allows identifying the reference date of the geography to be further searched by the external users. In this regard, these geographical time references embodied in the geospatial datasets and other territorial-based alphanumeric data enable managing the temporality of the geographies used for analysis and dissemination and retrieving the right data (one-only time reference) to display in the CE-SIG dissemination platform. This means that these time stamps of geographies are managed in the CE-SIG back-office application to select the correct reference/version for making data available in the platform in which data from only one-time reference can be visualised by the user, and not simultaneously from more than one.
- The indicators and associated metadata are recorded and managed in the in-house Integrated Metadata Management System[19] (SMI), the repository of the concepts, classifications, variables, data collection instruments and

[18] 'Carta Administrativa Oficial de Portugal' translated into Portuguese.
[19] 'Sistema de Metainformação' translated into Portuguese.

methodological documentation of SP (at the NSS and ESS levels). Metadata management concerning the indicators includes validating and updating related concepts such as sector, collective use facility, social services of general interest, types of service areas and catchment areas. Coding tables (categories) that cover designations, abbreviations and appearance order are also managed in SMI, such as classification, hierarchical classification, classification version (e.g., NUTS 2024), period version (e.g., administrative unit code), cumulative version (e.g., a hierarchy of Portugal, NUTS 1, 2 and 3, municipality and parish for the NUTS 2013). SMI also enables the organisation of the facilities and services by typology at multiple levels (there is a facility type, a service typology an association between them at the third level), namely for search purposes in the CE-SIG dissemination platform. After validating and registering the indicators in SMI - through a unique SMI code assigned to each indicator -, the output data are uploaded and updated in the DW, including the BDD. The SMI code for each indicator is used in the DW for identification and searching purposes. In the DW each indicator must have up to six dimensions of which the following two are mandatory: i) temporal (annual or supra-annual); and ii) geographical (mostly cumulative version for territorial indicators and at the levels of facility and service/catchment areas). The remaining dimensions depend on the indicator wherein each dimension addresses a version in SMI. The activities involved in this process aim to ensure the correct association between the (dissemination) indicators and the metadata for them to be prepared to be loaded and made available in the DW as well as the statistical outputs are ready to be displayed in the CE-SIG dissemination platform.

- The Territorial Statistics Unit selects the information to be displayed in the CE-SIG platform through the CE-SIG back-office application regarding the following functionalities: i) reference date of the geography (NUTS and CAOP); ii) active the reference frame (combination of the reference date of the geography, typology of facility and reference period); iii) typologies by sector (health, education, culture and civil protection), facility and service; iv) information boxes (pop-windows information windows containing the indicators and variables) that are grouped into three types: territory (context indicators from census data, sectoral indicators and accessibility and demand indicators), facility (description variables, indicators of the facility and associated accessibility and demands indicators) and areas (accessibility and demand geographies); v) specific variables of the facilities and services (specific to a certain typology of the facility depending on the sector); and vi) general variables of the facilities and services (common to all typologies of facilities, such as institutional nature). This back-office application addresses the GSBPM sub-process 7.1 (Update output systems) to manage data and metadata and make them ready to be disseminated and publicly available to the users, including a final check on formatting and cataloguing (although previously done) regarding the appropriate metadata (SMI).

- It is important to highlight that some IT systems and applications supporting the set-up of the dissemination components already existed or were internally developed by SP, such as login authentication, profile management, collection control paradata, autocomplete and reports for Excel, PDF and XML.
- In addition, other activities supporting the ongoing development and update of the CE-SIG dissemination platform as well as the management of its release can be carried out on a regular or new basis according to the identified user needs and requirements. It includes drafting and updating a user support manual (navigation help) and a technical note to assist the users while exploring the dissemination platform, using its functionalities, managing the mapping capabilities, accessing the indicators and extracting data. The user support manual and technical note are available for download in the CE-SIG dissemination platform and constitute relevant statistical and geospatial literacy documentation to share the methodology, concepts and other aspects related to data and metadata to support data interpretation by the users. Such activities are intended to be carried out through the collaboration between the Territorial Statistics and Dissemination Units.
- Activities concerning user support management can also be conducted to respond to user requests and clarifications about the release of statistical outputs, including information requests on access to microdata for research purposes or any other type of customised information request.

III.2. METHODOLOGY

The literature review from the previous chapters helped benchmark internationally agreed and recognised guidelines and best practices on statistical-geospatial data integration to support the methodology design, including the reference methodology structure and quality assessment checklist.

The methodology is divided into two operational and consecutive parts: i) Production Model; and ii) Assessment Matrix. These parts were inspired and partially built on some principles, conceptual models, methodological guidance, recommendations and other technical and non-technical considerations from the frameworks within the statistical-geospatial operating environment (e.g., GSGF, GSGF Europe, GeoGSBPM, CSPA, etc.). Moreover, these reference capabilities from the statistical-geospatial operating environment constitute a more comprehensive and extensive compilation of best practices and case studies from the statistical and geospatial communities at the national, regional and global levels. The following figure (Figure 10) provides a schematic representation of the developed methodology mapping an overview of the first operational part (Production Model) and a generic identification of the elements supporting the second operational part (Assessment Matrix).

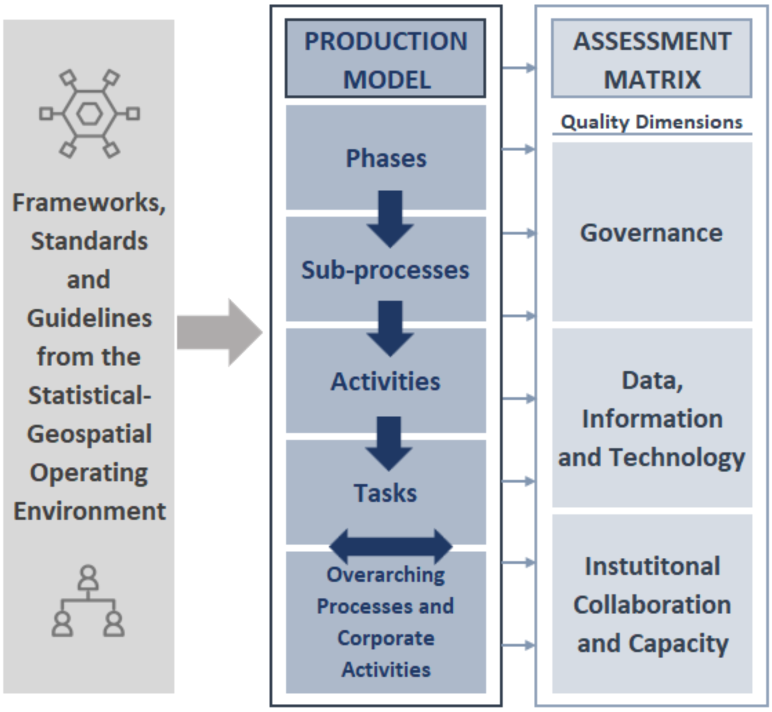

Figure 10. The methodology developed and the two operational parts: Production model and Assessment Matrix (source: author).

Furthermore, both the Production Model and Assessment Matrix do not intend to replace or enhance the existing frameworks - especially the ones well-established, widely implemented and peer-reviewed by the international statistical community - but rather complement them by combining several aspects from the reviewed topics (data, production, stakeholders, frameworks, standards, infrastructures and key elements). This methodology also aims to be used as a self-assessment tool for statistical organisations and other stakeholders from the respective countries to evaluate their capacity and maturity levels in terms of statistical-geospatial integration, mainstream the usage of geospatial data in the statistical production and improve consistent practices to produce geospatial statistics in a systematic manner. Lastly, it intends to help strengthen the international roadmap in statistical-geospatial integration and towards standardised geospatial statistics, particularly for policy-making.

The results from the methodological application will support the definition of generic guidelines, concrete recommendations and enhancement actions that are expected to address the issues identified in the SWOT analysis (subchapter 2.5) and overcome some of the shortcomings identified in the case study.

III.2.1. PRODUCTION MODEL

The Production Model was designed according to the GSBPM's operating structure (phases, sub-processes and overarching processes/corporate activities) to be aligned with the statistical business process and enable logical interlinkages to the assessment matrix. This will ensure consistent and harmonised implementation and evaluation of the respective production model and structuring elements (inputs, processes and outputs) following quality management activities with a strong statistical component that includes quality assessment, feedback and control mechanisms.

GSBPM as an international reference framework within the statistical community describing the statistical production model will enable a more streamlined production of standardised geospatial statistics through consistent input and output requirements, shared production pipelines and management activities and harmonised concepts. Hence, the generalised procedural matrix of GSBPM can constitute a conceptual and methodological foundation to rationalise the production, management and use of geospatial data and services, geospatial processes and statistical processes using geospatial content within the development of each statistical operation. The overarching processes and corporate activities can also align relevant statistical and geospatial standards to enhance interoperability from both communities, promote institutional collaboration and cooperation and develop a common quality approach embodying geospatial components in quality and metadata management.

Similar to the GSPBM, the developed Production Model outlines the same overall structure and organisation of the model breaking down into the phases and sub-processes (SP) that compose the production matrix. However, a varying number of geospatial-related and data integration activities (A) and tasks (T) are included in and assigned to each sub-process (inspired by GeoGSBPM). These activities and tasks also work as more operational building blocks that can be assembled in a consecutive or interactive sequence to support statistical production. This conceptual difference in the overall structure adds a new operational level to the production model by identifying more detailed actions and specific functional requirements supporting the development of a statistical operation producing geospatial statistics.

Only the sub-processes containing geospatial-related and data integration activities (A) and tasks (T) will be mapped in the Production Model whereas some of the activities do not encompass tasks as they already imply concrete actions or address a wide-ranging application. The Production Model comprises the following number of activities (A) and tasks (T) by GSBPM phase and sub-process, as follows the Table 9.

Table 9. Number of the activities and tasks of the Production Model by GSBPM phase and sub-process (source: author).

Stage	GSBPM Phases and Sub-process	Activities	Tasks
Design	**Specific Needs**	**24**	**39**
	1.1 - Identify Needs	3	7
	1.2 - Consult and confirm needs	7	14
	1.3 - Establish output objectives	3	1
	1.4 - Identify concepts	2	3
	1.5 - Check data availability	6	12
	1.6 - Prepare and submit business case	3	2

Stage	GSBPM Phases and Sub-process	Activities	Tasks
Design	**Design**	**20**	**29**
	2.1 - Design outputs	4	8
	2.2 - Design variable descriptions	3	7
	2.3 - Design collection	4	5
	2.4 - Design frame and sample	2	1
	2.5 - Design processing and analysis	4	5
	2.6 - Design production systems and workflow	3	3
	Build	**9**	**9**
	3.1 - Reuse or build collection instruments	3	6
	3.2 - Reuse or build processing and analysis components	2	1
	3.3 - Reuse or build dissemination components	4	2
Operation	**Collect**	**10**	**15**
	4.1 - Create frame and select sample	1	0
	4.2 - Set up collection	3	1
	4.3 - Run collection	3	9
	4.4 - Finalise collection	3	5
	Process	**17**	**19**
	5.1 - Integrate data	5	7
	5.2 - Classify and code	2	2
	5.3 - Review and validate	3	7
	5.4 - Edit and impute	3	5
	5.5 - Derive new variables and units	2	0
	5.7 - Calculate aggregates	2	3
	Analyse	**9**	**2**
	6.1 - Prepare draft outputs	4	0
	6.2 - Validate outputs	2	0
	6.3 - Interpret and explain outputs	2	0
	6.4 - Apply disclosure control	1	2
	Disseminate	**12**	**12**
	7.1 - Update output systems	3	3
	7.2 - Produce dissemination products	3	5
	7.4 - Promote dissemination products	4	3

Stage	GSBPM Phases and Sub-process	Activities	Tasks
	7.5 - Manage user support	2	1
Evaluation	**Evaluate**	**5**	**5**
	8.1 - Gather evaluation inputs	3	4
	8.2 - Conduct evaluation	1	1
	8.3 - Agree an action plan	1	0
Total		**106**	**130**

The sub-processes are systematically identified at the activity (A) and task (T) levels with the description of the activities and tasks involved attached to more comprehensive information and additional notes related to their development in the production cycle. However, not all geospatial-related and data integration activities (A) have a description and additional notes as the explanations provided at the task level support the analysis and understanding of the activity. The additional notes include extensive explanations and clarifications from theoretical and methodological perspectives to support the more effective implementation of the proposed production model. Since these additional notes contain a considerable amount of information, they were included in the production model matrix in Annexes (Annex 1).

This subchapter will only provide the identification of the sub-process alongside a summary description and outline the numbering and designation of the activities (A) and tasks (T). In case of having more descriptive and exhaustive information about a particular activity or task to gain a conceptual and methodological overview and facilitate its implementation, it is appropriate to consult the matrix in Annexes. In addition, the overarching processes and corporate activities are fully outlined and described in this subchapter involving statistical-geospatial collaboration, quality and metadata management and capacity building issues.

Furthermore, the Production Model does not have a fixed operating structure or mandatory sequence wherein the sub-processes, activities and tasks do not need to be performed in a rigorous chronological order throughout the production cycle of the statistical operation. Thus, the developed methodology also emphasises the nonlinearity, flexibility and interactivity dimensions of the production model for geospatial statistics by recognising the functional interdependencies between statistical and geospatial components and their diverse applicability degrees. Thus, a more organic and tailored sequence of the sub-processes, activities and tasks can be conducted according to the inputs, business processes and outputs. Examples of simultaneous sub-processes that do not follow the generally sequential order (may occur in parallel and be interactive) are traditionally carried out under the first phases of the model (Specify Needs, Design, Build and Collect) involving activities and tasks related to the identification of concepts and collection instruments and support. Nevertheless, each geospatial-related and data integration task (T) is identified with a number considering the assigned sub-process and activity and following the general sequential order of the GSBPM phases and sub-processes.

Specify Needs

(SP)1.1 - Identify needs

This sub-process comprises initial research and identification of what type of statistical outputs, products and services are necessary based on user needs, demands and requirements, including a new information request or organisational and institutional changes, such as budget cuts and legislative reform. This sub-process also includes consult and review guidelines and best practices among other statistical organisations and the statistical community at the national and international levels.

(A)1.1.1: Identify and assess current and future trends and drivers in geospatial data management.

> **(T)1.1.1.1:** Identify and describe the state-of-the-play of geospatial data creation, management and dissemination and statistical-geospatial data integration.
>
> **(T)1.1.1.2.** Examine and assess the existing business production process and the ability to accommodate and streamline geospatial components.

(A)1.1.2: Identify and describe the technical needs and requirements for statistical-geospatial data integration activities supporting the production of geospatial statistics.

(T)1.1.2.1: Explore and benchmark different examples of data integration and identify the most appropriate one according to the type of statistical output, product and/or service.

(T)1.1.2.2: Develop a preliminary needs assessment and gap analysis from the geospatial perspective.

(A)1.1.3: Investigate and consider successful best practices, business cases and case studies regarding statistical-geospatial data integration and geospatial statistics at both national and international levels.

(T)1.1.3.1: Consult the supporting documentation and materials from the main frameworks and standards within the statistical-geospatial operating environment.

(T)1.1.3.2: Review useful examples and positive experiences involving statistical-geospatial data integration practices and stories from different countries and organisations with various maturity and capacity levels.

(T)1.1.3.3: Consult both statistical and geospatial communities and map the task forces and working groups having activities related to statistical-geospatial data integration.

(SP)1.2 - Consult and confirm needs

This sub-process aims to consult with internal and external stakeholders and confirm the user needs for the statistical outputs, products and services to have a more comprehensive understanding and deeper insights on how, when, and why to deliver them.

(A)1.2.1: Map and consult national and international key stakeholders according to the needs and requirements regarding statistical-geospatial data integration.

(T)1.2.1.1: Identify the key stakeholders within the NSS and national statistical-geospatial operating environment.

(T)1.2.1.2: Consider non-authoritative/non-traditional data due to the rise and availability of emerging data sources, new data providers, innovative data creation methods, technological developments and changes in the geospatial industry.

(T)1.2.1.3: Identify other key stakeholders at the regional and global levels from contributing to the statistical-geospatial operating environment.

(T)1.2.1.4: Draft communication and engagement programs.

(A)1.2.2: Consult and involve users to help identify and confirm data needs and requirements for collection, production and dissemination purposes.

(T)1.2.2.1: Assess user needs with targeted users from various user communities.

(T)1.2.2.2: Develop user engagement strategies and actions promoting the value of statistical-geospatial data integration and geospatial statistics.

(A)1.2.3: Confirm the internal technical and non-technical resources, capabilities and capacities to integrate geospatial and statistical data and more broadly the capacity and maturity levels of data integration.

(T)1.2.3.1: Check the implementation of a unique identifier system within the technical infrastructure and production system to assess the chance to integrate geospatial data with statistical data.

(T)1.2.3.2: Diagnose the existing IT infrastructure and systems, production architecture and technologies.

(T)1.2.3.3: Check and review guidelines, good practices and methods necessary to address the technical/technological infrastructure needs and requirements.

(T)1.2.3.4: Assess and guarantee the necessary human resources to carry out the planned statistical-geospatial data integration activities.

(A)1.2.4: Assess the geospatial capacity and maturity levels within the statistical organisation.

(T)1.2.4.1: Reflect and identify what kind of geospatial data is needed for the modernisation of official statistics and what statistical purposes by the statistical organisation.

(T)1.2.4.2: Review licenses and protocols with GIS software providers and check potential technical capabilities and solutions that enable carrying out the geospatial dataflow and processes handling geospatial data and services.

(A)1.2.5: Confirm the type of geography (or geographical classifications) best fitted to user needs.

(T)1.2.5.1: Develop preliminary studies and exploratory approaches on geographies.

(T)1.2.5.2: Identify and check the existing geographies used for statistical production and dissemination in the national and international context.

(A)1.2.6: Confirm the smallest/lowest size of the geographical unit while balancing the associated opportunities and threats.

(A)1.2.7: Consult interoperability requirements and cross-cutting needs with internal and external stakeholders.

(SP)1.3 - Establish output objectives

This sub-process identifies the required statistical output objectives to meet the previously identified user needs (sub-process 1.2), considering the suitability between the expected outputs with users and ensuring quality measures (e.g., all users should have equal access to statistical releases at the same time with objectively and impartially and in a legible and accessible manner). Statistical confidentiality, data protection within the institutional environment, legal frameworks, and available resources need to be examined while carrying out this sub-process.

(A)1.3.1: Identify the most suitable output format according to the identified user needs, data requirements and technical/technological capabilities.

(A)1.3.2: Identify the geospatial products and/or services needed to support the statistical outputs (geospatial statistics) and the respective objectives and priorities.

(A)1.3.3: Evaluate data confidentiality methods and disclosure control techniques for high-resolution geospatial statistics.

(T)1.3.3.1: Assess data confidentiality methods and disclosure control techniques considering data protection legislation at the national and international levels (in the case of integrating a supranational regulatory system).

(SP)1.4 - Identify concepts

This sub-process identifies and clears up the required concepts to be measured from the users' perspective, whether they are or are not already aligned with existing statistical (and geospatial) standards. The definition of statistical and other types of concepts and variables to be used as well as their alignment with standards and frameworks should be done further ahead in the next phase (Design), more specifically in sub-process 2.2 (Design variable descriptions).

(A)1.4.1: Identify and review geospatial-related terminology in the scope of statistical-geospatial data integration.

(T)1.4.1.1: Consult and collect an agreed set of existing concepts, terms and definitions (vocabulary) in the scope of statistical-geospatial data integration.

(T)1.4.1.2: Review the identified concepts, terms and definitions (vocabulary) to assess if they are properly updated and if there is a need to be refined accordingly.

(T)1.4.1.3: Identify and evaluate conceptual and vocabulary gaps and potential connection points between existing statistical and geospatial standards and frameworks.

(A)1.4.2: Assess the differences in the classification of territorial units among different public institutions and governmental agencies at the national as well as outside the country and region.

(SP)1.5 - Check data availability

This sub-process checks and reviews current data sources from internal and external data providers that may meet user needs and requirements and be suitable for statistical purposes. It also includes identifying legal restrictions on data use and sharing (e.g., national data protection laws may prohibit or limit some NSO from obtaining administrative data from governmental agencies and public authorities, even when these data are digitally accessible) and gaps in the current legislation and regulatory context (data security and privacy) for easy access and sharing of data. The roles and responsibilities between statistical organisations and data providers (formal data agreements and institutional partnerships, including data and metadata requirements) are also established and assessed in this sub-process as well as the necessary information and technological resources.

(A)1.5.1: Inventorise and assess geospatial data availability and integrability.

> **(T)1.5.1.1:** Identify national geospatial data sources while considering legal, financial and quality issues.
>
> **(T)1.5.1.2:** Consult and compare the various origins of the geospatial data (e.g., from EO data, ground surveys or user-generated from mobile devices and social media applications).
>
> **(T)1.5.1.3:** Assess the capacities for potential NDSI datasets, especially among authoritative geospatial data sources, organisations and stakeholders providing official data at the local, regional and national levels.
>
> **(T)1.5.1.4:** Enlist the relevant geospatial data for statistical purposes, specifically needed for the geocoding infrastructure and needed to create or support the creation of statistical content and processes.

(A)1.5.2: Catalogue core and strategically relevant national and/or regional geospatial themes (and sub-themes) and datasets produced or maintained by national governmental agencies, public institutions and sub-national/regional or local authorities.

> **(T)1.5.2.1:** Search and examine the UN 14 Global Fundamental Geospatial Data Themes, UN-GGIM: Europe Core Data, INSPIRE geospatial data themes and authoritative geospatial datasets in the NSDI.
>
> **(T)1.5.2.2:** Audit and assess the situation of the 14 Global Fundamental Geospatial Data Themes for official statistics in the national context.
>
> **(T)1.5.2.3:** Check the geospatial data availability and key geospatial datasets from authoritative data sources in the European context.

(A)1.5.3: Check legal constraints on data collection, acquisition and use, especially regarding microdata and privately held data.

> **(T)1.5.3.1:** Evaluate the impacts that the existing data policies and national legislation (e.g., national statistical law) may have on conducting the statistical business process and producing the expected statistical outputs.

(A)1.5.4: Define the roles and responsibilities of potential organisations that can be geospatial data providers and custodians for data collection, production, management, maintenance and provision.

> **(T)1.5.4.1:** Prepare, draft, negotiate and sign formal agreements and protocols (e.g., georeferencing), including Memorandum of Understanding (MoU), to clarify mandates and obligations among the organisations of interest.

> **(T)1.5.4.2:** Establish a geospatial data governance committee embodying representatives from all authoritative data providers, administrative data custodians, academia and the private sector.

> **(T)1.5.4.3:** Formulate co-creation, co-management, co-maintenance and provision programmes and partnerships on geospatial data.

(A)1.5.5: Assess the capability to further aggregate geospatially enabled input data (statistical or administrative) according to the output format established for the dissemination product.

> **(T)1.5.5.1:** Define and establish clear roles and responsibilities on data custodianship, management and maintenance providing administrative and non-administrative (e.g., statistical) geographies for analysis and dissemination of geospatial statistics.

(A)1.5.6: Check geospatial access services and dissemination platforms.

(SP)1.6 - Prepare and submit business case

This sub-process records the outcomes and findings of the previous sub-processes in this phase into a business case for implementing the new or modified statistical business process according to certain requirements. In this regard, current or proposed statistical business processes are described in detail to produce current or new/revised statistical outputs, products and services, respectively. The assessment of cost-effectiveness and external constraints is also carried out in this sub-process.

(A)1.6.1: Consult with users on their needs in terms of geographies and internally assess related technical implications (e.g., cost, reliability, quality, etc.).

> **(T)1.6.1.1:** Assess the output format and its suitability for different GIS systems/environments.
>
> **(T)1.6.1.2:** Promote geospatial statistics and the potential of geospatial data, services and capabilities to improve and modernise the statistical production and its potential to address new and emerging user needs.

(A)1.6.2: Perform cost-effectiveness analyses from the outcomes of the previous sub-process, namely in terms of geospatial capacity, technical gaps and legal constraints.

(A)1.6.3: Prepare and draft a work plan to guarantee the geocoding infrastructure's sustainability and its long-term efficient management and maintenance.

Design

(SP)2.1 - Design outputs

This sub-process includes the detailed design of the statistical outputs, products and services to be produced and released, including development components, workflow and technical specifications (systems, tools, etc.) used for dissemination (Disseminate phase).

(A)2.1.1: Check and consult the broader stakeholder groups consuming geospatial data and services, including non-expert and expert users.

> **(T)2.1.1.1:** Take inspiration from dissemination solutions from the geospatial community.
>
> **(T)2.1.1.2:** Analyse the current situation of geospatial technology for dissemination purposes.
>
> **(T)2.1.1.3:** Consider the three main types of (geospatial) data analytics to create fit-for-purpose dissemination components tackling societal needs and promoting

the value of statistical-geospatial data integration for decision-making and policy lifecycle and forecasting.

(T)2.1.1.4: Design outputs contemplating used statistical and geospatial standards for data and services.

(T)2.1.1.5: Specify geospatial statistics output in terms of thematic content and spatial outputs while considering societal needs.

(A)2.1.2: Design outputs considering statistical-geospatial confidentiality issues to ensure statistical outputs are released with confidence and data protection guarantees.

(T)2.1.2.1: Acknowledge existing confidentiality specifications and challenges related to geospatial data, i.e. geocoded data.

(T)2.1.2.2: Check the existence of confidentiality policy or regulation at the national level that embodies geospatial data considerations.

(T)2.1.2.3: Examine existing guidelines and methodological materials to manage statistical-geospatial confidentiality throughout the statistical business process.

(A)2.1.3: Design outputs considering spatial visualisation capabilities.

(A)2.1.4: Consider both main types of geospatial standards: data/information standards for data and metadata models and technology standards for infrastructure and interfaces.

(SP)2.2 - Design variable descriptions

This sub-process establishes the variables to be collected (during collection through collection instruments) and any other variables that will be derived from them later (sub-process 5.5 - Derive new variables and units). In this sub-process, both statistical and geospatial classifications can be used, including geographical classifications for dissemination purposes, and preferably according to existing national and international standards.

(A)2.2.1: Design geographical variables (geographies) for the statistical unit level.

(T)2.2.1.1: Check existing geographical variables (geographies) based on the results from sub-process 1.5 (Check data availability) and ones already used for statistical purposes.

(T)2.2.1.2: Use the point-based location as the basic geospatial variable.

(T)2.2.1.3: Design methods and tools to create and manage unique identifiers or coding for geographical variables.

(T)2.2.1.4: Design geospatial services and applications for building and maintaining geographical variables (geographies), namely census geography and other statistical geographies.

(A)2.2.2: Use point-based location data and adopt the point-based approach for deriving geographical variables.

(A)2.2.3: Use regionally and globally comparable gridded geographies.

(T)2.2.3.1: Take advantage of Discrete Global Grid Systems (DGGS) reference systems.

(T)2.2.3.2: Consider the application of DGGS as a basis for statistical areas and data aggregation.

(T)2.2.3.3: Consider and assess the use of hexagon-based DGGS as an alternative geometry/shape to the traditional square grid cells.

(SP)2.3 - Design collection

This sub-process defines the most appropriate collection instruments and methods according to the type of data collection/acquisition (survey, sample survey, automatic data transfer, etc.), type of statistical unit collection (e.g., person, household, enterprise, etc.) and the available data sources (survey data, administrative registers, geospatial data, Big Data, etc.). The design of the collection instruments and methods depends on the modes and methods of data collection/acquisition and specifications of the data sources (e.g., CAPI for sample surveys, service interfaces for administrative data, or specialised technologies for geospatial data). This sub-process should also consider data and metadata requirements and quality mechanisms - especially when externally collected and processed -, meet user needs and demands (e.g., timeliness, type of geography, geographical granularity, etc.) and follow the outcomes from the previous sub-process (2.3 - Design variable descriptions).

(A)2.3.1: Design components for data collection and storage in a secure, standards-based and effective data management environment.

> **(T)2.3.1.1:** Ensure the quality of the collected and acquired geospatial data.

> **(T)2.3.1.2:** Validate the collected location data and metadata against authoritative data sources or a gazetteer can be a suitable option to add an extra validation step.

(A)2.3.2: Configure GIS tools, geocoding services and other geospatial capabilities for data collection according to the collection method specifications.

> **(T)2.3.2.1:** Consider unexpected and extreme scenarios.

(A)2.3.3: Attach geospatial data to the collected non-geospatial data (e.g., statistical and administrative data).

(A)2.3.4: Assess the need for fieldwork (e.g., field operation for a survey) in collecting and capturing geospatial data that is new, more accurate, and/or updated.

> **(T)2.3.4.1:** Design innovative procedures oriented to non-traditional data collection and acquisition methods and modes (e.g., sensor data, automatic, web-based transfer, multimode inputs, etc.).

> **(T)2.3.4.2:** Make use of open services for geospatial data and metadata collection.

(A)2.3.5: Promote public-private strategic partnerships and collaboration protocols to support data collection and acquisition.

(SP)2.4 - Design frame and sample

This sub-process is conducted when statistical production involves data collection based on sampling (e.g., sample statistical surveys) to identify the population of interest, sampling frame, including the type of geography or geographical classifications used, and to define the most suitable sampling criteria and methodology, namely based on the data sources.

(A)2.4.1: Design geosampling by taking advantage of geospatial data and selected geographical variables (geographies) to support the construction of frame and sample.

(T)2.4.1.1: Test and compare geosampling methods and techniques.

(A)2.4.2: Design geospatial services and capabilities supporting geosampling.

(SP)2.5 - Design processing and analysis

This sub-process designs the statistical methodological components for further processing and analysis (Process and Analyse phases), including dataflows, validation routines, disclosure control methods, design specifications and rules for coding, editing, imputation and matching, that vary based on the method and mode of data collection and data sources.

(A)2.5.1: Design geospatial processing, including dataflow, statistical processes using geospatial content, sharing and standardisation processes and services between both statistical and geospatial processes.

(T)2.5.1.1: Assess data processing and analytical capacity.

(T)2.5.1.2: Design processing and analysis methodologies and operational mechanisms handling geospatial data.

(T)2.5.1.3: Design matching and non-matching strategies in integrating statistical and/or administrative data with geospatial data as the location is used as a key variable for integration from various sources.

(T)2.5.1.4: Design other processing methodologies and geospatial services needed for geospatial data and spatial analysis.

(A)2.5.2: Design methods and tools for data aggregation.

(A)2.5.3: Design statistical-geospatial processing and analysis components taking into account standards agreed upon and extensively used in both statistical and geospatial communities.

(T)2.5.3.1: Use OGC standards, INSPIRE data models and SDMX for enhancing statistical-geospatial data integration, interoperability and accessibility conditions.

(A)2.5.4: Design geospatial analysis for producing more detailed geospatial statistics.

(SP)2.6 - Design production systems and workflow

This sub-process establishes the workflow from data collection to dissemination by mapping all the processes and services and their functional relationships required within the statistical production process and ensuring that they are fully integrated and streamlined under the business/enterprise architecture. A general outline of the dependency and complementarity relationships between statistical and geospatial processes and services should be designed in this sub-process to ensure they work together to provide a consistent production of statistical outputs, products and services (geospatial statistics).

(A)2.6.1: Design the production system and workflow for geospatial components.

(T)2.6.1.1: Describe and document the geospatial processes and activities.

(T)2.6.1.2: Describe and document guidelines, good practices and methods supporting the design of the geospatial production system and workflow.

(A)2.6.2: Design the geospatial production system and workflow to meet new and emerging user needs and demands for geospatial statistics.

(T)2.6.2.1: Adopt a data-driven and service-oriented technical infrastructure supporting the geospatial production system and workflow.

(A)2.6.3: Design business processes, activities and tasks supporting the geospatial production system and workflow.

Build

(SP)3.1 - Reuse or build collection instruments

This sub-process outlines the activities to build or reuse the collection instruments to be used during the Collect phase following the design specifications established in the Design phase and aligned with the internal business architecture. The collection instruments are prepared and tested, including functions, technology assessment and connection to the metadata systems.

(A)3.1.1: Check and consult modern collection instruments.

(T)3.1.1.1: Search and test already implemented and well-established geospatial services supporting data collection and provision from the geospatial community.

(T)3.1.1.2: Ensure the collection instruments embody geocoding tools establishing a direct connection to the metadata system of the statistical organisation.

(T)3.1.1.3: Provide and deliver geospatial datasets supporting data collection and fieldwork as a service by using web standards and GIS enterprise applications.

(A)3.1.2: Build the geocoded sampling frame by taking advantage of location data and geospatial datasets of geographies.

(A)3.1.3: Reuse or build geospatial services used for data collection based on the designed technical infrastructure and the data collection and storage components in the data management environment.

(T)3.1.3.1: Reuse existing collection instruments already used for gathering statistical data or administrative registers, namely through service interfaces.

(T)3.1.3.2: Adapt the data extraction routines according to the previously designed quality techniques and validation tools (e.g., point-of-entry validation) to improve the matching rate.

(T)3.1.3.3: Review existing geospatial services for data collection already used by the geospatial community (especially by the NGIA), administrative data providers and/or other data custodians by consulting the existing catalogue of services.

(SP)3.2 - Reuse or build processing and analysis components

This sub-process outlines the activities to reuse existing components or build new components required for processing and analysing (Process and Analyse phases), according to the established in the Design phase and aligned with the internal business architecture. It includes data integration functions, editing and imputing functions, data and metadata management services and geospatial services.

(A)3.2.1: Search and test already developed and well-established geospatial services used for processing and analysis from the geospatial community.

(A)3.2.2: Reuse or build geospatial processing, including dataflow, statistical processes handling geospatial content, sharing and standardisation processes and services between both statistical and geospatial processes to be carried out during the Process and Analyse phases.

> **(T)3.2.2.1:** Review existing geospatial services for processing and analysis already used by the geospatial community, including the NGIA and the GIS community.

(SP)3.3 - Reuse or build dissemination components

This sub-process outlines the activities to reuse existing components or build new components required for disseminating the statistical outputs, products and services (Disseminate phase), according to the established in the Design phase and aligned with the internal business architecture. Nationally and internationally agreed data and metadata standards (e.g., taxonomy and vocabulary) should be considered in this sub-process as well as user needs and requirements to ensure data is released in a findable, accessible and (re)usable manner.

(A)3.3.1: Assess and balance the options between reusing or building dissemination components taking into account the previously identified technical requirements and gaps, the available resources and capabilities and the user needs and demands.

(A)3.3.2: Embrace well-documented and standard-based services and API solutions while avoiding custom extensions and deprecating obsolete technologies and standards to improve discoverability and accessibility of data.

(A)3.3.3: Include and document geospatial metadata elements (e.g., using common taxonomy and vocabulary) in the dissemination components along with the data products and services.

> **(T)3.3.3.1:** Employ standardised metadata for geospatial data and services in compliance with internationally agreed standards, guidelines and best practices to increase the discoverability and accessibility of outputs for both internal and external users.

(A)3.3.4: Ensure data confidentiality and disclosure control methods that were established in the Design phase when reusing and building dissemination components.

> **(T)3.3.4.1:** Ensure the dissemination components embody technical capabilities considering privacy, confidentiality and legality issues and that the output data and information are released with confidence in which users can easily discover, access and visualise geospatial statistics.

Collect

(SP)4.1 - Create frame and select sample

This sub-process defines the frame and selects the sample from the collection according to the design specifications established in the previous sub-process 2.4 (Design frame and sample).

(A)4.1.1: Use geosampling alongside geographic criteria, spatial analysis techniques and mapping tools to create the frame and select the sample.

(SP)4.2 - Set up collection

This sub-process assures human resources, processes and technology - i.e. technological components of the collection instruments or the monitoring system - are functionally prepared to collect data and metadata, according to the design specifications established in the sub-process 2.3 (Design collection). It could be a very time-consuming sub-process since it includes the strategy, planning and training activities in preparation for the collection, especially when addressing a classic census operation based on survey data or a new statistical business process.

(A)4.2.1: Optimise the data collection strategy using geospatial data, technologies and capabilities.

(A)4.2.2: Test and configure GIS tools embodied in the collection instruments for the specific collection method.

> **(T)4.2.2.1:** Ensure that GIS tools are configured appropriately for the specific collection method while considering non-ordinary scenarios and extreme situations, such as technological breakdown and connectivity issues.

(A)4.2.3: Check the risk of a disclosure breach while data is being collected/transferred from primary and secondary data sources.

(SP)4.3 - Run collection

This sub-process involves the implementation of the collection wherein the different previously designed, reused or built are applied to collect or gather data and associated metadata from the microdata to the aggregated data levels. It can include fieldwork management (e.g., survey data), manual data entry at the point of contact, the initial contact with the providers, any follow-up or reminder actions (e.g., administrative data) and geocoding (e.g., GPS systems, GIS tools and web mapping services).

(A)4.3.1: Conduct geocoding during collection.

(T)4.3.1.1: Collect and document geocoding metadata.

(T)4.3.1.2: Register time stamps for geospatial metadata during collection.

(A)4.3.2: Geocode each collected statistical unit and, preferably at the most detailed level (microdata).

(T)4.3.2.1: Geocode according to a system of unique identifiers within the geospatial infrastructure.

(T)4.3.2.2: Collect and geocode data under statistical confidentiality and data protection guidelines, terms and instructions, especially in the case of microdata.

(A)4.3.3: Validate geospatial data during collection, preferably at the source.

(T)4.3.3.1: Conduct point-of-entry validation checks and other validation mechanisms handling geospatial data previously designed and built (2.3 Design collection and 3.1 Build or reuse collection instruments).

(T)4.3.3.2: Implement data entry procedures to ensure the management of temporality issues, i.e. data currency/timing.

(T)4.3.3.3: Encourage external data providers to provide consistency in terms of location data.

(T)4.3.3.4: Document the inaccuracies in geospatial data that are detected during data collection (e.g., field operation).

(T)4.3.3.5: Monitor geospatial data collection to make fieldwork more efficient.

(SP)4.4 - Finalise collection

This sub-process comprises the loading of the collected data and associated metadata into a fitting data management environment (within the internal business architecture) for further processing. It may include manual or automatic data capture, recognition tools to extract information from paper questionnaires, conversion of the formats of files, encoding the variables received from data providers for statistical purposes or pooling data by merging sample survey data from different collection cycles in one dataset. The metadata and paradata collected can be also analysed to ensure the collection activities have met the established requirements and standards, i.e. whether collected data are accompanied by sufficient and appropriate structural metadata (e.g., underlying concepts and definitions). Documentation activities can be also included, such as versioning and archiving of the software used in the collection instruments and errors detected during collection that require further data changes and updates in the data management environment and information systems.

(A)4.4.1: Verify and validate geospatial inaccuracies and inconsistencies detected during data collection, especially in the case of input statistical/administrative data from fieldwork.

(T)4.4.1.1: Assess the spatial precision (or positional accuracy) of the collected geographic coordinates.

(T)4.4.1.2: Measure the input records that have a unique ID and/or the geographically misclassified survey units detected during a sample survey, i.e. whenever the sample units of a certain are geocoded.

(T)4.4.1.3: Document and report the identified inaccuracies and inconsistencies in geospatial data in a standardised manner.

(T)4.4.1.4: Correct the detected errors and update the input data in the geospatial infrastructure.

(A)4.4.2: Check the geocoding matching rate.

(T)4.4.2.1: Establish a minimum matching rate.

(A)4.4.3: Use geospatial data loading and extracting services.

Process

From the geospatial perspective, this phase is the most demanding and arduous covering quality assessment procedures of the collected geospatial data and data integration between statistical or administrative data and geospatial data through geocoding, matching mechanisms or other data integration methods and techniques.

(SP)5.1 - Integrate data

This sub-process integrates data from one or more sources and combines the results of the previous Collect phase, from internal or external sources (e.g., respondents, administrative entities, private owners and other data providers) and different collection instruments, including statistical (e.g., surveys) and non-statistical data sources (e.g., administrative, secondary and other data sources, such as Big Data). This sub-process includes matching or record linkage routines (at micro and macro levels), data pooling, data fusion (reduction or replacement) and harmonisation procedures to create a set of linked data from different data sources and formats. Data integration can be conducted through modelling approaches or a mixture of these. Procedures differ based on the types of data sources, characteristics of datasets, detail level of data, the system of unique identifiers and the objectives for combining data (e.g., record linkage and statistical matching, etc.). In this sub-process, input data (basic statistics) are compiled and converted to create integrated data (integrated statistics, such as national accounts, in which the quality of the basic statistics and the quality of the integration are equally important for quality management.

(A)5.1.1: Combine geospatial data with statistical data or other non-statistical data.

(T)5.1.1.1: Ensure consistency in the matching process.

(T)5.1.1.2: Document matching methods and techniques.

(T)5.1.1.3: Use data integration services as business services joining geospatial and statistical data.

(A)5.1.2: Use of location as a matching key variable.

(T)5.1.2.1: Apply consistent geocoding and matching mechanisms.

(T)5.1.2.2: Use address linking services.

(A)5.1.3: Conduct geospatial data quality assessment.

(T)5.1.3.1: Implement assessment routines and approaches handling semantic and structural quality elements specific to geospatial data (that are conceptually and methodologically different from statistical data).

(T)5.1.3.2: Assess the uniqueness and consistency of identifiers (e.g., PID).

(A)5.1.4: Convert geospatial data according to national and/or international standards.

(A)5.1.5: Promote potential solutions for machine-to-machine mechanisms for data integration.

(SP)5.2 - Classify and code

This sub-process proceeds to classify and code the input data, including coding routines with different automation levels (manual, interactive, semi-automatic and automatic) to assign numeric and non-numeric codes based on pre-defined statistical and non-statistical concepts, variables and classifications (including geographical classifications). Statistical organisations should support this sub-process through a common repository of concepts, definitions of units and variables, and classifications and compliance with national, regional or international statistical and non-statistical standards (e.g., agreed standards by the geospatial community). This sub-process intends to facilitate data processing, namely new variable and unit derivation (sub-process 5.5 - derive new variables and units) and data aggregation (sub-process 5.7 - calculate aggregates) and ensures data consistency and comparability at national, regional and international levels.

(A)5.2.1: Geocode statistical and/or administrative data.

(T)5.2.1.1: Choose semi-automatic and/or automatic geocoding methods and techniques.

(T)5.2.1.2: Correct geocoded data at the source, if possible.

(A)5.2.2: Use geocoding services.

(SP)5.3 - Review and validate

This sub-process checks and validates input data to identify potential problems, errors, inaccuracies and discrepancies (e.g., outliers, miscoding, duplications and unstandardised data) and can be run iteratively, before and after data integration or after imputed and edited data (sub-process 5.4 - Edit and impute). It comprises review and validation routines for data from any source type using automatic or manual data inspection or editing procedures which can also occur during collection to validate and correct the identified errors directly at the source.

(A)5.3.1: Review and validate the location data.

> **(T)5.3.1.1:** Review and validate the quality of the location data and input geocoded data through quality assessment measures and quality control mechanisms handling geospatial data.
>
> **(T)5.3.1.2:** Review and validate the currency/timing of location data.
>
> **(T)5.3.1.3:** Review and validate the spatial accuracy of point-based geocoded data.

(A)5.3.2: Review and validate the address data.

> **(T)5.3.2.1:** Review and validate the input address registers in the preliminary verification round.

(A)5.3.3: Carry out quality assessment routines and procedures specifically oriented to geospatial data and its technical specifications which may differ from the traditionally applied to statistical/administrative data.

> **(T)5.3.3.1:** Check the topological consistency of geospatial data.
>
> **(T)5.3.3.2:** Check the geocode assignment.
>
> **(T)5.3.3.3:** Include temporal and geographical comparability and coherence within the quality assessment routines and procedures.

(SP)5.4 - Edit and impute

This sub-process is carried out when data are considered incorrect, missing, unstructured, unreliable or outdated, namely by replacing/changing erroneous values, inserting/adding new values, removing obsolete registers and editing variables. This sub-process comprises various methods, generally through rule-based methodological approaches and more recently, using (supervised) ML methods for imputating missing or incorrect data and defining imputation classes from Big Data sources. Recording or flagging of changes made in the datasets and metadata documentation on the editing and imputation process and methods used are included in this sub-process. The activities on metadata management, including documentation procedures, conducted during this sub-process should follow internal or external standardised metadata systems and be regularly updated in the scope of the statistical system and other statistical regulatory frameworks.

(A)5.4.1: Assess the quality and validity of geospatial data.

(T)5.4.1.1: Check geospatial consistency and coherence.

(A)5.4.2: Use geospatial information to edit and impute other variables.

(A)5.4.3: Edit and standardise address data.

(T)5.4.3.1: Edit address data.

(T)5.4.3.2: Standardise address data.

(T)5.4.3.3: Adopt geospatial standards and common technical specifications on address data.

(T)5.4.3.4: Use address standardisation services.

(SP)5.5 - Derive new variables and units

This sub-process derives data for variables and units that were not collected in the collection but are needed to deliver the required outputs according to the design specifications (sub-process 2.2 - Design variable descriptions). This sub-process is carried out by applying either arithmetic formulae or different model assumptions and can be iterative taking into account the correct order since some derived variables may themselves come from other derived variables. It also includes aggregation or split of the collection units and estimation methods to derive new units according to the statistical needs.

(A)5.5.1: Derive new geographical variables, especially from point-based location data provided at the unit record data level.

(A)5.5.2: Use geospatial services to support the derivation of new geographical variables.

(SP)5.7 - Calculate aggregates

This sub-process aggregates data from microdata or lower-level aggregated data, including summing data by pre-defined statistical and non-statistical classifications, such as demographic and geographical classifications. Data aggregation can also be carried out in this sub-process by determining average and dispersion measures (range, variance, standard deviation, etc.) and by applying the previously created weights in sub-process 5.6 (Calculate weights) for totals or representativeness purposes (e.g., sample surveys).

(A)5.7.1: Aggregate data by geographical classifications for statistical purposes.

(T)5.7.1.1: Maintain and update the metadata on the geographies.

(T)5.7.1.2: Develop and apply standardised mechanisms to support conversion between geographies.

(T)5.7.1.3: Adopt up-to-date and compliant geodetic datum reference, projection and coordinate system(s) at national and international levels.

(A)5.7.2: Use geospatial-based aggregation services.

Analyse

(SP)6.1 - Prepare draft outputs

This sub-process transforms the data from the sub-processes 5.7 (Calculate aggregates) and 5.8 (Finalise data files) into statistical outputs according to the design specifications and identified statistical and user needs, i.e. outputs are “fit for purpose” before dissemination. Other activities supporting the analysis of the statistical content are also included in this sub-process, such as recording of quality characteristics, metadata management (e.g., cataloguing and tagging) and geospatial draft outputs (e.g., spatial visualisation tools, GIS outputs and statistical-geospatial services).

(A)6.1.1: Prepare maps.

(A)6.1.2: Prepare GIS outputs.

(A)6.1.3: Prepare geo-statistical services.

(A)6.1.4: Consider semantic interoperability and metadata standards when preparing the analysis output.

(SP)6.2 - Validate outputs

This sub-process produces the quality validation of the outputs according to the (national or international) quality assurance framework, guidelines and user needs and involves activities related to accumulated knowledge and intelligence for sound statistical soundness, namely in terms of data and methodology. Methods and tools for quality assessment to check the validation and quality requirements (e.g., checklists) of the outputs are included in this sub-process. Moreover, other validation activities are included in this sub-process to enable a more informed analysis, such as comparing time series, checking metadata, measuring quality indicators, checking geospatial consistency of data, investigating inconsistencies and identifying discrepancies with user needs and expectations.

(A)6.2.1: Visualise spatially to validate the output data.

(A)6.2.2: Perform and compute a set of geospatial quality indicators, especially from the processing and analysing perspectives (e.g., geospatial methods).

(SP)6.3 - Interpret and explain outputs

This sub-process conducts an extensive interpretation and explanation of the outputs to gain a deeper understanding and insights, including thinking on the firstly identified needs and initial expectations, visualisation in multiple perspectives and in-depth statistical analysis (e.g. time series, consistency and comparability, revision, etc.). Any other analysis activities, methods and tools that help to better interpret and explain the outputs are included in this sub-process.

(A)6.3.1: Visualise spatially to interpret the output data.

(A)6.3.2: Consider semantic and technical interoperability and metadata standards to ensure data accessibility and usability.

(SP)6.4 - Apply disclosure control

This sub-process carries out all activities that ensure the data and metadata to be disseminated do not breach statistical confidentiality and data privacy/security rules according to the methodology defined (sub-process 2.5 - Design processing and analysis) and the policy and legal frameworks applicable to the statistical organisation or country. Statistical disclosure control methods, application of data suppression or perturbation techniques and output checking are examples of activities included in this sub-process, which may vary based on the different types of data sources and outputs. It may also involve activities that ensure protocols to safeguard data confidentiality are duly applied to users with access to microdata for research purposes and new statistical outputs, such as geospatial statistics and spatial visualisation of statistical indicators.

(A)6.4.1: Apply disclosure control methods handling geospatial data.

> **(T)6.4.1.1:** Assess constraints on data dissemination through the application of disclosure control methods handling geospatial data and capabilities.
>
> **(T)6.4.1.2:** Test and perform traditional statistical disclosure control methods and other methodologically sound methods that evaluate and deal with spatial disclosure risk.

Disseminate

(SP)7.1 - Update output systems

This sub-process manages the systems update, i.e. databases within data repositories and IT infrastructure, where data and metadata are stored and ready to be disseminated, including data and metadata formatting and loading into output systems and linking data to the relevant metadata (from the previously work on cataloguing and tagging). Although it is recommended that these activities be previously conducted, a final check should be carried out during this sub-process to make sure the necessary metadata are ready for dissemination (e.g., describing and discovering data, and providing appropriate metadata for helping users to interpret statistical outputs). This sub-process ensures metadata will be publicly and easily available and accessible to the users according to agreed standards.

(A)7.1.1: Keep the geographical classifications and related definitions, concepts and variables updated in the metadata system.

(T)7.1.1.1: Create and make publicly available hierarchical and tabular views of the geographical classifications' versions to external users.

(T)7.1.1.2: Provide correspondence tables (or versions correspondence) between old and current versions of the geographical classifications used for statistical purposes

(T)7.1.1.3: Cataloguing and tagging the geospatial statistics products.

(A)7.1.2: Explore and take advantage of the potential of available mapping tools and related digital geospatial technologies oriented to data representation and visualisation.

(A)7.1.3: Update in-house geospatial data dissemination services or search for new ones.

(SP)7.2 - Produce dissemination products

This sub-process produces the dissemination products, as previously designed in sub-process 2.1 (Design outputs) to meet user needs. This sub-process covers activities on the production of several types and formats of products for statistical dissemination, including the preparation of the product components and checking statistical dissemination norms and publication guidelines. Dissemination products can include both traditional forms (e.g., printed publications, press releases and websites) and more modern ones (e.g., interactive graphics and tables supported by maps, service-oriented products, LOD and applications for data analysis and with download capabilities). Statistical-geospatial data integration may occur only during this sub-process by combining statistical outputs or end products with geospatial data to produce statistical/thematic maps through cartographic data presentation methods.

(A)7.2.1: Adopt geospatial (metadata) standards and use open technologies and services to ensure easy access and discovery.

(T)7.2.1.1: Adopt modern encodings and technologies, OGC Web Services and RDF standards (open linked data).

(T)7.2.1.2: Implement open standards for geospatial content and services.

(T)7.2.1.3: Adopt web standards for building consistent and harmonious open web platforms, browsers and other software.

(A)7.2.2: Produce products with geospatial features and capabilities (e.g., thematic maps).

(T)7.2.2.1: Explore and check the different types of thematic maps.

(T)7.2.2.2: Consult and follow cartography and geovisualisation guidelines and considerations.

(A)7.2.3: Apply additional disclosure control from a geospatial perspective.

(SP)7.4 - Promote dissemination products

This sub-process addresses the active promotion of the statistical product from a specific statistical operation to reach a greater number of users from multiple user communities. This sub-process can be considered as an overarching process and includes activities related to management and web tools to better identify the potential users that will benefit from accessing and using the statistical products and improve communication with them.

(A)7.4.1: Foster the modernisation and innovation of dissemination products for geospatial statistics.

(A)7.4.2: Promote geospatial statistics products.

(T)7.4.2.1: Exploit the potential of GIS tools and capabilities to produce interactive and open statistical-geospatial products.

(T)7.4.2.2: Provide geospatial-related tutorial material to help users, especially non-GIS and geospatial experts.

(T)7.4.2.3: Provide spatially visual representations of increasingly detailed geospatial statistics.

(A)7.4.3: Promote geospatial statistics with a partner organisation.

(A)7.4.4: Promote geospatial awareness and literacy among the various user communities.

(SP)7.5 - Manage user support

This sub-process ensures user queries and requests for data, information and services are registered and that replies are provided within agreed deadlines (e.g., a certain number of working days), namely through a user support service or by filling in forms for more specific enquiries. The activities included in this sub-process are related to the overarching quality management process to evaluate new or changing user needs based on the regular review of the inquiries and requests, some of which may be publicly available to external users to reduce duplicate efforts on requesting similar requests.

(A)7.5.1: Support the users from a geospatial perspective.

(A)7.5.2: Include geospatial components to measure and evaluate user satisfaction on geospatial statistics.

(T)7.5.2.1: Develop and define user engagement and feedback mechanisms related to geospatial statistics.

Evaluate

(SP)8.1 - Gather evaluation inputs

This sub-process collects all evaluation inputs produced during this sub-process or from any other sub-process in statistical production in which most of the activities and tasks included in this sub-process are valid for most statistical operations.

(A)8.1.1: Consider the geospatial components in the production process.

(T)8.1.1.1: Harmonise statistical and geospatial metadata concepts.

(T)8.1.1.2: Match statistical quality dimensions and metrics with geospatial ones.

(A)8.1.2: Compile the outputs from the geospatial quality indicators from the previous sub-processes.

(A)8.1.3: Collect external and internal feedback and suggestions.

(T)8.1.3.1: Collect user feedback and contributions.

(T)8.1.3.2: Collect suggestions from the geospatial experts and GIS staff.

(SP)8.2 - Conduct evaluation

This sub-process analyses the evaluation inputs, compares them to the expected/target benchmarking results (when available) and summarises them into an evolution output (e.g., report or dashboard) for punctual or continuous improvements.

(A)8.2.1: Evaluate the statistical operation from a geospatial perspective.

(T)8.2.1.1: Develop procedures and methods to assess the geospatial outputs and functional relationships between statistical and geospatial production components.

(SP)8.3 - Agree an action plan

In this sub-process, an action plan based on the previously produced evaluation outputs is developed and agreed on to reinforce the necessary intervention at the leadership, corporate and high-level management levels.

(A)8.3.1: Produce an action plan from the geospatial perspective.

Overarching Processes and Corporate Activities

In addition to the operational building blocks addressing the data integration and geospatial-related activities (A) and tasks (T), a set of overarching processes and corporate activities are needed as cross-cutting cornerstones to support the eight phases of the production process in developing a statistical operation for geospatial statistics while also considering geospatial components throughout the production cycle. Whereas the overarching processes are in the scope of the GSBPM, the activities at the corporate level are included in GAMSO - a supplemental model of GSBPM outlining examples of activities related to capability development and corporate support - and support standardisation in production.

Overarching processes related to quality management should incorporate specifications related to geospatial data, processes and services, including implementing robust quality control, quality assurance procedures and metadata management (UNSD, 2021). Corporate activities address governance, institutional collaboration and other non-technical aspects that may influence the production process across the statistical organisation and contribute to the statistical-geospatial operating environment and overall data ecosystem (e.g., data custodianships, public-private agreements, coordination mechanisms, etc.). Corporate activities are fundamental to enhancing statistical-geospatial interoperability at multiple dimensions and promoting engagement and communication across and between the involved communities and users.

The overarching processes and corporate activities will be generally outlined and described - as opposed to what was previously mapped in the geospatial-related activities and tasks - since they are not necessarily assigned to a specific phase and sub-process when developing the statistical operation. The following overarching processes and corporate activities are required to produce geospatial statistics in a consistent and efficient manner and be complementary to some of the above-mentioned geospatial-related activities and tasks.

i. Design and implement a statistical-geospatial collaboration and cooperation strategy:

When statistical organisations define their vision and mission to understand the operational environment and emerging issues, they should establish and further achieve high-level goals and strategies related to collaboration, cooperation and coordination with the statistical, geospatial and administrative data communities. It includes activities to build, manage and maintain strategic collaboration and cooperation among stakeholders contributing to and enhancing the statistical-geospatial operating environment at the national and international levels. Collaboration and cooperation with external stakeholders from the national and international statistical systems should be prioritised in the activities related to statistical-geospatial data integration under such statistical systems should be organised and coordinated by the regulatory bodies (e.g., national governments, national statistical council, regional statistical office, etc.). The academic and user communities cannot be left out by adopting and fostering a societal cross-cutting collaboration landscape that continuously ensures innovation and methodological developments as well as monitoring of the current and emerging user needs and requirements.

ii. Identify, collect and use reference geospatial datasets from authoritative sources:

Reference geospatial datasets from internal and external data sources and providers framed under strategically relevant data themes of authoritative geospatial data sources need to be identified considering the current and emerging national and regional policy frameworks, data governance and legal issues. Geospatial datasets considered relevant to official statistics and modernising statistical production should be audited into a data inventory alongside descriptive metadata and data specifications (e.g., geographic coverage, attribute completeness, versioning, revision cycle and projection system, etc.). Minimum quality requirements and agreed dataset quality profiles regarding those core geospatial datasets need to be established to ensure the same standards on geospatial data, metadata, technologies and services are adopted by all stakeholders contributing to the statistical-geospatial operating environment. Robust institutional cooperation, geospatial capacity development and policies on sustainable and active data management and maintenance containing legal requirements and release guidelines are cornerstones for carrying out this activity. These prerequisites ensure the consistent provision, access and use of the datasets over time and by defining formal accountability of providers or custodians guarantee they keep data regularly updated and with adequate quality for users (e.g., statistical organisations) according to national public interests. This activity is also covered by the quality management overarching process.

iii. Develop and implement a unified approach to statistical-geospatial quality management:

A unified approach to statistical-geospatial quality management should be developed to encompass both statistical and geospatial quality dimensions, criteria and metrics to be applied at different stages to produce geospatial statistics in a systematic and consistent manner. A statistical-geospatial quality strategy towards a common quality framework is required to manage the quality of both statistical and geospatial data and services (source and product quality) and that statistical and geospatial processes are equally documented and monitored throughout the statistical production cycle (process quality).

Quality management activities that support a quality framework can be referred to and applied at two levels. Either the strategic/organisational level (e.g., policies and guidelines, etc.) or the operational level that is more related to the production chain (e.g., quality control and risk management procedures). Quality assurance tools and mechanisms (e.g., quality indicators and feedback mechanisms) for geospatial data need to be designed and carried out in the respective phases of the statistical operation development to assess geospatial quality considering the conceptual and technical specifications from the perspective of input, processing and output. Also, activities on quality documentation and process quality assessment and improvement need to evaluate statistical processes, geospatial processes and statistical processes handling geospatial contents as well as their functional relationships and interdependencies (e.g., when setting process quality targets). In this regard, quality feedback is required to assess the dependency flows between the statistical and geospatial processes (and respective production systems, workflows and infrastructures), what impact the outputs from a specific geospatial process can have on previous or subsequent statistical processes and to detect necessary changes in data to improve the quality. The statistical-geospatial quality management should also provide matching between statistical and geospatial quality terminologies and standards to ensure that (input) geospatial data is similarly interpreted and calculated than statistical data and enable the quality of geocoding and data integration methods as well as the accessibility and usability of geospatial statistics and related products (e.g., web-based mapping services).

Basic measures for geospatial data quality assurance should be conducted based on the ISO/TC 211 standards, especially the ISO 19157-1:2023 on data quality. This ISO standard can be used as a reference standard for geospatial data quality as other well-established standards under statistical quality frameworks are adopted by the international statistical community. It includes error and correctness indicators, and other measures based on methods to count errors or the number of correct values. Under the geospatial quality evaluation flow, the geospatial quality reporting mechanisms can be applied considering the following quality elements: i) completeness (commission and omission); ii) logical consistency (conceptual, domain, format and topological); iii) positional accuracy (absolute and relative); iv) thematic accuracy; v) temporal quality (accuracy of time measurement, consistency and validity); and vi) usability (based on user needs and requirements). These six quality categories can be used to assess the level of achievement on quality requirements.

Lastly, monitoring the geospatial developments is an important corporate activity related to quality management to ensure that corporate support (information, knowledge, methodologies and IT systems) for geospatial statistics is up to date and aligned with the trends of the statistical-geospatial operating environment and supporting communities. This activity is particularly important since the geospatial community and application fields are traditionally more industrial-oriented (e.g., technologies and standards) and keener to be ever-changing and fast-evolving due to digital and technological drivers and trends.

iv. Develop and implement a unified approach to statistical-geospatial metadata management:

It is important to design and implement metadata management requirements and activities - at both strategic/corporate and operational levels - for geospatial data and metadata to facilitate sharing, querying, accessing and using these components and capabilities (e.g., technologies, systems, etc.) throughout the statistical production process. Similar to statistical metadata, metadata management activities should focus on the creation/revision, updating, use and archiving of geospatial metadata in which geospatial data and metadata are equally considered as input for quality management and included in the consistent approach to metadata. Hence, the overarching metadata management process should formally cover the content and links between geospatial and non-geospatial information objects and processes.

In this regard, a statistical-geospatial metadata management strategy and the respective system should be designed considering a minimum amount of statistical and geospatial metadata, the preferred data description and formats and systematic geospatial metadata management procedures consistent with statistical metadata, including technical specifications and recommendations on collection, maintenance, validation and accessibility. This strategy and system can also contribute to aligning and streamlining statistical and geospatial data and processes and make them more metadata-driven (e.g., non-geospatial experts can easily discover geospatial conceptual models) while ensuring that both statistical and geospatial metadata are documented, updated and represented according to the established rules and norms. Moreover, connecting and harmonising geospatial data concepts and elements (data models, catalogues, vocabularies, etc.) with existing statistical metadata is crucial to either enabling staff to more efficiently integrate data or users to explore and exploit geospatial statistics easily. The previously mentioned SDMX 3.0 standard included technical specifications regarding geospatial metadata is a good example of a significant milestone regarding statistical-geospatial metadata management.

Geospatial-related considerations should be oriented to the following issues: i) registration (well-documented metadata and well-identified data); ii) the authoritative source (metadata elements from authoritative data sources); iii) capture at the source (capture geospatial metadata at their source, preferably in an automatic manner); iv) integrity (metadata-related activities fully embodied in the statistical business process); v) matching metadata (metadata generated in the statistical production process match the metadata available to users); vi) description of the metadata flow (geospatial metadata flow can be described alongside the statistical metadata flow); vii) exchange and use (geospatial metadata can easily be exchanged and used by machines and humans); viii) formats (ensure geospatial metadata is recognised by considering its specifications and different users of the geospatial data); and ix) availability (ensure geospatial data are readily available and usable taking into account user needs).

The statistical-geospatial metadata management strategy should also determine widely adopted core geospatial data and metadata standards (e.g., ISO 19115, ISO 19115-2 and GeoDCAT, etc.) by the geospatial community while considering emerging digital and technological trends (e.g., LOD, semantic web standards such as RDF, machine-readable formats, service-based mechanisms, etc.). Although emphasising standards describing, attributing and specifying the geospatial data (metadata standards), standards addressing geospatial data quality (standards related to geospatial data content standards that were outlined in quality management) should be used as complementary standards layer to improve data integration and standards interoperability. Considering open international standards promoting structured exchange format and flexible software solutions for encoding, it is highly recommended to expand the application range of data and technologies and adapt them to the national context in a more straightforward manner.

Lastly, high-level and corporate intervention is required to guarantee long-term vision and strategy on the quality of statistical and geospatial data, preservation of associated metadata and adoption of common standards, operating models and services to ensure that data structures and file formats continue to be acceptable, reliable and usable over the years and across organisations. These actions will improve statistical-geospatial (data) integration, interoperability and standardisation in statistical production and data ecosystem, especially on technical and semantic issues for discovery and accessing data via the NSDI and other national data catalogues/repositories.

v. Manage statistical-geospatial capability development and capacity building:

This overarching process includes all activities supporting the planning, development, monitoring and improvement of statistical-geospatial capability development and capacity building, including research and innovation activities for new statistical operations producing geospatial statistics or improving the efficiency of current ones. It addresses cross-cutting and continuous activities to assess, integrate and enhance required capabilities and capacities to produce geospatial statistics into the statistical production process and encompassing issues on business performance, methodologies, quality framework, information and technical infrastructure, IT systems and the human resources and their skills and knowledge (e.g., GIS expertise). It also includes activities and tasks that contribute to increasing the capability to integrate statistical and geospatial data, including the assessment of needs to be maintained (what works well) or reduced (inefficiencies), the identification of new activities, improvement actions and priorities to be incorporated in the next work programmes and statistical operations. These types of activities and tasks are relevant from the organisational/institutional view since they allow the detection of required changes, prioritisation of options, optimisation of resources and development of improvement programmes and actions. Lastly, this overarching process aims to define, maintain and evaluate detailed capability requirements related to geospatial data, processes, services, technologies, people, standards, frameworks and institutions supporting the planned statistical operations to effectively produce geospatial statistics.

III.2.2. ASSESSMENT MATRIX

The Assessment Matrix is a tool to evaluate the statistical-geospatial integration capacity and maturity levels of the statistical organisation. Some issues can also assess the performance of the case study in terms of applicability based on the sub-processes, geospatial-related and data integration activities (A) and tasks (T), mapped and described in the Production Model. This matrix aims to constitute a systematic evaluation material reinforcing the geospatial dimension within traditional statistical quality management processes and activities under the quality framework, established at the institutional level. Hence, the matrix supports the understanding and management of the quality of inputs, processes and outputs considering the most relevant aspects and requirements described in the Production Model.

In terms of structure, the Assessment Matrix is divided into the three key elements introduced and described in the previous chapter (governance; data, information and technology; and institutional collaboration and capacity). These key elements are assigned as quality dimensions representing a broader scope of statistical-geospatial integration and providing cross-cutting factors to evaluate related quality issues that can be identified and corrected throughout the development of statistical operation in the long term. Moreover, this set of quality dimensions aims to describe the statistical-geospatial quality and progress level within the overall statistical production of the statistical organisation, and sometimes specifically related to the case study, in a multidimensional and interconnected way by not covering one quality measure or establishing a ranking among them. Thus, it assesses and intends to assure compliance with both technical and non-technical quality requirements, guidelines and best practices for high-quality and standardised statistical-geospatial data, processes and outputs (geospatial statistics or other statistical-geospatial output and product).

Each quality dimension breaks down into a group of questions and requirements to assess its level of capacity and maturity regarding statistical-geospatial integration, and they are oriented to both policy and top management profiles, and staff from the technical and administrative levels. The questions and requirements built in the

Assessment Matrix provide a multi-level interdisciplinary approach ranging from the operational (expert) to the strategic and leadership level, while considering aspects related to production, organisational/corporate, governance and institutional environment. For each question and requirement, there is a design/implementation degree to be filled, containing the following three different qualitative metrics and respective descriptions:

1. **Not Designed/Not Implemented (ND/NI):** the question/requirement contains production components and non-technical elements not designed nor implemented to support the statistical production and the development of the statistical operation. The capabilities referred to in the question/requirement do not exist nor are formally established and applied in the statistical-geospatial operating environment of the country. In addition, the question/requirement can be in an early stage of implementation involving informal actions within and outside the statistical organisation (e.g., collaboration with external stakeholders) or preliminary awareness is recognised by some staff of the organisation.

2. **Partially Designed/Partially Implemented (PD/PI):** the question/requirement involves production components and non-technical elements that were partially planned and designed and/or partially implemented to support the statistical production and the development of the statistical operation. Development and design activities to define the inputs, processes, services and outputs (e.g., concepts and methodologies) were already mapped in the statistical operation as well as collection and processing components were built and tested via pilot projects, case studies or experimental statistics.

   Although part of the planned and designed production components and elements are implemented and running within the statistical production, they require more development work and improvement measures to increase quality, performance and efficiency (e.g. from small fixes to larger corrections) and be fully operationalised. Some of the capabilities referred to in the question/requirement might be being prepared or under improvement for a more corporate and mature implementation.

3. **Fully Designed/Fully Implemented (FD/FI):** the question/requirement includes production components and non-technical elements that were fully designed, operational and streamlined within the production pipelines (dataflows and workflows), providing a robust and well-established condition in the development of the statistical operation. This class also indicates the capabilities outlined in the question/requirement are present in all of the three main stages of the statistical business production process model, even though with different degrees of applicability: i) design of the statistical operation (first three phases); ii) operationalisation (from the collect to dissemination phase); iii) evaluation of the quality of the statistical operation (evaluate phase). Thus, those capabilities are consistently implemented across the organisation and play a structural role in its strategy, statistical work programme and statistical production.

Due to the generic structure of the assessment matrix that provides a simple diagnosis approach, the statistical organisation can comprehensively understand the current statistical-geospatial integration state and more easily identify development challenges related to existing and needed capabilities. Recognising the capacity and maturity levels in integrating and streamlining geospatial components within the statistical production will facilitate the implementation and operationalisation of GSGF (global or European versions) over small and assertive steps in more critical areas. In this regard, this assessment matrix can be used as a starting point to implement the framework or audit to progress of the framework at multiple levels (e.g., corporate, operational, etc.) as well as categories of capabilities (e.g., policy, legal, methodology, human resources and institutional).

Furthermore, the assessment matrix (Table 10) can also be used as a self-assessment tool by the statistical organisation (and the national mapping agency) for internal quality review and reporting to systematically check, review and manage the statistical-geospatial quality. Therefore, it provides an illustrative template for a checklist with a scoring system or a baseline for a more complete auditing questionnaire with the ability to register detailed notes and extensive considerations to evaluate compliance with best practices regarding statistical-geospatial data integration. The results can support quality management, monitor compliance (especially on frameworks and standards), identify capability gaps and needs for improvement, and define key action areas and capacity development priorities for a more mature implementation over time.

Table 10. Assessment Matrix (source: author).

Quality Dimension	Question/Requirement
Governance	**1.** Does the country have a national geospatial data strategy or a roadmap for geospatial data creation, management and dissemination?
	2. Does the country have an NSDI with formally designated data governance roles and responsibilities across the government on geospatial data production, management, maintenance and custodianship with clear and legally binding mandates to ensure the integrity and quality of authoritative geospatial datasets?
	3. Existence of a business model, funding resources and/or investment initiatives at the governmental and/or statistical system levels on geospatial data creation, management, maintenance and dissemination.
	4. The national policy and legal environment enable and ensure data protection, licensing and sharing between organisations and public institutions and facilitates its availability, accessibility and usability, namely through generic and/or specific-domain regulations and laws.
	5. Does the statistical organisation have a geospatial coordination unit that coordinates and carries out all activities and projects related to geospatial data collection and management and its integration with official statistics?
	6. Does the statistical organisation actively contribute to international activities, initiatives, projects and/or WG that promote statistical-geospatial (data) integration?
	7. Does the statistical organisation adopt internationally and/or nationally agreed statistical and geospatial standards and compliance mechanisms for effective implementation at the organisational and national levels?
Data, Information and Technology	**8.** Does the statistical organisation follow any international roadmap or strategy on reference geospatial themes and core geospatial datasets regarding common and harmonised geospatial data creation, management and dissemination at the local, national, regional and global levels?

Quality Dimension	Question/Requirement
Data, Information and Technology	**9.** The country has a national integrated statistical-geospatial data framework underpinning basic and authoritative statistical and geospatial data to support evidence-based policy-making and forecasting.
	10. Does the country have an official and standardised national register on addresses, buildings/dwellings and/or cadastral parcels?
	11. The statistical organisation is able to geocode/georeference each statistical unit (microdata) within most statistical domains (e.g., social, business and environmental statistics, etc.) to a reference location (e.g., geographic coordinates, address registers, small geographic area, grid system, etc.).
	12. What is the lowest geographical level in which statistical unit record data are collected (e.g., survey data and census operations) and/or administrative registers are acquired?
	13. Does the country have clear custodianship roles as well as production and maintenance mandates for boundary data underpinned by formalised coding systems linked to a common and agreed set of basic geographies (administrative, statistical and other types of geographies)?
	14. Does the statistical organisation extensively use the gridded geographies or any grid reference system (e.g., European ETRS89, DGGS, etc.) to support statistical production?
	15. The statistical organisation applies statistical disclosure control methods and (micro)data confidentiality and protection techniques and tools handling geospatial data.
	16. What is the extent of implementing geospatial data, information, technology, and other related technical capabilities in statistical production and statistical activities of the statistical organisation?
	17. The statistical organisation's data/information architecture and technical infrastructure (e.g., IT systems, technology applications, etc.) embodies a geospatial data repository or a geospatial information infrastructure for internally managing and providing geospatial data, technologies and services.
	18. The statistical organisation performs a series of geospatial services (business, application and modular services) to support data integration and geospatial-related activities and tasks.

Quality Dimension	Question/Requirement
Data, Information and Technology	**19.** The overall enterprise architecture of the statistical organisation ensures (dependency) functional relationships between the statistical and geospatial production components from the business, information, application and technology perspectives.
	20. The statistical organisation regularly conducts geospatial data quality assessment procedures, validation methods and/or feedback routines (e.g., point-of-entry validation, point-to-polygon, geometry coherence, homogenisation tools, etc.).
	21. Does the statistical organisation publish, share and provide access to geospatial data and geospatial statistics in an open format by using non-proprietary formats, API and standards-compliant services?
Institutional Collaboration and Capacity	**22.** Existence of more-or-less formal and bilateral co-creation, co-management, co-maintenance and provision programmes on geospatial data between the statistical organisation and external data providers.
	23. Does the statistical organisation have a robust working relationship, cooperation arrangement or a long-term strategic alliance with its geospatial counterpart (usually the NGIA) and other data providers, including administrative data custodians?
	24. Existence of internal guidance materials and methodological/technical documentation within the statistical organisation to support geospatial-related and data integration activities and tasks.
	25. The statistical organisation has a permanent staff team with geospatial expertise (e.g., GIS skills for geospatial data management, geocoding and spatial analysis) and trained personnel to handle statistical-geospatial data integration processes and capabilities (e.g., data integration techniques, production methods, confidentiality, standards, etc.).
	26. Do the statisticians and other non-geospatial experts within the statistical organisations have some level of geospatial awareness and literacy?
	27. Do you establish systematic consultations and develop feedback mechanisms with the broader user community regarding geospatial statistics?
	28. Development and implementation of action(s) plan(s) within the statistical organisation to enhance statistical-geospatial integration capacity and increase capabilities in this topic.

The following descriptions and explanations support the understanding of the questions/requirements and the assessment exercise in assigning the adequate implementation degree according to the current situation of the statistical organisation and the surrounding institutional environment and data ecosystem.

1. This question addresses national or sub-national geospatial policies and strategic programmes involving public institutions, inter-governmental bodies, local authorities and the private sector in which their roles, responsibilities and activities are defined following guidelines on development and implementation and a legal framework. This long-term strategy should underpin the value of geospatial data, its integration with statistical data and other data domains and the importance of geospatial capabilities, especially for good governance and digital ecosystem, towards a geospatial information infrastructure focusing on location-based knowledge (GKI concept).

2. This question aims to ensure that an NSDI is fully developed and operational with a clear set of custodianship mandates, data guidelines, technical standards and legal and institutional arrangements to provide an organised geospatial information infrastructure with authoritative reference geospatial datasets for geocoding and effective statistical-geospatial data integration.

3. This requirement ensures sustainability in the acquisition, management, maintenance and provision of geospatial data and regarding coordination actions and research, development and education activities. Financial partnerships between the public and the private sector, investment programmes with the academia, financial plans agreed upon at the policy level and common price data models for geospatial data are examples.

4. This requirement may include legal requirements for data creation, management and integration to ensure their quality and security (e.g., data protection and privacy laws) as well as policies on open data (e.g., open government data initiatives).

5. This question is important to ensure that the statistical organisation has a long-term vision regarding geospatial data and recognises their value to modernise statistical production and provide more geographically detailed statistics (e.g., sub-national/regional and local indicators).

6. This question also involves the (dis)alignment of the statistical organisation with related overarching policy and production frameworks (e.g., 2030 Agenda, EU Green Deal and Data Strategy, GSGF, IGIF, etc.).

7. This requirement assesses the organisational commitment to standards adoption and development of best practices as well as the level of governance coordination with other organisations and public institutions to foster organisational and legal interoperability in the country.

8. This question addresses the awareness and capacity to collect, manage and/or maintain priority geospatial themes/datasets with good quality for official statistics and data integration processes, particularly geocoding. It is recommended to conduct a data inventory and profile (metadata and quality descriptions) at the organisational and national levels. This exercise might require communication and joint work with other stakeholders, including public institutions such as the NGIA, government departments such as Ministry Offices, administrative data custodians, local authorities and other data providers that support the national data ecosystem. Examples are the UN Global Fundamental Geospatial Data Themes, UN-GGIM Core Data, INSPIRE Themes or Geodata files managed by GISCO (Eurostat), according to the EC geospatial data requirements.

9. This type of national data framework enables the users to link cross-domain geospatial to non-geospatial data (e.g., via data integration services), provides a common link between different information systems and applications and makes data accessible and (re)usable.

10. This question checks the availability of authoritative geospatial reference data for geocoding and the implementation of unique and persistent identifiers to facilitate the linkage between high-quality and standardised locations/geocodes (from spatial objects) to each statistical unit record data (e.g., microdata). Although such location data may be created and maintained at the sub-national, regional or local level in some countries, the records can be compiled into a single dataset. This national dataset should contain minimum attributes such as geographic location (geometry) and unique identifier. Additional attributes can be included to functionally and topologically link the datasets in a consistent and hierarchical manner.

11. An accurate, precise and consistent location addresses the direct position (x,y,z coordinates) rather than an indirect position (descriptive address register or a locality name). This requirement recognises a greater maturity level when high-precision location data is used for geocoding statistical/administrative data, especially microdata, and fully integrated into the data architecture and technical infrastructure of the statistical organisation. Thus, it implies that all statistical/administrative microdata are geospatially enabled through a common and consistent geocoding approach where location data is processed and stored only once (geospatial data repository) to improve dataflows and workflows across the data management environment.

12. This question aims to ensure that high-precision geocodes are collected and stored for each input data record, preferably microdata (statistical and administrative datasets, such as a person, household, business, etc.) and that geocoding is carried out at the most detailed level by prevailing the point-based approach (x,y,z coordinates). This will provide territorial flexibility for fit-for-purpose data aggregation, enhance spatial analysis and more adaptability to changes in geographies over time.

13. This question addresses the recommendation to establish a nationally common geographical base for flexible data aggregation and integration as well as for displaying, analysing and reporting statistical outputs in a comparable and consistent manner over space and time. It implies that such geographies are considered authoritative geospatial data (within the NSDI) and are produced, managed and/or maintained by officially assigned stakeholders (e.g., NSO and NGIA). The coding system is fully harmonised with other public institutions and governmental agencies and any code and territorial (geometry) changes are recorded (availability of historical geographies and boundaries) and duly reported and displayed. The use of standards-based conversion mechanisms, common methods and compliance with existing custodianship guidelines, technical specifications and metadata requirements are maturity prerequisites.

14. This question aims to inquire if the statistical organisation follows the generic trend of using gridded geographies to produce official statistics, namely to geospatially enabled statistical/ administrative data in a consistent manner, data comparison for spatial analysis and dissemination (output geographies of interest). Gridded geographies can also be used in the Design and Collect phases regarding geosampling.

15. This requirement ensures that, on one side, the statistical organisation certify data custodians release data considering privacy and confidentiality issues, and on the other side, data are managed and outputs are released based on privacy and confidentiality requirements, preferably aligned with existing national or international confidentiality policy and/or data protection regulation. It might be pre-tabular methods, such as targeted record swapping, and post-tabular methods, such as cell aggregation for grid data to overcome geographic differencing and regional breakdown issues. This requirement ensures privacy and confidentiality in managing geospatially enabled statistical/ administrative data (especially micro and sensitive data) and for dissemination of grid or small area statistics by developing, applying and sharing disclosure control methods tackling potential risks related to geospatial data. It may involve statistical disclosure control methods for tabular or microdata protection methods having geospatial considerations, namely for spatial visualisation (statistical maps).

16. The level of implementation is also interlinked with the type and number of statistical operations, programmes and domains wherein geospatial capabilities and related production components support statistical production and related business processes.

17. This requirement ensures that a geospatial-based technical infrastructure is incorporated into the overall business reference architecture and establishes logical connections with other data repositories (through a location-centred data architecture and well-defined interfaces) to standardise geospatial production components and streamline all dataflows and workflows. Such geospatial-based technical infrastructure might include interoperability tools, data/metadata integration and spatial visualisation services (e.g., on-the-fly mapping) or basic GIS applications for spatial analysis and geoprocessing, preferably following a service-oriented approach. A mature level assumes that geospatial data and related digital technologies, services and tools are open, easily shareable, accessed and used across all departments and units of the statistical organisation to run statistical production processes in an efficient manner as well as between organisations to view and use geospatial data produced internally.

18. This requirement might include services for geospatial data capture, acquisition, editing and management, integration and standardisation and validation, geocoding, more-or-less advanced spatial analysis, spatial visualisation and discoverability services.

19. This requirement ensures statistical-geospatial interoperability on data, metadata, methods, processes, technologies and services throughout the statistical production process and across organisations by adopting nationally or internationally adopted standards and good practices within the data and user communities. It might be considered a higher maturity level if a significant range of users and organisations within the national institutional environment and data ecosystem can easily and efficiently discover, access and use an extensive range of cross-domain data, applications and technologies. Moreover, a statistical organisation is more mature in matters of technical and semantic interoperability (mainly related to data and metadata) when its statistical production is standard-based by underpinning both statistical and geospatial models and standards and by developing new ones attending statistical-geospatial data integration needs and requirements.

20. This requirement claims that statistical organisations should validate input geospatial data, preferably at the source, through well-established validation routines and robust assessment mechanisms to ensure the quality of location data entering the data management environment to avoid inconsistencies, inaccuracies and errors from the spatial, temporal and attribute dimensions. Validation routines and assessment mechanisms should also encompass the consistency and temporality of the identifiers to enable correct linkage between location data and the statistical/administrative unit record data.

21. This question aims to ensure that good practices and standard-based approaches are taken in place to guarantee data and outputs can be freely distributed as well as easily discoverable, accessible and usable by the user communities. This question is important to improve compliance with FAIR principles as well as the understanding and awareness of the statistical organisation on geospatial metadata to facilitate the sharing of geospatial data and services. It might include developing simplified interfaces (e.g., SPARQL), LOD framework (e.g., for linking statistical data and metadata with geospatial-based web data), standard vocabulary and taxonomy applications (StatDCAT-AP and GeoDCAT-AP), OGC web services and API (WFS, WMS), among other semantic web and open standards (e.g., RDF, SMDX 3.0, etc.).

22. This requirement might include collaboration protocols, institutional arrangements, MoU and data agreements and public-private partnerships. External providers can involve local and regional authorities, specific-domain regulatory bodies, companies, academia and other institutions and organisations responsible for geospatial data provision.

23. This question assesses the political endorsement and policy-based institutional mandates between and across data providers and custodians within the country.

24. This requirement is relevant for promoting statistical-geospatial quality management within the statistical organisation by formally incorporating geospatial quality in the quality assurance framework. It is the extent of geospatial quality documentation to ensure geospatial quality assessment processes, control and evaluation mechanisms are systematically recorded, maintained and reported providing a structured documentation of geospatial production components within the statistical organisation and promoting their standardisation and continuous iteration. The development and availability of in-house appropriate documentation related to geospatial production components will also improve quality assessment and assurance and compliance with the quality dimensions from the geospatial perspective, especially methodological soundness) and facilitate knowledge management within the statistical organisation. It involves recommendations for geospatial reporting (e.g., a system of geospatial quality indicators), proofs-of-concept, technical tutorials, user manuals, compiling of good practices, methodological reports and other types of quality and guidance documentation about the geospatial production components.

25. This requirement ensures that statistical organisations recruit qualified staff with the required skills to support the implementation of the geospatial production components and related capabilities. It also implies that continuous technical development initiatives, regular professional training programmes or other capacity actions are carried out to enhance their methodological skills (up-skilling) and provide them with lifelong learning opportunities and more knowledge.

26. This question includes the ability of statisticians to acknowledge the value of geospatial data for official statistics and to analyse and disseminate geospatial data to support their production processes and tasks.

27. This question addresses the need to identify emerging user needs - especially geospatial needs - and meet their requirements and expectations to constantly improve existing statistical-geospatial products and geospatial statistics, develop new ones and expand the portfolio to a wider range of user groups and communities. The results from the regular consultations and the analysis of user needs will promote a more user-oriented approach to designing and building the outputs by making them more easily understood, accessible and usable. Requirements related to geospatial data sources, output geographies and formats (e.g., type of geography, spatial resolution, etc.), technical and semantic interoperability, data and metadata standards, confidentiality and quality should be included when conducting the consultation sessions and applying the feedback mechanisms as well as discussed with the users.

28. The action plan should address capacity building and capability development to integrate statistical and geospatial data, processes and services across statistical production and streamline geospatial production components in the generic business production model. It involves assessing the current situation on statistical-geospatial integration from different capability perspectives (e.g., data/information, processes, systems and technology, methods, human resources, standards and frameworks, and institutional environment), measuring user needs, identifying critical risks, design new activities and report improvement needs. The action plan can be a short-term action plan for quality improvements and implementation actions to reduce previously identified inefficiencies addressing the next production cycle of a statistical operation or a mid to long-term strategy, outlining a roadmap at both institutional and production levels (e.g., applying an international standard or adopting framework). It can also include work programmes for capability improvements and both small and large projects to plan, develop and monitor the organisational capabilities for multiple statistical business processes, including research and innovation activities (as described in GAMSO).

The above-described methodology encompassing both operational parts of the Production Model and Assessment Matrix will be applied in the selected case study (CE-SIG) in the next chapter as well as the respective results and discussion remarks supporting the final chapters on recommendations and conclusions.

III.3. APPLICATION

The previously described case study, providing a concrete example of statistical-geospatial data integration within statistical production for geospatial statistics, will be applied to the self-designed methodology, inspired by a critical assessment of the literature review and demonstrating an overview of the state of the art on the topic. The practical application of the case study is divided into two operational parts as the methodology layout: i) Production Model that outlines geospatial-related and data integration activities and tasks within the statistical business production process based on the existing frameworks, standards, guidelines and best practices within the operating environment; and ii) Assessment Matrix that aims to evaluate the capacity and maturity levels in terms of statistical-geospatial integration considering the outcomes from the production model.

Both operational parts of the methodology will produce qualitative results from descriptive analysis related to the case study characteristics associated with the development of the statistical operation and its production cycle and business processes. In the first part, a descriptive summary of each activity and task within the production model (sub-processes and overarching processes/corporate activities) will be presented regarding the applicability of the case study. Thus, only the data integration and geospatial-related activities and tasks that are applied to the production specifications of CE-SIG will be depicted, as the model was designed using a generic and standard structure. Thus, it is systematically enforceable to any statistical operation that integrates geospatial components/capabilities and aligns statistical and geospatial business processes. In the second part, one of the assessment metrics related to the design/implementation degrees - described in the previous subchapter - will be assigned to each question/requirement (Q/R).

An explanatory text will be documented to support the understanding and justify the design/implementation classification assigned to each question/requirement, particularly concerning the respective quality dimension. The examined qualitative results from applying the selected case study to the proposed methodological approach will provide a comprehensive review and insightful contributions supporting the following chapter on results, discussion and recommendations.

The following tables (Tables 11, 12 and 13) summarise the application of the two operational parts of the developed methodology (Production Model and Assessment Matrix) in the selected case study, the CE-SIG project. The analysis of the overarching processes and corporate activities from the methodological application regarding the Production Model will be outlined in a separate table (Table 12), below Table 11.

Table 11. Application of the Production Model in the CE-SIG project by Geospatial-related and Data Integration Activity (A) and Task (T) (source: author).

Activity (A) and Task (T)	Descriptive Summary
(A)1.1.1	A descriptive memory document was drafted in response to the call for application in the 2020 Technical Assistance Operation Program scope, highlighting the territorial dimension of cohesion reinforced by the EU to ensure people have access to public service, housing or employment opportunities regardless of the place of residence. The importance of the social and economic dimensions in supporting public policy design is also mentioned. In this document, the situation analysis for the development of the statistical operation was summarised, however, only the geospatial datasets within the data/information infrastructure of SP were identified to outline the fundamental ones to the implementation of the project.
(T)1.1.1.1	The descriptive memory document describing an overview of the project did not include any analysis of the current state of geospatial data creation, management and dissemination as well as statistical-geospatial data integration at the national and international levels. Nevertheless, a need was emphasised to enhance data integration, namely administrative data, and to provide training in geospatial data management and GIS tools.

Activity (A) and Task (T)	Descriptive Summary
(T)1.1.1.2	The descriptive memory document outlined the mandate of SP for the collection and integration of data/information to produce official statistics in the scope of existing inter-institutional relationships and the protocol for implementing the statistical operation. In this regard, the proposed implementation model aimed to ensure collaboration with public institutions for data provision and access, namely the NGIA (Directorate-General for Territory) and other technical units of sectoral ministries. However, no references to other geospatial stakeholders and geospatial components to be streamlined into the statistical production pipelines are made.
(A)1.1.2	SP compiled a series of case studies and best practices related to territorial accessibility indicators from projects of other countries and international organisations (e.g., Statistics Sweden, OECD, DG REGIO, etc.), including datasets, methods and tools used.
(T)1.1.2.1	The benchmarking exercise focused on providing a generic production framework for building territorial accessibility indicators rather than catalogue data integration methods and tools according to the technical needs of the statistical operation (e.g., the format of the statistical output).
(T)1.1.2.2	The descriptive memory document outlined some generic needs and requirements addressing geospatial capabilities to develop and implement the statistical operation. It covered the capacity to continuously update the geospatial database, IT hardware for storage, (geo)processing and programming, software supporting the webGIS application and routing/spatial analysis, training initiatives and specialised human resources (GIS experts). The planned budget and investment efforts were linked to the identified needs, mostly associated with the technological infrastructure (hardware and software). However, no more extensive needs assessment and gaps analysis were carried out from the geospatial perspective during this stage, except for the geospatial datasets and routing services (advantages and disadvantages). Although this project involved an outsourcing service to develop some of the application components (e.g. back-office and web dissemination platform), there was no need to use such external technical assistance to support data integration and geospatial-related activities.
(A)1.1.3	SP has researched and reviewed some working papers and statistical/accessibility studies involving the integration of statistical and geospatial data. However, no literature specifically oriented to data integration methods and tools from both concept and methodological perspectives was gathered and reviewed.

Activity (A) and Task (T)	Descriptive Summary
(T)1.1.3.1	No consultation and analysis were done on reference documentation and materials related to the frameworks and standards within the statistical-geospatial operating environment, namely GSGF Europe and GeoGSBPM.
(T)1.1.3.2	Useful examples and positive experiences were collected and reviewed regarding territorial accessibility indicators and studies (most from countries and organisations with high maturity/capacity levels), but not specifically oriented to data integration practices and implementation of reference frameworks.
(T)1.1.3.3	No consultations and discussions were previously conducted with the national statistical and geospatial communities before the development and implementation of the statistical operation, except the mandatory communication to the Statistical Council of Portugal in the context of the NSS activities and work programmes. Moreover, the collected literature and body of knowledge and experience acquired over the years with reports, statistical studies and participation in international actions (e.g., grants on sub-national statistics) provided a clear overview of related task forces, WG and coordination activities (e.g., OECD and Eurostat) supporting the relevance of the project.
(A)1.2.1	Only national stakeholders (potential data providers and administrative data custodians from different sectoral public domains) were mapped and consulted towards institutional collaboration and data agreements for the development of the project. The input data characteristics and requirements were discussed and further agreed on with the data providers/administrative data custodians.
(T)1.2.1.1	All potential national stakeholders (public institutions and technical units from governmental agencies and sectoral ministries) were identified to ensure bilateral institutional collaboration and cooperation for the development and implementation of the statistical operation, namely for data provision and access. It included the Directorate-General of Territory (DGT) - the NGIA -, the Agency for Development and Cohesion, the Directorate-General for Education and Science Statistics, the Shared Services of the Health Ministry and the National Authority for Medicines and Health Products, among other public institutions supporting data acquisition and transfer.
(T)1.2.1.2	For the aim of the project, only authoritative data (survey and administrative data) were considered, reinforcing institutional collaboration and work synergies within the NSS and with other intersectoral data providers from the public sector.

Activity (A) and Task (T)	Descriptive Summary
(T)1.2.1.3	No other key stakeholders at the regional and global levels were identified, including organisations responsible for standards, due to the national scope of the project. However, some outcomes from European and global stakeholders were used as input for developing and implementing the statistical operation.
(T)1.2.1.4	No communication and engagement program were drafted despite the fundamental institutional capabilities involving this project in which there was a clear need to engage other national stakeholders, especially outside the NSS, and effectively communicate the importance of the statistical operation for sectoral public policies, population and territory dimensions.
(A)1.2.2	No actions were carried out to identify current and emerging user needs and requirements, namely related to geospatial data, services and products that can be embodied in the web-based GIS dissemination platform.
(T)1.2.2.1	No specific-oriented user consultations and engagement initiatives were carried out to assess the needs of targeted user communities that might be interested in the project outcomes (e.g., policymakers, local authorities, research community, etc.).
(T)1.2.2.2	No user engagement strategy was designed nor actions promoting the value of statistical-geospatial data integration/geospatial statistics in the context of the project.
(A)1.2.3	The descriptive memory document identified and described the internal technical and non-technical resources, capabilities and capacities related to statistical-geospatial integration for the development and implementation of the statistical operation, namely for project financing/budget purposes. Focus was made on the in-house data/information infrastructure and technological capabilities that support the statistical production. An overview of the capacity to implement the project was described based on past experiences from standalone statistical studies and reports related to territorial accessibility indicators.
(T)1.2.3.1	It was concluded from an early stage that a new unique identifier system related to the facilities and services supporting BIES needed to be designed and implemented to develop the statistical operation since these types of statistical units were not part of the statistical production or registered in the metadata management system. This unique identifier system will relate each facility to the respective services (relational databases) and provide a logical and structured coding system in BIES.

Activity (A) and Task (T)	Descriptive Summary
(T)1.2.3.2	The descriptive memory document included an overview of the existing IT and production system conditions, including data architecture (data sources and relationships between data repositories) and available technological resources, and identified what needs to be externally implemented (outsourcing service) on this matter. SP has all the necessary geospatial technologies within its production systems to develop and implement this statistical operation (e.g., GIS software, storage capacity, mapping services, etc.).
(T)1.2.3.3	No guidelines, good practices or methods for data integration, geocoding and other issues related to geospatial data management were reviewed since most of these capabilities and technical resources are already taken into place in the statistical production of SP.
(T)1.2.3.4	The descriptive memory document outlined the necessary human resources from the involved units as well as the associated costs of the external staff required to develop and implement the statistical operation. SP already has sufficient human resources with geospatial expertise (e.g., GIS skills) to ensure the project implementation.
(A)1.2.4	No specific-oriented analysis or assessment study of the in-house geospatial capacity and maturity levels was carried out in the context of the project. However, some issues related to data/information, technologies and human resources were identified from the geospatial perspective, highlighting the supporting geospatial datasets and georeferenced data.
(T)1.2.4.1	The BGE (georeferenced point buildings dataset) was identified as one of the key data sources within the information infrastructure that is fundamental to the project and the first reference for georeferencing the facilities to expand and thematically diversifying the in-house geocoding infrastructure to support statistical production.
(T)1.2.4.2	The existing GIS software licenses and built-in capabilities and tools are suitable for ensuring the development and implementation of the project, including georeferencing, geospatial data management, network analysis, geoprocessing, and mapping services. Thus, no external services on GIS technology or IT software-related improvements were needed.

Activity (A) and Task (T)	Descriptive Summary
(A)1.2.5	Besides the traditional output geographies addressing local administrative units (municipality) and regional statistical units (NUTS), the accessibility and demand areas (geographies) were specifically developed and defined for this project in which statistical information (territorial indicators based on surface, dwellings and population) is aimed to be disseminated about these geographies and associated metrics. However, no assessment of the fitness level of such geographies to the user needs was carried out (e.g., check if the general public easily understands the accessibility and demand geographies when interpreting the indicators).
(T)1.2.5.1	The past accessibility studies and related reports on territorial indicators carried out by SP in different activities provided the body of knowledge and experience to establish the types of geographies on accessibility and demand and assess their advantages and disadvantages.
(T)1.2.5.2	The dissemination of the results from the statistical operation will focus more on traditional output geographies, such as the municipality and NUTS 3 territorial levels, as well as on the new geographical areas related to accessibility and demand. The possibility of sharing the territorial-based statistical indicators using gridded geographies (1 km^2 grid) is still being evaluated.
(A)1.2.6	Although SP uses multi-resolution gridded geographies for dissemination purposes (from 1 km^2 grid to 500 meters for the housing prices of certain cities) and recognises its advantages, this type of geography is still under evaluation in the context of the project, including operationalisation and confidentiality issues.
(A)1.2.7	Minimum attributes to be included in the transferred datasets from the data providers to SP were established to ensure a common ground in data structure and model and facilitate semantic and synthetic consistency issues. In addition, consistent metadata requirements were assured by following in-house well-established metadata management procedures. However, other aspects related to interoperability, namely technical and legal were not covered or applied.
(A)1.3.1	The selection of the output geographies and respective formats (from web mapping services) considered the usability and accessibility conditions (e.g., not overloading the platform when opening it) and addressed geospatial data that is publicly available and open to everyone in the NSDI.
(A)1.3.2	The geospatial services needed to support the statistical operation were identified and described, highlighting the built-in network/routing analysis service and the web mapping services to be displayed in the dissemination platform.

Activity (A) and Task (T)	Descriptive Summary
(A)1.3.3	No confidentiality methods and disclosure control techniques handling geospatial data and considering the spatial resolution of the statistical outputs (territorial indicators) were evaluated at this stage.
(T)1.3.3.1	No national and international legislation or regulatory framework related to data protection and confidentiality was specifically assessed for this statistical operation, even recognising the risks associated with geospatial data that involves sensitive data about the location of the individuals and other statistical units.
(A)1.4.1	Concepts describing the statistical operation, namely related to the accessibility/demand geographies, facilities and territorial indicators, were identified and reviewed based on the previous body of knowledge and experience in similar studies. It included 'physical distance', 'service area', 'catchment area', 'geographical location', 'service of general interest' and 'total area of territorial units'.
(T)1.4.1.1	More generic concepts and terms related to statistical-geospatial data integration, such as 'georeferencing', were already established and described in the metadata management system of SP. Most of these concepts and terms are allocated in the 'Territorial' thematic section. The identified sectoral concepts addressed the various sectors of the facilities and services, such as 'education', 'health' and 'culture, sports and recreation'.
(T)1.4.1.2	No review work and updates were conducted on the identified and existing concepts and terms related to statistical-geospatial data integration, such as 'georeferencing'. The existing concepts and terms that were related to the project were exclusively summarised to support the dissemination platform and associated applicational infrastructure, alongside the new concepts that were created specifically for the statistical operation.
(T)1.4.1.3	No conceptual and vocabulary evaluation was conducted, namely focusing on statistical and geospatial standards and frameworks to improve semantic interoperability between both domains.
(A)1.4.2	No assessment efforts were made regarding the classification of territorial units in the national and international context, even though some public institutions that provide data in the context of the statistical operation have sectoral-based functional geographies (e.g., health regions for regional administration of the national health policy).
(A)1.5.1	The geospatial datasets needed to support the development of the statistical operation were inventoried and assessed, particularly the input data used for the network/routing analysis.

Activity (A) and Task (T)	Descriptive Summary
(T)1.5.1.1	Geospatial data sources within the in-house data architecture and information infrastructure were identified and described to support the development of the statistical operation, highlighting the georeferenced buildings datasets (points of buildings) and the national dwellings file (addresses).
(T)1.5.1.2	Besides the georeferenced administrative data (points) of the facilities sent by the data providers, all geospatial data used in the project is created, managed and maintained internally based on both survey and administrative data.
(T)1.5.1.3	No capacity assessment for potential NSDI datasets to be used in the context of the project was conducted, giving priority to internal sources and administrative data with traditional data providers from the public sector under the established institutional framework.
(T)1.5.1.4	The exercise of identifying the geospatial data for statistical purposes, particularly enhancing the in-house geocoding infrastructure, exclusively covered the business production model of the statistical operation.
(A)1.5.2	The cataloguing work on key geospatial data themes and datasets in the context of the project was carried out when establishing institutional collaboration with the data providers that reported the geospatial data (and associated with administrative data) that could be made available to support the statistical operation.
(T)1.5.2.1	None of the mentioned reference data frameworks were searched and examined in the context of the project. In general, the buildings, population, road network and facilities/public services datasets overlap with such data frameworks.
(T)1.5.2.2	The Global Fundamental Geospatial Data Themes were not reviewed for the context of the project.
(T)1.5.2.3	No European authoritative geospatial datasets (e.g., GISCO/Eurostat) were checked. However, the georeferenced buildings dataset (according to INSPIRE technical specifications) was used in the context of the project and the HVD on reference parcels (statistical units for census operations and 1 km^2 grid) can be used in the future to support the statistical operation.
(A)1.5.3	Legal constraints on data collection, acquisition and use were considered, especially by the Administrative and Business Data Unit following the guidelines and recommendations of the institutional confidentiality policy, quality charter and the legal obligations from the national implementation of GDPR.

Activity (A) and Task (T)	Descriptive Summary
(T)1.5.3.1	A generic evaluation of the policy and legal impacts on data issues to the statistical operation was not conducted.
(A)1.5.4	The roles and responsibilities of the geospatial data providers (sectoral institutions from the public administrations) were defined. It is important to highlight that the external data providers deliver georeferenced data on facilities in administrative registers.
(T)1.5.4.1	Bilateral institutional protocols involving data agreements were drafted between SP and the data providers when data acquisition and provision mandates were not already in force. Some of the identified stakeholders already supplied and transferred administrative data on a regular basis (e.g., hospitals) or survey data were collected in the context of a specific statistical operation (e.g., museum surveys). In other cases, institutional collaboration and cooperation protocols were already established (e.g., with the NGIA). In addition, informal meetings and preliminary contact points with some potential stakeholders were carried out to present the project, create engagement opportunities and evaluate the possibility of formalising an institutional collaboration partnership with agreed commitments.
(T)1.5.4.2	No geospatial data governance committee was established in the context of the project. The Territorial Statistics Unit is responsible for managing the institutional collaboration and agreements with the involved data providers.
(T)1.5.4.3	No specific programmes and partnerships on geospatial data co-creation, co-management, co-maintenance and provision were established besides the above-mentioned institutional collaboration protocols and data agreements. In this regard, no financial, technical and human resources requirements related to the data commitments were evaluated.
(A)1.5.5	The capability to aggregate statistical/administrative data to the desired output format to be displayed in the dissemination platform is ensured since the main input data for the calculation of the territorial indicators covers point-based geocoded data (buildings with population data and facilities).
(T)1.5.5.1	The custodianship, management and maintenance issues of the analysis and dissemination geographies used in the statistical operation were already assured in which the NUTS and the accessibility/demand geographies were created/managed by SP and the administrative units (parish and municipality) were managed by the NGIA.
(A)1.5.6	Some mapping services, routing tools (with authoritative street data and ready-to-use solutions) and geospatial-related dissemination platforms were reviewed as inspiration for the development of the statistical operation.

Activity (A) and Task (T)	Descriptive Summary
(A)1.6.1	No specific consultation action was conducted to find out the needs of the users in terms of output areas to display the territorial indicators in the CE-SIG dissemination platform, namely to assess if the users can interpret some indicators at the accessibility/demand geographic levels or if they understand the relevance of the high-resolution outputs.
(T)1.6.1.1	The output format (multi-geographic levels via web mapping services) is fully supported by the in-house GIS software enterprise license and open-source GeoServer (compliance with OGC standards for sharing and publishing geospatial data). The CE-SIG dissemination platform - developed by an outsourcing service - also supported the technical specifications of the different output formats that were designed by SP.
(T)1.6.1.2	The descriptive memory document highlighted the importance of territorial-based statistics, i.e. geospatial statistics, and the need to have more thematically diverse and geographically detailed geospatial datasets to support innovative statistical operations that contribute to more territorially targeted public policies in different sectors, namely at the sub-regional and local levels.
(A)1.6.2	Cost-effectiveness analyses were only developed to assess the data sources used for the network databases and associated routing services, in particular measuring the advantages and disadvantages and evaluating the quality implications of using open-source solutions and in-house maintained street network data sources (e.g., reliability, comparability, easy-to-use, update frequency, etc.).
(A)1.6.3	No long-term work plan was drafted addressing the sustainability and maintenance/management requirements of geocoding infrastructure in the context of the project. Nevertheless, technical materials were developed outlining some topics related to the geospatial data repository and georeferenced data within the data management environment and information infrastructure (e.g., how the different data repositories operate with the geospatial data repository by describing dataflows and management processes).
(A)2.1.1	No consultation action with the various consumers of geospatial data and services was carried out in the context of the project.
(T)2.1.1.1	The standardised geospatial-based solutions developed by SP for statistical dissemination purposes (e.g. to share census data) were used as inspiration for the CE-SIG dissemination platform.

Activity (A) and Task (T)	Descriptive Summary
(T)2.1.1.2	The existing geospatial technological components of SP provided sufficient dissemination and visualisation capabilities to share the desired outputs, namely using API and web mapping services. However, the geospatial technology used for the dissemination of the statistical outputs is not very innovative regarding emerging trends in geospatial (data) visualisation technology to make the user experience more interactive.
(T)2.1.1.3	The territorial-based statistical indicators of the project incorporated a strong component of descriptive analytics (e.g., how much of the population lives within a certain distance and is served by a certain type of facility) and prescriptive analytics (support decision-making in sectoral public policies, such as in the management of health facilities).
(T)2.1.1.4	The web mapping services (e.g., WMS) embodied some interoperability issues but only from the geospatial domain. No formal statistical standards for data and services were implemented for the dissemination of the territorial-based statistical indicators and associated geographies.
(T)2.1.1.5	The descriptive memory document highlighted the integrated approach of the project involving better coordination of sectoral policies at all levels (from the local to the European level) and outlined that the detailed outputs enable supporting the decision-making processes for more rational policies and public investment to assess disparities within the same territorial context while promoting territorial cohesion. Thus, the project is mainly addressed to policy-makers without overlooking the usefulness of the information to the general population and the future need to adjust the geographical scale of analysis to a specific territory of interest.
(A)2.1.2	No specific confidentiality issues related to geospatial data were considered when designing the outputs (CE-SIG indicators overlapping the accessibility/demand geographies), including geographic differencing addressing the potential risks from the functional geographies related to accessibility/demand areas.
(T)2.1.2.1	Only traditional statistical confidentiality issues were taken into account, namely under the national legislation requirements and institutional confidentiality policy.
(T)2.1.2.2	No review exercise on national policy and legal framework regarding confidentiality considering geospatial data was carried out.

Activity (A) and Task (T)	Descriptive Summary
(T)2.1.2.3	Although some NSO already published some handbooks, technical reports and conference papers on this topic, no guidelines and methodological materials addressing issues related to statistical-geospatial confidentiality were analysed and reviewed in the context of the project.
(A)2.1.3	The use of web mapping services from open geospatial/web standards (WMS) enabled a more interactive and dynamic spatial visualisation experience for the users. However, the absence of available WFS in the CE-SIG dissemination platform compromises the data usability criteria and analytical capacity in which the user can not directly download the displayed geospatial data.
(A)2.1.4	Only technology standards from the geospatial domain, namely for sharing and displaying geospatial data (e.g., WMS), were implemented in the context of the project.
(A)2.2.1	The geographical variables were defined to support the CE-SIG dissemination platform, including the information boxes and description of the territorial indicators to facilitate interpretation.
(T)2.2.1.1	The input data geographical level (georeferenced points) was checked with the data providers and the analysis and dissemination geographies were confirmed within the in-house geospatial data repository and metadata management system (geographies used for statistical purposes).
(T)2.2.1.2	The input data on (residential) buildings and facilities are point-based georeferenced data (x and y coordinates) associating the statistical unit record of the facilities and services to a specific location in space.
(T)2.2.1.3	New methods were developed to design a unique identifier system related to the facilities and to manage the supporting database.
(T)2.2.1.4	The geospatial data repository is the only data management system supporting geographies, handling geographical division for statistical purposes and holding a manual versioning workflow (e.g., without API).
(A)2.2.2	The georeferenced point data on (residential) buildings and facilities are the basic input data for deriving new geographical variables (territorial indicators by accessibility/demand geographies) through geoprocessing and aggregation procedures.
(A)2.2.3	The plan is to use the 1 km^2 and 125 meters grid for processing (origin-destination matrix to generate accessibility/demand geographies) and dissemination purposes in the context of the project.

Activity (A) and Task (T)	Descriptive Summary
(T)2.2.3.1	This task does not apply to the context of the project.
(T)2.2.3.2	This task does not apply to the context of the project.
(T)2.2.3.3	The hexagon grid was not considered or assessed in the context of the project nor was it used for other statistical outputs and products.
(A)2.3.1	No new data collection and storage components were designed. The established data architecture and information infrastructure already streamlined in statistical production were used to support the development of the statistical operation, particularly the data collection and management activities.
(T)2.3.1.1	The Administrative and Business Data Unit implemented a set of quality assessment procedures (descriptive statistics and coherence analyses) for the input administrative data related to the facilities to ensure the acquired records from the data providers are aligned with the guidelines and technical specifications outlined by SP. In addition, the Geo-Information Unit validates the geographic coordinates during the georeferencing and geometric editing work (when necessary). The process of validating the address data (address description, locality, municipality and postal code) is still under development with the collaboration between the Administrative and Business Data Unit and the Information Infrastructure Unit.
(T)2.3.1.2	The location data acquired (geographic coordinates of the facilities for georeferencing) are validated with the in-house Geographic Building Database (there are facilities located in residential buildings) and in some cases validated one more time at the source directly with the data provider/administrative data custodian. The georeferenced data (points of facilities) were also validated using the official administrative boundaries to check if the parish/municipality sent was true or false based on the geographic criteria.
(A)2.3.2	No GIS tools, geocoding services or other geospatial capabilities for data collection were configured. The administrative data was acquired via transfer mechanisms and further georeferenced in a GIS environment.

Activity (A) and Task (T)	Descriptive Summary
(T)2.3.2.1	No extreme scenarios in data collection were considered since the administrative data transfers are ensured by the data providers/administrative data custodians and supported by institutional collaboration protocols and data agreements. Also, such administrative data records need to be duly collected, maintained, managed and updated by the assigned custodianship roles.
(A)2.3.3	The attachment (data integration) is already made and ensured by the data providers/administrative data custodian when sending the administrative records of the facilities in which statistical data (descriptive variables) are associated with the location data (x and y coordinates and address information).
(A)2.3.4	This activity does not apply to the context of the project.
(T)2.3.4.1	Although web-based transfer instruments could be designed to modernise and facilitate data collection processes, such related innovative acquisition modes were not developed in the context of the project. Some administrative datasets were sent using traditional mechanisms (e.g., Excel files sent via email) or received via the cloud to be further loaded into the in-house DW (Oracle database).
(T)2.3.4.2	No open services for geospatial data and metadata collection were used in the context of this project.
(A)2.3.5	Institutional collaboration protocols and data agreements were made mainly covering stakeholders from the public sector, including regulatory bodies and sectoral governmental agencies.
(A)2.4.1	This activity does not apply to the context of the project due to its national scope.
(T)2.4.1.1	This task does not apply to the context of the project due to its national scope.
(A)2.4.2	This task does not apply to the context of the project due to its national scope.

Activity (A) and Task (T)	Descriptive Summary
(A)2.5.1	The dataflow and workflow involving survey data, administrative data, geospatial data, geoprocessing and geospatial data visualisation services were systematised in internal technical documentation.
(T)2.5.1.1	The data processing and analytical capacity were assessed, namely in terms of software and hardware capacity for geoprocessing (spatial interpolation using point data, network analysis, routing, etc.) and geospatial services publishing. It was necessary to check if there is processing power (e.g., storage, number of requests, etc.) in the components of the different computers and within the existing information and communication technology system to ensure high and consistent performance in the processing and analysis processes, including in the calculation of the territorial indicators.
(T)2.5.1.2	Python scripts were designed to automate geoprocessing workflows and perform more advanced spatial analysis, including reprojection, network/routing analysis and spatial interpolation, among other GIS tools. One advantage is that Python is a free, open-source and cross-platform programming/scripting language supported by the GIS software used in SP that makes the work less time-consuming, more efficient and easily reusable for non-developers or non-programmers.
(T)2.5.1.3	Although most of the administrative data acquired already integrated associated location data (x and y coordinates and address information), ad hoc methods (manual and sometimes, case by case) were developed and applied when conducting geographic validation.
(T)2.5.1.4	Additional network/routing services were used to support the geoprocessing workflows and spatial analysis processes. These geospatial services are embodied in the GIS software enterprise package, which SP has contracted with a GIS company to ensure access to high-quality street data and navigation capabilities for retrieving the accessibility/demand geographies and providing intermediate information to calculate some territorial indicators (e.g., time-distance).
(A)2.5.2	No new methods and tools for data aggregation were designed in the context of this project since those used were already implemented in statistical production.
(A)2.5.3	This activity does not apply to the context of the project.

Activity (A) and Task (T)	Descriptive Summary
(T)2.5.3.1	OGC standards were adopted for the geospatial data visualisation services (web mapping services) as well as INSPIRE regarding the data model and technical specifications for the buildings dataset.
(A)2.5.4	The geospatial analysis components were able to handle the building points (national territory) considering the multi-size and geometry of accessibility/demand geographies (spatial intersection) in order to perform the geoprocessing workflow and spatial analysis processes at multiple analysis levels. The internal technical documentation outlined the geoprocessing and spatial analysis procedures within the development of the statistical operation and respective production cycle.
(A)2.6.1	The geospatial dataflow and associated workflow were designed and described in a specific internal technical documentation addressing the creation and management of geospatial input/output datasets and services and geoprocessing.
(T)2.6.1.1	An internal technical document was developed to outline the input geospatial datasets to support the geoprocessing workflow and spatial analysis processes and the output geospatial data resulting from those processes for operationalising the territorial indicators, to describe the set of steps and processes that support geoprocessing operations and other spatial analysis processes for calculating the indicators and the geospatial services to be displayed in the CE-SIG dissemination platform. This document includes the identification and description of the technical specifications related to the datasets, namely in terms of their characteristics, attributes and the schemas in which they will be stored in the geospatial repository, as well as addresses methodological issues linked to the creation, management and use of the geospatial datasets and services (e.g., metadata) to ensure harmonisation within the in-house geospatial information infrastructure throughout the production cycles. Also, this documentation manual is complementary to other technical documents that support the development and implementation of the project by providing summary explanations of the different production components of the statistical operation.
(T)2.6.1.2	Only the Georeferencing Protocol, which was developed in the context of the project, included guidelines and good practices related to georeferenced data.

Activity (A) and Task (T)	Descriptive Summary
(A)2.6.2	The geospatial production system and workflow were designed and related functional requirements were identified to meet the production needs of the statistical operation, including the operationalisation of the territorial indicators (geospatial statistics) and associated map visualisation. The accessibility/demand geographies were designed and further developed to be user-oriented output areas for the users to be able to easily understand and interpret the associated territorial indicators.
(T)2.6.2.1	A part of the production components and functional requirements related to the geospatial information infrastructure was designed to meet the production needs of the statistical operation, focusing on the existing data architecture characteristics and services/applications to support process, analysis and dissemination sub-processes (e.g., routing service for network analysis and CE-SIG back-office application).
(A)2.6.3	The Statistical Production Process Manual of SP maps and describes the business processes and tasks involving geospatial data and services, including the design of functional requirements and capacity building of the geospatial information infrastructure for extending the contribution of geospatial capabilities in the production process.
(A)3.1.1	No modern collection instruments were checked and no geospatial data collection instruments were designed and used in the context of the project.
(T)3.1.1.1	This task does not apply to the context of the project.
(T)3.1.1.2	This task does not apply to the context of the project.
(T)3.1.1.3	This task does not apply to the context of the project.
(A)3.1.2	This activity does not apply to the context of the project.

Activity (A) and Task (T)	Descriptive Summary
(A)3.1.3	This activity does not apply to the context of the project.
(T)3.1.3.1	This task does not apply to the context of the project.
(T)3.1.3.2	This task does not apply to the context of the project.
(T)3.1.3.3	This task does not apply to the context of the project.
(A)3.2.1	The geospatial service to perform the network/routing analysis supporting the calculation of the territorial indicators was tested before the development of the statistical operation, namely in other projects related to geocoding and accessibility studies.
(A)3.2.2	The analysis and processing components supporting the creation of the geospatial services (web mapping services) and calculation of the territorial indicators were reused. Geospatial dataflows were also reused based on the geospatial data repository technical specifications.
(T)3.2.2.1	Both enterprise and open-source applications and solutions embodying navigation and routing services to perform network analysis were reviewed in which the advantages and disadvantages were identified and assessed. One important criterion was the authoritative street data to guarantee consistency and reliability in the several production cycles.
(A)3.3.1	Excepting the geospatial data visualisation services (WMS), all the dissemination components were specifically designed and developed in the context of the project.
(A)3.3.2	The geospatial data visualisation services (WMS) to be displayed in the CE-SIG dissemination platform address OGC standard-compliant web mapping services that can be dynamically visualised in any web browser application. The WMS also complied with ISO 19128:2005 (Geographic information - web map server interface) which enables displaying any geospatial data as a digital image file on a computer screen.

Activity (A) and Task (T)	Descriptive Summary
(A)3.3.3	Only the versioning of the territorial units was registered in the statistical metadata management system considering the geographical time reference of the output areas.
(T)3.3.3.1	Standardised metadata for geospatial data and services were applied for the geospatial format (Esri shapefile as the most acceptable industry-standard geospatial data format by any GIS software) and web mapping services (WMS).
(A)3.3.4	Data confidentiality issues and disclosure control methods were only designed to handle statistical data.
(T)3.3.4.1	The national statistical confidentiality and data protection policy was ensured while overlooking the confidentiality issues related to geospatial data and associated risks.
(A)4.1.1	This activity does not apply to the context of the project.
(A)4.2.1	This activity does not apply to the context of the project.
(A)4.2.2	This activity does not apply to the context of the project.
(T)4.2.2.1	This task does not apply to the context of the project.
(A)4.2.3	When the administrative data is being transferred (usually via the cloud to be stored in the in-house DW), the Administrative and Business Data Unit only performs a preliminary validation through descriptive statistics and coherence analyses (based on key attributes to check missing or incorrect data). The received input data are also compared with the previous data available in the in-house dissemination database (e.g., time series analysis) to find suspicious incoherencies. It aims to identify differences between previously published data and the input data for further revision and check if the alphanumeric data are valid to calculate the territorial indicators. In this regard, no confidentiality and disclosure risk issues are assessed during this stage.

Activity (A) and Task (T)	Descriptive Summary
(A)4.3.1	Although the received administrative data contains location attributes, including geographic coordinates, this activity does not apply to the context of the project since no geospatial technologies and geocoding techniques were used as modes and methods for data collection (e.g., as it happened during census operations).
(T)4.3.1.1	This task does not apply to the context of the project.
(T)4.3.1.2	This task does not apply to the context of the project.
(A)4.3.2	The input administrative data was already geocoded (administrative code) by the data providers/administrative data custodians.
(T)4.3.2.1	Each administrative record (facility as a statistical unit) contains location attributes in which the municipality codes (fourth digit ID) are according to the national administrative coding system from the Official Administrative Division Map, annually produced by the NGIA (DGT). Regarding the geographic coordinates, they were further reprojected based on the official reference coordinate systems (Mainland Portugal: PT-TM06/ETRS89; Autonomous Region of Madeira and Autonomous Region of the Azores: PTRA08-UTM/ITRF93), established by the DGT.
(T)4.3.2.2	No guidelines and instructions on statistical confidentiality and data protection were consulted and applied concerning high-precision geocodes (geographic coordinates).
(A)4.3.3	The geographic validation during georeferencing from the received geographic coordinates was conducted after the preliminary data validation round by the Administrative and Business Data Unit. Thus, point-of-entry validation procedures were carried out to check if the point needs to be edited, relocated (when the facility point already exists in the BGE) or created (in case the facility point does not exist as a building in the BGE).

Activity (A) and Task (T)	Descriptive Summary
(T)4.3.3.1	The input geographic coordinates of the facilities were checked against the geospatial building dataset (BGE) to assess the edition efforts in which an editing classification was created to support this validation step. Thus, the georeferenced points are validated according to a type of edition, from 'no edited' to 'changes to the equipment building over the years', and a percentage of records for each editing class is calculated. This validation process also enables identifying the facilities that due to the change in the geographical location will result in differences in the production of the territorial indicators (e.g., change of municipality).
(T)4.3.3.2	The descriptive statistics calculated by the Administrative and Business Data Unit included temporality validation issues in which the different reference periods (in the education sector, school years) are compared with related statistical indicators already published in the dissemination portal and other statistical products (e.g., regional statistical yearbooks).
(T)4.3.3.3	The internal Georeferencing Protocol is intended to be shared with the data providers/administrative data custodians to ensure consistency in the delivered geographic coordinates, namely regarding data models and project systems. However, it is acknowledged that sometimes these stakeholders do not have the technical capacity to apply these guidelines. Furthermore, the corrected geographic coordinates (e.g., significant change in the location of the point) are further reported to the data providers/administrative data custodians for consideration in the next data delivery/transfer. In the future, a geographic file in GeoPackage format for the final coordinates in the official coordinate system will be made available for this purpose.
(T)4.3.3.4	This task does not apply to the context of the project.
(T)4.3.3.5	This task does not apply to the context of the project.
(A)4.4.1	The geospatial inaccuracies and inconsistencies from the location attributes of the input administrative data were verified and validated regarding the geographic coordinates (georeferencing). Validation procedures for address data will be further implemented.

Activity (A) and Task (T)	Descriptive Summary
(T)4.4.1.1	The Geo-Information Unit assessed the positional accuracy of the geographic coordinates when georeferencing in a GIS environment. An edition classification was designed to label the level of spatial precision of the received geographic coordinates based on the most accurate location (centroid of the building according to the Georeferencing Protocol).
(T)4.4.1.2	The geographical validation of the facilities' georeferenced points enabled cross-checking whether the delivered administrative code (municipality) was correct or incorrect and whether further changes needed to be made in the calculation of some territorial indicators (the ones supported by routing/network analysis and accessibility/demand geographies from the point-based georeferenced data).
(T)4.4.1.3	The detected inaccuracies regarding the geospatial data of the administrative records (geographic coordinates from the location attributes) are documented and reported in HTML files (by type of facility) for internal use, including the statistics (number and percentage of facilities) associated with the type of edition of the georeferenced points.
(T)4.4.1.4	The detected locational errors are corrected and duly updated in the geospatial buildings dataset (BGE) within the geospatial data repository. Updates concerning other databases within the data architecture (e.g., FNA) are also ensured from the collaboration between the involved units.
(A)4.4.2	The preliminary validation (from the attribute 'municipality') and summary statistics retrieved from the georeferencing and editing process (based on the type of edition classification) enabled checking the geocoding matching rate.
(T)4.4.2.1	No minimum matching rate was defined due to the requirement of having all administrative records (facilities) attached to a verified and accurate geocode (coordinate and descriptive address).
(A)4.4.3	No geospatial data loading and extracting services were used in the context of the project.
(A)5.1.1	The geospatial data (geocodes with the geographic coordinates, municipality and address description) were already attached to the administrative data records transferred by the data providers/administrative data custodians. Nevertheless, statistical-geospatial data integration applies in the project context in the next steps to join the descriptive information of the facility (e.g., services, institutional nature, etc.) with the georeferenced point data of the associated building in supporting the calculation of the territorial indicators and for display in the CE-SIG dissemination platform (pop-up information boxes).

Activity (A) and Task (T)	Descriptive Summary
(T)5.1.1.1	During the project's development, consistent practices were defined and documented to ensure the matching processes were carried out consistently, namely by designing and establishing a hierarchical and logical coding system of the facilities and respective services that support the master dataset (BIES).
(T)5.1.1.2	The geographic validation process identified and reported the cases of the facilities that were assigned with the incorrect geocode (e.g., incorrect municipality code due to inaccurate geographic location). These situations were documented and procedures to address them were included in the internal technical documentation related to the project.
(T)5.1.1.3	No data integration services were used in the context of this project.
(A)5.1.2	The ID of the facility is connected to the unique identifier of the building (BGE) in which location is one of the matching keys enabling establishing functional relationships with the other databases/data repositories of the in-house data architecture and information infrastructure, highlighting the master dataset of the project, BIES. Thus, the geospatial object (point geometry) can be linked to other non-geospatial data where the geospatial object ID is allocated to each record (e.g., the population database containing census data). The coding systems of the buildings and facilities enabled correct data linkage between location data/geospatial objects and the alphanumeric data of the facility and descriptive information of the respective services.
(T)5.1.2.1	This task was performed to check the consistency and validity of the location data attributes (geographic coordinates, address and municipality code).
(T)5.1.2.2	No address linking services were used in the context of the project.

Activity (A) and Task (T)	Descriptive Summary
(A)5.1.3	The geospatial data quality assessment routines were carried out iteratively throughout the Process phase, mainly during the geographic validation process (after the alphanumeric data validation). After georeferencing, the input point data of the facilities (geographic coordinates) were crosschecked with a polygon dataset on territorial units (administrative units) and compared to the territorial codes that have been assigned in the provided administrative records. Sometimes, reverse geocoding was also performed to get the address from the geographic coordinates and validate the one delivered by the data provider/administrative data custodian. The spatial accuracy was also validated using an edition classification to report quality statistics. The address validation process was under development. For some cases, time analyses based on location were also conducted to support coherence analyses regarding time series. The errors identified are reported to the respective data providers/administrative data custodians for correction and/or clarification purposes.
(T)5.1.3.1	The previously described geospatial data quality assessment procedures were implemented routinely to streamline the production cycle of the statistical operation, even when new data is occasionally sent due to feedback routines for reporting and correcting errors.
(T)5.1.3.2	All identifiers are assessed in terms of consistency and uniqueness, including the facility ID, the municipality code and the geographic coordinates (e.g., to check if the two facilities are located in the same building). The Administrative and Business Data Unit is responsible for the codes of the alphanumeric data whereas the Geo-Information Unit assesses the codes related to the location attributes.
(A)5.1.4	The georeferenced data (points of facilities) were reprojected according to the official reference coordinate systems (Mainland Portugal: PT-TM06/ETRS89; Autonomous Region of Madeira and Autonomous Region of the Azores: PTRA08-UTM/ITRF93), established by the DGT. The georeferenced data was also stored and managed in the Esri shapefile format within the geodatabase (feature classes).
(A)5.1.5	No machine-to-machine mechanisms for data integration were used in the context of the project.

Activity (A) and Task (T)	Descriptive Summary
(A)5.2.1	The facilities were geocoded at the parish level (six-digit ID) based on the results of the geographic validation process, i.e. after checking the spatial accuracy of the geographic coordinates and performing crosschecking with a polygon dataset on territorial units (administrative units). Also, when positional changes were required on the point's geographic location, a new high-precision geocode was assigned to the respective facility.
(T)5.2.1.1	No semi-automatic and automatic methods and techniques were used to geocode the administrative data records at the parish level.
(T)5.2.1.2	The high-precision geocodes (geographic coordinates) and municipality codes (area-based geocode) were validated. The identified errors and inconsistencies were further reported to the data providers/administrative data custodians for correction and/or clarification purposes (established feedback routines). It was also requested to include the changes made in the next data transfer, i.e. update their datasets.
(A)5.2.2	Geocoding services built into the GIS software were used for reverse geocoding during geographic validation. The reserve geocoding tool was used to obtain the address description from the provided and validated geographic coordinates (points of the facilities), turning a direct position of the georeferenced point data into an indirect position/descriptive location by using geolocators. This tool also aims to enhance the address data validation process, which is still under development.
(A)5.3.1	The location data (geographic coordinates and municipality codes) were reviewed and validated using the previously described geospatial data quality assessment routines. The address data validation is expected to involve geospatial and alphameric data treatment procedures from the Information Infrastructure Unit and the Geo-Information Unit.
(T)5.3.1.1	Some quality control mechanisms were implemented in the first round of data validation, including the coherence analyses to check suspicious differences between reference periods (from consecutive data transfer), missing information and other inaccuracies that would require revision. Comparing and validating the input data with the existing indicators in the Dissemination Database was also carried out to identify differences in indicator values at the municipality level (values provided vs values published).

Activity (A) and Task (T)	Descriptive Summary
(T)5.3.1.2	The data providers/administrative data custodians transfer the most recent data available, usually concerning the last reference year (in the case of the Education sector, the school year). The descriptive statistics conducted by the Administrative and Business Data Unit review some attributes over the different reference periods. In addition, the currency of the facilities was verified during the geographic validation, i.e. if the facility is still active/open, namely by consulting the website of the facility or directly contacting it. If necessary, questions regarding temporality issues of location data can be forwarded to the data provider/administrative data custodian.
(T)5.3.1.3	The spatial accuracy of the input georeferenced data (point locations of the facilities) was reviewed and validated using an edition classification, including changes at the building level, at the street level and a significant change in the point geographic location. The edition classification is supported by a coding system from 0 to 4 in which 2 and 3 represent a lower spatial accuracy. Code 4 is assigned when the facility building changes over the years, which is very common in nursery schools.
(A)5.3.2	The address validation process was under development.
(T)5.3.2.1	The address validation process was under development.
(A)5.3.3	The geospatial data quality assessment routines were carried out iteratively through the Process phase, mainly during the geographic validation process (after the alphanumeric data validation).
(T)5.3.3.1	The topological consistency of the georeferenced points of the facilities was assessed by matching the geographical location/position of the existing buildings in the BGE or creating new points according to the guidelines and technical specifications outlined in the Georeferencing Protocol (centroid of the building). The crosscheck with the polygon dataset of the administrative units also allowed checking the topological consistency to enable correct data aggregation for the calculation of the territorial indicators. Lastly, the topological consistency of the facilities (buildings) points was also validated to check overlaps between two georeferenced points and store the location (geometry and unique ID) only once in the geospatial data repository.

Activity (A) and Task (T)	Descriptive Summary
(T)5.3.3.2	The geocodes are assigned and validated during the geographic validation process based on the quality assessment procedures previously described. In case of mismatches between the geocode delivered by the data provider/administrative data custodian and the geocode assigned by SP, such situations are reported for data enrichment and future data transfer purposes.
(T)5.3.3.3	The descriptive statistics and coherence analyses performed by the Administrative and Business Data Unit in the first validation round embodied data time-series comparability issues whereas the Geo-Information Unit provided inputs concerning time versioning of the facilities location (life-span of each facility location). Such inputs are used to complete the edition classification during the geographic validation process.
(A)5.4.1	The geospatial data quality assessment routines were carried out iteratively through the Process phase, mainly during the geographic validation process (after the first round of alphanumeric data validation). In particular, regarding this activity, the erroneous values of the geographic coordinates and municipality codes were corrected based on the spatial crosschecking and editing work (spatial accuracy) and further reported to the data provider/administrative data custodians. This activity is relevant to ensure data aggregation from geoprocessing workflows in supporting the calculation of the territorial indicators, namely at the administrative unit levels.
(T)5.4.1.1	The spatial intersection of the georeferenced points (facilities and residential buildings) on the lines of the analysis geographies (e.g., polygon datasets of the accessibility/demand geographies) was checked to ensure the correct spatial allocation of the statistical units when calculating the territorial indicators. For instance, if a certain point spatially intersects the geometric contour of two adjacent geographies, it can lead to double counting and biased territorial indicators.
(A)5.4.2	The outcomes from the geographic validation were used to update the location attributes of the facilities and services dataset (BIES) and the dwellings dataset (FNA).
(A)5.4.3	The address validation process was under development.
(T)5.4.3.1	The address validation process was under development.

Activity (A) and Task (T)	Descriptive Summary
(T)5.4.3.2	The address validation process was under development.
(T)5.4.3.3	The address validation process was under development.
(T)5.4.3.4	The address validation process was under development.
(A)5.5.1	The accessibility/demand geographies were new geographical variables specifically derived for this project. They enabled the calculation of the customised territorial indicators at the level of these functional geographies (unit of analysis) from the georeferenced point data of facilities and population. The territorial indicators also derived new statistical variables registered and managed in the metadata management system (e.g., access time by foot and medium access by car). In total, 153 new geographical variables related to the project were registered in the in-house metadata management system (under the 'Territory' theme), including the ones associated with the territorial indicators (e.g., median access time by car, size class of time on foot, etc.) and typologies of facilities and services (e.g., museum, non-tertiary education facility, nurses, non-teaching staff in tertiary education, etc.).
(A)5.5.2	The statistical units related to the typologies of facilities were derived from georeferencing assigning a precise geographic location (geographic coordinates) to each facility. The accessibility/demand geographies as new geographical variables were created using a geospatial service performing routing/network analysis and with navigation capabilities using authoritative street data.
(A)5.7.1	Data aggregation was performed to calculate the territorial indicators (area and territory types) by the unit of analysis, including the accessibility/demand geographies (e.g., polygon rings by a time range). Thus, population census data at the microdata level was aggregated to sum data by the defined geographical classifications. The data aggregation was performed from the outputs of the geoprocessing workflows and spatial analysis processes, i.e. the buildings intersecting an isochrone were selected and retrieved and the sum operation was operationalised in database management.

Activity (A) and Task (T)	Descriptive Summary
(T)5.7.1.1	The metadata associated with the geographies used for data aggregation were registered, maintained and updated in the in-house metadata management system as geographical categories (with associated versioning) under the classification of the types of geographies included in the statistical operation.
(T)5.7.1.2	The different versions of the accessibility/demand geographies are stored in the geospatial data repository for historical register and visualisation purposes (the users can view the geographies from previous reference years). A coding system related to the accessibility/demand geographies and respective distance/time rings was developed to facilitate the identification and search, including a concatenated code of the year, facility code and distance/time code.
(T)5.7.1.3	The geoprocessing workflow and spatial analysis processes that supported data aggregation used reprojected geospatial data according to the official reference coordinate systems (Mainland Portugal: PT-TM06/ETRS89; Autonomous Region of Madeira and Autonomous Region of the Azores: PTRA08-UTM/ITRF93). The DGT establishes this set of geodetic/cartographic parameters for the national territory. The ETRS89 is the global reference system recommended by the European Reference Frame, a subcommittee of the International Association of Geodesy.
(A)5.7.2	No geospatial-based aggregation services were executed in the context of the project (e.g., spatial merge and dissolve operations).
(A)6.1.1	No traditional statistical or thematic maps were developed in the context of the project. In terms of spatial visualisation, only web mapping services were prepared to be displayed in the CE-SIG dissemination platform.
(A)6.1.2	Web mapping and visualisation services (WMS) were created to display the point facilities, accessibility/demand geographies and territorial units (administrative and statistical units). These services were reprojected to the WGS 84/Pseudo-Mercator, a projected coordinate system used for map visualisation in Google Maps and OpenStreetMap, Bing and Esri.
(A)6.1.3	It was decided to create the web mapping services for spatial data visualisation in a single layer (WGS84/Pseudo-Mercator) due to the lower effort and technical complexity and to guarantee better performance of the dataflow and workflow in the production process.

Activity (A) and Task (T)	Descriptive Summary
(A)6.1.4	The geospatial data outputs on the facilities (points) and accessibility/demand geographies (areas) displayed in the CE-SIG platform are linked to the information about the territorial indicators and associated metadata (with direct access to the metadata management system). The back-office application supports and maintains these production components involving semantic interoperability (e.g., facilities and services typology) and metadata management. However, the GeoDCAT-AP, as a geospatial extension to the generic DCAT, was not considered and applied to facilitate interoperability and ensure data description and maintenance.
(A)6.2.1	The geospatial data outputs displayed via the web mapping services were tested to check problematic overlaps for facilitating data visualisation and interpretation and assess the performance and response capacity of the map server considering multiple user requests. These tests aimed to increase response times experienced by users without compromising the overall performance of the CE-SIG dissemination platform.
(A)6.2.2	The territorial indicators that used the input geospatial dataset (e.g., georeferenced point data of buildings and accessibility/demand geographies) - highlighting the ones derived from the service areas - are validated using geoprocessing workflows (spatial intersection) to get minimum and maximum values and therefore, check if the results of the calculated indicators matched within the identified value intervals. Geoprocessing workflows also enabled the last check of the geospatial reference frame (of buildings) and the boundaries of the service and catchment areas (sample approach) to check if the number of facilities is correct based on the reference frame and date of the geography series. In addition, the access of the georeferenced points of the facilities to the navigation network is verified to ensure the accuracy of the routing and network analysis (e.g., distance by car and OD matrix).
(A)6.3.1	The geospatial data outputs (web mapping services) and territorial indicators (statistical information) to be displayed were visualised in the development format of the CE-SIG dissemination platform to check if the outputs are presented in an easy way to be interpreted by the users (e.g., detect no readable spatial overlaps in which the transparency of the data symbology need to be changed). The key spatial visualisation capability of the CE-SIG dissemination platform enabled addressing the erroneous interpretation of the geospatial statistics and making them clearer to the users.

Activity (A) and Task (T)	Descriptive Summary
(A)6.3.2	The back-office application and the in-house metadata management system ensure that the statistical and geospatial outputs to be displayed in the CE-SIG dissemination platform provide easy access to the statistical information (tables) and associated metadata to better interpret the geospatial statistics (territorial indicators).
(A)6.4.1	No disclosure control methods handling geospatial data and considering related theoretical and technical issues were applied in the context of the project. Also, no spatial disclosure risk issues were evaluated and put into practice for dissemination purposes.
(T)6.4.1.1	No dissemination constraints on data dissemination from the geospatial perspective were assessed, namely regarding the risk of privacy and statistical confidentiality breaches through geospatial data visualisation (e.g., geographic differencing issues by overlapping two or more geographical areas).
(T)6.4.1.2	No traditional statistical disclosure methods and techniques (e.g., perturbative and suppressive) were applied to geospatial data and addressing spatial disclosure risks (e.g., indirect personal data identification by location).
(A)7.1.1	All the (geographical) concepts, classifications and variables related to the project were introduced, managed, maintained, updated and made available in the in-house metadata management system and linked to the CE-SIG back-office application.
(T)7.1.1.1	The versions of both geographical and non-geographical classifications related to the statistical operation were registered and available in the in-house metadata management system. The versioning of such classifications is managed in the same system. They were also publicly available for users in an external environment through direct links provided in information pop-up boxes displayed in the CE-SIG dissemination platform.
(T)7.1.1.2	Versioning, change tables and correspondence tables on geographical divisions related to the project were managed and available in the in-house metadata management system. It included administrative divisions (parishes and municipalities) and NUTS classification (breaking down into municipalities) used for analysis and dissemination purposes.
(T)7.1.1.3	No extensive cataloguing and tagging work involving both statistical and geospatial data domains was carried out to make the outputs in the CE-SIG dissemination platform easily discoverable, accessible and usable by all users. Some preliminary catalogue work related to the in-house geoportal (for internal use) was under development.

Activity (A) and Task (T)	Descriptive Summary
(A)7.1.2	Web mapping services (WMS) - an open and standard-based protocol/interface - were developed to display map images in the CE-SIG dissemination platform (web technology) from the geospatial datasets stored in the in-house geospatial data repository via a server/service provider (no local data storage required in which users do not have to store such data in their devices). Alongside the interoperability benefits, WMS provide advantages regarding spatial visualisation and representation of geospatial data, such as dynamic and on-fly view (users can zoom in and out according to their needs and requests, focus on the specific area of interest, view selected and combined data, etc.) and georeferencing accuracy in which the mapped data is aligned with the real-world geographical locations. Another visualisation advantage is the capacity to process large stored geospatial datasets and deliver them in the form of maps to several users at the same time.
(A)7.1.3	Web mapping services (WMS) are extensively used in other statistical operations and to support the statistical production process at SP, namely for dissemination purposes. The web mapping services are updated according to the level of update of the geospatial data stored in the geospatial data repository.
(A)7.2.1	Web mapping services (WMS) were created for users to visualise the output geospatial data in the form of dynamic maps (as opposed to printed maps). The geospatial data visualisation web services were created using GeoServer, an open-source service designed for geospatial data sharing and supported by the open geospatial community (e.g., voluntary developers) for continuous improvements. Moreover, GeoServer is commonly used to publish geospatial data across the web, including WMS and other web-based geospatial services, since it complies with OGC standards ensuring interoperability across (GIS) applications, systems/platforms, devices and other servers. SP has its own GeoServer institutional account (enterprise use) to enable the connection and access to the datasets in the in-house geospatial data repository in a safe and cost-effective manner (e.g., integration with the security infrastructure of the statistical organisation and permissions management). In this regard, all updates and changes in the geospatial datasets will be displayed in the published maps. No geospatial metadata standards were adopted, such as the 3.0 version of SMDX (introducing issues related to geospatial data and attributes)

Activity (A) and Task (T)	Descriptive Summary
(T)7.2.1.1	OGC web mapping services (WMS) were used to display geospatial data as non-proprietary formats to provide more flexibility and usability of the CE-SIG dissemination platform (more user-oriented) and to support the interpretation of the territorial indicators and associated statistical information. No LOD technology (for classifications) was implemented.
(T)7.2.1.2	The created web mapping services (WMS) follow OGC open standards to ensure interoperability and integration with other systems, applications and devices.
(T)7.2.1.3	The web mapping services (WMS) complied with reference web standards, most of them were developed and established by OGC, the geospatial standardisation organisation, to ensure these services can be integrated with different systems and applications and used by multiple users across the web for sharing, publishing and visualising geospatial data. The ISO 19128: 2005 (web map server interface), the Hypertext Transfer Protocol, the Coordinate Reference Systems (to ensure mapping data is consistently and accurately projected) and the XML and Geography Markup Language (for metadata and requests management) are some examples of the main standards followed by these web geospatial services.
(A)7.2.2	The CE-SIG dissemination platform embodied geospatial data representation and visualisation components, including dynamic maps of the accessibility/demand geographies, search by geographical criteria, layer querying/requesting, zoom by geographic scope (mainland Portugal, Madeira and Azores) and selection of different types of base maps.
(T)7.2.2.1	Only dynamic maps were displayed in the CE-SIG dissemination platform, and not statistical/thematic maps, mostly used in traditional statistical publications and studies.
(T)7.2.2.2	Some cartography and geovisualisation guidelines and good practices were considered based on the know-how and professional experience of the in-house staff with geospatial expertise (e.g., determining the symbology in terms of size, colours, etc.) to ensure map readability.
(A)7.2.3	No additional disclosure control methods and techniques handling geospatial data were applied during this sub-process, nor recommendations measuring disclosure risks and addressing confidentiality issues related to geospatial data were followed.
(A)7.4.1	The CE-SIG dissemination platform is a web-based platform focusing on dynamic spatial visualisation and facilitating the exploration of geospatial statistics, namely the territorial indicators that are supported by user-friendly mapping capabilities.

Activity (A) and Task (T)	Descriptive Summary
(A)7.4.2	The CE-SIG dissemination platform provides dynamic spatial visualisation of mapped output data (via WMS) with mapping capabilities and pop-up information to make the user experience more interactive and customised based on their data needs and geographical areas of interest. Thus, the user will have a better navigation experience and the understanding and interpretation of the outputs will be easier due to the dynamic visual representation of the geospatial data (e.g. identify spatial patterns). Also, the CE-SIG platform is an open, publicly available and accessible web-based statistical product.
(T)7.4.2.1	A technical support document and a user guide are under development and will be available in the CE-SIG dissemination platform via online reading or download. The technical support document addresses a step-by-step tutorial to support the user while exploring the platform, especially for first-time users. The user guide is more guidance material providing conceptual and methodological explanations and summary descriptions of some project production components which aim to help the user interpret the outputs displayed in the platform, including accessibility/demand geographies, and understand how the territorial indicators were calculated. Whereas the technical support document is more oriented to the CE-SIG dissemination platform functionalities and usage, the user guide focuses on the project's conceptual and methodological framework.
(T)7.4.2.2	The CE-SIG dissemination platform displays territorial indicators (geospatial statistics) at the parish level (the lowest administrative unit) and at the level of the smallest service area/catchment area in terms of surface.
(T)7.4.2.3	Although institutional collaboration partnerships with some public institutions and governmental bodies for data provision were established, the promotion of the CE-SIG dissemination platform will be conducted unilaterally by SP.
(A)7.4.4	The user guide embodies some issues related to geospatial literacy to enhance the capacity of the users to spatially visualise and analyse the output data displayed in the CE-SIG dissemination platform. These issues can also strengthen (in the medium to long term) geospatial awareness about the importance of location (geospatial knowledge) and the usefulness of geospatial data and capabilities for official statistics while acknowledging the associated benefits for multiple user communities.
(A)7.5.1	The documentation addressing the technical support and a user guide aims to support the users, namely from the geospatial perspective.

Activity (A) and Task (T)	Descriptive Summary
(A)7.5.2	No geospatial components on the accessibility and usability of the CE-SIG platform were included in user satisfaction mechanisms, such as the responsive level of the mapped data when activated and adjustability in mobile devices. Moreover, no user satisfaction mechanisms related to the project were developed.
(T)7.5.2.1	No user engagement and feedback mechanisms related to the CE-SIG dissemination platform were developed to check and monitor if the product meets identified user needs and requirements.
(A)8.1.1	The geospatial production components supporting the development of the statistical operation (throughout the statistical production process) were not formally assessed nor used as evaluation inputs under the quality management of SP.
(T)8.1.1.1	No statistical and geospatial metadata conceptual elements (including web semantic standards, such as the SMDX 3.0) were aligned and harmonised under the metadata management in the context of the project.
(T)8.1.1.2	No geospatial quality assurance mechanisms and geospatial quality indicators for quality reporting were developed and applied regarding the geospatial production components in the context of the project. Thus, no alignment between statistical and geospatial quality dimensions was done.
(A)8.1.2	No geospatial quality indicators were designed and computed in the context of the project.
(A)8.1.3	No external and internal feedback and suggestions were collected in the context of the project.
(T)8.1.3.1	No user feedback and contributions, namely from quality reports and user satisfaction surveys, focus group discussions or other user-oriented quality assessment exercises, were collected in the context of the project. In addition, no follow-up mechanisms with the users were carried out to regularly collect information about their perceptions for improvement purposes.
(T)8.1.3.2	No suggestions were collected for the different user groups, including geospatial experts and GIS staff. Thus, no customised user feedback mechanisms (e.g., a specific type of user survey) were developed to collect information feedback and measure the satisfaction of the geospatial community.

Activity (A) and Task (T)	Descriptive Summary
(A)8.2.1	No evaluation of the statistical operation was carried out from a geospatial production perspective, and since no geospatial quality reporting is formally established and streamlined in the statistical production of SP (e.g., lack of geospatial quality indicators in quality management).
(T)8.2.1.1	No systematic geospatial quality evaluation procedures and methods were implemented in the context of the project, besides the geospatial data quality assessment routines carried out in the Analyse and Process phases. Also, the outcomes from these quality assessment routines to support the validation of the territorial indicators were not incorporated into quality reports within the institutional quality management framework.
(A)8.3.1	No action plan from the geospatial evaluation perspective that included improvements actions addressing geospatial dataflows, processes and services was produced in the context of the project.

Table 12. Application of the Production Model (Overarching Processes and Corporate Activities) in the CE-SIG project (source: author).

i) Design and implement a statistical-geospatial collaboration and cooperation strategy	The project fosters institutional collaboration between SP and external data providers, including regular data providers - some for survey operations - (e.g., education, health and culture sectors) and new administrative data custodians (e.g., civil protection). However, these collaboration and cooperation relationships in terms of data provision and supply are more addressed to the development of the statistical operation and in the scope of the National Data Infrastructure (focusing on administrative data repositories and having the census based on administrative data as key research line) rather than based on a statistical-geospatial strategy involving the statistical and geospatial communities, and additionally the administrative data community. In this regard, only fit-to-purpose data protocols and formal agreements focusing on data access and delivery issues were established to ensure the execution of the project.
ii) Identify, collect and use reference geospatial datasets from	The project enables the thematic expansion of the NSDI by georeferencing the facilities/services from different sectors by the authoritative sources that are formally assigned to hold, manage and maintain the datasets (data custodianship mandates). It covers a data theme that is increasingly relevant for evidence-based policy-making (data-driven policy lifecycle), for effective resource allocation and

authoritative sources	management where public budgets of governments are getting scarcer and to address economic, social and environmental challenges related to sustainable development. The input authoritative datasets supporting the development of the statistical operation provide basic information, including the location of the facilities and services of public interest (Utility and Government Services, as it is called Theme 6 of INSPIRE). The identified external data providers/administrative data custodians as public institutions and governmental agencies ensure the timeliness and quality of the data (and metadata), facilitate the data supply and access/delivery mechanisms and guarantee responsible data management and maintenance, including preservation, security, privacy and confidentiality issues. In addition, authoritative street data was used to perform the routing/network analysis that provided inputs to calculate the territorial indicators and comprised a type of geospatial data that is only used to produce statistical content, i.e. not directly to geocode statistical/administrative data. Using this enriched authoritative dataset from trusted and authenticated sources (Esri street/traffic data from commercial, community and government suppliers) ensured the quality of the outputs, namely in terms of timeliness and geographic coverage.
iii) Develop and implement a unified approach to statistical-geospatial quality management	No statistical-geospatial quality management framework is streamlined throughout the entire statistical production process as an overarching activity. Thus, the quality management framework adopted in SP does not formally include statistical-geospatial integration capabilities, including quality feedback mechanisms between statistical processes handling geospatial data, services and processes and evaluation of statistical-geospatial services. The Statistical Production Process Manual, an internal reference quality document identifying and systematically describing the phases, sub-processes and main tasks of statistical production, only partially reflects the geospatial-related activities, namely concerning the functional requirements of the IIG for the task execution. This document transposes an operational approach to conduct the various tasks and assign their respective responsibility in the course of the development and operationalisation of the statistical operation. However, the current version of the manual does not include the statistical operation related to the CE-SIG project in which the associated tasks are not reviewed, identified and mapped in the production process matrix and the respective responsibilities are not formally assigned.

iv) Develop and implement a unified approach to statistical-geospatial metadata management	The in-house metadata management system of SP (SMI) only partially supports a common framework to manage both statistical and geospatial metadata elements, including standard taxonomy and vocabulary for ensuring semantic interoperability between the two domains. Nevertheless, the repository encompasses geospatial-related concepts, classifications and variables, including correspondence tables between different types of geographies, versions of the territorial units (e.g., administrative and statistical) and cumulative/aggregation tables that are tagged in the 'Territory' theme. Although some components of the metadata management system are integrated through harmonisation and integration rules, no full alignment between the geospatial and statistical metadata elements is ensured. In the context of the project, around 90 concepts were registered in the metadata management system (in which around 20 are tagged in the 'Territory' theme), more than 15 new classifications (e.g., size classes of distance, type of catchment areas of facilities/services, typology of museums, etc.), 5 new correspondence tables (e.g., level of education, types of hospitals - institutional nature, etc.), 153 new geographical variables and 330 indicators.
v. Manage statistical-geospatial capability development and capacity building	The Geo-Information Unit and the Territorial Statistics Unit provide all necessary technical capacities and knowledge to carry out data integration and geospatial-related activities to support the development of the statistical operation, including georeferencing, GIS programming and mapping skills. Data/information, methodological and technological requirements were also fulfilled in most production components, highlighting the geographic validation of the input data, the systematic dataflows/workflows (geoprocessing) and GIS software solutions for processing and analysing. However, more actions need to be taken to improve capability development in statistical-geospatial integration. It may involve automating (or semi-automating) some dataflows and workflows that still adopt a manual approach (e.g., feedback mechanisms on data quality issues, data integration and management procedures within the information infrastructure) and the use of standard services to efficiently run statistical-geospatial processes and better mainstream respective activities and tasks in the statistical production. In addition, new knowledge and skills in confidentiality issues when integrating statistical and geospatial data and releasing geospatial statistics need to be acquired to properly address related theoretical and technical challenges.

Table 13. Application of the Assessment Matrix in the CE-SIS project (source: author).

Quality Dimension	Q/R	Implementation Degree ND/NI	PD/PI	FD/FI	Explanatory Text
Governance	1		X		At the national level, two main bodies are responsible for the strategic guidelines regarding geospatial data, the Guiding Council of the National Geographic Information System (CO-SNIG) and the Cartography Coordinating Council (CCC). Both bodies are supported by a national legal framework establishing their mission, competencies and responsibilities, regulating activities and transposing EU community law to the national context (e.g., INSPIRE to NSDI). The first body is oriented towards developing the National Geographic Information System (SNIG), namely in designing the general objectives, ensuring good coordination between the members, approving work programmes and providing inputs on national technical standards related to geospatial data (e.g., guidelines for reporting and access). The NGIA (DGT) is assigned as the chair of this strategic body. The second body coordinates the activities of public authorities and services that are legally qualified to produce authoritative cartography. It is composed of 20 public institutions and governmental services, including the NGIA, SP (NSO), Portuguese Environment Agency, Institute for Mobility and Transporte, National Energy and Geology Laboratory, Army Geospatial Information Centre, Regional mapping services of the Autonomous Regions of Madeira and Azores, National Association of Portuguese Municipalities and the five Regional Coordination and Development Commissions, among other. Although formal guidelines, generic objectives and activities related to SNIG, there is no clear, agreed, politically or legally binding national geospatial data strategy with a strong political and institutional commitment going beyond the NSDI vision and its specific geospatial data creation, maintenance and dissemination requirements. Having a national geospatial strategy should not be restricted to the development of the NSDI or European data policy, but rather cover other non-technical aspects related to geospatial data (e.g., innovation and education initiatives).
	2			X	The SNIG is the NSDI and is strategically coordinated by the CO-SNIG and operationally coordinated by the NGIA. SNIG aims to provide an infrastructure to register, search and access geospatial data through visualisation and download services (Geoportal) and ensure harmonised cataloguing and publication of metadata through non-proprietary

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					technological solutions and open standards. Geospatial data are produced and maintained by both public and private organisations in Portugal, although public organisations are obliged to register the geospatial datasets that they create and maintain as well as the associated metadata according to technical guidelines and standards (e.g., ISO 19115 and INSPIRE Directive, etc.). SNIG has been supported by a robust legal framework since its creation in 1990 (Decree-Law No.54/90, 13 February) that established all public institutions are obliged to document all produced or maintained geospatial datasets regarding the national territory or water bodies under national law. Over the years, new legislation has been created to attend to the needs arising from the evolution of SNIG, namely when INSPIRE entered into force. In this regard, Decree-Law No.180/2009 (7 August) transposed the INSPIRE Directive (2007/2/EC) to the national context, created the CO-SNIG and established the National Geographic Data Registry and general rules for setting up SDI in Portugal. The Decree-Law No.84/2015 (21 May) changed the composition of CO-SNIG - defined in the previously mentioned Decree-Law - in order to include new organisations, namely those responsible for cartography and geospatial data in the Autonomous Regions of Madeira and Azores, and allow the participation in meetings by prestigious entities. From this point, the strategic vision of SNIG modifies to an NSDI based on an open data policy that ensures the sharing of geospatial data produced by public institutions and bodies of the government administration at the national, regional and local levels. This vision will lead to the release of a new SNIG Geoportal in 2019 providing direct access to the INSPIRE Geoportal in the process of INSPIRE monitoring until 2020. The Decree-Law No.29/2017 (16 March) changed Decree-Law No.180/2009 to tackle the gaps and deficiencies identified by the EC in the legal transposition of INSPIRE in Portugal improving its implementation in the national context and the functioning of SNIG. The composition of CO-SNIG is also extended to all entities responsible for the production and maintenance of geospatial data. Afterwards, Law No.68/2021 (26 August) transposes the European Directive on open data and reuse of public sector information (EU Directive No.2019/1024, 20 June) into national law and sets the guidelines for the next evolution stage of SNIG, highlighting

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					the HVD and the Implementing Regulation 2023/138 that establishes the provisions of their publication and reuse. Moreover, this Implementing Regulation identifies the HVD, namely in the thematic category 'Geospatial' that includes datasets within the scope of some of the INSPIRE data themes (e.g., administrative units, buildings and cadastral parcels). In this regard, SP and SNIG identified addresses, reference parcels (census geographies and grid) and toponomy as HVD and made them available via web download services. Regarding the case study, none of the geospatial datasets addressing the facilities used in the CE-SIG project is embedded or available in the NSDI, however only geospatial datasets on the administrative units are publicly displayed on the NSDI website for download and with INSPIRE-compliant metadata. The NSDI only provides facilities for some regions and municipalities that do not achieve the full geographical coverage for the entire country, nor thematic and attribute completeness since each public authority sets its specific data model and range of facilities, i.e. exclusively public and not private.
	3		X		The Decree-Law No.29/2017 (16 March) - updating the previous Decree-Law No.180/2009 (7 August) - outlines that CO-SNIG, the body responsible for strategic coordination of SNIG, should approve the financing plans and share of costs for each integrated service in the work programming for the operationalisation of SNIG (Article 5). Also, the NGIA, as the body responsible for setting up, developing and maintaining the SNIG, can propose to CO-SNIG the financial plans, alongside collaboration protocols, when applicable (Article 6). The charging fees for accessing geospatial data and corresponding services, especially having large volumes and being duly maintained and frequently updated by the public authorities, as well as the profits from e-commerce services and licences can also be used to ensure the financial sustainability of the NSDI. The actions supporting the development and operationalisation of SNIG and other SDI in the national context can use grant funding options at the European, national, regional or local levels. The EU financial programs provide funding opportunities for projects that enhance capacity building and tackle problems and gaps related to the national implementation of INSPIRE. Under the statistical system, ESSnet grants and funded projects can contribute to

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					the production, management, maintenance and dissemination of geospatial data in the context of the modernisation of official statistics, namely related to activities enhancing statistical-geospatial data integration. The CE-SIG project was funded by the 2020 Technical Assistance Operational Program for a given execution period, which compromises the financial sustainability in the medium and long term from the perspective of the statistical production lifecycle and continuous development of the statistical operation.
	4			X	Regarding geospatial datasets and services, Decree-Law No.29/2017 (16 March) - updating the previous Decree-Law No.180/2009 (7 August) - states that access to such data services should be public, by any appropriate means for telecommunication (mainly Internet), user-friendly and considering the user requirements while the organisations responsible to produce and share them should ensure interoperability with others SDI (Article 17). Moreover, public authorities shall guarantee that geospatial data and corresponding services (search, visualisation, download, etc.) are publicly available (free of charge), however technical options can be taken to prevent their reuse for commercial purposes (Article 18). Article 20 summarises the restrictions of public access to geospatial data and services via specific (e.g., transformation services) and e-commerce services if the dissemination namely undermines the confidentiality of personal data (or files related to a natural person), commercial/industrial information and intellectual property rights, among other aspects. The outlined restrictions are allowed in specific cases and are always based on a restrictive interpretation to ensure the public interest prevails while balancing transparency and the need to protect sensitive interests. EU directives and regulations on data sharing and protection, namely concerning open data and conditions of reuse and dissemination of data from the public sector, have been transposed into the national legislation and policy framework. Under the Personal Data Protection Action, the GDPR - Regulation (EU) 2016/679 (27 April) -, is an EU regulation establishing strict rules on the protection of personal data and its free circulation, and has a direct application in all MS, including Portugal. This regulation repeals the Directive 95/46/EC (24 October) that had a national transposing by Law

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					No.67/98 (26 October). Later, GDPR was transposed to the national legislation by Law No.58/2018 (8 August) ensuring the implementation of the EU regulation in the national legal system, namely by creating a national supervisory authority (National Data Protection Commission). As an independent administrative body, this Commission monitors and supervises compliance with the GDPR and all national legal and regulatory frameworks on the protection of personal data. In addition, other examples were already mentioned in the previous questions/requirements, namely concerning the Open Data Directive and its national legal application. Bilateral protocols and customised data agreements support the transferring and acquisition of input data supporting the CE-SIG project that comes from the other institutions of the public administration, however such institutional instruments are not legally binding under a regulation/law in force nor formally under a specific data policy. Moreover, all public institutions must have a Data Protection Officer following one of the obligations imposed by the National Data Protection Commission (Article 12 of Law 58/2019). In this regard, SP has its designated Officer to ensure a rigorous approach to protecting personal data when processing it, guaranteeing the integrity, security and confidentiality of the information and ensuring its exclusive use for statistical purposes.
	5			X	SP has a Geo-Information Unit in the Department of Methodology and IT that involves a team of experts in geography and geospatial knowledge with GIS skills. This unit has the following competencies and duties in its organisational structure: i) aims to research and implement innovative solutions, particularly related to remote sensing data and VGI into the in-house geospatial information infrastructure and geospatial integration of administrative data; ii) ensure the development, maintenance and management of the geospatial information infrastructure to support the statistical production process; iii) Design and define technical specifications and develop application components within the scope of geospatial technologies; iv) provide geospatial data services accordingly in INSPIRE for the topic which SP is responsible, and promote and coordinate the implementation of the Directive in collaboration with other national authorities and organisations; v) develop and manage the address data within the National Data Infrastructure and set up the official

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/NI	PD/PI	FD/FI	
					national address database through the use of administrative data sources; v) implement the Geoportal of SP and incorporate mapping capabilities into the various components of the main portal; vi) implement the GSGF in cooperation with the NGIA, NSO, UNECE and UN-GGIM, in the context of the ESS and towards the implementation of the geospatial component in the national data infrastructure; and vii) represent SP in the national and international organisations in the geospatial data domain and related fields, participate in projects promoting the development of the geospatial information infrastructure and contribute to national vision and guidelines for geospatial data. The Geo-Information Unit closely works with other units and departments, especially the Methodology Unit regarding geosampling, the Information Infrastructure Unit for the management of the data repositories and the Territorial Statistics Unit for the development of geospatial statistics. The Geo-Information Unit strongly contributes to the CE-SIG project involving processes and tasks related to georeferencing, geographical validation, management of geospatial datasets and development of geospatial services to be displayed in the dissemination platform.
	6		X		SP has actively participated in several projects and activities to integrate statistical and geospatial data in the last two decades and contributed to several initiatives in the international data integration agenda, especially in the European regional context. SP has participated in the four editions of the GEOSTAT projects (2010 to 2022), embedded into the long-term ESS strategy in the form of ESSnet projects to increase statistical-geospatial capacity in the European context. In this regard, SP strongly contributed to the development and implementation of GSGF Europe, particularly in providing good practice cases supporting the framework structure and common technical guidelines (e.g., grid statistics). SP is also an active member of the EFGS steering committee contributing to the activities of the voluntary organisation, namely the organisation of the annual conferences and producing new materials and methodological documentation. SP is also involved in the actions and tasks under the UN-GGIM: Europe SDG line of work to develop methodologies to calculate the SDG indicators that have a territorial dimension (geospatial data analysis) and share recommendations. SP is involved in the UNECE's task force that aims to produce international guidelines and recommendations for the next

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					2030 Census round regarding geospatial data, GIS technology and small area statistics and promote data integration to support census operations (e.g., disseminate more geographically disaggregated census data). In general, SP has been engaged in regional activities to foster greater integration of statistical and geospatial data by funded projects related to geospatial statistics. However, there is a long journey ahead for SP to fully implement GSGF Europe and other surrounding frameworks, such as IGIF and GeoGSBPM, in which continuous participation in projects and initiatives promoting statistical-geospatial integration within statistical production will be necessary.
	7		X		SP fully adopts the GSBPM in its statistical production process through an internal manual embodying a model matrix at a more detailed level (the task), systematically following the development and operationalisation of statistical operation. CSPA is also partially adopted through the in-house metadata system (a statistical service for managing and consulting statistical metadata), the dissemination database (an interface for accessing structured statistical data ready for dissemination) and the use of the SMDX to exchange data and metadata with Eurostat. GSGF Europe is partly implemented by SP regarding Principles 1 and 2 on reference geospatial information infrastructure, geocoding and data management environment. In addition, Principle 3 is more maturely implemented due to the ESS context related to harmonised dissemination geographies (e.g., NUTS and TERCET territorial typologies). SP also extensively uses OGC standards related to web services (e.g., WMS and WFS) and geospatial data storage (e.g., GeoPackage), namely to disseminate census data at small geographic areas, download the geospatial HVD and support the NSDI (SNIG). Although ISO/TC 211 standards on geographic information are not directly adopted within the geospatial information infrastructure to support statistical production, some are embodied in INSPIRE technical specifications and implemented by SP in the geospatial datasets/services setting up the NSDI (e.g., HVD). Nevertheless, SP is committed and engaged to continuing the implementation of reference frameworks such as GSGF Europe and adopting both statistical and geospatial standards, focusing on open standards and non-proprietary mechanisms for interoperability and dissemination (e.g., machine-readable

Quality Dimension	Q/R	Implementation Degree ND/NI	PD/PI	FD/FI	Explanatory Text
	9				formats and via API). The CE-SIG back-office and dissemination platform are good examples of this goal, particularly in implementing web standards that handle statistical and geospatial data.
Data, Information and Technology	8		X		SP (in collaboration with the NGIA) follows the INSPIRE data themes on Statistical units, Geographical names, Addresses and Buildings as well as the HVD in the geospatial thematic category (national address base, census geography, 1 km^2 grid and places), in accordance the related legislation. These datasets support the NSDI and are publicly available in the SNIG Portal whose strategy endorses compliance with the HVD and EU Directive No.2019/1024 (20 June) on open data and the reuse of public sector information. However, SP does not directly nor binding follow the UN 14 Fundamental Geospatial Data Themes under the 2030 Agenda in which most of the work is carried out by the UN-GGIM Core Data WG in identifying requirements and providing recommendations. Nevertheless, SP is committed to supporting the EC geospatial data requirements (aligned with UN-GGIM: Europe Core Specifications), particularly addresses, postal codes, and utility and governmental services.
	9	X			Portugal does not have an official statistical-geospatial data framework duly implemented by the key national stakeholders, namely the NSO, NGIA and administrative data providers, and used by policymakers. SP is the only public institution committed to implementing the GSGF Europe to produce detailed and harmonised geospatial statistics to support census operations and foster data-driven decision-making and evidence-based policy-making in the context of regional and global policy frameworks. However, the implementation of the statistical-geospatial framework is still at a very preliminary stage in the national scope, highlighting gaps in institutional collaboration, capacity and legislation. To effectively implement this type of framework the active participation, involvement and communication between key stakeholders from the national statistical, geospatial and administrative data communities is a fundamental prerequisite which the NSDI (SNIG) is a good example to follow in the future.
	10		X		Portugal has an authoritative national register on cadastral parcels produced, managed, maintained and released by the NGIA, under the legal framework for the land registry (Decree-Law No.72/2023 of 23 August that came into office on 21

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					November 2023). The NGIA works closely with several public institutions from the central governmental administration, local authorities (municipalities) and topography/GIS companies to support the land registry operations and maintain the cadastral parcels duly updated. All cadastral parcels have an associated geometry with the corresponding ID, area, municipality, parish, section reference and building number. Portugal does not have an official and unique national address register with geocoded records, however some public institutions and companies have their address databases and related management systems. To highlight the Portuguese Tax Authority that manages an administrative address database of all individual taxpayers (singular people cadastre) in the country and it is useful for collecting the municipal property tax for both rustic and urban properties. The address data contains a textual description of the address alongside the geographic coordinates but has several null records. SP has an internal address and buildings/dwellings datasets supporting the census operations and overall statistical production that are managed and maintained according to in-house guidelines and updated from both survey and administrative data. The National Building Database provides alphanumeric records at the dwelling and building levels regardless of whether they are intended exclusively for housing or other residential purposes, i.e. an integrated file of buildings and dwellings in which the buildings contain the register of all respective dwellings. This master file establishes a 1:1 functional relationship with its geospatial data repository at the building level, the Geographic Building Database, through the building code (geospatial object ID). This geospatial data repository provides the geometries of buildings (geocoded points) whereas the master file on dwellings/buildings contains a set of general variables and specific attributes. Moreover, SP made available the 2018 National Address Database in the NSDI as a vector database containing addresses from FNA for mainland Portugal and in compliance with the Annex I Theme Addresses from INSPIRE. This dataset is also part of the HVD identified according to the Implementing Regulation 2023/138 of Directive (EU) 2019/1024 on open data and the reuse of public sector information, Lastly, the Portuguese Postal Service is the private company

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					responsible for assigning, collecting and managing postal codes, individual postal boxes and boxes of postal offices at the national level and has an address dataset containing an exhaustive list of the postal codes and a good data model for addressing. However, this dataset has several issues, including textual gaps in the description of the address registers (e.g., invalid/incomplete street names and door numbers), records only containing locality associated with the postal code and the absence of geographic coordinates. In summary, there is no public institution or administrative data custodian formally assigned to collect, manage and maintain authoritative data on buildings, dwellings and addresses that enables harmonised location data with standardised and consistent identifiers/geocodes built on a national coding system.
	11		X		SP can fully geocode buildings, dwellings and population data records to a point location and consequently assign and aggregate such data to higher geographical levels (grid, census geography, administrative units, etc.) via geoprocessing procedure. These statistical units contain high-precision geocodes enabling the smooth linkage processes. It means any unit record data of population, buildings and dwellings can be georeferenced/geocoded based on the in-house location data framework that spatially puts a specific person/household in the corresponding dwelling and those dwelling in the respective building through a matching codes system. In addition, public facilities and services (e.g., schools, hospitals, fire brigades, etc.) have been georeferenced and geocoded from geographic coordinates and address registers to support specific statistical products and social statistics, expanding the role of geospatial data in more statistical domains. However, this full capacity to geocode/georeference statistical microdata mainly addresses population data from census operations and administrative data sources, and no other statistical/administrative data from the remaining statistical domains. Moreover, some of the received administrative records only contain the administrative unit code at the parish level (e.g., income and house rental data) or do not provide adequate quality requirements (e.g., geographic coverage, completeness, etc.) that enable geocoding/georeferencing to a precise location. In this regard, there is still a lot of room for development and improvement to fully geocode/georeference statistical unit records for other statistical domains (e.g., all

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					business locations for more accurate business statistics and workplace locations for labour statistics). During the development of the CE-SIG project, SP demonstrated the capability to consistently and accurately georeference the input data related to the facilities (points) (according to the in-house Georeferencing Protocol) and apply georeferencing to other statistical domains.
	12			X	Statistical unit record data (microdata) are collected and maintained via both survey and administrative data (e.g., social surveys and administrative data provided by the municipalities on construction works) at the point-based level, i.e. x and y geographic coordinates of buildings location. The points of the buildings have the associated address, however the coordinate point of the address at the door level is not collected or acquired in order to build a more integrated and hierarchical location data framework for statistical purposes. In the CE-SIG project, the input data related to the facilities provided by the specific-domain institutions from the public administration are georeferenced at the point level in the centroid of the (main) building footprint.
	13	X			According to Article 2 (Mission and Duties) of Regulatory Decree No.30/2012 of 13 March, the DGT must promote, in coordination with other entities, the cartographic coverage of the national territory, the preparation and maintenance of the CAOP (administrative geography) for cartographic and cadastral purposes. This Map constitutes the delimitation of the administrative boundaries of the country into parishes and municipalities and is annually published and publicly available based on the legislation approved by the Assembly of the Republic that has the power to amend and set administrative boundaries. The last versions of the administrative map have been done following the ISO standards (coding list and identifiers) and a specific data model. The last version already incorporates the corresponding NUTS. As the NSO, SP is responsible for producing and/or maintaining statistical geographies, namely enumeration areas and small statistical areas for census operations, NUTS (EU Regulation 1059/2003) and other geographies supporting statistical production, namely for analysis and dissemination (e.g., EU Regulation 2017/2391 on the territorial typologies). The legal references previously mentioned in the first questions/requirements endorse the production and maintenance of SP regarding boundary data (e.g., INSPIRE and

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					HVD).
	14	X			SP uses the geospatial dataset containing 1 km^2 grid cells provided by GISCO (Eurostat) as the reference gridded geography to support statistical production. Gridded geography is used in the design and creation of the frame and selection of the sample for data collection and dissemination purposes, namely web mapping applications displaying grid statistics with higher spatial resolution for specific territories and study areas (e.g., 500 meters in cities). In this regard, SP has been increasingly using statistical grids as output geographies for dissemination alongside the traditional administrative and statistical geographies (the last ones for census/population data reporting). SP publicly provided the GRID ETRS89 LAEA 1 km^2 of Portugal (2024) via WMS and for download (data visualisation and extraction services) in the NSDI, constituting an HVD. This dataset also comes from the participation of SP in the GEOSTAT projects to allow the dissemination of geospatial statistics associated with new global geographic references and enable comparative cross-border analysis between the various EU countries. At this stage, none of the CE-SIG indicators is calculated and disseminated at the grid level, however it is expected to include this type of geography for statistical analysis and dissemination purposes.
	15		X		Although SP has no extensive background in applying statistical disclosure control techniques in grid data and small statistical areas, such methods for statistical confidentiality and data protection handling geospatial data have been mainly focused on census data (demographic data) and derived grid statistics. According to international recommendations and best practices, SP uses targeted record swapping, a pre-tabular (perturbative) method traditionally applied to microdata, and the cell key method, a post-tabular (suppressive) method commonly applied to the table cells, by adopting a mixed approach on confidentiality methods. In the 2021 Census, the statistical method 'targeted record swapping' was applied to protect the privacy of individuals and ensure that descriptive information is not obtained from the data disclosed. This method guaranteed that the population and its main characteristics remain unchanged for the different geographic

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/NI	PD/PI	FD/FI	
					levels. However, methods managing geographic differencing issues have not yet been formally implemented in the disclosure control business processes of SP as they are difficult to operationalise (e.g., grid cells versus administrative units) due to computing and complexity considerations. The random noise cell-key method (post-tabular method) already used by other statistical organisations from the geospatial perspective (i.e. modify the number of persons in grid cells) for demographic grid data can be a reliable option. In addition, investing time in research and development in ML methods for automatic confidentiality processes, namely for small area statistics, to efficiently reduce disclosure risks, can be an improvement action. Furthermore, it is a methodological and IT challenge to manage and find an appropriate trade-off between an acceptable level of confidentiality and privacy risk and the usability of the data, while avoiding differences between data for the same territorial units in order to maintain data consistency and minimise significant loss of information.
	16		X		The geospatial technical capabilities (data, information, technology and other geospatial technical components) are broadly used throughout the statistical production of SP mainly supported by the IIG. Geospatial data, services and processes enable assessing and designing new statistical products, sampling design and methods, data collection, data processing (e.g., data aggregation and spatial analysis), data integration and data visualisation/dissemination. The centralised storage of geospatial data and corresponding technology provides an operating model for the production, management and sharing of geospatial data, metadata and services across the organisation to produce geospatial statistics. At SP, geospatial-related and statistical-geospatial data integration activities are performed in 15 out of 44 GSBPM sub-processes (34.1%) highlighting the Design, Collect, Analyse and Disseminate phases. Moreover, geospatial data and functional requirements and components of the IIG are present in most of the production process phases in the context of a survey. 47 of the 129 tasks (36%) outlined in the task matrix of the internal statistical production process manual include geospatial technical capabilities for the development of some statistical operations. However, geospatial technical capabilities can be more extensively used by other units,

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					statistical operations and activities supporting the statistical programmes, namely for configuring workflows, reviewing and validating data, validating outputs and applying disclosure control methods.
	17	X			The centralised storage of geospatial data (geospatial data repository) and the corresponding technological environment (geosystem design architecture) provides an operating model for the production, management and sharing of geospatial data, metadata and services across SP to support the statistical production and development of certain statistical operations and products. This geospatial data repository is the cornerstone of the overall IIG of SP. The statistical organisation's staff can use geospatial services internally to help them with their business processes. For instance, the experts from the Methodology Unit can easily visualise geospatial data (e.g., grid cells and territorial units) to assist them in the frame and sample design and during data collection for sample frame management through an interactive mapping application. In addition, SP also provides a customised web mapping application for external users, usually administrative data providers (e.g., Energy Agency), to enable them to visualise some geospatial datasets (e.g., building points) supporting their activities.
	18		X		The Geo-Information Unit provides a series of internal geospatial services to support statistical production and corresponding business cases, highlighting sampling design and data collection for all survey operations (GEOINQ web platform with a set of functionalities). Mapping services and dashboards for operating and dissemination purposes (e.g., environment and education) in several statistical domains are also provided to support some statistical activities, namely to visualise geospatial data (map viewer). Besides the application services extensively used in SP, such as web mapping services (e.g., WMS), download services (OGC GeoPackage) and business services built-in in the GIS software (e.g., coordinate transformation, geocoding, aggregation, routing, etc.), there are still many types of geospatial services to be developed and implemented to support statistical production. This is particularly important for modular services for geospatial quality reporting and feedback between statistical and geospatial business processes, for disclosure control and for address data linking, editing, and standardisation wherein an official national address repository is primarily required. A

Quality Dimension	Q/R	Implementation Degree ND/NI	PD/PI	FD/FI	Explanatory Text
					geospatial service to manage, display and visualise all historical administrative, statistical and functional geographies used in official statistics and track the boundary changes (geographical boundaries timeline) is also needed for internal and external users. In this regard, more types of geospatial services need to be designed and built on the existing geospatial information infrastructure of SP and further implemented to effectively support several phases and streamline both statistical and geospatial business processes within the statistical production. Priority should be given to open, interoperable and standardised geospatial services widely agreed upon and adopted by the national statistical and geospatial communities.
	19	X			The IIG of SP, which is underpinned by the in-house geospatial repository, is connected to the population, dwellings and business registers via the ID of the building. However, either geocoding/georeferencing and (dis)aggregation (spatial analysis) capabilities through modular/shared services are not designed and implemented within the production process to enable effective functional relationships between geospatial and statistical processes and related infrastructure (information, IT systems, technologies, etc.). Not all the connections between statistical processes (using geospatial content) and geospatial processes are streamlined nor standardised, namely regarding quality management and dataflows and workflows (input-output schemas). In this regard, more automation and interoperability between in-house statistical and geospatial data architectures and supporting infrastructures is needed to ensure that a change in a geospatial object is directly reflected in the corresponding statistical object, and vice versa.
	20		X		SP conducts some geospatial data quality assessment procedures and validation routines, especially regarding new point buildings collected from surveys between census operations, however they are not properly streamlined in statistical production and/or are duly documented. In this regard, there is no systematic geospatial data assessment approach extensively implemented within a comprehensive geospatial quality management and reporting framework but rather ad hoc corrections to handle erroneous data for a specific purpose. The main quality control routines are conducted under the in-house geospatial information infrastructure for editing the

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					acquired and stored data and are mostly performed by GIS-based processes and tools allowing the identification of topological and attribute errors for census geography (statistical subsections, sections and localities) and buildings. Systematic spatial cross-check and data verification procedures are also implemented in the context of the Indicators System of Urban Operations that consists of administrative data on buildings and dwellings permits and completed construction works, dynamically maintained and provided by the municipalities. These regular validation procedures evaluate the location data (x and y coordinates) of the completed buildings' registers sent by the municipalities, which includes assessing them against the in-house addresses database from FNA, before inserting the new building point in the final spatial dataset (BGE) or updating existing ones. When erroneous coordinates are detected, such records are back sent to the respective municipalities reporting the errors to be further corrected at the source ensuring feedback routines between SP and the data providers/custodians. Thereafter, this system enables the continuous update of the spatial dataset on residential buildings between census series and ensures that the stock of information on buildings and dwellings is timely for the census operations and other statistical surveys, namely by adding newly constructed buildings and dwellings and excluding the demolitions. Additionally, the address data (descriptive information) supporting sample frames and surveys are validated (along with the georeferenced points) via the geoportals of some municipalities, an internal tool to consult postal codes, GIS capabilities and open mapping applications. These validation routines also support the management and maintenance of FNA regarding address data updates. In addition, point-to-polygon operations are performed to check the spatial intersection of point buildings within administrative and statistical geographies for coding validation and record of the historical series. For instance, this could be applied to check if a certain building point spatially intersects the boundary line between two areas or to have its geographical background based on the overlapping of the different versions of geographies.
	21		X		SP publicly shares and publishes geospatial data concerning the census geography with the main statistical variables from the census population data, namely an OGC GeoPackage file

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					and QGIS plugin to download reference data from the 1991 to 2021 census at the national, regional (NUTS) and local (municipalities) levels. These census dissemination products alongside the web mapping application 'GeoCensos', which enables exploring data, visualising census geography, creating thematic maps or extracting shapefiles, were internally developed by SP considering a standard-based and open data access approach towards a more user-friendly perspective. However, not all geospatial data supporting statistical production and geospatial statistics are fully disseminated and broadcasted in the same accessible manner as previously described, namely for confidentiality issues. Moreover, SP does not have a Geoportal (or a geospatial statistics portfolio) - as happens with other statistical organisations - that could display statistical information via geospatial services (e.g., data extraction and integration, thematic map visualisation, etc.), publish geospatial datasets for download and make available geospatial statistics products for the broader audience from the several user communities in a more dynamic and innovative way.
Institutional Collaboration and Capacity	22			X	SP has three key institutional protocols concerning geospatial data access and sharing with public institutions and regulatory bodies. Firstly, SP has a collaboration protocol with the NGIA (DGT), supported by the MoU that establishes an agreement on the sharing and exchange of both statistical and geospatial data relevant to the activities of both organisations. It includes datasets such as the administrative units, INSPIRE spatial data themes on toponomy, addresses, statistical units and buildings, human health and safety and population distribution, some publicly available in the NSDI. Data sharing and exchange between both institutions are aligned with the legislation in force regarding statistical confidentiality and personal data protection, following Law No.22/2008 of 13 May on the NSS, the GDPR and Law No.58/2019 of 8 August, that ensures the national legal implementation of the Regulation (EU) 2016/679 of the European Parliament and the Council of 27 April 201, i.e. GDPR. Secondly, SP has a cooperation protocol with the National Authority of Communications, an independent administrative and regulatory body in the communications sector, including electronic and postal communications. This protocol aims to foster the synergies between both organisations regarding a national framework for building georeferencing to support the

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					geographical survey of the network coverage and follow best practices in sharing information of public interest, particularly in terms of data security, integrity and confidentiality. In this regard, SP provides the BGE (georeferenced dataset of buildings) - preferably less than semesterly - with certain attributes, namely the unique code, x and y coordinates for the different project systems, address with toponomy and the number of dwellings. This geospatial dataset is sent in either GeoPackage or shapefile format, outlining the intention of sharing such data via WFS in the future. The National Authority of Communications should ensure the update of the respective geospatial data, whenever it has relevant information, and collect up-to-date information from the operators on residential buildings, commercial and industry facilities and agricultural holdings to further share with SP. Thirdly, SP establishes a collaboration protocol with the Energy Agency, a legal entity with a public utility that promotes and carries out activities in the field of energy and builds connections with policies from other domains, namely regarding the efficient use of water and energy efficiency in mobility, industry and buildings. In this regard, the administrative data about the buildings' certification provided by the municipalities to SP after construction works, i.e. completed buildings (Indicators System of Urban Operations) thematically overlaps with the Building Energy Certification System and processing of associated statistical data, managed by the Energy Agency. Thus, sharing data and information on this topic is fundamental among parties, especially in the scope of Portugal's Digital Agenda. This protocol aims to establish a technical partnership between both organisations regarding the access and sharing of data and information on energy certification of buildings, mapping, georeferencing, and description of buildings and dwellings, as well as promote their joint use and other related collaboration actions. SP should provide the georeferenced building points via web mapping services with the identification of the internal building code, census building code, INSPIRE building code, x and y coordinates according to the reference projection systems (defined by the NGIA) and the conformal project of WGS 1984 Web Mercator Auxiliary Sphere (used by Google and other big companies for mapping). In addition, SP should provide a list of descriptive variables of the buildings and dwellings, including some address attributes, through a web service

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					(electronic data transfer via HTTP in XML format). These protocols and collaboration actions involve clear duties and responsibilities in creating, managing, maintaining and sharing the identified geospatial datasets and associated statistical data while preserving the principles of integrity and confidentiality and considering their public value for the respective institutional missions and activities and several sectoral policies. Nevertheless, more collaboration protocols involving institutional agreements for the provision and sharing of data and information should be promoted and established with other external organisations and data providers, especially administrative data custodians, while extending to other core geospatial data themes and datasets that are relevant for statistical purposes (e.g., geocoding) and public use (e.g., HVD). Lastly, SP also has a long-lasting commercial partnership with a GIS service provider for software license maintenance and technical support.
	23			X	SP has a long-term institutional partnership with the NGIA (DGT) from both strategic and operational perspectives, supported by consecutive collaboration protocols, data agreements and technical support actions addressing the common work synergies and activities of both institutions. Such strategic and technical alliance has been mostly underpinned by the sharing of data, information and resources and policy development in the context of governance issues. The institutional collaboration and cooperation between the NSO and NGIA have been mainly about authoritative and core geospatial and non-geospatial datasets from different thematic areas for statistical and cartographic purposes (e.g., address registers, georeferenced data on facilities, statistical indicators, etc.). The lines of action under the established institutional collaboration framework between both institutions have been underpinned by EU policy frameworks (e.g, Green Deal, HVD, etc.), statistical programmes (e.g., census round and land use/land cover statistics) or international initiatives on geospatial data management (e.g., UN-GGIM). SP and the DGT work together in the development of the NSDI (SNIG), as part of the strategic coordination group, towards an open data policy that guarantees the free supply of geospatial data produced or held by public administration entities as well as in the maintenance of the IIG to support statistical

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					production and statistical operations of the NSS. In this regard, it is important to highlight the bilateral work in the implementation of and compliance with the INSPIRE Directive in the development of the national infrastructure in the field of geospatial data, particularly in the components that are relevant for both institutions. The follow-up of the activities related to statistical-geospatial data integration has also been playing a guiding role in the collaborative actions between both institutions, namely regarding the implementation of GSGF (Principle 1), development of new statistical operations and production of territorial indicators for monitoring spatial and urban planning. Moreover, harmonising concepts, methods and procedures in the statistical and geospatial production and respective business processes (e.g., standards), setting up specific project teams, establishing data exchange routines and service integration (cooperative data and service sharing chains) and providing technical training for modernisation and addressing capacity needs and gaps have been other lines of action under the institutional collaboration and technical cooperation efforts between the NSO and NGIA. Towards the vision of a national data infrastructure and a more efficient census model (use of administrative data), SP has fostered a more multi-disciplinary institutional collaboration and cross-sector partnerships with administrative data custodians and providers of other sources within the broader NSS over the last few years. These institutional arrangements with public administration entities, aim to support research and development initiatives in exploiting emerging data sources for statistical purposes and provide microdata and administrative records for more timely and disaggregated statistical outputs and innovative products, namely by fully or partially replacing survey collection. It will also enable reducing costs of the census and survey operations, minimising the statistical burden and increasing the frequency of information while following international guidelines regarding population and housing censuses. The work carried out related to experimental statistics (StatsLab - Statistics in Development) and the future Resident Population Database are good examples to demonstrate the relevance of establishing such strong partnerships. The first example addresses statistical products from ongoing projects that use new data sources and methodologies (e.g., AI and ML

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					techniques) are disseminated. It includes topics that are relevant for economic and social analysis, such as social economy, income, mobility, employment and transport, among others. The second example constitutes the main project of data integration to go from the traditional census model to the register-based model for producing annual census statistics based on administrative data from different sources and custodians (e.g. civil register, social security register, tax register, private employment register, social protection for public servants, etc.).
	24		X		SP, particularly the Geo-Information Unit, has some internal guidance materials and methodological/technical documentation regarding databases available in the geospatial repository and functional relationships with other data repositories, INSPIRE data themes, methodological procedures for certain projects, spatial analysis tools and technical information about the geospatial services (e.g., mapping services in the GIS server). However, these guidance materials and documentation on methodology addressing the geospatial component in statistical production are not produced and maintained in a consistent and systematic manner (e.g., absence of business activities related to methodical documentation), nor are formally incorporated into the quality management processes of the statistical organisation. It underlines a lack of internal benchmarking work to produce documentation for the production and dissemination of geospatial statistics, namely covering input data quality, metadata, geocoding, spatial analysis, disclosure methods, standards, and reference coordinates/projection systems, among others. In this regard, there needs to be more practical documentation outlining and describing geospatial dataflows and workflows supporting statistical production to standardise processes and make the performed tasks more consistent and easily replicated, saving time and increasing productivity. Having the geospatial capabilities for the statistical business processes well-documented with methodological detail, such as step-by-step tutorials and technical reports, will ensure that the outputs will be coherent, comparable and with higher quality while avoiding errors, operational disruptions and knowledge loss over time. Also, drafting guidance notes, summary documents, methodological papers and handbooks containing technical information relevant to the statistical/geospatial

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					processes are valuable instruments to validate and verify data, monitor processes and services and assign responsibility and accountability on the different tasks (e.g., geocoding). Nevertheless, in the CE-SIG several guidance materials and methodological documentation were produced and updated during the development of the project. This documentation collection addresses the background of the project (e.g., goals, institutional environment, concepts, main functional components, etc.), workflow and dataflow for validation, processing and integration data sources (administrative data), geospatial dataflow and validation (Georeferencing Protocol and technical specifications on geospatial datasets and services), descriptions of the databases and their functional relationships, and procedures for calculation of indicators, among others topics related to the production process.
	25		X		SP has a permanent staff team with geospatial expertise in the Geo-Information Unit, and some other staff members with similar skills in other units. This staff have the skills, knowledge and competencies required to undertake geospatial-related and data integration activities and tasks supporting statistical production, namely in terms of GIS software, geocoding statistical/administrative data, geoprocessing tools, geospatial database management and maintenance, metadata registration (INSPIRE), spatial analysis and statistical cartography (dissemination of geospatial statistics, including mapping web services). However, there are some expertise gaps related to geospatial frameworks, standards, confidentiality methods for handling geospatial data and advanced EO data and processing. These identified gaps require further domain-oriented professional training initiatives to up-skilling and gain experience in specific areas, collaboration with related experts and actions to improve such capabilities (e.g., investment in specialised software and IT) and endure capacity and education development (e.g., enhance know-how in emerging geospatial technologies).
	26	X			In general, most of the statisticians and other non-geospatial experts at SP do not have a minimum or acceptable level of geospatial awareness and literacy, highlighting the lack of willingness to embrace the concept of integrating statistics and geospatial data and the general mindset of only acknowledging geospatial capabilities for census operations and statistical cartography. Nonetheless, the senior statisticians who closely work with the Geo-Information Unit

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					present some level of awareness of the geospatial potential to produce official statistics, namely regarding the capabilities for geosampling, spatial analysis and visualisation of statistical outputs (e.g., web mapping services and applications). This type of internal collaboration and working relationships in daily activities and tasks enable non-geospatial experts to be acquainted with some geospatial concepts and practices and acknowledge their benefits and usefulness for statistical production, which consequently facilitates bilateral work making it more inclusive, streamlines relationships between statistical and geospatial pipelines and enhances statistical-geospatial integration. In the context of the CE-SIG, this type of gap in geospatial awareness and literacy was verified across the different staff teams and units working on the development of the project, especially during technical meetings and decision-making processes. It is appropriate to outline insufficient and ineffective organisational support and understanding of geospatial data, processes, services, technologies and other related capabilities due to the lack of knowledge in these topics, receptiveness to have learning experiences and/or motivation for training initiatives.
	27		X		SP has permanent and occasional user satisfaction assessment initiatives to identify their needs. Such initiatives follow SP's Quality Policy (Principle 4 - Commitment to Quality in the Quality Charter regarding the relationship with users) and the mission and duties of the Statistical Council of Portugal (Article 8 on Statistical accessibility of the Law No. 22/2008, 13 May regarding the NSS in which paragraph 2 states that 'official statistics are a public good and must meet the needs of users in an efficient manner'). In this regard, some tools and activities are carried out addressing users under the Quality Management System of SP, including users' satisfaction surveys concerning the available statistical dissemination products, provided services and practices of revisions (according to the ISO 10004:2008 - Customer satisfaction) and services for compliments, suggestions and follow-up to data/information consumers and general users. Satisfaction surveys can occasionally occur for specific user groups and other types of activities related to the satisfaction of users can be carried out, whenever appropriate (e.g., more informal focus group discussions). In addition, tailor-made responses to specific user requests (customer

Quality Dimension	Q/R	Implementation Degree			Explanatory Text
		ND/ NI	PD/ PI	FD /FI	
					support and personalised service) and general user support are available regarding customised geospatial data provision and fit-to-purpose spatial analysis and statistical cartography for research purposes (while ensuring confidentiality and security issues). These evaluation mechanisms and satisfaction assessment tools from the user perspective aim to identify and measure new and emerging needs in users in order to meet their current and future needs. The results ensure statistical quality, produce statistical outputs more relevant to civil society, develop convenient and innovative ways for statistical data and outputs to be easily and equally accessed and visualised by the public in general, draft improvement actions and adopt management measures.
	28	X			SP does not formally have an action plan focusing on statistical-geospatial integration to plan, develop, improve and monitor capabilities and build capacity in this area, namely capability priorities and solutions as well as improvement measures to more efficiently streamline statistical and geospatial data, processes and services in statistical production. Neither is a long-term action plan in the form of a normative document outlining the vision, mission and strategic goals of the statistical organisation concerning this topic and expressing the institutional commitment to adopt a reference framework, apply a standard or follow specific guidelines and recommendations needed to support statistical production from the geospatial perspective. Although SP participated in several international projects related to statistical-geospatial integration over the years and has played an active role in modernising the production of geospatial statistics in the ESS context, an action plan assessing the in-house statistical-geospatial capabilities and/or for the implementation of a framework, highlighting the GSGF, is yet not available at the organisational level.

IV. RESULTS

The application of the two operational parts of the methodology provided an extensive analysis of the case study concerning the statistical production process underpinning the development and operationalisation of the statistical operation (Production Model) and the overarching statistical-geospatial strategy, maturity and quality framework (Assessment Matrix) which extends beyond the institutional environment of the statistical organisation.

The output information generated through the methodological application and corresponding analytical exercise and its interpretation enabled the formulation of summary notes and remarks concerning the development and modernisation levels, key gaps and improvement areas. The first part of the results related to the Production Model are summarised by statistical production phase and overarching processes/corporate activities (GSBPM) to facilitate their analysis and understanding as well as support a more targeted implementation of the improvements and recommendations, as detailed below:

Specify Needs

- The drafted documentation (descriptive memory document and surrounding technical materials supporting the functional framework of the project) provided an overview of the identified and examined statistical and geospatial production needs and background to endure the development of the statistical operation, highlighting the relevance for territorial-based sectoral public policies. However, the statistical needs have been identified in more detail than the user needs related to geospatial data, services and products.
- No consultation actions were carried out with external stakeholders from the statistical and geospatial communities within the national operating environment, in which only the potential data providers and administrative data custodians were called in to discuss data/metadata availability, requirements and delivery issues. No consultation and engagement actions oriented to the users were conducted in the project context.

- Only authoritative data (survey data and administrative data) were considered for the context of the project, including georeferenced data (administrative data with location attributes, including geographic coordinates and addresses). However, the guidelines, technical requirements and recommendations of global or European geospatial data frameworks were not reviewed nor considered when checking new data themes and datasets to support the development of the statistical operation.
- The input data characteristics and requirements were discussed and agreed on with the data providers/administrative data custodians when needs were consulted and confirmed, and available data was checked.
- Different types of geographies for analysis and dissemination were discussed and identified, including administrative and functional geographies based on accessibility metrics and specifically created for the statistical operation. Gridded geographies (1 km^2 grid) will be used to support the calculation of some territorial indicators (e.g., centroids of the grid cells to generate origin-destination matrixes by car).
- No confidentiality and data protection issues were considered during this stage.
- New concepts and definitions were identified to improve metadata management (e.g., data elements, documentation and information needs, etc.) and facilitate the outputs' interpretation and understanding by the end users. New statistical data, measurement units and (geographical) classifications and variables required new associated concepts to ensure alignment with existing standards in the organisation.
- The action plan of the project described the business case, including the production components that already exist in statistical production, the technical and non-technical capabilities that can be provided by the statistical organisation to produce the outputs, the timetable and the budget.

Design

- The drafted documentation provided the description of the design activities supporting the development of the statistical operation, including dataflow and workflow configurations (e.g., geoprocessing workflow), production systems (information infrastructure and IIG) and design elements supporting the CE-SIG dissemination platform, such as information needs and metadata. The existing technical capabilities were also identified, including those related to the available in-house datasets and geospatial technological components (complied with technology standards from the geospatial domain). The experience from past accessibility studies conducted at the national and European levels provided methodological soundness to design the business processes supporting the production of similar outputs, i.e. territorial indicators based on accessibility and demand metrics.

- The design of the outputs only considered traditional statistical disclosure control methods in terms of confidentiality and privacy concerns, particularly for the facilities having a private institutional nature.
- The statistical unit level addressing the facility and respective services has a point-based location based on the geographic coordinates in the administrative registers and survey data. The input georeferenced data (points of facilities and residential buildings) enabled the creation of new and flexible variables based on the accessibility/demand geographies and the calculation of the indicators based on area and territory.
- Geographical validation routines and geospatial data assessment procedures were designed, including reporting the errors directly to the data provider/administrative data custodian and validating the delivered data at the source. In this scope, a Georeferencing Protocol was developed for internal use and shared with the external stakeholders who were delivering/transferring the data.
- No geospatial services were designed to collect the geospatial data. The survey data are collected and administrative data are delivered with location attributes already attached to the alphanumeric data, containing the geographic coordinates, address information (street, postal code, locality) and administrative code of the municipality.
- The processing and analysis components were designed and described in the technical documentation, highlighting the dataflow of the statistical and geospatial data (validation routines and functional requirements), geoprocessing

workflows and geospatial data visualisation services. These production components were discussed and agreed upon by the involved units and departments throughout several general and bilateral technical meetings.

- Some production components were designed and developed externally by an outsourcing service, such as the CE-SIG back-office application and dissemination platform.

Build

- No modern geospatial data collection and loading were built or reused to support data collection and acquisition processes of the statistical operation.
- Geospatial services were extensively built and reused to support geoprocessing workflows, spatial analysis and geospatial data visualisation. Some of these geospatial services were identified from previous benchmarking exercises and comply with internationally agreed OGC geospatial standards for mapping and sharing georeferenced data online (e.g., WMS). In-house GIS technology and geospatial technical capabilities from the IIG were used.
- The components of the dissemination platform overlooked confidentiality issues related to geospatial data and associated risks on the outputs released (e.g., geographic differencing by regional breakdowns).

Collect

- Preliminary validation routines regarding alphanumeric data were conducted and internally reported via HTML files, including descriptive statistics and coherence analyses with the previous data from the dissemination database (e.g., time series analysis). Such HTML files can be reported to and openly shared with the data providers/administrative data custodians for future improvements.
- Geographical validation procedures were carried out to identify errors and inconsistencies related to the location attributes (geographic coordinates, address information and administrative code) and were directly reported to the source. Traditional spatial analysis techniques and approaches to assess input geospatial data were performed. The corrections were inserted and updated in the geospatial data management environment. The geocoding matching rate was

calculated when conducting geographical validation and the editing process in a GIS environment.

- No geocoding services were used to collect input georeferenced data. The acquired data already contains high-precision geocodes (geographic coordinates) assigned to each administrative unit record. Official reference coordinates systems were adopted.
- The geographic database of residential buildings stored in the in-house geospatial repository (IIG) was used to verify and validate the input georeferenced data and to make sure the location is only stored one time (avoiding overlapping points). If the facility already exists in a building represented in the BGE no point is created, whereas if the facility is located in a building that does not exist in the BGE (usually isolated buildings with exclusive non-residential use, such as hospitals and schools) a new point is created.
- The Georeferencing Protocol was a key guidance material during the acquisition of the geocoded administrative data to ensure consistency and quality over the consecutive data deliveries/transfers from the data providers/administrative data custodians.

Process

- The integration of statistical and geospatial data was ensured by the data providers/administrative data custodians in which the validation and matching processes were conducted by SP with close collaboration with the external stakeholders. In this project, statistical-geospatial integration focused more on joining the descriptive information of the facilities and services with their spatial representation (georeferenced point data) for calculating the territorial indicators and displaying information in the CE-SIG dissemination platform (pop-up information boxes).
- A new hierarchical and logical coding system of the facilities and corresponding services was developed to guarantee functional relationships between the different data repositories within the data management environment and the information infrastructure (e.g., BIES and BNE) feeding the dataflow and

workflow of the project. Location data is stored at the statistical unit record level with unique and standardised identifiers (BGE), i.e. building ID is linked to the unique code of the facility based on the sector, hierarchical level of the facility and service.

- Geospatial data quality assessment routines were applied throughout the Process phase iteratively with other units and departments involved in the project, including feedback mechanisms to report and correct errors in the location attributes compared to the previous period. An edition classification was established to assess the level of spatial accuracy of the input georeferenced data points.
- Geographical validation was carried out before performing the geoprocessing workflows and spatial analysis for calculating the territorial indicators (accessibility/demand geographies) to ensure the correct geographical position of the geospatial objects on the facilities and avoid biased outputs.
- No data integration and location data linking services were used in the context of the project, including machine-to-machine and automatic solutions.
- Beyond the validation of the received geocodes (geographic coordinates, addresses and municipality code), the geocode of the administrative unit at the parish level was assigned after geographical validation (first validation round alongside alphanumeric data validation). Geocoding services built into the GIS software were used to ensure accurate geocoding.
- The address validation process was under development and will require enhancements in the future.
- The georeferenced points of the facilities and residential buildings (BGE) - attached to census population data - enabled deriving the new geographical variables in a flexible and straightforward manner, highlighting the accessibility/demand geographies generated from a routing service. In total, 153 new geographical variables related to the project were registered in the in-house metadata management system (under the 'Territory' theme), including the ones associated with the service and catchment areas, the territorial indicators and

typologies of facilities and services. These geographical variables were also used to enrich the information boxes displayed in the CE-SIG dissemination platform and facilitate the interpretation of the territorial indicators and other mapped outputs.

- Data aggregation was partially executed from the outputs of the geoprocessing workflows and spatial analysis (spatial intersection) and in database management for the calculation of the territorial indicators based on census population data. No geospatial-based services for data aggregation were operationalised since the data was not grouped in a GIS environment.

Analyse

- Last geospatial validation checks are carried out using geoprocessing workflows to confirm the geospatial reference frame (of buildings) and the boundaries of the service and catchment areas. The access of the georeferenced points of the facilities to the navigation network was also verified.
- Validation and quality requirements agreed upon with the external stakeholder (the company that provided the outsourcing service) that developed the BIES (facilities and services database), CE-SIG back-office application and dissemination platform were checked and reported for correction and improvements.
- Web mapping and visualisation services following OGC standards were created to display and share geospatial data in the CE-SIG dissemination platform (e.g., points of facilities, administrative units, accessibility/demand geographies). Dynamic maps were configured instead of the traditional static cartographic/thematic maps to display statistical outputs.
- Some elements of the semantic interoperability are managed in the CE-SIG back-office, however the geospatial extension of the generic DCAT was not considered and applied to facilitate statistical-geospatial interoperability. The CE-SIG back-office and the in-house metadata management system enabled the statistical and geospatial outputs to be duly displayed in the CE-SIG dissemination platform with

the right associated descriptive information and metadata (pop-up information boxes).

- When preparing the web mapping services and dynamic maps, data visualisation and interpretation of the mapped outputs as well as the performance and responsiveness of the map server were assessed to ensure a good interactive experience and usability of the geospatial data displayed in the CE-SIG dissemination platform.
- Some of the territorial indicators that resulted from the geoprocessing workflows were additionally validated during this sub-process as a final validation round.
- No disclosure control methods handling geospatial data were formally applied, except the confidentiality rules on the private institutional nature of the facilities (e.g., hospitals) and the minimum number of released facilities (statistical units) per territorial unit (suppressive method on the given number of observations to be disclosed). This traditional confidentiality approach follows legal and privacy requirements of the national confidentiality regulation (statistical confidentiality law) and methodological criteria (choice of thresholds for dissemination).

Disseminate

- The in-house metadata management system and the CE-SIG back-office were two key production components to support the outputs to be displayed in the CE-SIG dissemination platform, both territorial indicators and associated statistical information as well as geographical variables and classifications. Versioning, change tables and correspondence tables on geographical divisions were managed and available in the metadata management system. The CE-SIG back-office and the in-house integrated metadata management system also ensured that output data and information were properly loaded into and visible in the CE-SIG dissemination platform, duly linked to the associated metadata, and formatted in a standardised manner according to internal rules. In a nutshell, SMI supported the data dissemination activities in the project context, namely to ensure standardisation and timeliness in concepts, variables and classifications.

- No extensive cataloguing and tagging encompassing both statistical and geospatial data domains were applied to improve the discoverability and usability of the outputs displayed in the CE-SIG dissemination platform. Just preliminary work on this matter was under development within the IIG, particularly addressing the in-house geoportal for internal use.
- The internal staff ensured the geospatial technological development supporting the CE-SIG dissemination platform and took advantage of the existing technical capabilities provided in the IIG, such as web mapping services and open and interoperable solutions for sharing geospatial data online. Some cartography and spatial visualisation guidelines and good practices for communicating statistics on maps (statistical cartography) were implemented.
- No additional disclosure control from the geospatial perspective was applied to the dynamic maps and other geospatial components available in the CE-SIG dissemination platform.
- A technical support document and a user guide will be made available for online reading and download to assist users and promote the project to the general public, in which the user guide will embody some issues related to geospatial literacy for better map readability and spatial interpretation of the outputs.

Evaluate

- The geospatial production components supporting the development of the statistical operation were not formally assessed nor related evaluation inputs were gathered under the quality management of the statistical organisation, including geospatial quality assurance mechanisms and geospatial quality indicators for reporting.
- No user feedback initiatives were carried out as well as no suggestions from the in-house geospatial experts were collected to identify gaps and needs for improvements.
- No evaluation of the statistical operation was carried out from a geospatial production perspective, besides the geospatial data quality assessment routines carried out in the Analyse and Process phases. The outputs of the geographical

validation iterations were not incorporated into quality reports.

- No action plan from the geospatial evaluation perspective was produced for future production cycles of the statistical operation.

Overarching Processes and Corporate Activities

- Institutional collaboration between the statistical organisation and the data providers/administrative data custodians regarding new datasets (e.g., fire brigades) was exclusively implemented for the project aim via customised data agreements that mainly address data requirements and access/delivery conditions. The survey data that were regularly collected for other statistical operations were already based on previously established partnerships (e.g., Education and Health Ministries). The collaboration and cooperation actions carried out were more addressed to the National Data Infrastructure development framework embodying the vision of SP rather than a concrete statistical-geospatial strategy at the institutional or national levels.
- Some reference geospatial datasets from authoritative data sources were used either as input or as complementary for creating statistical content to support the development of the statistical operation. It included geospatial datasets internally produced and managed by SP (e.g., the BGE that constitutes an HVD and is part of the NSDI), georeferenced data from the data providers/administrative data custodians from the public sector (facilities) and authoritative street data. The validation routines carried out by SP regarding the georeferenced data of the facilities also enriched the quality of registers, especially in terms of location attributes for geocoding purposes.
- Although the Statistical Production Process Manual of SP, used as an internal quality reference document, maps and describes the business processes and tasks involving geospatial data and services, the current version still does not include the statistical operation related to the CE-SIG project. Thus, the geospatial-related tasks, associated descriptions and respective responsibilities were not yet reviewed and identified to streamline the geospatial production components and capabilities more systematically in the course of the

development and operationalisation of this statistical operation under the generic production model.

- Although the metadata management system of SP includes several metadata geospatial elements (e.g., new geographical variables and classifications, correspondence tables, indicators, etc.) tagged in the 'Territory' theme, there is no full alignment between those elements with the respective statistical elements.
- Both units with geospatial experts and GIS staff demonstrated all the necessary capacity, skills and knowledge to carry out the data integration and geospatial-related activities supporting the development of the statistical operation. It is important to highlight the GIS programming competencies - which usually require more advanced training and experience (e.g., in Python) - that enabled the autonomous creation of development tools and customised applications that used geospatial data and web mapping services to support the CE-SIG dissemination platform. Although an external company (outsourcing service) mostly designed the CE-SIG dissemination platform, the Geo-Information Unit fully ensured the geospatial programming part. However, capacity gaps in automating dataflows and workflows and in geospatial confidentiality issues were identified and need to be addressed in the next production cycle.

The second part of the results addresses the output information based on the Assessment Matrix filling in which key analysis points are outlined below:

- SP has an enduring institutional collaboration and cooperation partnership with the NGIA (DGT) as some of the producers and contributors of authoritative geospatial datasets from different data themes that feed the NSDI. However, there is no national geospatial data strategy that goes beyond the objectives and activities of the body responsible for the NSDI development, maintenance and coordination, including the INSPIRE roadmap and the HVD guidelines. The legal framework of the NSDI ensures its financial sustainability through financial plans and collaboration protocols. Projects related to geospatial data can be funded by the national programmes for the application of EU funds towards the national development and grants be budgeted by the ESS to increase the geospatial capacity of statistical organisations (e.g., reference geospatial datasets for statistical purposes). Regarding other geospatial data frameworks, SP does not follow the UN 14 Fundamental Geospatial Data Themes. Moreover, Portugal does not formally implement an official and binding statistical-geospatial framework that widely involves key stakeholders in the national operating environment. The National Data Infrastructure promoted by SP can be a starting point towards a data framework that encompasses the statistical, geospatial and administrative data communities in the national context.

- European laws and regulations on open data, sharing and reuse conditions of public information and personal data treatment and protection - the most illustrative being the GDPR and Open Data Directive - are being transposed into the national legislation and policy framework to ensure suitable implementation into the national context. The National Data Protection Commission supervises and monitors compliance with European regulations on data protection and other related national legal and regulatory provisions by public institutions and private organisations. SP designated a Data Protection Officer to ensure the implementation of the legal obligations regarding personal data protection throughout statistical production following the National Data Protection Commission's mandatory guidelines for the public sector.
- SP has a Geo-Information Unit that has been consolidated in its organisational structure over the last decades, mostly oriented to ensure the development, maintenance and management of the geospatial information infrastructure to support statistical production, highlighting census operations for the creation of small statistical areas. It has a cross-cutting action across the statistical organisation establishing work synergies with other units and departments. The Geo-Information is responsible for ensuring compliance with the INSPIRE and HVD Directives concerning the data themes and datasets having statistical relevance. It also represents SP in the NSDI Council (CO-SNIG) and it has participated in several activities and projects related to statistical-geospatial integration in the European context over the years. Although the permanent staff has the adequate skills, knowledge and competencies required to undertake

geospatial-related and data integration activities and tasks supporting statistical production, some gaps are identified addressing standards, confidentiality and EO data and processing issues.

- In addition, SP has been actively involved in the SDG geospatial roadmap through the UN-GGIM: Europe WG focusing on statistical-geospatial data integration to address the SDG indicators. It also participates in the UNECE's task force for the upcoming 2030 Census round concerning geospatial data and small area statistics for international guidelines and recommendations.
- SP fully adopted the GSBPM in its production model through its Statistical Production Process Manual which identifies and systematically documents the various business phases and sub-processes of a statistical operation, but adding a new operational level to the production process matrix, the tasks and their associated responsibilities. SP is also committed and engaged in implementing the GSGF (Europe) since its participation in the GEOSTAT projects over the years has underpinned its roadmap towards enhanced statistical-geospatial integration in statistical production and standardised geospatial statistics. It involves improving capabilities on the geospatial infrastructure and geocoding within the generic information infrastructure (Principles 1 and 2), developing harmonisation and interoperability solutions (Principle 4) and continuing to use open standards and non-proprietary formats for dissemination (Principle 5).
- Concerning location data, cadastral parcels are produced, managed and maintained by the NGIA underpinned by a legal mandate whereas there is no unique and official address register at the national level with a clear custodianship role (third-party authoritative address data). Thus, several public institutions (e.g. the tax authority and municipalities) and organisations from the private sector have their address datasets with their technical specifications and follow their fit-for-purpose standards. SP completely collects - from survey and administrative data - and manages its own building and address data within its information infrastructure to support statistical production, adopting an "in-house" approach in geocoding. Although the Geographic Building Database provides point georeferenced data of all residential buildings (geocoded

population data), SP still has to improve its capacity to geocode/georeference statistical/administrative unit record data (microdata) and expand its location data framework, highlighting address data and buildings at the door level. The National Address Base alongside the census geography (statistical subsections and sections), 1 km^2 grid and places (toponomy) are HVD's geospatial datasets that SP provides for free download (WFS and WMS) with full national geographic coverage.

- The national administrative geographies (municipalities and parishes) are managed and updated by the NGIA following standards on the coding list and identifiers whereas SP is responsible for producing, maintaining and publishing statistical geographies, such as NUTS (ESS) and small statistical areas (for census operations) as well as functional geographies used for statistical purposes (e.g., gridded geographies).
- Although SP has no extensive background in applying statistical disclosure control techniques in grid data and small statistical areas, such methods for statistical confidentiality and data protection handling geospatial data have been mainly focused on census data and derived grid statistics. Confidentiality breaches from geographic differencing issues are not considered when disseminating geospatial statistics.
- Geospatial technical capabilities are broadly integrated throughout the statistical production of SP mainly supported by the IIG, highlighting the centralised geospatial data repository integrated within the information infrastructure, geoprocessing workflows and mapping services for sample design and data collection. However, more extensive geospatial services and modular services to support statistical production can be used, namely for geospatial data validation, disclosure control, the cataloguing of published geospatial data and services and quality feedback between statistical and geospatial business processes.
- SP conducts some geospatial data quality assessment procedures and validation routines, particularly to verify topological and attribute errors regarding new georeferenced point buildings from the survey and administrative data. However, such geospatial quality control routines are not properly streamlined in statistical

production and/or are duly documented.

- SP publicly shares and publishes freely downloadable geospatial datasets concerning the census geography via non-proprietary formats (OGC GeoPackage) and an open-source plugin (QGIS). However, the statistical organisation does not have a Geoportal to broadcast and make geospatial data, services and geospatial statistics products accessible to the users in one centralised web hub for discovery, visualisation and download.
- SP establishes three main institutional and cooperation protocols, including MoU, concerning geospatial data access and sharing with public institutions (NGIA) and regulatory bodies (National Authority of Communication and Energy Agency). These stakeholders highly use the geospatial dataset of the residential buildings to support their activities and products. In collaboration partnerships and institutional environment, the NGIA is the long-term partner of SP from both strategic and operational perspectives to address common and specific needs, highlighting the work carried out related to authoritative geospatial datasets for the NSDI, and more recently in the HVD roadmap.
- SP has some methodological guidance regarding databases available in the geospatial repository and functional relationships with other data repositories, INSPIRE data themes, methodological procedures for certain projects, spatial analysis tools and technical information about the geospatial services. However, such documentation needs to be formally integrated into the in-house quality management system. It is recognised a need for more practical and up-to-date documentation describing geospatial production components (e.g., dataflows and workflows) is to promote statistical-geospatial capacity development.
- Most of the statisticians and other non-geospatial experts at SP do not have a minimum or acceptable level of geospatial awareness and literacy due to the lack of willingness to embrace statistical-geospatial integration in their tasks and the general mindset of only acknowledging geospatial data for census operations and statistical maps. An action plan on statistical-geospatial integration is not available to enhance statistical-geospatial capacity and capability development and to be used as an operational roadmap to streamline geospatial production

components into the development of a statistical operation and increase the efficiency and modernisation in the production of geospatial statistics and other geospatial-statistical outputs.

IV.1. FINDINGS AND DISCUSSION

The interpretation and comprehensive understanding of the results provided a set of research findings and discussion points aligned with the theoretical and methodological framework of this study. The research findings and discussion points are summarised in the Table below (Table 14).

Table 14. Research Findings and Discussion Points (source: author).

Research Findings	Discussion Points
• The Statistical Production Process Manual of SP provided operational guidance to streamline geospatial-related activities and geospatial production components and capabilities across statistical production, contributing to a more integrated statistical-geospatial quality management approach. It showcases that GSBMP is widely spread (awareness) and implemented in SP as the business processes are mapped and conducted consistently across the statistical organisation at the management level to improve production capabilities and standardisation. Lastly, it demonstrates that GSBPM can more-or-less incorporate geospatial production components compared to other statistical production models. • A good institutional environment promoted by the statistical organisation with the data providers and other national stakeholders can expand data themes applied in statistics and avoid duplicated datasets and efforts. Having an institutional background and formalised protocols with data providers and administrative data	• SP has a mature implementation level of GSBPM from the statistical perspective but a more basic one from the geospatial perspective, especially regarding systematic documentation (input data, processes and outputs) and quality management. • The geospatial needs related to data requirements and methodologies need to be equally identified and evaluated alongside the user needs. • The usefulness of following guidelines and recommendations from international data frameworks may depend on the extent of data audit and description the statistical operation requires (during data availability). • The existing technical and non-technical capabilities supporting statistical production, especially involving geospatial production components, need to be harnessed and maximised from an efficiency and modernisation perspective, i.e. reduce inefficiencies while investing in research and innovation in what works well. Rather improving current activities supporting statistical production than designing and building new production components and configuring workflows

Research Findings	Discussion Points
custodians provides better conditions for using authoritative data and harmonising input data, and facilitates data quality feedback routines (peer-reviewed datasets). • The collaboration with the new data providers/administrative data custodians (e.g., civil protection and pharmacies) is not so straightforward since sometimes it is very difficult for them to meet the data requirements and minimum quality profile for the statistical needs and efficiently showcase the relevance of the outputs. The databases managed and maintained by these external stakeholders are usually oriented to support their activities and quality issues are not taken into account from a statistical-geospatial perspective. • When identifying needs and designing production components and workflow, it is important to look for best practices and use cases outside the statistical organisation (benchmarking). This is fundamental since most of the time 'the wheel is already invented' and just needs to be adapted or systematically replicated. • Good articulation and systematic workflow synergies between alphanumeric and geographical data validations (preliminary validation routines) are fundamental in assessing and verifying the quality and consistency of the collected data from both statistical and geospatial (location) perspectives before storage. It is also very important for temporal analysis, database regular updates (geospatial repository and linked statistical databases) and correct frame selection for further processing. • Although SP has an adequate geocoding infrastructure, its data architecture and	from scratch (e.g., improved dataflow or a change in methodology). • Data collection instruments can be more automatic using machine-to-machine solutions, including API and data extraction/loading services. However, it also depends on the technical expertise and technological maturity of the data providers/administrative data custodians. Ready-to-use and user-oriented geocoding services could be also shared or made available to the data providers/administrative data custodians to ensure good quality geocoded data (e.g., precise geographic coordinates). • The geographical validation should include address data assessment routines to ensure a more extensive implementation and maintenance of point-based geocoding infrastructure and improve the quality and accuracy of location data. Address validation processes will enhance overall geographical validation and contribute to a more complete data quality feedback and reporting. • The systematic data validation routines and quality checks should be more automatic and streamlined in statistical production since they were conducted in a very manual and on-demand manner. • Although a comprehensive in-house metadata management system that includes versioning, change tables and correspondence tables of geographical variables and classifications, semantic interoperability solutions on statistical and geospatial data need to be designed and incorporated. In the same way that SP uses SDMX for reporting statistical information to Eurostat, the new version embodying geospatial

Research Findings	Discussion Points
information infrastructure are not location-centred as location data for geospatially enabling statistical and/ or administrative data does not play an integral part in the data lifecycle and information management workflows. In other words, location (and geospatial data and spatial entities) are not an integral part of the data architecture. Only some of the data repositories establish functional relationships based on location as a matching key (buildings and dwellings) which undermines (location) interoperability. In addition to this issue, it was verified that address data has different structures/fields and descriptions across internal datasets, including BGE and BNE, which lead to data incompatibilities within the same data architecture. • Point georeferenced data associated with census population data and the functional geographies (accessibility/demand geographies) generated from a geospatial service provided territorial indicators that otherwise would not be possible to calculate by using area-based data and traditional geographies, such as statistical and administrative geographies. It would not be possible to have these types of indicators (area, facility and territory) if it were not available high-precision geocodes of the facilities (point locations) to enable more territorially flexible aggregations and derive new geographical variables for dissemination. This type of statistical-geospatial product provides new insights and analytical opportunities for policy-makers to target their sectoral public policies in the territories where they are most needed based on population and accessibility metrics and better allocate	metadata specifications needs to be adopted in the institutional metadata management approach. • Regardless of the adoption of open geospatial standards on data and services (data formats for storage extraction, and web mapping services for spatial visualisation and sharing), the path towards the use of harmonised statistical and geospatial standards to support data integration activities and enable interoperability is still unclear. • A strict statistical point of view in managing confidentiality and personal data protection issues can undermine the capacity and capability development to apply disclosure control methods handling geospatial data and considering geographic principles. The experimental culture level and geospatial awareness of the statistical organisation are key criteria for developing and implementing innovative approaches to reduce disclosure risks that arise from location (locationally identifiable data), especially in both static and dynamic statistical maps and other statistical-geospatial products focusing on spatial visualisation. • It is reasonable to claim that no evaluation processes were carried out because the outputs have not been made public to the users. However, this situation does not prevent drafting an action plan, namely from a geospatial perspective, to improve future production cycles. This is also applicable to the review of the Statistical Production Process Manual of SP which did not map the tasks and assigned responsibilities in the business production model. • The extent how which a statistical-geospatial strategy can fit and support

Research Findings	Discussion Points
resources supported by evidence-based and data-driven decision-making. • Interoperability from standardised methods and workflows (alignment between statistical and geospatial processes) and following international frameworks is more consolidated rather than technical and semantic interoperability involving data/metadata, technologies and services. This particularly applies to the absence of compatibility and coherence of data received from external stakeholders since they do not meet the defined standards, highlighting the different coordinate systems, datums or projections of the delivered geographic coordinates and lack of common identifiers. Moreover, the organisational commitments to adopt geospatial standards are not equally established and prioritised compared to statistical standards, especially data and metadata standards. • The staff of the statistical organisation ensured all required capacity, technical skills and knowledge (e.g., GIS) to carry out the geospatial-related activities and tasks on statistical-geospatial integration supporting the development of the statistical operation. However, training and capacity building in geospatial confidentiality are necessary. • The documentation strategy, from descriptive memory to methodological guidance for internal use, was a very useful approach in assigning responsibilities and roles in the data flow and workflow, and in guaranteeing the development of the statistical operation (various breaks over the years). The outputs from the internal documentation provided the inputs for the support user documentation.	the development and implementation of the National Data Infrastructure should be assessed. The HVD roadmap and other international geospatial data frameworks (e.g., IGIF) can provide guidelines and recommendations for designing a geospatial ecosystem within the National Data Infrastructure. • The NSDI and National Data Infrastructure from authoritative data sources can evolve into a national roadmap for data integration supported by standards and interoperability principles with clear strategic leadership and cooperation mechanisms via official networks from the public and private sectors. • SP has an intermediate statistical-geospatial maturity level regarding the GSGF Principles 1 and 2 with the need for improvements related to address and data lifecycle within the data management environment (temporal issues and automatic dataflows). SP already has some experience in using common geographies for analysis and dissemination (Principle 3) mainly due to the ESS context. The extensive use of open geospatial standards for visualising and sharing geospatial data and the release of non-proprietary formats for data extraction also showcase a good level of expertise (Principle 5). However, capabilities enhancing statistical-geospatial interoperability (Principle 4) and geospatial confidentiality issues (Principle 5) need to be further developed, implemented and improved. • Geospatial awareness is more likely to be acquired by non-IT experts rather than staff working in information systems, data architecture, applicational development or

Research Findings	Discussion Points
• The trade-off between modernisation, feasibility and quality is often very tenuous and difficult to measure. Timetable constraints in terms of deadlines and communication shortcomings with external stakeholders (lack or delay in response), including data providers and outsourcing services, are critical obstacles weighing this trade-off balance. • A national high-quality, authoritative and standardised address register with the physical locations is extremely needed alongside a legally assigned data custodian responsible for the creation, management, maintenance and delivery of such data. The binding custodianship roles and responsibilities should be underlined by established national guidelines, standards and compliance mechanisms that follow international recommendations and best practices. • Geospatial data and other related capabilities streamlined in statistical production contribute to data-driven and evidence-based policy-making, allowing to better design, manage and apply public policies in a spatially measurable way (better policy-making and decision-making issues from data/information that are accurately measured in space).	technological infrastructure. The units and departments from the statistical domains (e.g., economic, demographic, social, agriculture, etc.) are more open to embrace geospatial capabilities and statistical-geospatial integration processes to support their activities. This is related to the technical challenges regarding geospatial data with IT support, workflows, and technological/software solutions as well as organisational issues that compromise the streamlining of geospatial-related activities in statistical production. Moreover, other units from the IT department - where the Geo-Information unit is allocated - do not have very high knowledge of geospatial technology and services, especially in mapping tools and geospatial standards for spatial visualisation.

The findings and discussion contributed helpful insights for identifying existing needs and understanding potential gaps in order to support the elements of the next sub-chapter that will clarify what the statistical organisation should do to enhance its statistical-geospatial integration capacity and improve related capabilities. It also determined the key strengths and what sub-processes, activities and tasks involving geospatial capabilities work well and need to be preserved in statistical production.

In a nutshell, the research findings and discussion points enabled transposing needs assessment and gap analysis into operational implementation through action-oriented recommendations for future enhancement and improvements.

IV.2. IMPROVEMENTS AND RECOMMENDATIONS

The interpretation of the results of the methodology and subsequent findings provided guidance and an operational basis for identifying gaps and areas for development and improvement to strengthen the statistical-geospatial capacity and related capabilities of the statistical organisation. Thus, it was possible to identify enhancement measures, improvement actions and recommendations to increase the capabilities to integrate statistical and geospatial data, processes and services in the course of statistical production. Although the generic results were addressed to a specific statistical operation, several statistical organisations with different statistical-geospatial capacity and maturity levels can implement these proposed improvements and recommendations to streamline geospatial production components and increase related efficiency in their business models, technical infrastructure and organisational architectures.

The targeted enhancements and continuous improvement actions aim to at increasing the efficiency and effectiveness of statistical-geospatial integration capabilities in a multi-dimensional perspective, including both technical and non-technical issues, from data and standards to quality management, institutional environment and policy. Thus, the defined enhancement measures, improvement actions and specific recommendations are assigned to the previously defined three key elements (governance; data, information and technology; institutional collaboration and capacity), according to types of activities, aspects and intervention levels, in order to support and facilitate their implementation. Ita also ensures both conceptual and thematic alignment with the literature review (SWOT analysis) and the designed methodology (quality dimensions embodied in the Assessment Matrix).

Governance:

- A clear organisational leadership and high-level commitment aligned with the statistical-geospatial integration agenda and the GSGF (Europe) implementation roadmap are needed to harness opportunities for funding, modernisation and innovation, and participation in national and international development efforts.

- The statistical organisation compliance with UN-enforced frameworks and production models for statistical-geospatial integration, such as the GSGF and the GeoGSBPM, starts at the leadership level involving strategic interest and long-term organisational commitment. The NSO can take the lead in fostering the national implementation of GSGF (Europe) - for instance, via a steering committee, council or WG - by establishing the collaboration design, communication strategy and management model, coordinating supporting activities and ensuring guidelines are fulfilled and goals are achieved.
- Active leadership at the high-level management and corporate levels of the statistical organisation can set the strategic direction of the network of key stakeholders in the national statistical-geospatial operating environment, namely by assuming coordinating roles and responsibilities and promoting data-based collaboration and cooperation activities (e.g., data sharing programmes and expert networks).
- A pioneer and consensus leadership in the field of statistical-geospatial data integration that promotes communication, agreements and mutual understanding will also identify common gaps and opportunities, avoid duplication of data and work, and overlap of activities and outcomes, such as standards, guidelines and methodological resources. Also, only a leadership fully aware of the strategic value of geospatial data in official statistics (geospatial statistics) and the key role of geospatial standards for location interoperability can overcome organisational and data/information silos issues (e.g., data managed in closed systems, licensing restrictions and different requirements) that trigger constraints in statistical and geospatial data integration and exchange.
- The allocation of more financial and human resources will also influence the openness of the NSO to participate in regional and international projects, grants and research initiatives related to statistical-geospatial integration and

geospatial statistics, namely to develop new methodologies, tools and applications.

- Establish a national policy and legal framework on geospatial data incorporating compliance mechanisms and establishing mandatory responsibilities and duties for geospatial data collection, management, maintenance and delivery.
    - This framework should contain binding instruments and legal obligations and define common guidelines and standards to create, manage and distribute geospatial data, prioritising the relevant data themes and datasets for national policy. Both binding and non-binding mechanisms under this policy and legal framework should promote accountability among the stakeholders (preferably, authoritative data sources) and establish guidelines on geospatial data governance that go beyond the NSDI.
    - The NSO should identify the geospatial data themes that are important for official statistics, endorsing the core datasets on location data for geocoding and statistical content (Tier 1 and Tier 2), such as addresses, postal codes, buildings, dwellings and cadastral parcels. A more cohesive national geospatial data ecosystem will enable this type of data to be more easily streamlined within the statistical production, i.e. directly integrated into the data architecture and information infrastructures of the NSO tackling data quality and incompatibility issues.
    - This recommendation is particularly important concerning address data for national implementation due to the absence of a national regulatory framework that mandates a specific public institution or governmental agency the data custodianship rights and responsibilities on this type of data. Also, it would solve complexity issues of address data by overcoming existing gaps in the definition and adoption of common standards and regulations and the lack of an understanding between stakeholders about data characteristics (each one creates its own data model for its specific application and purpose).

- Embody the geospatial data roadmap into the data strategy of the statistical organisation and/or national government to foster data integration, diversify data sources and create societal value through location.
    - Overarching policy and legal frameworks related to open data, such as the HVD and their national implementation, can be a feasible direction and motivational driver to ensure a data governance environment with coherent policies, laws and regulations to increasingly integrate geospatial data within the national data ecosystem and align with EU directives, policies and strategies.

Data, Information and Technology:

- Document the geospatial business processes (dataflows and workflows) within the statistical production model (GSBPM) in a systematic manner for consistent quality reporting and to establish a common language with the statistical pipelines in the course of the production process.
    - Identifying and describing the business processes supported by geospatial production components (input, process and output) and defining the respective roles and responsibilities promotes accountability and internal organisational interoperability by aligning both statistical and geospatial processes and sharing common goals. It will also improve production model issues between the unit responsible for geospatial data management and cartography and the other units and departments that have working relationships and share common activities and tasks.
    - Documenting statistical and geospatial business processes and their functional relationships (e.g., validation routines, quality feedback, data enrichment, etc.) will promote more standardised methodologies, operating models, and consistent approaches for integrating statistical and geospatial data. Also, mapping both statistical and geospatial pipelines and having a common description enables a common understanding via shared terminology and facilitates a step-by-step operational approach. The documentation of the geospatial processes also enhances their monitoring

and assessment by improving the process requirements (e.g., how inputs are transformed into outputs) and the outputs, and supporting future improvements related to process quality.

    - This is particularly important between the geospatial team unit and the units/departments of IT systems, data architecture and information infrastructure to improve semantic and technical interoperability between statistical and geospatial data - and metadata -, services and technologies. No units/departments can better diagnose and recognise the statistical organisation's existing information, technological capabilities and production models to design and implement new solutions to optimise and improve such capabilities and processes.

- Develop a location-centred data architecture and configure IT systems to accommodate a geospatial repository within the data management environment.
    - The technical infrastructure should accommodate a central repository of geospatial data, coding systems, correspondence tables (geographic references) and conceptual models that establish functional relationships with other non-geospatial data repositories from different statistical domains throughout the data lifecycle.
    - This architecture should ensure geospatial dataflows are fully streamlined across the statistical business processes and the geospatial data conceptual model scheme (relationships between geospatial objects, geometric objects and statistical objects, involving semantics and metadata) guarantees technical interoperability in statistical production. The conceptual model is compatible with the overall data architecture logic which enables geospatial data streaming and smooth and standardised data linkage/integration via location between the geospatial objects stored in the geospatial data repository and the statistical objects stored in other data repositories.
    - The Geospatial Reference Architecture of Statistics Finland - its supporting conceptual model design, centralised geospatial data repository and data streams schema - is a good practice case to follow when configuring and

implementing a location-centred data architecture in the statistical organisation.

- Review global or regional data frameworks (e.g., UN Fundamental Geospatial Data Themes and UN-GGIM: Europe Core Data) when checking data availability and inventorying geospatial datasets to address data gaps supporting the development of a statistical operation.
    - The guidelines, technical specifications and recommendations of these data frameworks are useful for defining the characteristics and requirements of the geospatial datasets for statistical purposes and supporting NSDI development. The definition of the data characteristics and requirements should respond to new and emerging user needs and demands, identified and assessed in the preliminary stage (Design phase), through focus groups, targeted surveys and user engagement actions involving all data and user communities, without overlooking the VGI community for open data.
    - Audit the authoritative geospatial datasets available in the NSDI, identify the responsible organisation and respective sector/domain (e.g., health governmental agency or private transportation office), classify them according to the priority degree for statistical-geospatial data integration and assess their application in official statistics and their linkage to other geospatial datasets. The outcome of the previously mentioned classification should be scattered between 'core/key geospatial datasets' and 'auxiliary/supplementary datasets'. The first ones are essential to supporting the geocoding infrastructure of the statistical organisation, such as location data on buildings and addresses, and the second ones support the statistical production, especially in the Analysis and Process phases (e.g., EO data, road network and land use/land cover data). The use of authoritative geospatial datasets available in the NSDI is also important since it is more likely such datasets employ standardised metadata and quality requirements compliant with national and/or international standards.
    - It is also important to identify the geospatial data object type and model (vector data vs raster data) and the respective formats and size units, i.e. in

the case of vector data, point, line and polygon, and in the case of raster data, the available spatial resolution (e.g. 20 meters).

- Ensure the follow-up and reinforcement in compliance with national and international standards, highlighting the INSPIRE Directive and the HVD (European Data Strategy and Open Data Directive) as cornerstones of the statistical-geospatial integration agenda.
    - Work efforts in data and metadata standards aim to address shortcomings regarding poor semantic and technical interoperability and harmonisation between different data sources from the statistical and geospatial domains in the national context, including syntactic barriers. The EIF can provide general guidance to ensure both technical and semantic interoperability in collecting and exchanging location data/information in a way that is understood across the public sector by using dominant and comparable methods, formats, concepts and metadata following recent digital technology trends.
    - In the context of the INSPIRE and HVD national implementation process, the NSO should be an active stakeholder in the NSDI development and implementation support, particularly in ensuring the standards compliance and maintenance of the geospatial data themes for which it is thematically responsible (e.g., population distribution, statistical units, buildings, etc.).
    - Consider the Global Geodetic Reference Frame to get precise geographic locations in location datasets/layers such as addresses, buildings and settlements, geographical names, cadastral parcels and functional areas (administrative and statistical geographies). This is important to ensure a common geographic reference framework to consistently assign any object on the Earth's surface (there are multiple reference systems) and avoid errors when performing transformations between coordinate systems or in mapping projections, i.e. displaying the geospatial data in a digital environment.
    - The advantages of and knowledge about standards should be widespread within the statistical organisation in a multi-level perspective, from high-level

management and corporate support to the expert and operational involving business processes, technology and methodology issues. In this regard, research and IT development efforts may be required.

- Establish agreed data profiles, structures and formats, including descriptive metadata and data models, and minimum data quality requirements (e.g., mandatory attributes, revision cycle and description of quality).
    - Get to know about the data collection practices of the data providers and administrative data custodians concerning georeferencing/geocoding and the adopted geospatial standards (if applicable). Since different organisations from the public and private sectors usually have their geospatial data in different coordinate systems and projections it is important to move forward to standardised approaches in georeferencing methodologies and geocoding techniques. In this regard, it is very helpful to describe and document internal georeferencing/geocoding practices, including technical guidelines and terminology, to be able to share with the data providers and administrative data custodians and assist them (e.g. SP Georeferencing Protocol, manuals and handbooks, etc.).
    - It is also relevant to make these organisations aware of these technical issues and geospatial data specifications because most of them treat this type of data in the same way that statistical data, do not have the expertise or skills to deal with the storage and management of location references/attributes or are not engaged in combining and integrating data.
    - This is particularly important for standardising address data through common address concepts, agreed addressing components and a shared data model at the organisational and/or national levels. A common address data format will facilitate data integration from external data sources and enhance data interoperability and usability of location data within the in-house data architecture and information infrastructure (e.g., with buildings and road network datasets).

    - The FAIR principles can be used as guidelines when establishing standardised data formats in order to be easily interoperable and compatible between different software and hardware systems of the NSO and external stakeholders. International standards in data distribution and service development from both statistical and geospatial domains should be also used for data encoding, access, processing, visualisation and metadata, such as XML, comma separated values (CSV), GeoJSON and GeoPackage.
- Invest in research and development on confidentiality issues arising from geospatial data and related to statistical and geospatial data integration.
    - Benchmarking, identifying the state-of-the-art and further reviewing and testing best practices, methodologies and disclosure control techniques applied by other statistical organisations can be an adequate option in the short run. During this process, a collaborative working environment should be established between the geospatial unit and the methodology unit to effectively exchange knowledge. Whereas the geospatial unit provides geospatial expertise (e.g., GIS skills and computing solutions) to deal with geographic differencing and other risks, the methodology unit has more experience in designing and applying disclosure control in statistical production, namely regarding microdata and dissemination. Also, the methodology might be more aware of the policy and legal framework regarding confidentiality in force at the institutional and/or national level. In this regard, it is important to assess the available hardware and software capabilities since computing statistical disclosure control techniques handling geospatial data may require more machine and technological efforts due to the size and complexity, especially when trying to detect every single potential disclosure risk in grid data or small administrative units.
    - Issues related to the usability of the outputs and eventual loss of geographic detail need to be carefully evaluated against the increasing user demands to have more disaggregated statistical data and information for providing a higher granularity of the geographic context (smaller scales) of the phenomenon. If such issues are not considered the released data and

information can be damaged and unreadable, and lead to misinterpretations when spatially visualised, as well as create conditions to go back to when statistical data were released at larger levels, such as supra-regional. In addition, it is important to ensure data can be viewed and accessed through safe mechanisms and that all visualisation capabilities guarantee confidentiality and data protection.

    - Lastly, the challenges that emerged from geospatial-related confidentiality issues should not demotivate nor fail to persuade the use of grid geographies as a key complement to traditional geographies for statistical analysis and dissemination, without overlooking their advantages, highlighting stability, comparability and flexibility.

- Assess the suitability for standards implementation by considering the in-house data architecture (datasets, repositories and relationships) and the technical infrastructure, including IT systems and technology inventory.
    - It is important to look for existing and potential internal bureaucratic processes, communication and collaboration gaps across the statistical organisation, funding resources, software and hardware requirements, data/information security and licensing, and expertise in using international. These key issues are important to acknowledge when implementing common standards in statistical production and across the data lifecycle that may hamper the process by posing both technical and non-technical obstacles and challenges (e.g., lack of data compatibility and high IT costs).
- Adopt open statistical and geospatial standards embedded in the IT infrastructure to increase interoperability between both domains and respective systems and technologies throughout statistical production.
    - An illustrative example is connecting and indexing the statistical information to the corresponding geographical areas (e.g., geospatial data on administrative boundaries stored in Oracle Spatial) to be queried, retrieved and visualised online by the users in a web-based mapping application via non-proprietary servers (e.g., GeoServer), web mapping services (OGC) and

API for (spatial) querying operations.

    - Explore cloud technology solutions to implement mapping as a service to deliver geospatial data, base maps with satellite imagery, and statistical maps across the statistical organisation based on user requests, production needs and workflow requirements from other units and departments.

Institutional Collaboration and Capacity:

- Map potential data providers and administrative data custodians from the public and private sectors and develop institutional partnerships and collaboration alliances with them. These institutional partnerships should be preferably supported by an institutional and/or legal framework, namely data protocols and formal agreements outlining the agreed data requirements (e.g., attributes, revision cycles, delivery specifications, etc.) and definition of roles and responsibilities.
    - Data providers and administrative data custodians operating within the NSS are more open to engaging with the NSO since the process of trust building, institutional collaboration and general arrangements are already well-established and their roles and responsibilities in the statistical operating environment are usually well-defined by law. Also, the institutional collaboration between the NSO and the NGIA can contribute to reaching out to the geospatial data providers and foster cooperation with them, especially if they are part of the NSDI.
    - When establishing the first contact with potential data providers and administrative data custodians on geospatial data for statistical purposes it is important to share reference documentation that will help them achieve the objectives and meet defined requirements, namely regarding data characteristics, conceptual models and related technical specifications. Reference documentation can include internal methodological guidance and technical materials used to assist business processes, such as geocoding practices and metadata requirements.
- Establish institutional arrangements for the provision and delivery of geospatial data, including the design of the structure, transmission formats and integration

requirements. The service levels should be also defined as requirements and criteria supporting quality management of data collection/acquisition and reception and facilitating quality control and reporting routines and quality assurance procedures.

  - These institutional arrangements should be based on the outcomes from the identification of the data sources and check of data availability within the national geospatial data ecosystem (highlighting the NSDI) and network of authoritative stakeholders. Moreover, having a continuous and constructive dialogue and strengthening the partnership with trustworthy data providers/administrative data custodians outside the NSS and NSDI have proven to be an advantage when requesting revisions and clarifications on the received data.

- Develop solutions oriented to data alliances/partnerships between public and private sectors and NGIA agreements across governmental agencies and public institutions for access to and use of authoritative geospatial datasets and related technical expertise.
  - Timely and open access to authoritative geospatial data attached to fundamental datasets under core data themes is essential to support governmental bodies in the policy lifecycle (from design to monitoring) and oversight.
- Promote joint international workshops, initiatives and activities bringing together both statistical and geospatial communities and other key stakeholders from across all sectors and organisations to share ideas, innovations, lessons learned, common challenges and inspiration for future work.
  - These measures intend also to promote networking for establishing new collaborations and partnerships and improve cooperation and communication in the field of statistical-geospatial data integration.
- Effective communication and collaboration strategies and actions embodying marketing skills need to be targeted to strategy/leadership and high-level management levels to ensure the long-term sustainability of data integration activities, particularly related to financial partnerships and funded projects.

- Raise geospatial awareness across the different units and departments of the statistical organisation through information sessions, workshops and other engagement initiatives.
    - These actions aiming to increase geospatial awareness - including the need for more integration of statistical and geospatial data - should be designed based on previously identified work synergies (when documenting statistical and geospatial business processes) and showcase the benefits of the geospatial domain for official statistics, highlighting geospatial standards and technologies.
    - Raising geospatial awareness within the statistical organisation and welcoming more cross-cutting expertise on board can facilitate the integration of geospatial production components into statistical production by making ongoing geospatial processes more efficient and new ones more effectively implemented.
- Promote professional training initiatives on geospatial programming skills within NSO, including GIS programming and algorithm design skills.
    - It emphasises Python script programming which has been showcased to be a good tool - as opposed to desktop and web tools with customised parameters - for managing statistical and geospatial data management, such as streamlining data loading, geocoding, data aggregation and creating interactive statistical maps. GIS programming skills enable building and performing automated, reproducible, replicable, expandable, and more importantly more efficient workflows for advanced geospatial (big) data analytics and spatial visualisation. Advancements in language models and the growing combination of GIS technology and AI (GeoAI) have created new possibilities to automate repetitive tasks. It covers tasks related to exploratory analysis from large amounts of geospatial data (e.g., geospatial AI models and GeoAI techniques), feature extraction/object detection via deep learning methods and creation of maps without using a user interface or manually adjusting symbology.

- Workflow-based technologies can be an alternative more user-oriented solution to create, use and share workflows for executing and repeating multiple geospatial-related tasks efficiently in one digital platform. In short, these capacity building initiatives will enable higher production efficiency by increasing automation and reducing time-consuming and resource-intensive, especially when processing large volumes of geospatial data.
- GIS programming skills are also important for extensively using both commercial GIS software and free licensed technologies or open-source solutions, which usually require these competencies. Also, such programming skills allow internal geospatial personnel not to depend on hardware and software inventories, available technological infrastructure, and budgetary and resource constraints which is a major issue as statistical organisations face difficulties in recruiting and retaining staff with expertise and knowledge, especially in data management and technology.
- In addition, participation in technical workshops or intensive training courses on geospatial programming skills can foster innovation and modernisation in the medium term to support data integration and geospatial-related activities and tasks in the course of statistical production. It can also contribute to the upskilling (skills development) to deal with the emerging trends in (Geospatial) Big Data, Data Science, Location Analytics and ML methods handling geospatial data and spatial models.
- Although not directly related to the case study, it is also important to promote training and research initiatives on EO data and processing in official statistics to increase the geospatial capacity of NSO in addressing SDG monitoring and reporting, namely in enhancing the accuracy, timeliness and comparability of geospatial statistics across institutions, regions and countries. Investments in EO training are also relevant to statistical organisations to reduce costs and the excessive statistical burden on respondents during data collection.

V. CONCLUSIONS AND FUTURE RESEARCH

Theoretical and methodological frameworks for statistical-geospatial integration have matured and spread over in the last decade as key stakeholders in the global and regional governance and institutional environments have increasingly recognised the need and value of data integration and geospatial statistics for policy-making, particularly for sustainable development. Global efforts and extended regional activities in this field have been driven and applied through strategic and work synergies within the UN 2030 Agenda and surrounding policy agendas and strategies, and in line with the modernisation of official statistics and trends in geospatial data management. The integration of statistical and geospatial data has proven to be an appropriate approach to provide more detailed, timely, harmonised and high-quality data to address data gaps that affect the measurement and monitoring of the SDG, and to meet the requirements of the indicators at the regional and local levels. This line of work on data integration also contributed to improving geospatial (data) capacity and redesigning statistical production and related statistical systems to support multi-level policy instruments and decision-making processes (e.g., using EO data in official statistics).

High-level global organisations, regional bodies and interest groups from the statistical and geospatial communities (e.g., UNECE, UN-GGIM, Eurostat and EFGS) have been engaged in fostering statistical-geospatial integration and advancing in the production of standardised geospatial statistics through implementation guidance for reference frameworks. Collaboration and cooperation initiatives, coordination actions and funded projects around statistical-geospatial data integration (e.g., GEOSTAT projects in the ESS context) have been fundamental in building a methodological foundation and enhancing the capacity of organisations and countries to produce geospatial statistics. They covered various development work activities related to geospatial statistics, covering data/information, processes, technological and organisational issues, from practical use cases to more general guidelines and recommendations. The conceptual and operational body of knowledge in integrating statistical and geospatial data has been particularly consolidated by the national statistical and geospatial communities (NSO and NGIA) by strategically following and implementing production frameworks, guidelines and standards as well as exchanging information and experiences, and sharing resources and best practices.

The GSGF and its European version (GSGF Europe) have been decisive in setting the future roadmap for the integration of statistical and geospatial data and in providing opportunities to improve statistical and geospatial production chains at the national level. These internationally recognised and endorsed frameworks defined how the statistical and geospatial communities should position themselves in their own national and regional operating environments and established a consistent and systematic production model for harmonised, accessible and usable geospatial statistics (e.g. point-based geocoding infrastructure). The GSGF (Europe) Principles and Key Elements provide generic guidance to assist organisations and countries in the overcharging implementation of the framework encompassing policy, legal and institutional issues, whereas the supporting methodological resources break down to more operational guidelines for a more concrete application by showcasing a series of good practices, successful cases studies and experiences. In particular, the GSGF Europe has a set of Requirements and Recommendations, linked to each of the five Principles, and underpinned by good practice cases embodying specific actions or steps to enable a more targeted implementation of the framework.

The GeoGSBPM, the geospatial version of the GSBPM – a reference production framework in official statistics - has also pushed forward the integration of geospatial data, processes, and services in statistical production. GeoGSBPM identifies and describes geospatial-related activities within the statistical production process to systematically streamline geospatial capabilities throughout the business model, underlining a conceptual overlap between the GSBPM and GSGF.

The literature review supported the overview of the current situation of statistical-geospatial data integration and getting a comprehensive understanding and insight of the progress through a SWOT analysis by identifying internal and external factors across three key elements: i) Governance; ii) Data, Information and Technology; and iii) Institutional Collaboration and Capacity. The situational analysis indicated that non-technical issues related to governance (policy and legal alignment and funding) and institutional collaboration and capacity (awareness and human resources) are the main obstacles compromising the future progress of data integration activities and enabling a greater capacity in integrating statistical and geospatial data. Moreover, technical issues related to interoperability (technical and semantic), harmonisation and standards are also shortcomings to be noted due to the use of different data formats, collection methods, business processes, technologies and solutions supporting technical infrastructures and operating models.

Leadership and governance mechanisms can be decisive in the statistical-geospatial capacity and capability development of organisations and countries. An active and agreed leadership promoting inclusive and cooperative governance models will strengthen institutional engagement and collaboration between the statistical and geospatial communities (e.g. data sharing agreements), secure long-term financial resources (e.g., development work, training programmes and implementation of frameworks) and foster innovation (e.g., research projects in methods and technology). Thus, strategic guidance and coordination actions of high-level key stakeholders will strongly influence the methodological progress of national stakeholders in modernising their statistical-geospatial operating environments, by helping them improve capabilities related to geocoding, technical infrastructures, data interoperability and standards.

Following the literature review, the GeoGSBPM, together with the GSGF (Europe) conceptualisation and the surrounding guidance and methodological materials, constituted the core benchmarks for the development of the first part of the methodology (Production Model), linked to the structure of the statistical business production process. In addition, the Requirements and Recommendations (GSGF Europe) strongly helped to define the second part of the methodology related to the Assessment Matrix to evaluate both national and organisational statistical-geospatial maturity and capacity levels according to three quality dimensions based on the aforementioned key elements.

The CE-SIG project developed by SP was the selected case study to identify and describe the data integration and geospatial-related activities and assess the performance of the case study on the streamlining of geospatial production components into statistical production, and the implementation degree of statistical-geospatial capabilities by the statistical organisation. It is acknowledged that despite the application exhaustiveness and thematic extent, the methodology can be used at the activity level or by a customised approach in which only parts of the statistical business process model are considered based on the characteristics of the statistical operation or requirements of the geospatial statistics product.

It was concluded that although SP has a mature implementation level of GSBPM from a statistical perspective, embodied in its internal production manual, which provides systematic operational guidance on the development of a statistical operation, similar documentation regarding geospatial production components is still missing. Nevertheless, the GSBPM and the layout of the production manual will easily incorporate current geospatial processes and capabilities across statistical production, such as geospatial data validation and quality feedback routines to the sources.

The ongoing good collaboration between SP and data providers/administrative data custodians, highlighting the NSDI network, can facilitate future development on external organisational interoperability and cross-domain data interoperability to ensure greater harmonisation and integration within the national data ecosystem. From the internal scope, the geocoding infrastructure supported by point-based reference data is adequate, without overlooking the need to optimise data management and maintenance routines through consistent and automated procedures, especially focusing on address data standardisation across different data repositories (e.g., unique data format and structure). However, the data architecture and information infrastructure do not have a location-centred design in which only some data repositories are functionally linked to a geospatial repository and data management practices are not oriented to geocoded data. As the geospatial infrastructure of SP (IIG) is still very much designed for the census operations pipelines (basic statistical infrastructure), more governance alignment and technical integration between the IIG and the data management environment is required to produce geospatial statistics from administrative data and other emerging data sources at the microdata level.

Technical and semantic interoperability as well as statistical-geospatial confidentiality are two key issues for development work and future enhancement in SP, the first encompassing metadata management considering the in-house system and the second one involving methodological research and benchmarking on best practices. The staff of SP has the suitable geospatial expertise and technical skills to perform data integration and geospatial-related activities in the course of statistical production and by establishing good work synergies with other units and departments - focusing on IT-related - can foster geospatial awareness and literacy across the organisation.

In summary, SP has an intermediate statistical-geospatial maturity level on the GSGF Principles 1 and 2 with some gaps related to address data and data management practices, it extensively uses common geographies for analysis and dissemination (Principe 3) and open geospatial standards for mapping and sharing geospatial datasets to the users. However, more modernisation and capability development efforts need to be carried out concerning Principle 4 on statistical-geospatial interoperability both inside and outside the organisation, namely through common (data and metadata) standards, agreed guidelines, aligned business processes and shared best practices.

The results of the methodology supported the definition of improvements and recommendations to strengthen the statistical-geospatial capacity and develop new or improve existing capabilities related to statistical-geospatial data integration. The proposed improvement actions and recommendations can be used by statistical organisations and countries at different development stages to improve their statistical-geospatial (data) infrastructures and respective operating environments, and to support both national and organisational implementation of the GSGF.

Clear organisational leadership and high-level commitment around statistical-geospatial integration are needed to foster strategic guidance and coordination activities among the key stakeholders, especially regarding the implementation of reference frameworks. This type of leadership and political commitment will be a prerequisite to effectively establishing a robust national policy and legal framework on geospatial data to align the geospatial data roadmap with the broader data strategy and development agenda (e.g., SDG 17 on partnerships to facilitate global and national access to geospatial data). In addition, high-level engagement is required to unlock governance and institutional barriers and for decision-makers, such as deputy directors and top managers of statistical organisations, to formally acknowledge geospatial as a core business of statistical production and not to overlook it in the statistical programmes and activities supporting statistical development, including strategic and budget planning.

Location-centred data architecture, data standardisation (especially in address data), geospatial business processes documentation, including quality reporting, and research on statistical-geospatial confidentiality issues are key action areas that need development work in SP. Methodological guidance, best practices, and case studies on statistical-geospatial integration, coupled with an experimental approach and openness to innovation, can be useful tools for implementing some specific recommendations.

In the scope of institutional collaboration and capacity development, it is recommended to map data providers and administrative data custodians in order to further establish institutional collaboration partnerships based on data protocols and formal agreements. Active participation and engagement of key stakeholders from both statistical and geospatial communities in joint networks and data integration projects will help to build mutual understanding, identify synergies and common gaps and share opportunities. In this regard, visionary leadership and high-level coordination are crucial to avoid duplication of data and work, double funding situations and overlap of activities and outcomes.

Raising geospatial awareness across the different units and departments within the statistical organisation will facilitate designing and running data integration and geospatial-related activities in the development of a statistical operation and streamline geospatial production components within statistical production. Moreover, professional training in geospatial programming skills should also be included in the education and training planning of NSO, embodying geospatial technology advancements to automate geospatial workflows and address the potential of EO data in complementing statistical production and enhancing statistical outputs.

The CE-SIG project highlighted the value of integrating detailed georeferenced data with statistical data by providing territorial indicators on spatial and population coverage of cross-sector facilities and services with public interest. Geography underpinned the conceptual foundation of CE-SIG as geospatial data and services were the operational cornerstones shaping most workflows of the business production model and supporting the visualisation capabilities of the dissemination platform. Therefore, the quality and management of location data are so fundamental for the development of the statistical operation that changes in the geographic references affect the inputs and outputs of the business processes. This is evident in the georeferencing and resulting location accuracy of the georeferenced point data that will impact the next geoprocessing workflows for generating functional geographies and calculating the territorial indicators, i.e. a poorly positioned point will require new network analysis and consequently, a recalculation of the territorial indicators associated with those derived geographies.

Meaningful insights resulted from statistical-geospatial integration capabilities embodied in statistical production. Key features include granular location data (population and facility point data) integrated with statistical/administrative data into a unified database within the data architecture, advanced spatial analysis, statistical information based on traditional and newly defined functional geographies and mapping services for spatial visualisation. Integrating geospatial production components into statistical production enabled knowing that a hospital has a certain population living within a specific distance or time range based on accessibility metrics and cross-referencing those values with useful statistical information on capacity and services, such as available number of beds and medical specialities. At the same time, this geospatial statistics product enabled assessing the covered school-age population living near a primary school by considering the precise location of their households and based on acceptable access criteria and norms.

These insights underline the geospatial analytical potential of this statistical-geospatial product, allowing the detection of local and sub-regional territorial and population asymmetries in terms of supply and demand for general interest facilities and governmental services. This is very convenient for making public resources management more efficient and service delivery more effective by better allocating them where are most needed and optimising their distribution considering morphological (physical characteristics of the territory), population characteristics and accessibility needs. It is also helpful to support processes of planning and monitoring public administration facilities and government services (e.g., choosing the appropriate location and size for a new facility), which are usually explained by needs assessment studies, including population analyses with demographic and socioeconomic descriptions.

Georeferenced population data that are automatically updated from census data and administrative registers and integrated with other datasets from different sources (e.g., for data enrichment) can address these preliminary analyses by identifying emerging demographic changes and complex migration patterns (urban-rural flows) and conducting spatio-temporal analysis of the population distribution to understand population trends. Effective use of updated location data (e.g., real-time data) and geospatial technology (e.g., routing services) across statistical production, alongside consistent data management and integration practices will enable regular release of the territorial indicators to be used as auxiliary information to policy formulation in several sectors, such as health and education. In addition, the use of point-based population data provides a direct and flexible production approach to calculate policy-relevant indicators for any geography of interest (e.g., functional territorial typologies) for analysis and dissemination, as no data disaggregation or mixed methods are required.

CE-SIG can also be a potential tool supporting a monitoring system for policy-makers to better track territorial and capacity gaps in facility provision and service delivery and further design more accurate place-based policy responses throughout the policy lifecycle. Thus, geospatial data contribute to data-driven policy-making, helping to better address public policy issues by accurately measuring what is happening in space.

From a governance perspective, these territorially targeted policy interventions will support local and sub-regional territorial development and promote territorial cohesion as more pragmatic and tailored political decisions consider changing requirements and needs, both from a data and a user perspective. More granular, timely, integrated and enriched data and analysis will ground a data-driven foundation for a spatially evidence-based policy approach providing a cross-sectorial policy integration that addresses different types of territories and action scales. This is particularly important for structurally vulnerable and less developed regions, such as rural and mountain areas, that usually lack access to basic services and do not have higher transport connectivity and accessibility conditions due to physical geographic barriers, population ageing and depopulation phenomena and weak economic and social development. In this regard, by accessing these integrated data analysis and geospatial statistics, regional bodies and local authorities have key information to actively contribute to the policy agenda setting endorsed by the national government and design their policies and strategies regarding public facility provision and service delivery.

Future research activities and innovative development work should address the identified gaps and areas for development and improvement to enhance the statistical-geospatial integration capacity and capabilities of the statistical organisations and support the implementation of GSGF (Europe) to produce harmonised geospatial statistics. At the highest level, political endorsement and visionary leadership are necessary to highlight the role of geospatial data for national strategic and policy framework and decision-making, put statistics-geospatial integration on the agenda and build synergies and commitments between the key stakeholders from statistical and geospatial communities. Reviewing policy and legal issues is required for a more integrated data ecosystem in which a collaborative data governance model, custodianship guidelines and maintenance mandates on geospatial data production should be established for both public and private sectors and aligned with global and regional data frameworks. Priority should be also given to location data used to produce official statistics (e.g., address data that usually present more data gaps) and to key geospatial data themes that are statistically relevant to support policy frameworks and development agendas (e.g., land use/land cover, transport network and basic services).

The academic community can be a relevant stakeholder in cooperating with statistical organisations to conduct joint research projects for innovative methodologies and techniques on statistical confidentiality that handle statistical-geospatial confidentiality issues, namely by providing theoretical foundations and proposing robust methodologies. Partnerships with academia and the private sector can foster the development of cutting-edge technical and technological solutions to implement a location-centred data architecture, improve IT systems, modernise geospatial services and deliver efficiency in processes involving data integration. These partnerships can also enable skills upgrading and the development of geospatial expertise through training initiatives that address digital technology and business trends in the field.

Finally, focus should be given to interoperability and standards issues as key prerequisites for facilitating statistical and geospatial data harmonisation, integration, access and use. It involves establishing guidelines and designing technical and non-technical compliance mechanisms for (authoritative) data providers and administrative data custodians to fully adopt common data standards and to align their business production processes. It is important to configure standardisation approaches and streamline business cases across the data lifecycle (e.g., unique identifiers and consistent geocoding methods) as more organisations from the public sector produce their geospatial datasets in silos and private stakeholders deliver data products without regulatory compliance and quality requirements. In addition, technical capabilities concerning geospatial metadata models aligned with international geospatial standards (e.g., ISO 19115) should also be improved according to what works well in the metadata management (e.g., description, discovery, exchange, etc.) to ensure statistical-geospatial harmonisation and interoperability throughout statistical production.

REFERENCES

Ariza-Lópes, F. J., Rodríguez-Pascual, A., Lopez-Pellicer, F. J., Vilches-Blázquez, L. M., Villar-Iglesias, A., Masó, J., Díaz-Díaz, E., Ureña-Cámara, M.A., & González-Yanes, A. (2021*). An Analysis of Existing Production Frameworks for Statistical and Geographic Information: Synergies, Gaps and Integration*. ISPRS Int. J. Geo-Inf. 2021, 10(6), 374. doi: 10.3390/ijgi10060374.

Australian Bureau of Statistics (2018). *Statistical Spatial (SSF) Guidance Material - Geocoding Unit Record Data Using Address and Location*. Consulted at https://www.abs.gov.au/statistics/statistical-geography/statistical-spatial-framework-ssf/statistical-spatial-framework-guidance-material/ssf-guidance-material-geocoding-unit-record-data-using-address-and-location#purpose (accessed on 5 May 2022).

Bernard, L., Kanellopoulos, I., Annoni, A., & Smits, P. (2005). *The European geoportal - one step towards the establishment of a European Spatial Data Infrastructure.* Computers, Environment and Urban Systems 29(1), 15-31. https://doi.org/10.1016/j.compenvurbsys.2004.05.009.

Bousquin, Justin. (2021). *Discrete Global Grid Systems as scalable geospatial frameworks for characterizing coastal environments*. Environmental Modelling & Software. 146, 1-14. https://doi.org/10.1016/j.envsoft.2021.105210.

Coetzee, S., Gould, M., McCormack, B., Mohamed, Z. S., Scott, G., Kmoch, A., Alameh, N., Strobl, J., Wytzisk, A., & Devarajan T. (2021). *Towards a sustainable geospatial ecosystem beyond SDIs*. doi: 10.13140/RG.2.2.22555.39203. Consulted at https://ggim.un.org/meetings/GGIM-committee/11th-Session/documents/Towards_a_Sustainable_Geospatial_Ecosystem_Beyond_SDIs_Draft_3Aug2021.pdf

Committee on Earth Observation Satellites (2018). *Earth Observation Handbook: Satellite Earth Observations in Support of the Sustainable Development Goals*. Consulted at https://eohandbook.com/sdg/files/CEOS_EOHB_2018_SDG.pdf (accessed on 12 March 2022).

Costemalle, V. (2019). *Detecting geographical differencing problems in the context of spatial data dissemination*. Statistical Journal of the IAOS, 35(4), 559-568. doi: 10.3233/SJI-190564.

Craglia, M. (2015). *Spatial Data Infrastructures*. In: Wright, J.D., International Encyclopedia of the Social & Behavioral Sciences (Second Edition), Elsevier, 130-135. doi: 10.1016/B978-0-08-097086-8.72058-7.

Cressie, N. A. C. (1993). *Statistics for spatial data*. 1st Edition, New York: John Willey & Sons.

Daas, P. J. H., Puts, M. J., Buelens, B., & van den Hurk, P. A. M. (2015). *Big Data as a Source for Official Statistics*. Journal of Official Statistics, 31(2), 249–262. http://dx.doi.org/10.1515/JOS-2015-0016.

Dangermond, J., & Goodchild, M. F. (2020) *Building geospatial infrastructure*. Geospatial Information Science, 23(1), 1-9. doi: 10.1080/10095020.2019.1698274.

Dékány, K. (2019). *New approaches in georeferenced statistical data and related confidentiality issues.* Economic et statistique, 108.

Do, L., Trinh, T., Aryan, P., Wetz, P., Kiesling, E., & Tjoa, A. M. (2015). *Toward a statistical data integration environment: the role of semantic metadata*. In Proceedings of the 11th International Conference on Semantic Systems (SEMANTICS '15). Association for Computing Machinery, 25–32. https://doi.org/10.1145/2814864.2814879.

Elwood, S., Goodchild, M. F., & Sui, D. Z. (2011). *Researching Volunteered Geographic Information: Spatial Data, Geographic Research, and New Social Practice.* Annals of the Association of American Geographers, 102(3), 571–590. https://doi.org/10.1080/00045608.2011.5956.

European Commission (2020). *Commission Implementing Regulation (EU) 2023/138 of 21 December 2022 laying down a list of specific high-value datasets and the arrangements for their publication and re-use*. Consulted at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32023R0138 (accessed on 6 September 2023).

European Committee of the Regions (2019). *A territorial approach for the implementation of the SDGs in the EU – The role of the European Committee of the Regions*. Commission for Economic Policy. Consulted at https://data.europa.eu/doi/10.2863/11396 (accessed on 20 February 2022).

European Committee of the Regions (2021). I*ntegration of Geographic and Statistical data for better EU policy making.* Commission for Territorial Cohesion Policy and EU Budget. Consulted at https://data.europa.eu/doi/10.2863/512257 (accessed on 4 February 2022).

European Statistical System (2014). *The ESS Vision 2020*. European Statistical System Committee (ESSC). Consulted at https://ec.europa.eu/eurostat/documents/10186/756730/ESS-Vision-2020.pdf/8d97506b-b802-439e-9ea4-303e905f4255 (accessed on 1 March 2022).

European Statistical System Committee (2013). *Scheveningen Memorandum on Big Data and Official Statistics*. 99th DGINS Conference. 25 September. Consulted at https://ec.europa.eu/eurostat/documents/13019146/13237859/Scheveningen-memorandum-27-09-13.pdf/2e730cdc-862f-4f27-bb43-2486c30298b6?t=1401195050000 (accessed on 14 June 2022).

European Statistical System Committee (2018). *Bucharest Memorandum on Official Statistics in a Datafied Society (Trusted Smart Statistics).* 104th DGINS Conference. 10-11 October. Consulted at https://ec.europa.eu/eurostat/documents/13019146/13239158/The+Bucharest+Memorandum+on+Trusted+Smart+Statistics+FINAL.pdf/59a1a348-a97c-4803-be45-6140af08e4d7?t=1539760880000 (accessed on 14 June 2022).

European Union (2007*). Directive 2007/2/EC of the European Parliament and of the*

Council of 14 March 2007 on Establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). Consulted at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32007L0002&from=EN (accessed on 31 May 2021).

European Union (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation).* Consulted at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&amp (accessed on 31 May 2021).

European Union (2019). *Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector Information.* Consulted at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019L1024 (accessed on 7 June 2022).

Eurostat (2017). *The European Statistics Code of Practice*. Consulted at https://ec.europa.eu/eurostat/documents/4031688/8971242/KS-02-18-142-EN-N.pdf/e7f85f07-91db-4312-8118-f729c75878c7?t=1528447068000 (accessed on 23 February 2022).

Eurostat (2019a). *The integration of statistical and geospatial information – a call for political action in Europe*. Consulted at https://ec.europa.eu/eurostat/documents/4031688/10158240/KS-03-19-423-EN-N.pdf/c8f75ee1-2181-288c-1efa-1622c5abb980 (accessed on 5 June 2021).

Eurostat (2019b). *Merging statistics and geospatial information — Experiences and observations from national statistical authorities, 2012-2015 projects.* 2019 edition. Consulted at https://ec.europa.eu/eurostat/documents/7870049/9809882/KS-FT-19-004-EN-N.pdf/aec76d11-967e-4d67-aad2-49c52b264748?t=1559812901000 (accessed on 18 May 2023).

Eurostat (2020). *European Statistical System Handbook for Quality and Metadata Reports. Manuals and Guidelines*. Consulted at https://ec.europa.eu/eurostat/documents/3859598/10501168/KS-GQ-19-006-EN-N.pdf/bf98fd32-f17c-31e2-8c7f-ad41eca91783?t=1583397712000 (accessed on 20 February 2022).

FAIR (2016). *FAIR Principles*. Consulted at https://www.go-fair.org/fair-principles/ (accessed on 20 March 2023).

Gelfand, A. E., Diggle, P., Guttorp, P., & Fuentes, M. (2010). *Handbook of Spatial Statistics*. 1st Edition. CRC Press. https://doi.org/10.1201/9781420072884.

Geospatial World Forum & United Nations Statistics Division (2020). *Advancing Role of Geospatial Knowledge Infrastructure in World Economy, Society and Environment.* Discussion Document. Consulted at https://geospatialmedia.net/pdf/GKI-Discussion-Document-Ver1.0.pdf (accessed on 16 October 2022).

GEOSTAT 1A (2012). *GEOSTAT 1A – Representing Census data in a European population grid*. ESSnet GEOSTAT 1A project. EFGS and Eurostat. Consulted at https://www.efgs.info/wp-content/uploads/geostat/1a/GEOSTAT1A-final-report.pdf (accessed on 20 May 2021).

GEOSTAT 1B (2013). *Production Procedures for a Harmonised European Population Grid - Aggregation Method*. ESSnet GEOSTAT 1B project. EFGS and Eurostat. Consulted at https://www.efgs.info/wp-content/uploads/geostat/1b/GEOSTAT1B-Appendix1-production-procedures-bottom-up.pdf (accessed on 22 May 2021).

GEOSTAT 1B (2014). *GEOSTAT 1B Final Report*. ESSnet GEOSTAT 1B project. EFGS and Eurostat. Consulted at https://www.efgs.info/wp-content/uploads/geostat/1b/GEOSTAT1B-final-technical-report.pdf (accessed on 23 May 2021).

GEOSTAT 2 (2016). *Spatialising Statistics in the ESS - Results from the 2015 GEOSTAT 2 survey on geocoding practices in European NSIs*. ESSnet GEOSTAT 2 project. EFGS and Eurostat. Consulted at https://www.efgs.info/wp-content/uploads/2017/03/GEOSTAT2-Spatialising-statistics-in-the-ESS-2015.pdf (accessed on 25 May 2021).

GEOSTAT 2 (2017). *A Point-based Foundation for Statistics - Final report from the GEOSTAT 2 project*. ESSnet GEOSTAT 2 project. EFGS and Eurostat. Consulted at https://www.efgs.info/wp-content/uploads/2017/03/GEOSTAT2ReportMain.pdf (accessed on 30 May 2021).

GEOSTAT 3 (2019a*). GSGF Europe – Implementation guide for the Global Statistical Geospatial Framework in Europe*. Proposal from the GEOSTAT 3 project. ESSnet GEOSTAT 3 project. EFGS and Eurostat. Consulted at https://www.efgs.info/wp-content/uploads/geostat/3/GEOSTAT3_GSGF_EuropeanImplementationGuide_v1.0.pdf (accessed on 3 June 2021).

GEOSTAT 3 (2019b). *Automated Linking of SDMX and OGC Web Services - A Feasibility Study by the GEOSTAT 3 project on the implementation of principles 4 and 5 of the Global Statistical Geospatial Framework*. ESSnet GEOSTAT 3 project. EFGS and Eurostat. Consulted at https://www.efgs.info/wp-content/uploads/geostat/3/GEOSTAT3_WP1_Automated_Linking_SDMX_OGC_v1.0.pdf (accessed on 9 June 2021).

GEOSTAT 4 (2021a). *GSGF Europe: GSGF and Frameworks Environment*. ESSnet grant project GEOSTAT 4. Consulted at https://www.efgs.info/wp-content/uploads/2022/02/GSGF_and_FrameworksEnvironment.pdf (accessed on 5 April 2022).

GEOSTAT 4 (2021b). *GSGF Europe*. ESSnet grant project GEOSTAT 4. Consulted at https://www.efgs.info/wp-content/uploads/2022/03/GSGF_Europe.pdf (accessed on 10 April 2022).

GEOSTAT 4 (2021c). *GSGF Europe: Reference Architecture*. ESSnet grant project GEOSTAT 4. Consulted at https://www.efgs.info/wp-content/uploads/2022/02/GSGF_Europe_ReferenceArchitecture.pdf (accessed on 18 April 2022).

GEOSTAT 4 (2021d). *GEOSTAT 4: Recommendations for Geospatial Quality Reporting.* ESSnet grant project GEOSTAT 4. Consulted at https://www.efgs.info/wp-content/uploads/2022/03/GEOSTAT-4_Recommendations-for-Geospatial-Quality-Reporting.pdf (accessed on 22 April 2022).

GEOSTAT 4 (2021e). *GSGF Europe: Managing Confidentiality in Geospatial Statistics*. ESSnet grant project GEOSTAT 4. Consulted at https://www.efgs.info/wp-content/uploads/2022/02/ManagingConfidentiality_in_GeospatialStatistics.pdf (accessed on 30 April 2022).

GEOSTAT 4 (2021f*). GSGF Europe: A Point-based Foundation for Statistics*. ESSnet grant project GEOSTAT 4. Consulted at https://www.efgs.info/wp-content/uploads/2022/02/A_Point-basedFoundation_for_Statistics.pdf (accessed on 3 May 2022).

GEOSTAT 4 (2022a). *GSGF Europe: Terms and Definitions*. ESSnet grant project GEOSTAT 4. Consulted at https://www.efgs.info/wp-content/uploads/2022/02/Terms_and_Definitions.pdf (accessed on 6 May 2022).

GEOSTAT 4 (2022b). *GEOSTAT4: Table Joining Service Proof-of-Concept*. ESSnet grant project GEOSTAT 4. Consulted at https://www.efgs.info/wp-content/uploads/2022/02/81_GEOSTAT4-TJS-PoC-_v1.0.pdf (accessed on 11 May 2022).

GEOSTAT 4 (2022c). *GSGF Europe: Requirements and Recommendations*. ESSnet grant project GEOSTAT 4. Consulted at https://www.efgs.info/wp-content/uploads/2022/02/RequirementsAndRecommendations-1.pdf (accessed on 20 May 2022).

GEOSTAT 4 (2022d). *GSGF Europe: Geospatial Statistics According to Data Collection Method – Survey Data*. ESSnet grant project GEOSTAT 4. Consulted at https://www.efgs.info/wp-content/uploads/2022/02/Geostat4-SurveyData.pdf (accessed on 22 April 2022).

GEOSTAT 4, (2022e). *GSGF Europe: Geospatial Statistics According to Data Collection Method – Big Data*. ESSnet grant project GEOSTAT 4. Consulted at https://www.efgs.info/wp-content/uploads/2022/02/Geostat4_BigData.pdf (accessed on 22 April 2022).

GEOSTAT 4, Eurostat (2020). *Results from the GEOSTAT 4/GISCO survey*. GEOSTAT 4 Webinar, 26-28 May. Consulted at https://www.efgs.info/wp-content/uploads/2020/06/GEOSTAT4_WP2_webinar_May2020.pdf (accessed on 23 January 2023).

Goodchild, M. F (2010). *Twenty of progress: GIScience in 2010*. Journal of Spatial Information Science (1): 3-20. doi: 10.5311/JOSIS.2010.1.2.

Goodchild, M. F., Duckham, M., Worboys, M. (2003). *Foundations of Geographic Information Science*. CRC Press.

Griffith, D. A., (2003). *Spatial autocorrelation and spatial filtering: Gaining understanding through theory and scientific visualization.* Springer-Verlag. doi:

10.1007/978-3-540-24806-4.

Guptill, S. C. (2001). *Spatial Data*. In International Encyclopedia of the Social & Behavioral Sciences, Pergamon, Editors: Neil J. Smelser, Paul. B. Baltes, 14775-14778. ISBN 9780080430768. doi: 10.1016/B0-08-043076-7/02508-0.

Hadley, C. (2018). *The Global Fundamental Geospatial Data Themes Journey*. Presentation in the International Workshop on Global Fundamental Geospatial Data Themes for Africa, Addis Ababa, Ethiopia. Consulted at https://ggim.un.org/meetings/2018-Addis_Ababa/documents/1.10.Clare_Hadley.pdf (accessed on 15 April 2023).

Haldorson, M., & Mostrom, J. (2018). *Implementing the Statistical Geospatial Framework at Statistics Sweden*. Draft National Report, Statistics Sweden. Consulted at https://www.efgs.info/wp-content/uploads/geostat/3/Impementing-the-GSGF-at-Statistics-Sweden v0.96_Draft_for_GEOSTAT.pdf (accessed on 4 November 2022).

Haldorson, M., Zaccheddu, P. G., Fohgrub, B., & Petri, E. (2016). *Geospatial information management in Europe - Responding to the user needs*. Statistical J. of the IAOS, 32(4), 481–487.

Hardwood, A., & Mayer, A. (2016). *Big data and semantic technology: A future for data integration, exploration and visualisation*. Statistical Journal of the IAOS, 32(4), 613-626. doi: 10.3233/SJI-160989.

Hassani, H., Saporta, G., & Silva, E.S. (2014). *Data mining and official statistics: the past, the present and the future.* Big Data, 2(1), 34–43. doi: 10.1089/big.2013.0038.

Instituto Nacional de Estatística (2024*). Sistema de Metainformação (SMI).* Consulted at https://smi.ine.pt/ (accessed on 31 May 2024).

Instituto Nacional de Estatística. (2022). *Policy of Statistical Confidentiality*. Consulted at https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_inst_confestatistica&xlang=en (accessed on 14 June 2024).

International Organization for Standardization (2002). *ISO 19108:2002, Geographic information - Temporal schema.* Edition 1. Consulted at https://www.iso.org/standard/26013.html (accessed on 10 November 2022).

International Organization for Standardization (2013). *ISO 19157:2013, Geographic information — Data quality.* Edition 1. Consulted at https://www.iso.org/standard/32575.html (accessed on 10 November 2022).

International Organization for Standardization (2014). *ISO 19115-1:2014, Geographic information - Metadata.* Edition 1. Consulted at https://www.iso.org/standard/53798.html (accessed on 10 November 2022).

International Organization for Standardization (2016*). ISO 19119:2016, Geographic information - Services.* Edition 2. Consulted at https://www.iso.org/standard/59221.html (accessed on 10 November 2022).

International Organization for Standardization (2019a*). ISO 19107:2019, Geographic* *information - Spatial schema.* Edition 2. Consulted at https://www.iso.org/standard/66175.html (accessed on 10 November 2022).

International Organization for Standardization (2019b). *ISO 19111:2019, Geographic information - Referencing by coordinates.* Edition 3. Consulted at https://www.iso.org/standard/74039.html#amendment (accessed on 10 November 2022).

International Organization for Standardization (2019c*). ISO 19112:2019, Geographic information - Spatial referencing by geographic identifiers.* Edition 2. Consulted at https://www.iso.org/standard/70742.html (accessed on 10 November 2022).

International Organization for Standardization (2023). *ISO 19157:2023, Geographic information — Data quality (Part 1: General requirements).* Consulted at https://www.iso.org/standard/78900.html (accessed on 10 November 2022).

Julião, R. P. (2001). *Tecnologias de Informação Geográfica e Ciência Regional: Contributos Metodológicos para a Definição de Modelos de Apoio à Decisão em Desenvolvimento Regional.* Faculdade de Ciências Sociais e Humanas, Universidade Nova de Lisboa.

Kadmon, R. (2001). *Remote Sensing and Image Processing*. In Encyclopedia of Biodiversity. Elsevier, 121-143. https://doi.org/10.1016/B0-12-226865-2/00232-7.

Kraak, M., Roth, R. E., Ricker, B., Kagawa, A., & Le Sourd, G. (2020). *Mapping for a Sustainable World*. United Nations. New York. Consulted at https://digitallibrary.un.org/record/3898826/files/MappingforaSustainableWorld20210603.pdf?ln=en (accessed on 10 October 2023).

Loonis, V., de Bellefon, M., Costemalle, V., Fontaine, M., Ardilly, P., Audric, S., Buron, M., Durieux, E., Eusébio, P., Favre-Martinoz, C., Floch, J., Fontaine, M., Genebes, L., Le Gleut, R., Lardeux-Schutz, R., Lévy, D., Le Saout, R., Merly-Alpa, T., Renaud, A., Sémécurbe, F., Bouayad-Agha, S., Marcon, E., Puech, F., Le Gallo, J., Védrine, L., Bouche, P., & Zhu, W. (2018). Handbook of spatial analysis. Theory and practical application with R. Insee Méthodes 131. Consulted at https://ec.europa.eu/eurostat/documents/3859598/9462709/INSEE-ESTAT-SPATIAL-ANA-18-EN.pdf/c4f87d5b-b508-4aff-ad7d-264da463077e?t=1545319662000 (accessed on 6 March 2024).

Magnani, M., & Montesi, D. (2010). *A Survey on Uncertainty Management in Data Integration*. Journal of Data and Information Quality, (2)1, 1-33. doi: 10.1145/1805286.1805291.

Maguire, D. J, & Longley, P. A. (2005). *The emergence of geoportals and their role in spatial data infrastructures.* Computers, Environment and Urban Systems, 29, 3–14. doi: 10.1016/j.compenvurbsys.2004.05.012.

Man, W. H. E. (2013). *Thinking Outside the Disciplinary Box in Coping with Dilemmas in Geoinformation Management for Public Policy*. Transactions in GIS, 17(3), 452-462. https://doi.org/10.1111/tgis.12008.

Masser, I, (2020). *Beyond SDI – Current trends in spatial data infrastructures - not one* *but many SDIs*. European Umbrella Organisation for Geographic Information. UN-GGIM, Solving the Puzzle: Understanding the Implementation Guide, IGIF – Global Consultation Draft. Consulted at https://eurogi.org/beyond-sdi-ian-masser/ (accessed on 22 June 2022).

Masser, I. (2005). *The future of spatial data infrastructures*. ISPRS Workshop on Service and Application of Spatial Data Infrastructure, XXXVI (4/W6), Oct.14-16, Hangzhou, China. Consulted at https://www.isprs.org/proceedings/xxxvi/4-w6/papers/7-16IanMasser-A001.pdf (accessed on 20 June 2022).

Neves, C., & Moreira, F. (2017*). How to integrate statistical and geospatial data using administrative data on construction works - The case of Indicators System of Urban Operations (SIOU).* Statistical Journal of the IAOS, 33, 627–635. doi: 10.3233/SJI-170382.

Organisation for Economic Co-operation and Development (2007). *Glossary of Statistical Terms.* OECD Publishing, Paris. https://doi.org/10.1787/9789264055087-en.

Organisation for Economic Co-operation and Development, International Labour Organization, International Monetary Fund & International Statistical Committee of the Commonwealth of Independent States (2002). *Measuring the Non-Observed Economy: A Handbook.* OECD Publishing, Paris. https://doi.org/10.1787/9789264175358-en.

Partnership in Statistics for Development in the 21st Century (2022). *Digital Transformation of National Statistical Offices.* OECD Publishing, Paris. https://doi.org/10.1787/ee4b1b85-en.

Partnership in Statistics for Development in the 21st Century & Statistics Sweden (2021). *Guide on Geospatial Data Integration in Official Statistics.* Consulted at https://www.paris21.org/sites/default/files/inline-files/Geospatial_Data_Integration_in_Official_Statistics_0.pdf (accessed on 13 October 2022).

Rajabifard, A., Feeney, M.E., Williamson, I., & Masser, I. (2003). *Developing Spatial Data Infrastructures: from Concept to Reality.* 1st Edition. CRS Press. https://doi.org/10.1201/9780203485774.

Reimsbach-Kounatze, C. (2015*). The Proliferation of "Big Data" and Implications for Official Statistics and Statistical Agencies: A Preliminary Analysis*. OECD Digital Economy Papers, No. 245, OECD Publishing, Paris. http://dx.doi.org/10.1787/5js7t9wqzvg8-en.

Scott, G., & Rajabifard, A. (2017). Sustainable development and geospatial information: a strategic framework for integrating a global policy agenda into national geospatial capabilities. Geo-Spatial Information Science, 20(2), 59–76. doi: 10.1080/10095020.2017.1325594.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H. Y., Milèinski, G., Nikšieč, M., Painho, M., Podör, A., Olteanu-Raimond, A. M. R., & Rutzinger, M. (2016). *Crowdsourcing, citizen science or volunteered geographic information? The current state of* *crowdsourced geographic information.* ISPRS Int. J. of Geo-Info., 5(5), 23.

Sun, K., Zhu, Y., Pan, P., Hou, Z., Wang, D., Weirong Li, W., & Song, J. (2019). *Geospatial data ontology: the semantic foundation of geospatial data integration and sharing.* Big Earth Data, 3(3), 269-296, doi: 10.1080/20964471.2019.1661662.

Tait, M. (2005). *Implementing geoportals: Applications of distributed GIS.* Computers, Environment and Urban Systems, 29(1), 33-47. doi: 10.1016/j.compenvurbsys.2004.05.011.

United Nations (2014). *Fundamental Principles of Official Statistics.* UN General Assembly Resolution 68/261, United Nations. Consulted at https://unstats.un.org/unsd/dnss/gp/FP-New-E.pdf (accessed on 4 March 2022).

United Nations Committee of Experts on Global Geospatial Information Management (2015). *A guide to the Role of Standards in Geospatial Information Management*. Prepared cooperatively by the Open Geospatial Consortium (OGC); The International Organization for Standards (ISO) Technical Committee 211 Geographic information/Geomatics; and the International Hydrographic Organization (IHO). Consulted at https://ggim.un.org/documents/standards%20guide%20for%20unggim%20-%20final.pdf (accessed on 16 September 2022).

United Nations Committee of Experts on Global Geospatial Information Management (2018). *Determination of global fundamental geospatial data themes*. Eighth session. New York, 1–3 August 2018. Consulted at https://ggim.un.org/meetings/GGIM-committee/8th-Session/documents/E_C.20_2018_7-global_fundamental_E.pdf (accessed on 5 July 2022).

United Nations Committee of Experts on Global Geospatial Information Management (2019). *The Global Fundamental Geospatial Data Themes*. Consulted at https://ggim.un.org/meetings/GGIM-committee/9th-Session/documents/Fundamental_Data_Publication.pdf (accessed on 16 September 2022).

United Nations Committee of Experts on Global Geospatial Information Management (2020). *Future Trends in Geospatial Information Management: The Five to Ten Year Vision*. Third Edition. Consulted at https://ggim.un.org/meetings/GGIM-committee/10th-Session/documents/Future_Trends_Report_THIRD_EDITION_digital_accessible.pdf (accessed on 21 September 2022).

United Nations Committee of Experts on Global Geospatial Information Management (2020). *Integrated Geospatial Information Framework (IGIF)*. Consulted at https://ggim.un.org/IGIF/overview/ (accessed on 30 September 2022).

United Nations Committee of Experts on Global Geospatial Information Management (2021). *The SDGs Geospatial Roadmap.* Consulted at https://ggim.un.org/meetings/GGIM-committee/11th-Session/documents/The_Geospatial_SDGs_Roadmap_WGGI_IAEG_SDGs_20210804.pdf (accessed on 25 September 2022).

United Nations Committee of Experts on Global Geospatial Information Management (2022). *Development of Contextual Understanding, Information, and Analytics towards Determining the National Geospatial Information Ecosystem.* Discussion Paper. Consulted at https://ggim.un.org/meetings/GGIM-committee/12th-Session/documents/Future_National_Geospatial_Information_%20Ecosystem_Jul2022.pdf (accessed on 7 January 2023).

United Nations Committee of Experts on Global Geospatial Information Management (2024). *Policy and legal frameworks, including issues related to authoritative data.* Fourteenth session. New York, 7–9 August 2024. Consulted at https://documents.un.org/doc/undoc/gen/n24/138/46/pdf/n2413846.pdf (accessed on 23 April 2022).

United Nations Committee of Experts on Global Geospatial Information Management: Europe (2015). Data Integration - Definition of priority user needs for combinations of data. Version 1.1. Work Group B. First delivery. Consulted at https://un-ggim-europe.org/wp-content/uploads/2018/11/2015-10-23_UN-GGIM-Europe_Report-from-SWG-B1-on-Priority-User-Needs-ver-1.1-1.pdf (accessed on 10 March 2023).

United Nations Committee of Experts on Global Geospatial Information Management: Europe (2016). *Core Data Scope*. Version 1.2. Working Group A - First Deliverable of Task 1.a. Consulted at https://un-ggim-europe.org/wp-content/uploads/2018/11/UN-GGIM-Europe-WGA-Core_Data_Scope-v1.2.pdf (accessed on 16 March 2023).

United Nations Committee of Experts on Global Geospatial Information Management: Europe (2019). *The Territorial Dimension In SDG Indicators: Geospatial Data Analysis and Its Integration with Statistical Data*. Working Group on Data Integration. Consulted at https://un-ggim-europe.org/wp-content/uploads/2019/05/UN_GGIM_08_05_2019-The-territorial-dimension-in-SDG-indicators-Final.pdf (accessed on 5 April 2023).

United Nations Committee of Experts on Global Geospatial Information Management: Europe (2021). *Data Integration Methods - Analysis of future trends in geospatial data capture, creation, maintenance and management and recommendations for amplified use of good practices*. Version 1.1. Working Group on Data Integration. Consulted at https://un-ggim-europe.org/wp-content/uploads/2021/11/UNGGIM-Europe_WG_DataIntegration_Subgroup-II_DataIntegrationMethods.pdf (accessed on 11 March 2023).

United Nations Economic Commission for Africa (2007). *Geoinformation in Socio-Economic Development: Determination of Fundamental Datasets for Africa*. Consulted at https://repository.uneca.org/bitstream/handle/10855/21282/Bib%2056965.pdf?sequence=1&isAllowed=y (accessed on 16 September 2022).

United Nations Economic Commission for Europe (2014). *Big data - an opportunity or threat to official statistics?*. Sixty-second plenary session. Conference of European Statisticians. Paris, 9-11 April 2014. Consulted at https://unece.org/fileadmin/DAM/stats/documents/ece/ces/2014/32-

Eurostat-Big_Data.pdf (accessed on 20 August 2022).

United Nations Economic Commission for Europe (2015). *Conference of European Statisticians - Recommendations for the 2020 Censuses of Population and Housing*. Consulted at https://unece.org/DAM/stats/publications/2015/ECECES41_EN.pdf (accessed on 16 May 2023).

United Nations Economic Commission for Europe (2016*). A common international conceptual framework for geospatial and statistical data acquisition, data management, and data use: goals and barriers*. Conference of European Statisticians. Sixty-fourth plenary session, Paris. Consulted at https://unece.org/fileadmin/DAM/stats/documents/ece/ces/2016/mtg/ECE_CES_2016_24-1602220E.pdf (accessed on 13 November 2022).

United Nations Economic Commission for Europe (2017). *A Guide to Data Integration for Official Statistics*. Version 2.0. High-Level Group for the Modernisation of Official Statistics (HLG-MOS). Consulted at https://unece.org/sites/default/files/2024-07/HLG-MOS%20Guide%20to%20Data%20Integration%20for%20Official%20Statistics.pdf (accessed on 20 July 2022).

United Nations Economic Commission for Europe (2018). *Common Statistical Data Architecture (CSDA).* Version 2.0. Consulted at https://statswiki.unece.org/display/DA/CSDA+2.0 (accessed on 22 July 2022).

United Nations Economic Commission for Europe (2019). *In-depth review of satellite imagery/earth observation technology in official statistics*. Conference of European Statisticians. 67th plenary session. Geneva, 26-28 June 2019. Consulted at https://unece.org/DAM/stats/documents/ece/ces/2019/ECE_CES_2019_16-1906490E.pdf (accessed on 30 May 2022).

United Nations Economic Commission for Europe (2019a*). Generic Statistical Business Process Model (GSBPM)*. Version 5.1. Consulted at https://statswiki.unece.org/display/GSBPM/GSBPM+v5.1 (accessed on 17 July 2022).

United Nations Economic Commission for Europe (2019b). *Generic Activity Model for Statistical Organisations (GAMSO).* Version 1.2. Consulted at https://statswiki.unece.org/display/GAMSO/GAMSO+v1.2 (accessed on 19 July 2022).

United Nations Economic Commission for Europe (2021a). *Machine Learning for Official Statistics.* Consulted at https://unece.org/sites/default/files/2022-09/ECECESSTAT20216.pdf (accessed on 21 May 2022).

United Nations Economic Commission for Europe (2021b). *Common Statistical Production Architecture (CSPA).* Version 2.0. Consulted at https://statswiki.unece.org/display/CSPA/CSPA+v2.0 (accessed on 24 July 2022).

United Nations Economic Commission for Europe (2021c). *Geospatial view of Generic*

Statistical Business Process Model (GeoGSBPM). Version 1.0. Consulted at https://statswiki.unece.org/display/GSBPM/GeoGSBPM (accessed on 11 October 2022).

United Nations Economic Commission for Europe (2024a). *Building capacity in statistical and geospatial data integration across the United Nations Economic Commission for Europe for Europe region.* Conference of European Statisticians. Seventy-second plenary session. Geneva, 20-21 June 2024. Consulted at https://unece.org/sites/default/files/2024-05/ECE_CES_2024_08_E.PDF (accessed on 6 March 2023).

United Nations Economic Commission for Europe (2024b) *Generic Statistical Information Model (GSIM).* Version 2.0. Consulted at https://github.com/UNECE/GSIM-2.0/blob/main/README.md (accessed on 26 July 2022).

United Nations Economic Commission for Europe (2024c). *INGEST Task Force on Standards Issues - A path towards the use of common standards to support the INtegration of GEospatial and STatistical information.* Consulted at https://unece.org/sites/default/files/2024-08/INGEST%20Task%20Force%20on%20Standards%20Issues_UNECE%20Working%20Paper_combined.pdf (accessed on 23 July 2024).

United Nations Statistical Commission and United Nations Committee of Experts on Global Geospatial Information Management (2019). *The Global Statistical Geospatial Framework (GSGF).* Consulted at https://unstats.un.org/unsd/statcom/51st-session/documents/The_GSGF-E.pdf (accessed on 2 February 2021).

United Nations Statistics Division (2000). *Handbook on geographic information systems and digital mapping.* Consulted at https://unstats.un.org/unsd/publication/seriesf/seriesf_79e.pdf (accessed on 14 July 2022).

United Nations Statistics Division (2017). *Earth Observations for Official Statistics - Satellite Imagery and Geospatial Data Task Team report.* Consulted at https://unstats.un.org/bigdata/task-teams/earth-observation/UNGWG_Satellite_Task_Team_Report_WhiteCover.pdf (accessed on 15 September 2022).

United Nations Statistics Division (2019). *United Nations National Quality Assurance Frameworks Manual for Official Statistics - Including recommendations, the framework and implementation guidance*. Consulted at https://unstats.un.org/UNSDWebsite/nqaf/Attachment%20B-UNNAQF%20Manual%20in%20PDF.pdf (accessed on 20 April 2022).

United Nations Statistics Division (2021). *Handbook on Management and Organization of National Statistical Systems*. Third Edition. Consulted at https://unstats.un.org/capacity-development/handbook/index.cshtml (accessed on 5 May 2022).

Van Halderen, G., Minchin, S., Brady, M., & Scott, G. (2016). *Integrating statistical and geospatial information, cultures and professions: International developments* *and Australian experience.* Statistical Journal of the IAOS, 32(4), 457–470. doi: 10.3233/SJI-161018.

World Wide Web Consortium (2014). *Resource Description Framework (RDF).* Consulted at https://www.w3.org/RDF/ (accessed on 17 March 2023).

Yung, W., Karkimaa, J., Scannapieco, M., Barcarolli, G., Zardetto, D., Sanchez, J. A. R., Braaksma, B., Buelens. B., & Burger, J. (2018). *The use of machine learning in official statistics.* UNECE Machine Learning Team. Consulted at https://unece.org/sites/default/files/2024-07/HLGMOS%20The%20use%20of%20machine%20learning%20in%20official%20statistics.pdf (accessed on 4 June 2022).

Zhang, G., & Zhu, A. (2018.) *The representativeness and spatial bias of volunteered geographic information: a review*. Annals of GIS, 24(3), 151-162, doi: 10.1080/19475683.2018.1501607.