A Work Project presented as part of the requirements for the Award of a Masters Degree in Management from the NOVA – School of Business and Economics

Responsible Data Practices in the Social Economy Sector in Portugal:

A Code of Conduct for Data Collection, Management, and Sharing

MAFALDA SOFIA CARVALHAL ABREU

54680

A Project carried out under the supervision of:

Advisor: Leid Zejnilović

Co-advisor: Lénia Mestrinho

20-12-2023

Abstract

As various sectors undergo digital transformation, the significance of ethical and responsible data management has become increasingly prominent. The social economy emerges as a particularly vital ecosystem due to its inherently social nature. Acknowledging this importance, a need to establish a community of practice and support data production, management and sharing within the social economy sector has been identified. This thesis aims to develop a code of conduct for data collection, management and sharing, with a focus on Portuguese social economy entities. Anchored in legal frameworks, a quantitative research approach has been employed to glean insights into the current data and identify key factors that can enhance the code's utility and guidance. The results reveal several gaps in law compliance.

Keywords: Social Economy organizations; Data practices; General Data Protection Regulation; Code of Conduct

Table of Contents

1. Introduction 3
2. Background 5
   2.1. Portuguese Social Economy 5
   2.2. Legal Frameworks 6
3. Literature Review 7
   3.1. Data Landscape 7
4. Code of Conduct 11
5. Research and Data Analysis Methodologies 12
6. Research Findings 14
7. Discussion of Results 17
8. Application of the Code of Conduct 20
9. Conclusion 21
10. References 23
11. Appendices 27

1. Introduction

In an era where digital transformation has become the catalyst for sweeping change, the importance of digital transition cannot be emphasized enough. Its processes have showcased the capacity to develop and implement sustainable solutions for addressing several issues such as poverty and unequal opportunities (Rosário and Dias, 2022), also playing a pivotal role in facilitating the achievement of long-term carbon neutrality goals, providing an effective contribution to the creation of more and better jobs, the internationalization of companies, and the modernization of society in general (Diário da República, 2020).

While the roots of this phenomenon can be traced back to the 1950s (Rosário and Dias, 2022), the ongoing digitalization is set to further expand and deepen (OECD, 2017). This holds true for the European Commission, which proposed co-creating transition pathways for 14 industrial ecosystems, given that both green and digital transitions have been established as political priorities. One such pathway focuses on the proximity and social economy industrial ecosystem as parts of the social economy contribute to transitions at large by providing sustainable goods and services while addressing the digital divide. At the same time, the commission believes that reinforcing sustainable and digital practices will build long-term resilience and empower the various stakeholders to be active levers of this transformative process. Nonetheless, social economy organizations generally have a low level of digitalization, primarily concerning digital skills, investment in digital infrastructure, and leveraging the potential offered by the platform economy (European Commission, 2022).

A key action area to tackle these issues and drive the digital transition involves the development of a code of conduct for data collection, management and sharing in the social economy. This objective is also recognized by Nova SBE Data Science Knowledge Center (Nova SBE DSKC), whose main mission is to advance knowledge about data-driven

decision-making and its application in society. Through its Social Sector Database project, the Nova SBE DSKC is actively working on a public and easily accessible platform to map the social ecosystem and address the challenges of limited and disjointed data sources. Recognizing the need to support data collection and production in social organizations, this entity perceives that a guide on the use and management of its data and indicators can help address it. In this regard, this thesis is conducted in collaboration with the Nova SBE DSKC as a Field Lab, with the overarching goal of developing a code of conduct for data collection, sharing and management in the social economy and explore how to best utilize it. The research question to be addressed is: 'Which factors require consideration in a data processing code of conduct to ensure its utility and guidance for social organizations in Portugal?'. This encompasses considerations such as vulnerabilities, potentials, and imperatives within the realm of data practices, whether it involves evaluating if organizations adopt specific practices, thus, shedding light on potential vulnerabilities, analyzing the types of data handled, which can reveal imperatives such as legal requirements or identifying areas and capabilities that have the potential to enhance the social economy.

Hence, the thesis' structure unfolds as follows. Firstly, the background and theoretical foundation, covering national social economy and legal frameworks at both the national and European Union levels, is presented, as well as an analysis of the data landscape in the sector. Subsequently, the establishment of the best practices from a legal standpoint is undertaken to formulate a benchmark code of conduct. The research section regarding current data practices of Portuguese social impact organizations then unfolds, covering the research and data methodologies, research findings and a thorough discussion. Finally, suggestions on how to improve the application of the code of conduct, encapsulating the culmination of the theoretical foundation and empirical research will be provided.

2. Background

2.1. Portuguese Social Economy

Social economy, as defined by the Portuguese law, *Lei de Bases da Economia Social – Lei Nº 30/2013*, May 8th, encompasses all socio-economic activities freely carried out by the entities that are covered by Portuguese legal regulations. These comprise a wide array of organizations: cooperatives, mutual associations, mercies, foundations, private social solidarity institutions (IPSS) not falling within the aforementioned categories, associations with altruistic purposes operating in the cultural, recreational, sports, and local development fields, entities covered by community and self-management subsectors, incorporated, according to the Constitution, in the cooperative and social sector and other legal entities that adhere to the guiding principles of the social economy and are listed in the social economy database. The set of fundamental guiding principles, which serve as the cornerstone of social economy organizations' activities, is as follows: a) the primacy of people and social objectives; b) free and voluntary membership and participation; c) democratic control; d) conciliation between the interests of members, users, or beneficiaries and the general interest; e) respect for the values of solidarity, equality, non-discrimination, social cohesion, justice, fairness, transparency, individual and shared social responsibility, and subsidiarity; f) autonomous and independent management from public authorities and any other external entities to the social economy; and g) the allocation of surpluses for the pursuit of the goals of social economy entities in line with the general interest (Diário da República, 2013).

According to the Portuguese *Social Economy Satellite Account 2019-2020*, published in July 2023, there were 73,851 social economy entities in 2020, 0.4% more than in 2019. These entities generated 3.2% of national GVA (Gross Value Added), 5% of compensation of employees, 5.2% of total employment and 5.9% of employees (expressed, in both cases, in full-time equivalent units - FTE) (INE, 2023).

In summary, Portugal's social economy substantially impacts the national economy, encompassing diverse legal and organizational structures. However, despite these variations in structures, social impact organizations frequently face similar challenges in their development and expansion (OECD, 2022) and legal frameworks can assist them in overcoming these challenges. This sets the stage for the subsequent introduction of regulations that shape the sector and its relationship with data.

2.2. Legal Frameworks

Legal and regulatory frameworks lay the groundwork for organizations by establishing clear definitions, objectives, and guiding principles, which can help them address challenges across various domains. Their influence extends to provide visibility, recognition, credibility and facilitate growth for social economy organizations (OECD, 2022), effectively defining the contours of their operational landscape.

Amidst these regulatory considerations, it's crucial to introduce data protection laws, such as the General Data Protection Regulation (GDPR) and the *Lei da Proteção de Dados Pessoais* (Portuguese Data Protection Law), especially given that many Portuguese organizations have reported the collection of both highly confidential and highly sensitive data (Zejnilovic and Mestrinho, 2020).

The GDPR, enforced in May 2018, represents a seminal milestone in the realm of data privacy and protection, with profound implications for organizations operating within the European Union. It harbors a twofold purpose: to enhance the data protection rights of individuals and to improve business opportunities by facilitating the free flow of personal data in the digital single market (Council of the European Union, 2012). Non-compliance with this regulation can result in significant financial penalties, rendering it an indispensable legal tool within the landscape of data privacy and protection.

Transitioning to a country-level context, the *Lei da Proteção de Dados Pessoais* serves an important role in aligning the nation with the GDPR, while simultaneously offering context-specific nuances on data protection within the country. It lays down rules and guidelines for data processing, security measures, and the appointment of data protection officers within organizations, with oversight provided by the data protection authority, *Comissão Nacional de Proteção de Dados* (CNPD) (Ministério Público, 2019). In the case of social economy organizations based in Portugal, this law is of utmost importance, as it stipulates how they should handle personal data and ensures that they are in line with both European and national data protection standards.

As organizations navigate the complex terrain of data privacy and protection, compliance with these laws is, as observed, indispensable. However, only about 76% of Portuguese social economy entities claim to be in "total compliance" with the GDPR. This may be explained by a lack of full awareness about its specifics and the incorrect belief that nonprofits are exempt from its stipulations (Zejnilovic and Mestrinho, 2020). Thus, there is a pressing need to bolster data protection legislation, prompting the call for the development of a comprehensive code of conduct in data processing within this sector. To further explore factors such as these that merit contemplation in the code, a closer look at the current data landscape becomes essential, coupled with an understanding of how handling data can yield benefits for the sector.

3. Literature Review

3.1. Data Landscape

In recent years, data has become an integral part of operations for organizations across various sectors. The data movement is in full swing (Patil, 2015), with a growing emphasis on the importance of data management and governance (Enders, 2018). Consequently, data is

now widely recognized as a pivotal factor in the long-run growth of any modern economy (Cong et al., 2022), highlighting the importance of adopting responsible practices. This is particularly important for the social economy, where most organizations often interact with individuals in vulnerable circumstances.

To develop its "Transition pathway for Proximity and Social Economy" (2022), the European Commission consulted with stakeholders to assess the data landscape. One key finding was a lack of basic awareness about data - what it encompasses, how it can be utilized, which risks may occur and what advantages it may offer in improving and expanding business activity. This highlights the need for a basic understanding and awareness of concepts, responsibilities and legal obligations. Data maturity levels were also found to be generally quite low in this ecosystem. In particular, social economy entrepreneurs, namely those operating small and micro-enterprises, face a challenge in managing data and establishing protocols for storage, ownership, sharing, and monetization of data. (European Commission, 2022). Although these smaller organizations, rarely, if ever, use complex databases, the need for effective management of low-volume simplistic data remains vital to ensure accuracy, security, and usability (Hernandez-Hall, 2021). Prominent barriers faced by organizations include difficulties in identifying meaningful information, a lack of technical expertise and external influences that can disrupt operations (Mayer and Fischer, 2023). This absence of practical knowledge is corroborated by the Digital Equity Survey conducted by Connect Humanity (2022), which indicates that only 37% of Portuguese organizations provide digital training to their workforce and/or the individuals they serve.

A lack of a comprehensive and standardized data governance framework further aggravates the situation, contributing to poor data standards and quality, while also introducing concerns regarding data security. These technical, administrative and trust-related challenges limit the

ability of cross-border data access and sharing (ECTAA, 2023). Therefore, having a guiding code of conduct for data processing is crucial to help address them and for organizations to responsibly leverage the various benefits that data can offer.

Beyond the financial benefits derived from boosting performance and improving the bottom line (Brahmaiah and Sreekrishna, 2020), data analytics also holds a profound social value in addressing pressing societal issues. Civic hackathons, for instance, invite the general public to solve societal problems by forming teams and developing solutions within a time-compressed setting. Governments and NGOs have used them to encourage the use of available data sets for innovative problem-solving and to extract public value (Yuan and Gasco-Hernandez, 2021; Mergel, 2015). An example of a successful hackathon is the #WirVsVirus project in Germany, which mobilized thousands of participants to generate numerous project ideas to tackle complex societal challenges that emerged in light of the COVID pandemic. This initiative was grounded in the concept of Open Social Innovation (OSI), serving as a testament to the role of openness in two key aspects. It facilitated data sharing among all stakeholders by deploying well-known and user-friendly digital tools and mobilized supporters to contribute resources to the teams participating in the hackathon (Gegenhuber et al., 2023). In alignment with this spirit, the European Commission has actively pursued open data initiatives by supporting the creation of EU-wide common data spaces, which aim to "overcome existing legal and technical barriers to data sharing and, as such, unleash the enormous potential of data-driven innovation" (European Commission, 2022). They also expressed that a code of conduct for social impact organizations could act as the initial stride toward engagement in these common data spaces by promoting awareness of the challenges and opportunities in Business-to-Business, Business-to-Government, and Government-to-Government contexts (European Commission, 2021). In this context, discussing data interoperability is relevant as it emerges as a fundamental requirement for

enabling data exchange among entities. In the European Interoperability Framework, it is defined as “the ability of different organizations to interact towards mutually beneficial goals, involving the sharing of information and knowledge between its entities through the business processes they support” (European Commission, 2017). Stakeholders consulted by the European Commission highlighted the importance of educating social entrepreneurs about the interplay between data management, interoperability, and broader advancements in data technologies, “including the potential of artificial intelligence and Machine Learning for the improvement of services” (European Commission, 2022). Thus, promoting data interoperability is relevant. Nonetheless, a certain degree of data maturity is needed among social economy organizations, which is currently mostly lacking.

Another factor to take into account concerns the financial considerations in implementing data practices. Expenses for training personnel on updated data procedures and security protocols may involve hiring trainers, creating training materials or enrolling employees in external training programs. Monitoring of data practices may require investing in auditing tools and personnel. Upgrading or implementing new systems, databases and tools may also include significant financial commitment. Portuguese organizations share a unanimous concern about their biggest challenge – funding difficulties. While public funding serves as a primary income source, complemented by private donations, the perceived decline in public funding is noted by almost all organizations (Monteiro et al., 2015). Moreover, donors may impose funding restrictions, such as not allowing infrastructure spending, thereby constraining how nonprofits manage their technology and their data (Hernandez-Hall, 2021).

These key factors that characterize the social economy data landscape must be considered in developing and applying a code of conduct within this sector. Such code should serve as a comprehensive guide, enhancing awareness of data concepts, prioritizing practical guidance

on data management protocols and standards and promoting responsible data utilization and reutilization. Thus, it can boost data maturity levels, further ensure compliance with legal and ethical obligations, and enable responsible data sharing.

4. Code of Conduct

By analyzing the GDPR, the *Lei da Proteção de Dados Pessoais* and various data protection codes of conduct and policies, such as those by UNICEF, Banco Alimentar and International Red Cross and Red Crescent Movement, a code of conduct has been designed to tackle the widespread lack of knowledge regarding data and data practices identified within the sector. It intends to gather the best practices from a legal standpoint, also making data handling easily understandable for organizations through the use of plain language, icons, and the inclusion of practical examples. As such, it can serve as a benchmark for intervening with social economy entities and addressing their particular needs.

The code of conduct, which can be found in appendix 1, encompasses considerations for both personal and non-personal data and is by no means a binding document for organizations. It covers data collection, management and sharing and features a glossary of important definitions. As organizations frequently handle personal data, topics such as consent, data subject rights, data retention periods and purpose limitation were considered. This extends to the handling of confidential information, prompting the inclusion of data security practices, which encompass data backups, restrictions on data access and pseudonymization. Upon reviewing the guidelines on personal data breach notification by the European Commission (2018) and also made available by the CNPD, it was considered relevant to incorporate details on what information should be conveyed to this authority in the case of a data breach. Simultaneously, acknowledging the European Commission's emphasis on fostering data sharing among different entities, the proposed code of conduct integrates insights into secure

data-sharing practices such as ensuring that entities accessing the data are reputable and encouraging traceability for data transfers and modifications.

The code was continuously refined throughout the entire duration of this Work Project. To guarantee its compliance with the legislation, it underwent validation by a lawyer, whose review focused on verifying whether it appropriately conveyed legal obligations and ensured that every suggested practice aligns with current law.

5. Research and Data Analysis Methodologies

To delve deeper into the current data landscape of the social economy sector in Portugal, a quantitative research initiative was undertaken. Its main objectives were to assess and characterize data practices implemented by social economy organizations and to determine their views on the need for a code of conduct in data processing. By doing so, the research aligns with the broader aim of this thesis by identifying key factors that can be considered in offering appropriate guidance for data practices within this sector. This assessment was guided by a dual commitment, which involved evaluating the compliance of data practices with the current legislation (appendix 2).

With this purpose in mind, a survey was conducted in the Portuguese language to target organizations within this ecosystem. This survey was made through Google Forms and can be found in appendix 3. To assess respondent comprehension and ensure the survey's validity and quality, a pre-test was conducted with 4 senior researchers from Nova SBE DSKC, all of whom possess significant expertise in the fields of social economy and data. The survey is composed of 21 questions, encompassing various response formats, such as short answer, single-choice questions, checkbox selection, traditional five-point Likert scale questions (ranging from "Strongly disagree" to "Strongly agree") and a linear scale question.

The survey was structured into 7 sections, with the initial section focusing on characterizing the participating organizations. The second section introduced a glossary defining key data-related terms such as confidential data, sensitive data, and pseudonymization and was dedicated to the characterization of the data held by these organizations, also encompassing an assessment of the existence of a Data Protection Officer (DPO). Subsequently, the survey progressed through three sections exploring various facets of data practices. The first of these sections inquired about data collection (section 3), the second about data management (section 4), and the third about data sharing practices (section 5). In cases where organizations did not share data with third parties, they were guided to the subsequent section (section 6), which aimed to gauge their perspectives on the potential implementation of a comprehensive code of conduct regarding data collection, management, and sharing. Lastly, in the final section, organizations were asked about the existence of a code of conduct within their organization and had the option to share their email address for potential future contact related to the research. The chosen sampling technique was convenience sampling for its practicality and ease of implementation. The survey was sent to over 4,100 organizations listed in Nova SBE DSKC's social database and it was featured in the October edition of the Social Database Newsletter to increase its visibility and reach among the target audience. Ultimately, the sample yielded 110 responses, which were stored in a CSV file for further analysis.

Clustering was employed as the preferred data analysis methodology given that the segmentation of social economy organizations can help assess the data landscape within the sector and identify gaps in organizations' data practices. Thus, the provided code of conduct can be applied with a targeted focus to intervene with various identified groups of social impact organizations.

The input variables were converted from categorical to numerical, with some being scaled and normalized. Appendices 4 and 5 contain details about the data preparation process. To help determine the most effective approach, Agglomerative clustering, K-means and Spectral clustering were employed and their respective average silhouette scores and Dunn index values were computed. As evident from the plot presented in appendix 6, there is a prevailing trend of low scores in all the clustering methods. Notably, in scenarios with 2 or 3 clusters, Agglomerative clustering either equals or outperforms the other algorithms. However, K-means stands out with higher scores, particularly peaking at around 0.22 with the number of clusters set to 5. This suggests that, in this specific context, K-means achieved a better balance of cohesion within clusters and separation between them. Regarding the Dunn index values, Spectral clustering exhibits higher values overall, with K-means closely approaching in the 5 clusters scenario (appendix 7). Considering these metrics, the algorithm of choice was K-means and the selected number of clusters was 5, which aligns with the findings derived from the elbow rule method (appendix 8).

6. Research Findings

Within the surveyed entities, the predominant legal structures are associations (65%), followed by foundations (9%) and cooperatives (5%). The remaining are entities such as mercies, IPSS (not covered by the clauses in Article 4 in the *Lei de Bases da Economia Social*) and others. Regarding statutory classifications, most organizations fall under the designation of IPSS, with the second most prevalent being the status of Public Utility, acquired by 24% of respondents. The majority of respondents represent small to medium-sized organizations, with 59% having fewer than 50 staff members, and 30% with 51 to 250 staff members. 54% of all respondents deal with data on a daily basis and 29% engage with data on a monthly basis. Meanwhile, 11% reported processing data annually. They indicated a mix of digital and paper formats in handling their information, with only 12 adopting a fully

digital approach and 3 exclusively relying on paper. Additionally, a majority of organizations claimed they handle sensitive and confidential data. 54% lack a designated DPO, while 46% have one in place. Regarding data practices, 70% stated that they always obtain consent before collecting personal data, followed by 21% who mentioned that this happens frequently. A minority reported situations where consent is obtained sometimes (7%), rarely (1%), or never (1%). Roughly half of organizations claimed they share data with third parties, with 59% of them having established procedures to address data breaches, while 41% lack such protocols. Further details regarding the survey results are in appendices 11, 12 and 13.

Following the application of the K-means algorithm, scatterplots have been generated using T-SNE and UMAP (appendices 9 and 10), depicting all instances along with each cluster's centroid. The cluster's names (Pseudo-Compliant Practitioners, Data Skeptics, Nearly Compliant Sharers, Poor Data Practitioners and Nearly Compliant Handlers) were assigned based on the data analysis, which will be elaborated upon later. Henceforth, these labels will be referred to as the groups of instances constituting a specific cluster.

The plot in appendix 14 reveals that Nearly Compliant Handlers predominantly attain the highest individual silhouette scores, whereas Pseudo-Compliant Practitioners tend to exhibit the lowest scores, with only a few surpassing the average silhouette score. Notably, this former cluster encompasses 26 elements, establishing itself as the largest cluster, while the latter comprises 16 elements, positioning it as the smallest (appendix 15). In terms of the centroid coordinates (Table 1) for the Pseudo-Compliant Practitioners cluster, it's noteworthy to emphasize that, dimensions "12.informed_withdraw_consent" and "16.share_data" both have a coordinate value of 1. Furthermore, "20.own_CC_yes" stands at 0.88. In the Data Skeptics cluster, research findings show a low "16.share_data" and a high "20.own_CC_no" value, while the Nearly Compliant Sharers cluster exhibits high values for the

"17.procedures_breach_notify_immediately" dimension. Conversely, the Poor Data Practitioners cluster has the highest value among all clusters for the "17.procedures_breach_none" set at 1 and for having the lowest for the "9.dpo" coordinate set at 0.21. Lower values are observed across various dimensions derived from question 15, encompassing "15.3.regular_training", "15.4.contingency_plans" and "15.6.test_efficiency". Finally, the Nearly Compliant Handlers cluster centroid has the highest values for both "12.informed_withdraw_consent" (0.81) and "20.own_CC_yes" (0.96) dimensions in comparison to the other clusters.

Table 1: Centroid coordinates of K-means clusters

Clusters	Pseudo-Compliant Practitioners	Data Skeptics	Nearly Compliant Sharers	Poor Data Practitioners	Nearly Compliant Handlers
Dimensions					
9.dpo	0.69	0.4	0.46	**0.21**	0.54
12.informed_withdraw_consent	**1**	0.64	0.79	0.37	**0.81**
12.informed_rectification	0.5	0.4	0.79	0.21	0.77
12.informed_erasure	0.62	0.44	0.79	0.32	0.77
12.informed_none	0	0.32	0	0.42	0.04
15.1.priority_protection	0.8	0.76	0.81	0.7	0.82
15.2.clear_policies	0.72	0.66	0.73	0.42	0.81
15.3.regular_training	0.5	0.44	0.58	**0.25**	0.6
15.4.contingency_plans	0.36	0.36	0.56	**0.2**	0.58
15.5.pseudonymization	0.45	0.33	0.54	0.33	0.56
15.6.test_efficiency	0.31	0.35	0.54	**0.2**	0.5
16.share_data	**1**	**0**	**1**	**1**	**0**
17.procedures_breach_none	0.31	0	0	**1**	0
17.procedures_breach_notify_immediately	0	0	**1**	0	0
17.procedures_breach_notify_CNPD	0.25	0	0	0	0
17.procedures_breach_notify_in_time_limit	0.44	0	0	0	0
17.procedures_breach_not_applicable	0	1	0	0	1
20.own_CC_no	0.06	**0.92**	0.25	0.68	0
20.own_CC_dont_know	0.06	0.08	0.12	0.16	0.04
20.own_CC_yes	**0.88**	0	0.62	0.16	**0.96**

The decision tree (appendix 17), created to help in interpreting the clusters, comprises a total of 6 decision nodes. Progressing from right to left along the branches, the root is "17.procedures_breach_notify_immediately" followed by "16.share_data", "17.procedures_breach_none" and "12.informed_rectification". The left child node of “16.share_data” is “20.own_CC_yes", whose own child node is "20.own_CC_no". Furthermore, the right branch is associated with a "false" condition set by the decision node and the left with a “true” condition. To analyze clusters, both centroid coordinates and the decision tree were taken into consideration.

7. Discussion of Results

As the European Commission (2022) noted, “basic knowledge and awareness about responsibilities and legal obligations regarding data management (mainly GDPR) is indispensable”. The findings of the research substantiate this need (appendices 11, 12 and 13). While 50 of the surveyed organizations claimed that they convey information regarding the right to withdraw consent, rectification, and erasure to data subjects, as stipulated in Articles 7 (Conditions for Consent), 16 (Right to Rectification), and 17 (Right to Erasure) of the GDPR (and included in the topic 1.4. of the proposed code of conduct), 60 organizations do not provide information on all three rights. Within this group, 17 claimed to not notify data subjects about any of the aforementioned rights entirely. Regarding data breaches, 41% of organizations sharing data with third parties lack established procedures, failing to adhere to the requirements specified in Article 33, which calls for the notification of a personal data breach to the supervisory authority, and Article 40, which emphasizes communication of the affected data breach to the data subject (both addressed in topic 2.6. of the code). Furthermore, a minority also engages in data validation through pre-ticked boxes, a practice that, according to the GDPR, does not constitute consent. This is clarified in section 1.3. of

the proposed code. Furthermore, the majority of organizations acknowledge the advantages associated with implementing a code of conduct for data practices. They recognize that embracing such a code is useful for establishing good data practices, for fostering ethical standards and responsibility in data processing and that it signifies a long-term investment in ensuring the security and integrity of data. Nonetheless, 49% of the organizations claim that the existing codes and regulations are sufficient to guide them.

In the analysis of the five clusters generated by the K-means algorithm, Pseudo-Compliant Practitioners are noted for sharing data while lacking an immediate notification procedure for all affected parties in the event of a data breach involving shared data. Notably, a majority has some form of procedure in place, either involving notification to the CNPD or the affected parties within a specified time frame. This is evident from the decision tree (appendix 17), where 11 out of the 16 instances did not choose the option indicating no procedures in place. Moreover, they are characterized by providing information about the data subject's right to withdraw consent, while for other data subject rights, there is not a distinct grouping of organizations that consistently inform or refrain from informing about them, especially concerning rectification of data (refer to Table 1). Most organizations within this cluster seem to have their own code of conduct, as indicated by lower values for "20.own_CC_no" and "20.own_CC_dont_know" coordinates, contrasted with higher values for "20.own_CC_yes". On the other hand, Data Skeptics are characterized by not sharing data and lacking a code of conduct, as 23 out of the 25 do not have it. Meanwhile, Nearly Compliant Sharers stand out for having a procedure in place for all affected parties in case of a data breach involving shared data, a feature that sets them apart from all other instances. Poor Data Practitioners exhibit a diminished focus on data security, as indicated by most coordinates consistently registering values below 0.5. None of them have established procedures for addressing data breaches related to shared data, and a majority do not appear

to have a designated DPO. Lastly, Nearly Compliant Handlers do not share data and 25 out of 26 of them are equipped with a code. This cluster also includes more organizations that inform data subjects about their rights and have practices in place regarding data security.

With that in mind, Nearly Compliant Handlers are labeled as such because they form the cluster with the highest coordinate values across several data practices, particularly in terms of data security, with the majority having a code in place. However, there is still room for improvement in the implementation of these practices, which justifies the use of the term "nearly". This differs for Poor Data Practitioners, as they exhibit the lowest adherence to most data practices. Nearly Compliant Sharers have the closest coordinate values to Nearly Compliant Handlers, both in terms of informing about data subjects' rights and data security practices, but also engage in data sharing and have a procedure for data breaches. Data Skeptics, identified by their lack of data sharing and a code of conduct, appear less involved with data overall. Pseudo-Compliant Practitioners are the most likely to have a DPO and are characterized by having a code of conduct, a feature shared with Nearly Compliant Handlers. Despite this similarity, they do not actively participate in data practices to the same extent as Nearly Compliant Handlers or Nearly Compliant Sharers, particularly concerning the implementation of contingency plans, pseudonymization, and assessing the effectiveness of data security practices.

In summary, the data analysis reveals a general lack of compliance with legal regulations. While most organizations acknowledge the necessity of a code of conduct, they perceive their existing codes as sufficient. By segmenting them into clusters, some variations emerge in terms of their commitment to data sharing, procedures for handling data breaches, and the presence of a code of conduct. However, varying levels of implementation of data practices

were observed, which made it challenging to identify patterns. This leads to the conclusion that the social economy exhibits a highly diverse data landscape.

8. Application of the Code of Conduct

In conversations with the Nova SBE DSKC regarding their activities for social economy entities, it has come to light that one of their approaches involves conducting workshops that aim to educate participants on different data-related topics, including data storytelling and the utilization of tools such as Excel. With the organizations segmented into clusters, Nova SBE DSKC can engage with these identified groups through targeted approaches. For instance, they could organize workshops on data security practices tailored for Poor Data Practitioners or even on the implementation and importance of a code of conduct for Data Skeptics. The proposed code of conduct could then be utilized for a better understanding of data practices among the targeted groups.

It is also important to acknowledge the existing disparities in data practices implementation. Despite gaining insights into the data landscape, the collected data remains limited, while a more extensive dataset holds the potential to unveil patterns in organizations' data practices, fostering a clearer and more nuanced comprehension of organizations' needs. In this setting, creating a tool to offer guidance could also be a valuable proposition.

Organizations would be presented with targeted questions derived from the survey to pinpoint deficient or absent practices. These include aspects such as whether they engage in specific practices, the types of data they handle, and their compliance with legal requirements. As the code of conduct is organized into modular topics addressing different practices, the objective is to align identified gaps with the specific topics in the proposed code. Organizations would then be presented with these topics as guidance, allowing them to address their specific needs. These could be organized from mandatory to encouraged, with information provided about the penalties for non-compliance with the mandatory ones. Additionally, they could be

arranged from simpler to implement to more challenging. For instance, practices like creating data backups or restricting data access are simpler and more affordable data security measures to implement, while pseudonymization, which requires a higher level of technical expertise, falls into the category of more complex practices, but also more effective. Essentially, social economy entities seeking to enhance their data practices could answer these questions and receive guidance from the code of conduct in a clear and easily understandable manner to also overcome the challenges posed by the intricacies of legal language. Simultaneously, information can be gathered, contingent upon obtaining their consent, to enhance the clustering methodology. This refinement in segmenting organizations can prove advantageous, empowering Nova SBE DSKC to intervene more effectively within the identified clusters.

9. Conclusion

Several key factors have been identified to ensure that a code of conduct for data processing is a guiding force among social economy entities. These encompass a lack of basic awareness about data and data practices, insufficient knowledge on protocol development, limited technical expertise and a lack of financial resources. The sector displays significant diversity in terms of practices, with several gaps in law compliance, such as insufficient information provided by the organizations about data subjects' rights and a lack of procedures for handling data breaches.

In conclusion, to establish an effective and instructive code of conduct for this sector, it is essential to provide clear explanations of data and explicitly outline the current legal obligations applicable to the social economy entities, specifically in terms of data protection laws. Emphasizing the importance of adhering to these data practices reveals to be equally vital, as well as the segmentation of organizations into clusters, which intends a more targeted

application of the code. Given the diverse range of actors in the sector, each at varying levels of implementation of data practices, the code aims to serve as a tool to actively engage with organizations, seeking to address areas where they may be lacking and fostering the gradual development of a community of practice within the sector.

10. References

Banco Alimentar. 2023. “Política de Privacidade e de Proteção de Dados Pessoais”. Accessed December 14, 2023. https://www.bancoalimentar.pt/politica-de-privacidade-e-protecao-de-dados/

Brahmaiah, Madamanchi and Prof. Talluri Sreekrishna. 2021. “Survey on Growth of Business using Data Analytics for Business Intelligence in Real-Time world”. *Gorteria,* Vol. 33, 12 (9): 407-415. Research Gate.

Cong, Lin Willian, Wenshi Wei, Danxia Xie and Longtian Zhang. 2022. “Endogenous growth under multiple uses of data”. *Journal of Economic Dynamics and Control* Vol. 14. https://doi.org/10.1016/j.jedc.2022.104395

Connect Humanity. 2022. “Data Dashboard”. Connect Humanity. Accessed December 17, 2023. https://datadashboard.connecthumanity.fund/data

Council of the European Union. 2012. “Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)”.

Diário da República. 2013. *Lei de Bases Da Economia Social* nº 30/2013, de 2013-05-08. Nº 88: 2727-2728

Diário da República. 2020. *Resolução do Conselho de Ministros* n.º 30/2020. Série I de 2020-04-21. Nº 78: 6-32

ECTAA. 2023. “Code of Conduct on Data Sharing in Tourism”.

Enders, Tobias. 2018. “Exploring the Value of Data – a Research Agenda”. In Lecture Notes in *Business Information Processing*, 2-25

European Commission. 2017. “New European Interoperability Framework: Promoting seamless services and data flows for European public administrations”.

European Commission. 2018. “Guidelines on Personal data breach notification under Regulation 2016/679”.

European Commission. 2021. “Join the committee for drafting a Code of Conduct for data management and sharing in the social economy: Call for Expression of Interest”. Accessed December 14, 2023. https://single-market-economy.ec.europa.eu/calls-expression-interest/join-committee-drafting-code-conduct-data-management-and-sharing-social-economy-call-expression_en

European Commission. 2022. “Staff working document on data spaces”. European Commission. Accessed December 14, 2023. https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces

European Commission. 2022. “Transition Pathway for Proximity and Social Economy Ecosystem”.

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)”. *Official Journal of the European Union.* L 119: 1-88

Gegenhuber, Thomas, Johanna Mair, René Lührsen and Laura Thäter. 2023. “Orchestrating distributed data governance in open social innovation. Information and Organization”. *Information and Organization*, Vol. 33, 1. https://doi.org/10.1016/j.infoandorg.2023.100453

Hernandez-Hall, Ashley. 2021. “Nonprofit Data Management: A Stage Model”. A dissertation submitted in partial fulfillment of the requirements for the Doctor of Philosophy, University of Nevada. https://digitalscholarship.unlv.edu/cgi/viewcontent.cgi?article=5154&context=thesesdissertations

INE. 2023. “Social Economy Satellite Account 2019-2020”. *Destaque - Instituto Nacional de Estatística.*

International Red Cross and Red Crescent Movement Family Links Network. 2015. “Code of Conduct on Data Protection”.

Mayer, Duncan J. and Robert L. Fischer. 2023. “Exploring data use in nonprofit organizations”. *Evaluation and Program Planning*, Vol. 97. https://doi.org/10.1016/j.evalprogplan.2022.102197

Mergel, Ines. 2015. “Opening Government: Designing Open Innovation Processes to Collaborate With External Problem Solvers”. *Social Science Computer Review*, Vol. 33, 5: 599-612. https://doi.org/10.1177/0894439314560851

Ministério Público. 2019. *Lei da Proteção de Dados Pessoais* n.º 58/2019 de 2019-08-08. Nº151: 3-40.

Monteiro, Elisabete, Filipe Pinto, Leonor Rodrigues, Marisa Tavares, Rosário Pereira Faria, and Rosário Silva. 2015. “Diagnóstico Das ONG Em Portugal, Fundação Calouste Gulbenkian.”

OECD. 2017. *Key Issues for Digital Transformation in the G20.* Berlin: OECD

OECD. 2022. “OECD Recommendation on the Social and Solidarity Economy and Social Innovation”. OECD. Accessed December 14, 2023. https://www.oecd.org/cfe/leed/social-economy/social-economy-recommendation/

Patil, DJ and Hillary Mason. 2015. *Data Driven: Creating a Data Culture*. Sebastopol: O’Reilly

Rosário, Albérico and Joana Dias. 2022. “Sustainability and the Digital Transition: A Literature Review”. *Sustainability*, Vol. 14(7) . https://doi.org/10.3390/su14074072

Satopaa, Ville, Jeannie Albrecht, David Irwin, and Barath Raghavan. 2011. "Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior." In 2011 31st International Conference on Distributed Computing Systems Workshops: 166-171. https://doi.org/10.1109/ICDCSW.2011.20

UNICEF. 2020. “UNICEF Policy on Personal Data Protection”.

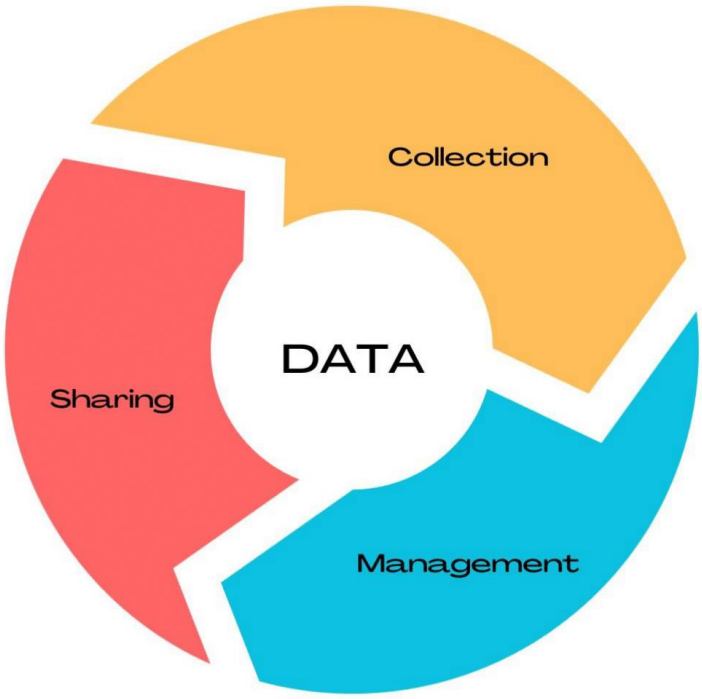
Yuan, Qianli and Mila Gasco-Hernandez. "Open innovation in the public sector: creating public value through civic hackathons". *Public Management Review* Vol. 23: 523-544. https://doi.org/10.1080/14719037.2019.1695884

Zejnilovic, Leid and Lénia Mestrinho, ed. 2020. *Data-driven Decision-making in the Portuguese Social Sector.* NOVA Data Science Knowledge Center.

11. Appendices

Appendix 1: Code of Conduct for Data Collection, Management and Sharing

Code of Conduct for Data Collection, Management and Sharing



Portuguese Social Economy

Table of Contents

Introduction 2

Glossary 3

1. Data Collection 5

1.1. Purpose limitation 5

1.2. Data Retention Period 5

1.3. Consent 5

1.4. Data Subject Rights 6

2. Data Management 8

2.1. Data Protection Officer (DPO) 8

2.2. Data Access 9

2.3. Data Backups 10

2.4. Pseudonymization 10

2.5. Regular Training and Awareness 10

2.6. Personal Data Breach 10

2.7. Data Accuracy 11

2.8. Compliance with Data Retention Period 11

3. Data Sharing 12

3.1. Third Parties 12

3.2. Purpose Limitation in Data Sharing 12

3.3. Data Protection in Data Sharing 12

3.4. Data Access, Transfers and Modifications 12

3.5. Data Interoperability 12

Introduction

The present Code of Conduct is focused on both personal and non-personal data and addressed to any individual or organization actively engaged in the portuguese social economy sector.

Its overarching goal is to:

- Advocate for transparency and responsibility in data governance, ensuring that it can be utilized and shared in compliance with the law.
- Establish clear lines of accountability for data-related activities and cultivate a culture of ongoing improvement by regularly monitoring and updating practices.
- Enhance data security measures to safeguard against unauthorized access, breaches, and potential threats.
- Support collaborative initiatives that aim to share data for the greater benefit of the portuguese social economy sector.

This code covers data collection, data management, including practices on protection and storage, and data sharing. Additionally, it features a glossary containing key concepts.

The following icons can be found throughout the code:

Indicates if a practice is an explicit requirement in the General Data Protection Regulation.

Indicates valuable insights, advice, or warnings for better comprehension and adherence.

Indicates practical examples, aiding in a better understanding concepts and practices.

Social economy entities under portuguese jurisdiction are invited to publicly commit to this Code of Conduct. Compliance with the code is voluntary.

Glossary

Data processing

Any operation or set of operations that are performed on data, whether or not by automated means, such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

GDPR – Article 4

Personal Data

Any information relating to an identified or identifiable person such as name, identification number, location data, an online identifier and factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity.

GDPR – Article 4

Confidentiality and Integrity Principle

The principle which requires that personal data be processed in a manner that ensures its security, including protection against unauthorized or unlawful processing, as well as against its loss, destruction, or accidental damage. Appropriate technical and organizational measures must be implemented to prevent unauthorized access and use of the data by unauthorized individuals.

GDPR – Article 5

Sensitive Data

A special category of personal data that is protected by EU legislation and can only be processed by organizations if specific guarantees exist. The subsequent personal data is considered ‘sensitive’ and is subject to specific processing conditions:

- personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs;
- trade-union membership;
- genetic data, biometric data processed solely to identify a human being;
- health-related data;
- data concerning an individual’s sexual life or sexual orientation.

GDPR - (10), Articles 4 and 9

Metadata

Information that we generate, store, and share to describe various things, essentially representing "data about data".

Pseudonymization

Processing of personal data in such a manner that it cannot be directly linked to a specific individual without additional information. Such information is stored separately and safeguarded through technical and organizational measures to prevent the personal data from being attributed to any identified or identifiable person.

GDPR – Article 4

Comissão Nacional de Proteção de Dados

(CNPD)

The CNPD is the national authority responsible for overseeing and enforcing compliance with GDPR and *Lei da Proteção de Dados Pessoais*, in strict respect for human rights and the freedoms and guarantees stated in the Constitution and the law.

Lei da Proteção de Dados Pessoais – Chapter IV, Article 22

1. Data Collection

1.1. Purpose limitation

Personal data shall only be processed for **clearly defined and limited purposes**, established before the data is collected. Any further processing beyond the initially specified purposes is permissible under the following circumstances:

- with the individual's consent;
- for public interest, statistical, historical or scientific ends, which are considered to be compatible with the initial purposes.

GDPR – (50), Article 5 and Article 7

⚠ **Avoid collecting excessive or irrelevant data** by clearly defining the specific purpose for collection! Time, personnel, and storage can be, therefore, utilized more effectively.

1.2. Data Retention Period

Time limits should be set by the organization for erasure or a periodic review to prevent retaining personal data longer than necessary.

GDPR - (39)

1.3. Consent

Consent is to be obtained through **transparent and affirmative action**, signifying the voluntary, well-informed, and unequivocal agreement of individuals concerning the processing of their personal data. This includes:

- ticking a box when visiting an internet website
- choosing technical settings for information society services or another statement or conduct that clearly indicates in this context individuals' acceptance of the proposed processing.
- written statements
- oral declarations

Silence, pre-ticked boxes or inactivity should not constitute consent.

GDPR - (32)

1.4. Data Subject Rights

Individuals must be informed about the following rights:

- The right to information: Individuals are entitled to receive specific information from the organization. This encompasses insights into the purpose of data collection, the legal basis for processing, interests pursued by the organization, recipients of the data, and retention periods, which enables them to make informed decisions about their data.

  GDPR - Articles 13 and 14
- The right of access: Individuals have the right to receive from the organization confirmation regarding whether their personal data is currently being processed or not.

  GDPR - Article 15
- The right to rectification: Individuals are entitled to prompt rectification of any inaccuracies in their personal data. Considering the processing purposes, the individual also has the right to have any incomplete personal data completed, which may involve providing additional information through a supplementary statement.

  GDPR - Article 16
- The right to withdraw consent: Individuals have the right to withdraw their consent at any time.

  GDPR - Article 7
- The right to erasure: Individuals have the right to obtain the erasure of personal data when: it is no longer necessary, consent has been withdrawn, the personal data have been unlawfully processed, the individual objects to the processing, no legitimate reason for continued processing exists and data erasure is necessary for compliance with a legal obligation.

  GDPR - Article 17
- The right to restriction of processing: Individuals are entitled to obtain restriction of processing when:
  a) data accuracy is in question and verification is pending;
  b) processing is deemed unlawful (yet the individual opts for restriction instead of erasure);

c) the data is no longer necessary for the organization but the individual wishes to preserve it for legal claims; and
d) the individual objects to processing, leading to temporary restriction while the controller verifies whether legitimate grounds outweigh the objections.

GDPR - Article 18

- The right to data portability: Individuals have the right to receive the personal data concerning them from the organization and to transmit this data to another entity when data processing is legally grounded, either through explicit consent or out of necessity for contractual performance.

GDPR - Article 20

- The right to object: Individuals are entitled to object to the processing of their personal data under specific circumstances, contingent upon the purpose and legal basis for processing. They have the right to cease the processing of their data for direct marketing and retain the option to object to the processing when it is used for scientific, historical research, or statistical purposes unless such processing is necessary for a task in the public interest.

GDPR - Article 21

- The right not to be subject to a decision based solely on automated processing: Individuals are entitled not to be subject to a decision based solely on automated processing, if such decisions have legal effects or significantly impact them. This right does not apply if the decision is necessary for a contract, authorized by applicable laws with measures to protect rights or based on explicit consent. In cases where exceptions apply, the organization must implement measures to protect the individual's rights, including the right to human intervention, expressing their viewpoint, and contesting the decision. Additionally, decisions should not rely on special categories of personal data unless specific conditions are met, and adequate measures are in place to safeguard the individual's rights and interests.

GDPR - Article 22

The individuals' rights are ensured upon a **written request addressed to the Data Protection Officer or any external personnel or consultants with responsibilities related to personal data protection** (Check *2.1.3. Alternatives*).

2. Data Management

2.1. Data Protection Officer (DPO)

2.1.1. Responsibilities

The DPO assists the organization in all matters related to the protection of personal data. The DPO should, specifically:

- Inform and advise the organization, as well as their employees, about their respective obligations under data protection law.
- Monitor the organization's compliance with all legislation related to data protection, including conducting audits, awareness activities, and training for personnel involved in data processing operations.
- Provide advice whenever a Data Protection Impact Assessment (DPIA) has been carried out and monitor its implementation.
- Act as the point of contact for individuals' requests regarding the processing of their personal data and the exercise of their rights.
- Collaborate with CNPD and serve as the point of contact for them on issues related to data processing.

CNPD, GDPR – Article 39

2.1.2. Mandatory vs Voluntary

The appointment of a DPO is mandatory in three specific situations:

a) Whenever the processing is carried out by a **public authority or body**;
b) Whenever the **core activities**[1] of the organization consist of processing operations that require regular and systematic monitoring of individuals on a large scale; or

[1] "Main activities" can be understood as the essential operations necessary to achieve the objectives of the organization. However, the interpretation of "main activities" should not exclude activities in which data processing is an inseparable part of the organization's activities.

c) Whenever the core activities of the organization consist of **large-scale processing**[2] **of special categories of data** or **data relating to criminal convictions and offenses.**

GDPR – Article 37

Even in the absence of a legal obligation, an organization can voluntarily opt to appoint a DPO.

European Parliament - Guidelines on Data Protection Officers ('DPOs')

2.1.3. Alternatives

An organization not legally obligated to appoint a DPO and choosing not to designate one voluntarily can seek external personnel or consultants with responsibilities related to personal data protection. In this case, it is crucial to ensure that there is no confusion regarding their role, status, position, and duties. It must be clear in all communications within the company and with CNPD, individuals, and the general public what is the role of this employee or consultant.

2.2. Data Access

The organization should establish clear guidelines outlining **which individuals** within the organization are **granted access** to the collected data and **delineate the circumstances** under which such access is permissible. It should also ensure that any individual who has access to data and acts under its instructions will not process it except in a manner compliant with the current law.

⚠ Ensure that the responsibilities of each individual and entity involved in processing personal data are clearly allocated and are reflected in appropriate contractual clauses.

GDPR – Article 25

[2] The GDPR does not define what constitutes large-scale processing, although Consideration 91 provides some guidance. According to the mentioned consideration, it would notably include "processing operations that involve a large amount of personal data at a regional, national, or supranational level, may affect a considerable number of data subjects, and are likely to entail a high risk.".

2.3. Data Backups

The organization is encouraged to regularly back up critical data, and ensure that the backup process is tested, reliable and safeguarded against potential breaches.

2.4. Pseudonymization

The organization is encouraged to use pseudonymization as a measure to enhance security and protect the rights of individuals.

GDPR – Articles 4, 32, 40

When sending Excel sheets containing sensitive data via e-mail, the sender and receiver may be authorized personnel within the organization. However, the IT support team could also have access to these emails, even if they are not explicitly considered authorized personnel. In such cases, it is recommended to employ pseudonymization as a precautionary measure.

2.5. Regular Training and Awareness

The organization is encouraged to educate employees about data protection principles. A well-informed workforce is less likely to engage in risky behaviors that could compromise the security of personal data.

2.6. Personal Data Breach

A regulation on personal data breaches should be instituted by the organization, defining proper reporting channels, conducting reviews or investigations of incidents, implementing technical responsive measures, and issuing notifications to individuals and relevant stakeholders.

The organization shall notify the CNPD without undue delay and, where feasible, no later than 72 hours after becoming aware of it, unless the breach is not likely to result in a risk to the rights and freedoms of individuals.

Mandatory information to report to CNPD:

- **Nature of the Breach**: Describe the incident, explaining how the breach occurred.

- **Categories and Number of Data Subjects Affected**: Specify the groups of individuals whose personal data has been compromised.
- **Categories and Approximate Number of Personal Data Records Involved:** Quantify the extent of the breach by indicating the number of records affected and the categories of personal data compromised.
- **Contact Information for DPO or Alternative Contact:** Communicate the name and contact of the DPO or another contact point where additional information can be obtained.
- **Likely Consequences of the Breach:** Outline the potential impact on individuals and any foreseeable risks to their rights and freedoms.
- **Measures Taken or Proposed:** Describe the actions already taken to address the breach and mitigate its impact. Additionally, outline any proposed measures to prevent similar incidents in the future.

Guidelines on Personal data breach notification under Regulation 2016/679, GDPR – Article 33, 34

2.7. Data Accuracy

Efforts to process personal data accurately and promptly should be ensured through **periodic reassessment of retained information**. The frequency of these reviews should be determined by considering factors such as the **relative time sensitivity** of the data. It is essential to substantiate and document the rationale for determining the reassessment frequency.

Personal data held in archives is exempt from reassessment, correction, or the obligation to be kept current.

GDPR – Article 5

2.8. Compliance with Data Retention Period

Personal data shall be kept and preserved only for the predefined timeframe established (*Check 1.2. Data Retention Period*), following which it will be subject to deletion.

GDPR - (39)

3. Data Sharing

3.1. Third Parties

The organization should ensure that other entities accessing the data are **reputable** and provide the highest level of guarantees in this regard.

3.2. Purpose Limitation in Data Sharing

The organization and respective third parties should mutually agree on the specific purpose for data usage and guarantee not to use data for **unlawful purposes or take advantage of it to speculate or for other such purposes.**

3.3. Data Protection in Data Sharing

The organization should ensure that **third parties commit to protecting data, preventing loss**, **theft**, **unauthorized access**, and **unauthorized alterations.**

3.4. Data Access, Transfers and Modifications

Data access, whether in read-only or fully editable modes, must undergo rigorous **auditing**. Data transfers or modifications must be **traceable**, encompassing the inclusion of metadata containing information about the author and the specific nature of the modification.

Suppose a partner organization updates financial data related to a shared initiative. Rigorous auditing, including metadata with details about the author and the specific nature of the modification, ensures accountability. This transparency not only safeguards data integrity but also fosters trust among collaborating entities, enhancing the effectiveness of social initiatives.

3.5. Data Interoperability

The organization and its third-party collaborators should actively communicate and collaborate to ensure that data exchange mechanisms are consistent and compatible across systems.

Appendix 2: Table with sources for the survey

Question Nr.	Question/Statement	Source
2.	What is the legal structure of your organization?	***Article 4*** *in* *Decreto Lei no 58/2013 de 8 de maio do Presidente da República*. Diário da República: I série, No 88 (2013)
9.	Does your organization have a DPO (Data Protection Officer)?	***Articles 9, 11, 12, 13*** *in* *Chapter III - Encarregado de proteção de dados. Lei da Proteçao de Dados Pessoais no 58/2019 de 8 de agosto.*
10.	Does your organization obtain consent before collecting personal data?	***Recital 32*** in General Data Protection Regulation
11.	How does your organization validate the collection of personal data?	***Recital 32*** in General Data Protection Regulation
12.	Data subjects are informed on:	***Articles 7, 13, 14, 15, 16, 17, 18, 20, 21, 22*** *in* Sections 2, 3 and 4 of Chapter III in the General Data Protection Regulation
14.	Who has access to data inside your organization?	***Article 25*** *in* Section 1 of Chapter IV in the General Data Protection Regulation
15.5.	Your organization pseudonymizes personal data.	***Articles 25, 32*** *in* Sections 1 and 2 of Chapter IV in the General Data Protection Regulation
17.	In the event of a data breach involving shared data, what procedures does your organization have in place to notify affected parties, both within and outside the organization?	***Articles 33, 34*** in Section 2 of Chapter IV in the General Data Protection Regulation

Appendix 3: Survey to the Portuguese Social Economy Entities

Questionário às Entidades da Economia Social Portuguesa

Em conjunto com o Data Science Knowledge Center (DSKC) da Nova School of Business and Economics (SBE), no âmbito do projeto Base de Dados Social, a estudante de Mestrado em Gestão, Mafalda Carvalhal, está a desenvolver uma tese que visa a criação de um Código de Conduta para a recolha, gestão e partilha de dados nas entidades da economia social portuguesa.

O nosso objetivo é compreender as práticas atuais de recolha, gestão e partilha de dados das organizações, bem como identificar desafios, oportunidades e necessidades específicas relacionadas com o tratamento de dados. Para isso, contamos com a sua valiosa contribuição para obtenção de resultados significativos!

Qualquer pessoa ligada a uma organização social pode responder ao inquérito. No entanto, dado o seu tema, gostaríamos que este fosse preferencialmente respondido por **alguém que esteja familiarizado com as práticas de tratamento de dados da organização**. Se não é o seu caso, pedimos, por favor, que reencaminhe para a pessoa que considere mais adequada.

O tempo estimado de resposta é de **10 minutos**. Todas as respostas serão tratadas de forma totalmente confidencial e as informações recolhidas serão usadas exclusivamente para fins estatísticos no âmbito deste estudo, sendo processadas de forma anónima e agregada. Pode interromper a sua participação em qualquer momento, se assim o entender. Quaisquer perguntas sobre este inquérito podem ser colocadas através do e-mail **54680@novasbe.pt**.

Agradecemos antecipadamente a sua colaboração e contribuição para o avanço do conhecimento nesta área crucial. O seu envolvimento é vital para o sucesso deste projeto. Obrigado!

Organização

Qual a função que desempenha na sua organização? *

A sua resposta

Qual o formato legal da sua organização? *

- ◯ Associação
- ◯ Cooperativa
- ◯ Fundação
- ◯ Misericórdia
- ◯ Associação Mutualista
- ◯ Outra:

A sua organização possui algum dos seguintes estatutos? *

- [ ] IPSS
- [ ] ONGA
- [ ] ONGD
- [ ] ONGPD
- [ ] Utilidade Pública
- [ ] Nenhum dos anteriores
- [ ] Outra: _____

Quantas pessoas integram a sua organização? *

Inclui trabalhadores e voluntários

- ◯ Menos de 50
- ◯ 50-250
- ◯ 251-500
- ◯ 501-750
- ◯ 751-1000
- ◯ Mais de 1000

Dados na Organização

Neste questionário, é fundamental a compreensão clara dos termos utilizados nas perguntas que se seguem. Assim, segue-se um pequeno glossário sobre os mesmos:

- **"dados pessoais"** incluem qualquer informação, de qualquer natureza e independentemente do suporte, incluindo som e imagem, que se relaciona com uma pessoa singular identificada ou identificável, também conhecida como o 'titular dos dados'. (Lei de Proteção de Dados Pessoais)
- **"tratamento de dados"** refere qualquer operação ou conjunto de operações sobre dados, efetuadas com ou sem meios automatizados, tais como a recolha, o registo, a organização, a conservação, a adaptação ou alteração, a recuperação, a consulta, a utilização, a comunicação por transmissão, por difusão ou por qualquer outra forma de colocação à disposição, com comparação ou interconexão, bem como o bloqueio, apagamento ou destruição. (RGPD - Regulamento Geral de Proteção de Dados)
- **"dados sensíveis"** são uma categoria especial de dados pessoais (RGPD). Estes estão protegidos pela legislação da UE e só podem ser tratados pelas organizações se existirem garantias específicas. Os seguintes dados pessoais são considerados "sensíveis" e estão sujeitos a condições específicas de processamento:
  - Dados pessoais que revelem origem racial ou étnica, opiniões políticas, crenças religiosas ou filosóficas;
  - Filiação a sindicatos;
  - Dados genéticos, dados biométricos processados exclusivamente para identificar um ser humano;
  - Dados relacionados com a saúde;
  - Dados relativos à vida sexual de uma pessoa ou orientação sexual. (RGPD)

- **"dados confidenciais"** são dados que devem ser tratados de forma a garantir a sua segurança, incluindo a proteção contra o tratamento não autorizado ou ilícito, bem como contra a sua perda, destruição ou danificação acidental. (a partir do RGPD)
- **"pseudonimização",** o tratamento de dados pessoais de forma que deixem de poder ser atribuídos a um titular de dados específico sem recorrer a informações suplementares, desde que essas informações suplementares sejam mantidas separadamente e sujeitas a medidas técnicas e organizativas para assegurar que os dados pessoais não possam ser atribuídos a uma pessoa singular identificada ou identificável. (RGPD)

Com estas definições em mente, avançaremos para as perguntas relacionadas com a caraterização dos dados recolhidos pela sua organização e com o processo de tratamento de dados.

Com que frequência a sua organização procede ao tratamento de dados? *

- ◯ Diariamente
- ◯ Mensalmente
- ◯ Anualmente
- ◯ Raramente

Os dados armazenados pela organização encontram-se num formato: *

	1	2	3	4	5	
Exclusivamente digital	◯	◯	◯	◯	◯	Exclusivamente físico

A sua organização: *

	Discordo completamente	Discordo	Não discordo nem concordo	Concordo	Concordo completamente
Vê necessidade de migrar para o formato digital.	◯	◯	◯	◯	◯
Planeia migrar para o formato digital.	◯	◯	◯	◯	◯
Lida com dados sensíveis.	◯	◯	◯	◯	◯
Lida com dados confidenciais.	◯	◯	◯	◯	◯

Em relação a quais entidades ou indivíduos a organização mantém informações? *

- [ ] Utentes
- [ ] Voluntários
- [ ] Funcionários
- [ ] Outra: _____

A sua organização tem um Encarregado de Proteção de Dados? *

- ◯ Sim
- ◯ Não

Práticas de Recolha de Dados

A organização obtém consentimento antes de recolher dados pessoais? *

- ◯ Sempre
- ◯ Frequentemente
- ◯ Às vezes
- ◯ Raramente
- ◯ Nunca

Como a sua organização valida a recolha de dados pessoais? *

- [ ] Caixa de opção explícita (Através do website marcar uma caixa específica com, por exemplo, "Aceito" ou "Concordo")
- [ ] Caixa de opção Implícita (No website uma caixa específica está pré-selecionada com, por exemplo, "Aceito" ou "Concordo")
- [ ] Formulários de Consentimento (geralmente assinado pelo utilizador como prova de aprovação)
- [ ] Consentimento Oral
- [ ] Outra: _____

Os titulares dos dados são informados sobre: *

- [ ] O direito de retirar o seu consentimento a qualquer momento.
- [ ] O direito de obter a retificação de dados pessoais incorretos.
- [ ] O direito de obter a eliminação de dados pessoais que lhes digam respeito.
- [ ] Nenhum dos anteriores

Os titulares dos dados dão os seus dados pessoais sem constrangimentos? *

- ◯ Sempre
- ◯ Frequentemente
- ◯ Às vezes
- ◯ Raramente
- ◯ Nunca

Práticas de Gestão de Dados

Quem tem acesso aos dados dentro da sua organização? *

- ☐ Os responsáveis designados pela proteção de dados
- ☐ A Direção
- ☐ Todos os trabalhadores
- ☐ Ninguém dentro da organização
- ☐ Outra: _____

A sua organização: *

	Discordo Completamente	Discordo	Não discordo nem concordo	Concordo	Concordo completamente
Prioriza a segurança e a proteção de dados em todas as operações.	◯	◯	◯	◯	◯
Tem políticas claras de gestão de dados confidenciais.	◯	◯	◯	◯	◯
Fornece formação regular sobre segurança de dados para os trabalhadores ou voluntários.	◯	◯	◯	◯	◯
Tem planos de contingência para incidentes de segurança de dados.	◯	◯	◯	◯	◯

Pseudonimiza os dados pessoais.	○	○	○	○	○
Testa e regularmente avalia a eficácia das medidas técnicas e organizacionais para garantir a segurança do processamento de dados.	○	○	○	○	○

Partilha de Dados

Partilha dados com entidades externas? *

p.e. parceiros, partes interessadas, entidades governamentais

- ○ Sim
- ○ Não

No caso da violação de dados envolvendo dados partilhados, que procedimentos a organização tem em vigor para notificar as partes afetadas, tanto dentro como fora da organização? *

- ○ Notificar imediatamente as partes afetadas
- ○ Notificar as partes afetadas dentro de um prazo especificado
- ○ Notificar a Comissão Nacional de Proteção de Dados
- ○ Não existem procedimentos em vigor.

A organização já enfrentou desafios relacionados com a partilha de dados, como conflitos sobre a propriedade de dados ou disputas com partes externas sobre a utilização de dados? *

- ○ Sim
- ○ Não

Implementação do Código de Conduta

No que respeita à criação e implementação de um **Código de Conduta para a recolha, gestão e partilha de dados nas entidades da economia social portuguesa:** *

	Discordo completamente	Discordo	Não discordo nem concordo	Concordo	Concordo completamente
A implementação do Código de Conduta requer mais esforços do que traz benefícios.	◯	◯	◯	◯	◯
A implementação do Código de Conduta conduz a burocracia desnecessária.	◯	◯	◯	◯	◯
O Código de Conduta deve ser adaptado às necessidades e valores específicos da organização para ser verdadeiramente eficaz.	◯	◯	◯	◯	◯
Os códigos/ regulamentos já existentes são suficientes para orientar a organização.	◯	◯	◯	◯	◯
O Código de Conduta é uma ferramenta valiosa para promover a ética e a responsabilidade no tratamento de dados.	◯	◯	◯	◯	◯

A implementação do Código de Conduta é um processo complexo e dispendioso.	◯	◯	◯	◯	◯
A implementação do Código de Conduta é um investimento a longo prazo na segurança e integridade dos dados.	◯	◯	◯	◯	◯
A criação de um Código de Conduta seria útil para a implementação de boas práticas de gestão de dados na minha organização.	◯	◯	◯	◯	◯

A sua organização possui um código de conduta? *

- ◯ Sim
- ◯ Não
- ◯ Não sabe

Deixe-nos o seu e-mail se gostaria de ser contactado/a para futuras questões sobre o tema.

A sua resposta

Agradecemos sinceramente a sua participação neste inquérito. O seu contributo é essencial para o sucesso do nosso projeto e para a criação de um Código de Conduta que possa beneficiar as organizações da economia social em Portugal. Acreditamos que os resultados deste estudo serão valiosos para melhorar as práticas de recolha, gestão e partilha de dados. Mais uma vez, muito obrigado pelo seu tempo e esforço.

Appendix 4: Variables' names and respective survey questions/statements

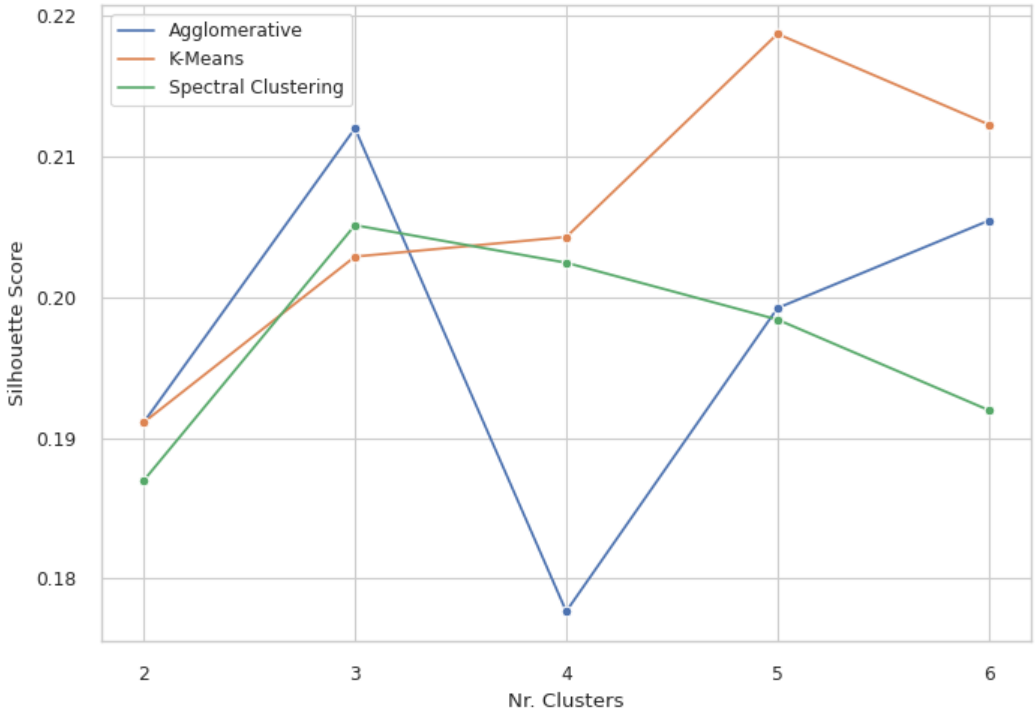
Variable Name	Question Statement	Type of Question
9.dpo	Does your organization have a DPO (Data Protection Officer)?	Binary
12.informed_withdraw_consent 12.informed_rectification 12.informed_erasure 12.informed_none	Data subjects are informed on:	Checkbox
15.1.prioritize_protection	Your organization prioritizes data security and protection in all operations.	Likert Scale
15.2.clear_policies	Your organization has clear policies for managing confidential data.	
15.3.regular_training	Your organization provides regular data security training for employees.	
15.4.contingency_plans	Your organization has contingency plans for data security incidents.	
15.5.pseudonymization	Your organization pseudonymizes personal data.	
15.6.test_efficiency	Your organization tests and regularly assesses the effectiveness of technical and organizational measures to ensure data security.	
16.share_data	Does your organization share data with third parties?	Binary
17.procedures_breach_none 17.procedures_breach_notify_immediately 17.procedures_breach_notify_CNPD 17.procedures_breach_notify_in_time_limit 17.procedures_breach_not_applicable	In the event of a data breach involving shared data, what procedures does your organization have in place to notify affected parties, both within and outside the organization?	Single Choice
20.own_CC_no 20.own_CC_dont_know 20.own_CC_yes	Does your organization have its own Code of Conduct?	Single Choice

Note: From all variables derived from the questions in the survey, 20 were selected for generating clusters (the ones listed in the table above). This selection process involved experimenting with various combinations of variables and identifying those that resulted in the highest average silhouette scores. Additionally, centroid coordinates were considered to help determine the most relevant variables in the clustering process.

Appendix 5: Scales from survey questions

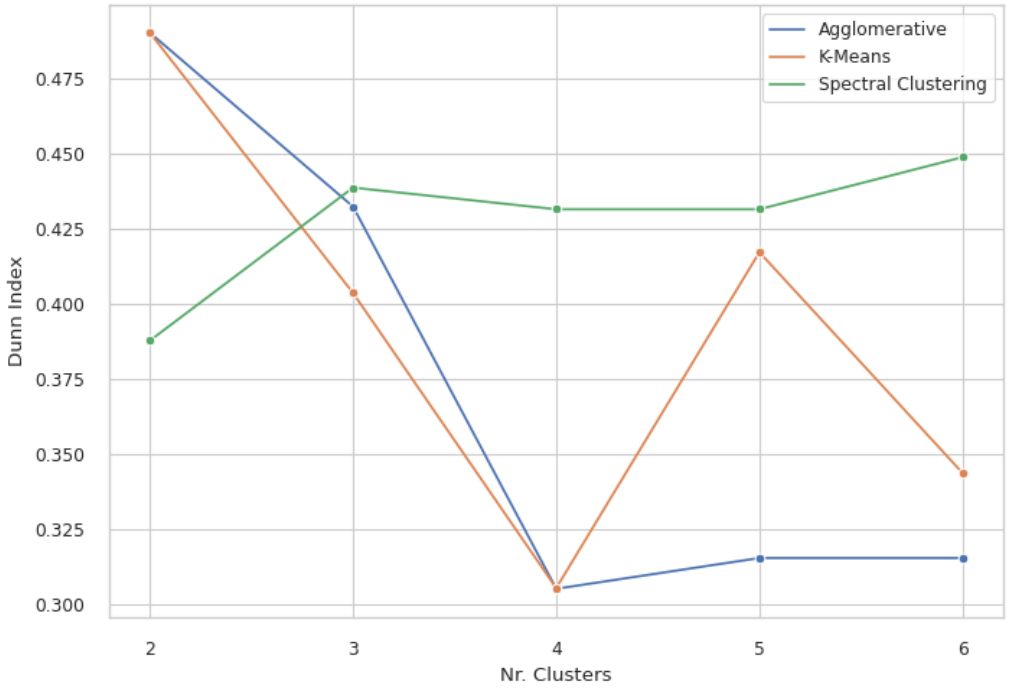
Normalized Scale	Likert Scale
0	Strongly disagree
0.25	Disagree
0.5	Neither disagree nor agree
0.75	Agree
1	Strongly agree

Appendix 6: Average silhouette scores of Agglomerative clustering, K-Means and Spectral clustering



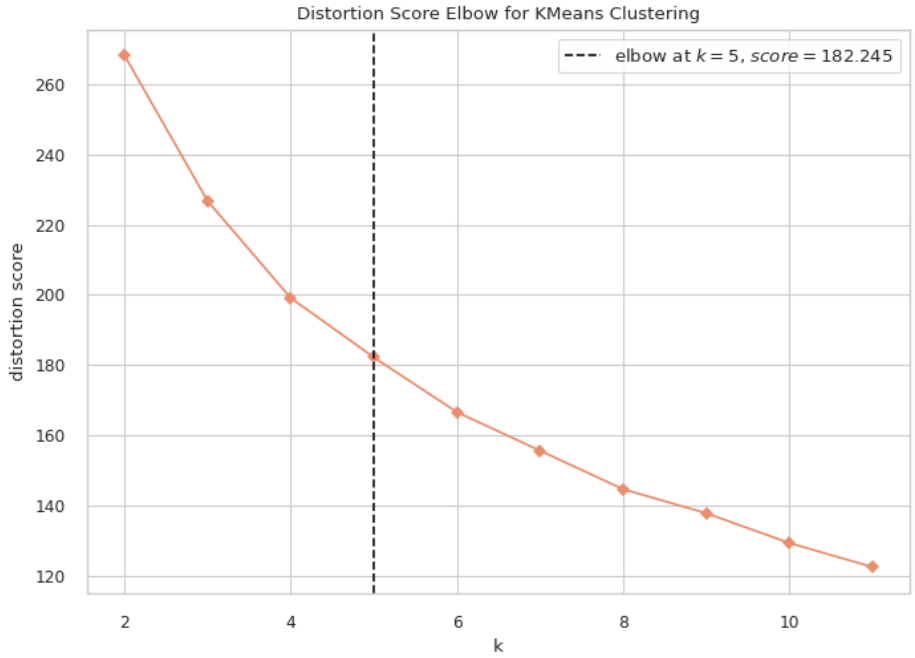
	Silhouette Score		
Model	Agglomerative	K-Means	Spectral Clustering
Nr. Clusters			
2	0.19	0.19	0.19
3	0.21	0.20	0.21
4	0.18	0.20	0.20
5	0.20	0.22	0.20
6	0.21	0.21	0.19

Appendix 7: Dunn index values of Agglomerative clustering, K-Means and Spectral clustering



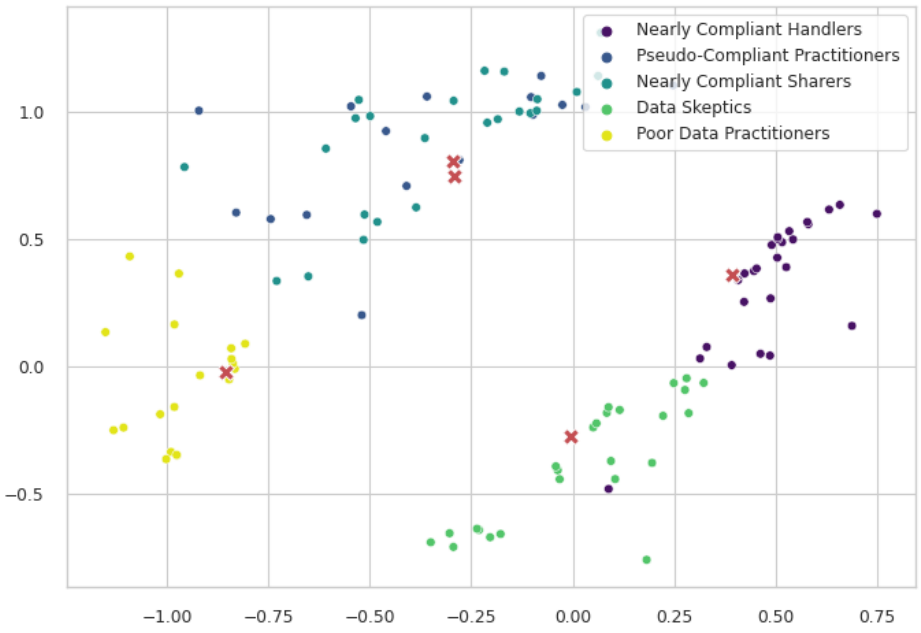
	Dunn Index		
Model	Agglomerative	K-Means	Spectral Clustering
Nr. Clusters			
2	0.49	0.49	0.39
3	0.43	0.40	0.44
4	0.30	0.31	0.43
5	0.32	0.42	0.43
6	0.32	0.34	0.45

Appendix 8: Distortion score elbow for K-means clustering

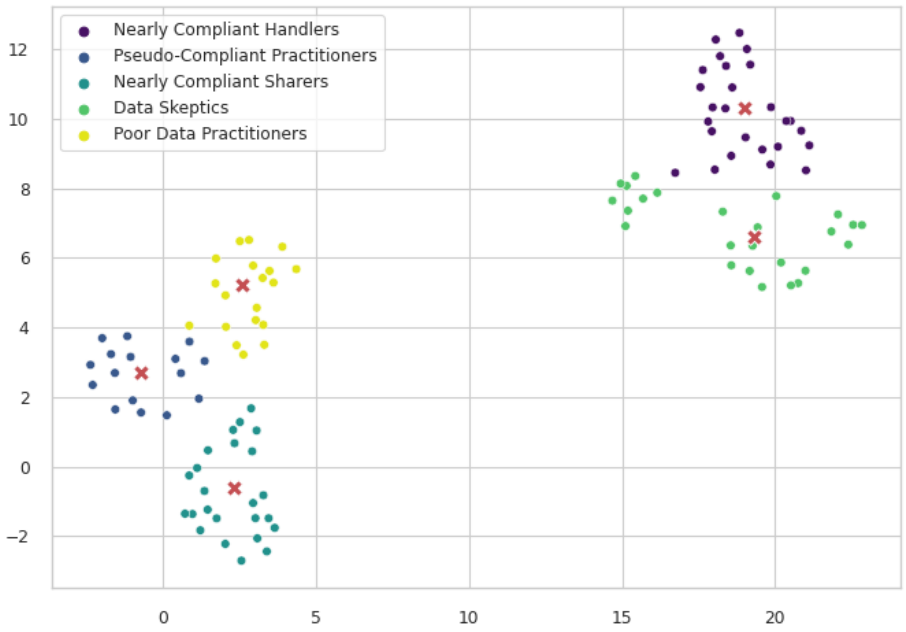


Note: The Kneedle Algorithm described in the article "Finding a "Kneedle" in a Haystack: Detecting Knee Points in System Behavior"(2011) was used to detect the "elbow".

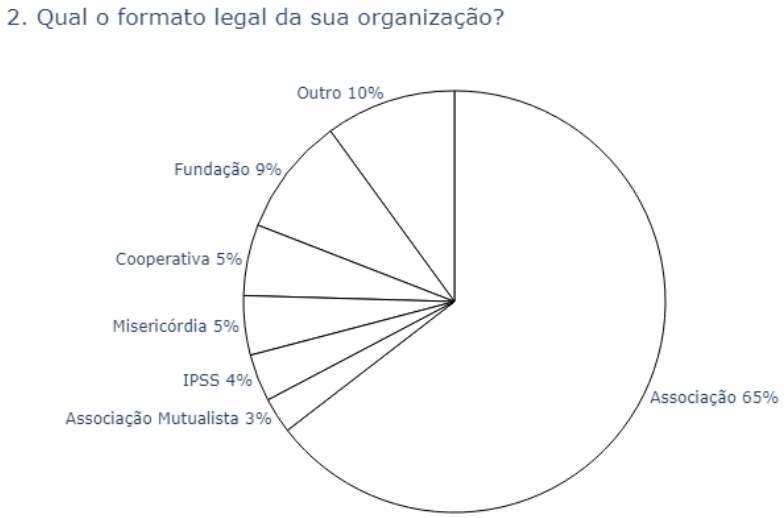
Appendix 9: Scatterplot of K-Means Clustering with 5 Clusters (including centroids) generated by T-SNE



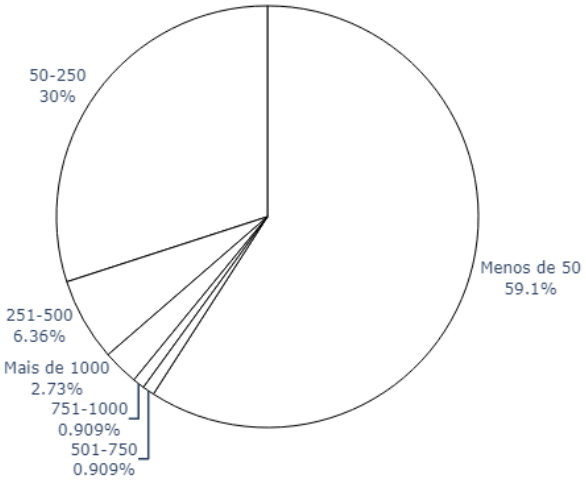
Appendix 10: Scatterplot of K-Means Clustering with 5 Clusters (including centroids) generated by UMAP



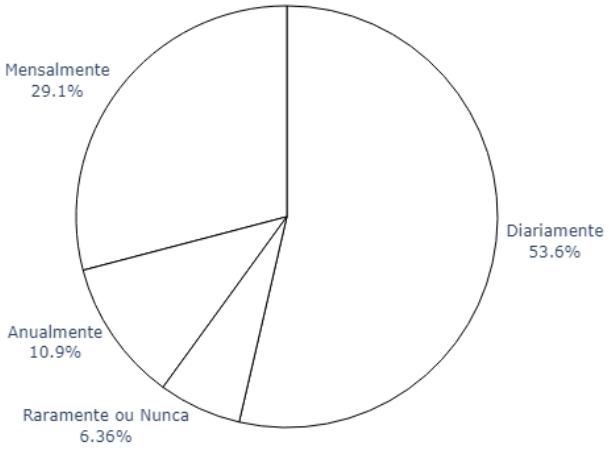
Appendix 11: Single-choice Questions (Descriptive Statistics)



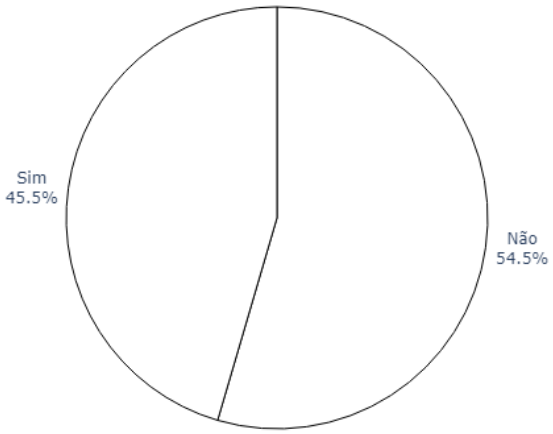
4. Quantas pessoas integram a sua organização?



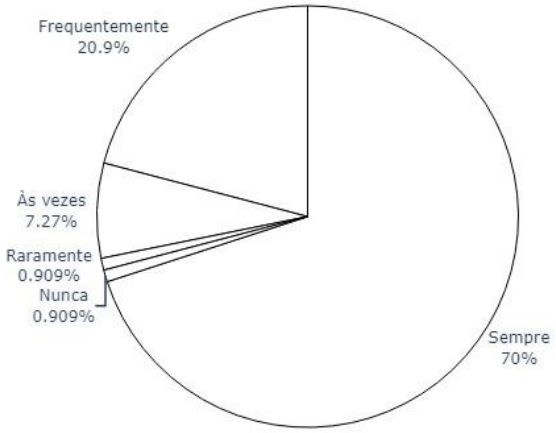
5. Com que frequência a sua organização procede ao tratamento de dados?



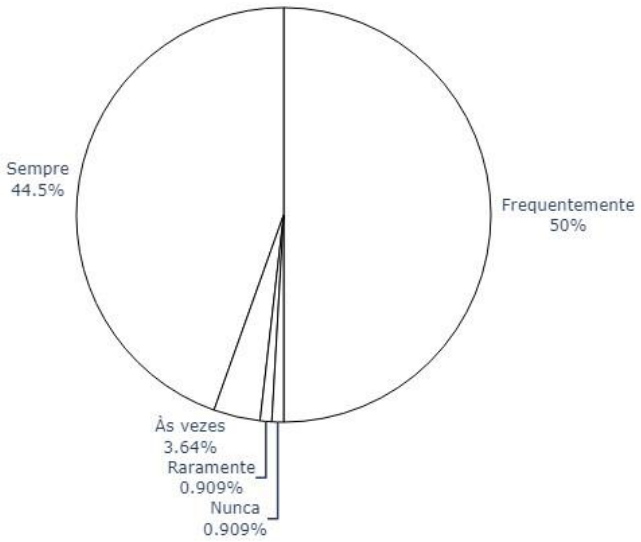
9. A sua organização tem um Encarregado de Proteção de Dados?



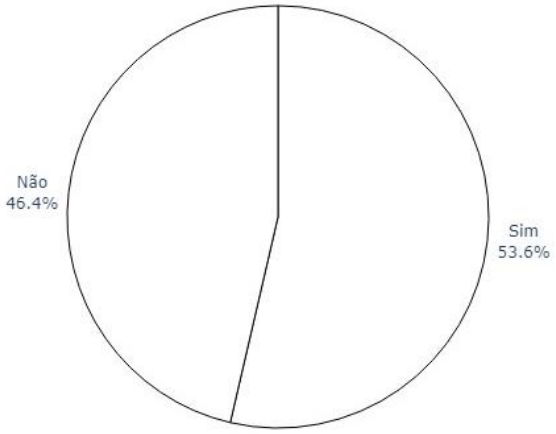
10. A organização obtém consentimento antes de recolher dados pessoais?



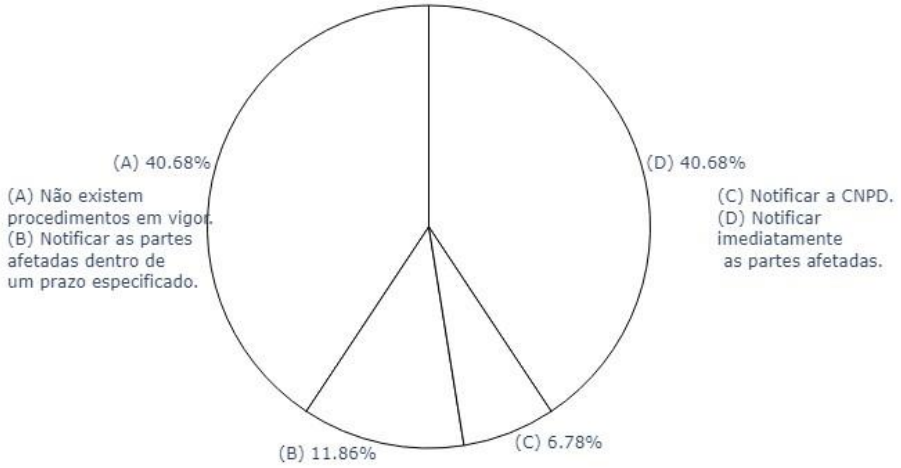
13. Os titulares dos dados dão os seus dados pessoais sem constrangimentos



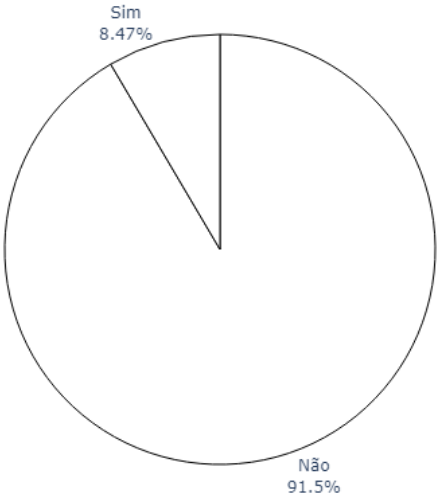
16. Partilha dados com entidades externas?



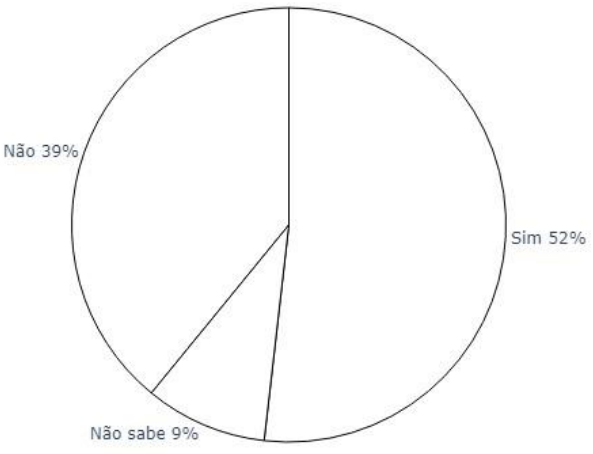
17. No caso da violação de dados envolvendo dados partilhados, que procedimentos a organização tem em vigor para notificar as partes afetadas, tanto dentro como fora da organização?



18. A organização já enfrentou desafios relacionados com a partilha de dados



20. A sua organização possui um código de conduta?

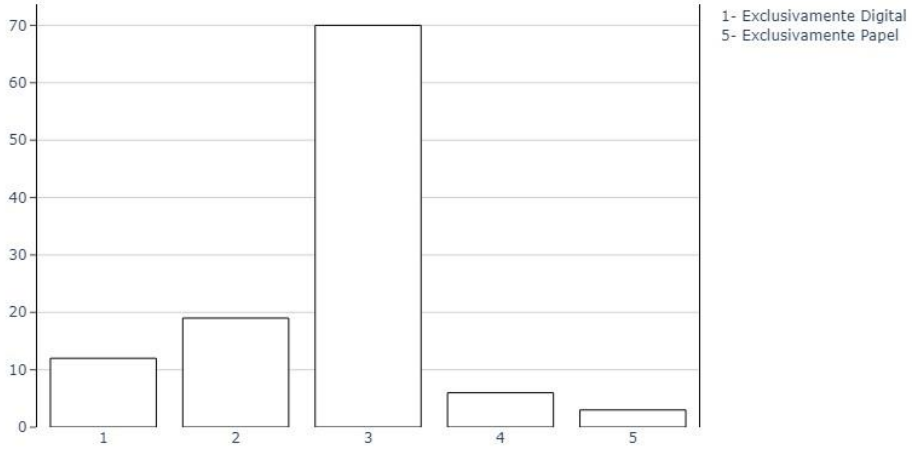


Appendix 12: Checkbox Selection Questions (Descriptive Statistics)

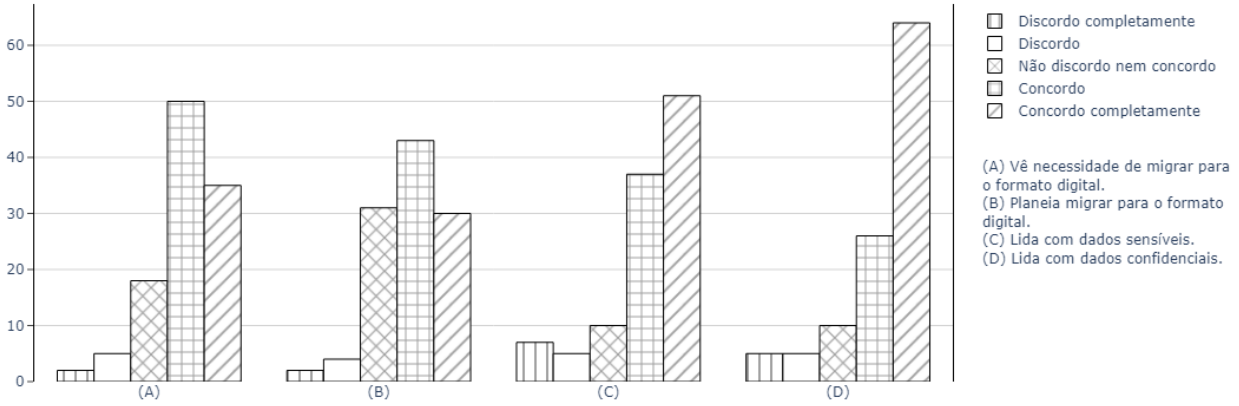
Questions	Option	Count	Relative Percentage (%)	Absolute Percentage (%)
3. A sua organização possui algum dos seguintes estatutos?	ONGA	3	2.26	2.73
	ONGPD	8	6.02	7.27
	ONGD	7	5.26	6.36
	Nenhum dos anteriores	17	12.78	15.45
	Outro	1	0.75	0.91
	IPSS	71	53.38	64.55
	Utilidade Pública	26	19.55	23.64
8. Em relação a quais entidades ou indíviduos a organização mantém informações?	Utentes	101	36.33	91.82
	Outros	16	5.76	14.55
	Voluntários	64	23.02	58.18
	Funcionários	97	34.89	88.18
11. Como a sua organização valida a recolha de dados pessoais?	Caixa de Opção Explícita	37	21.76	33.64
	Consentimento Oral	32	18.82	29.09
	Formulários de Consentimento	89	52.35	80.91
	Outro	3	1.76	2.73
	Caixa de Opção Implícita	9	5.29	8.18
12. Os titulares dos dados são informados sobre:	O direito de obter a eliminação de dados pessoais que lhes digam respeito.	66	29.6	60
	Nenhum dos anteriores	17	7.62	15.45
	O direito de retirar o seu consentimento a qualquer momento.	79	35.43	71.82
	O direito de obter a retificação de dados pessoais incorretos.	61	27.35	55.45
14. Quem tem acesso aos dados dentro da sua organização?	Todos os trabalhadores	11	6.79	10
	Outro	12	7.41	10.91
	A direção	68	41.98	61.82
	Os responsáveis designados pela proteção de dados	71	43.83	64.55

Appendix 13: Scale Questions (Descriptive Statistics)

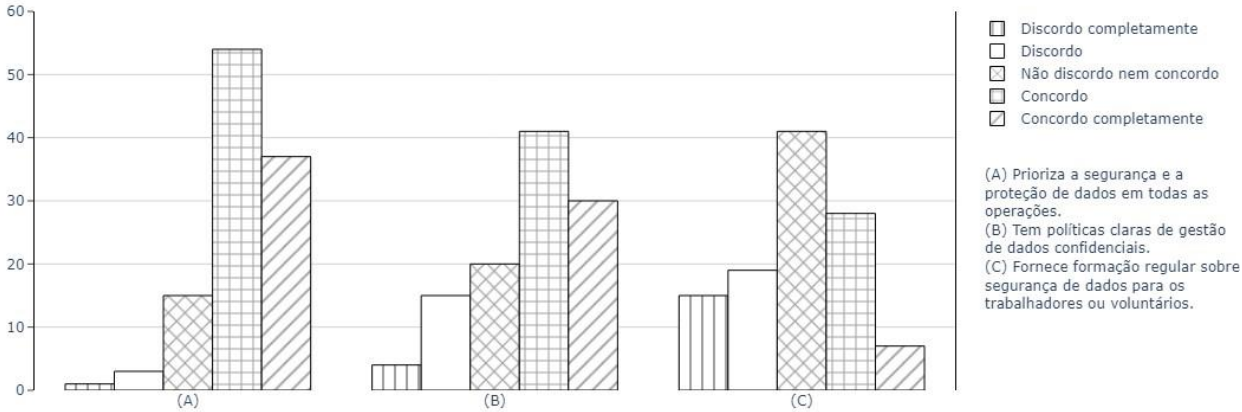
6. Os dados armazenados pela organização encontram-se num formato:

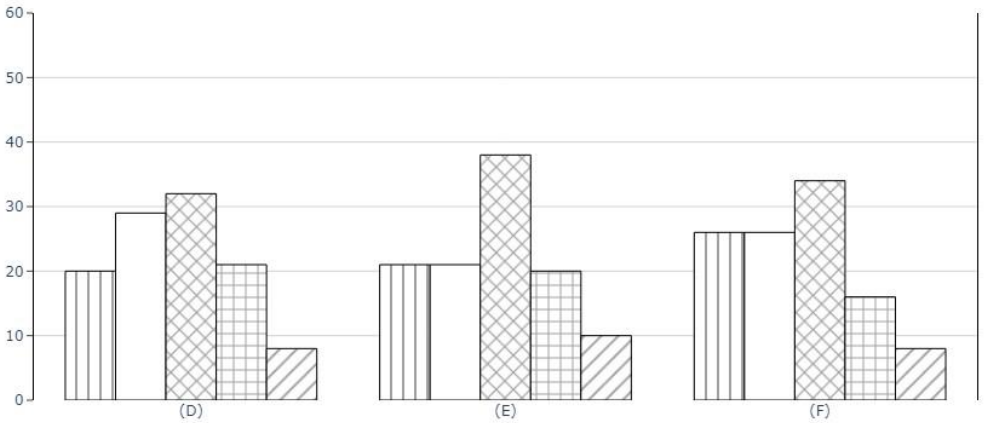


7. A sua organização:



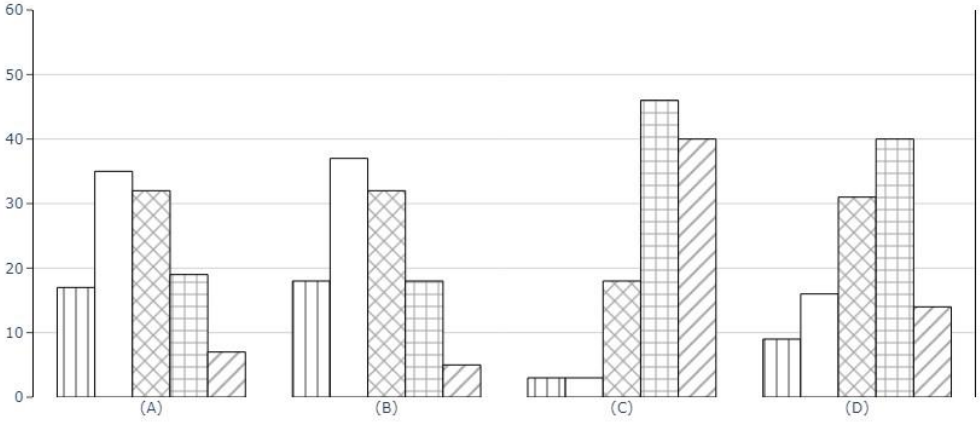
15. A sua organização:



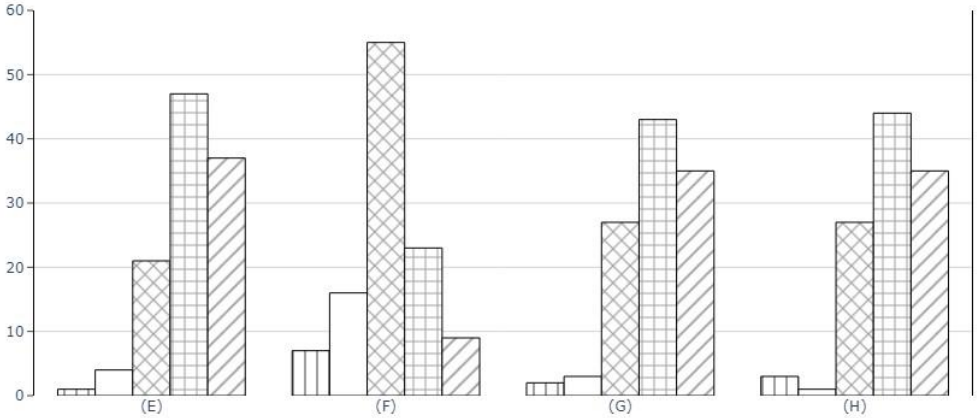


(D) Tem planos de contingência para incidentes de segurança de dados.
(E) Pseudonimiza os dados pessoais.
(F) Testa e regularmente avalia a eficácia das medidas técnicas e organizacionais para garantir a segurança do processamento de dados.

19. No que respeita à criação e implementação de um Código de Conduta para a recolha, gestão e partilha de dados nas entidades da economia social portuguesa:

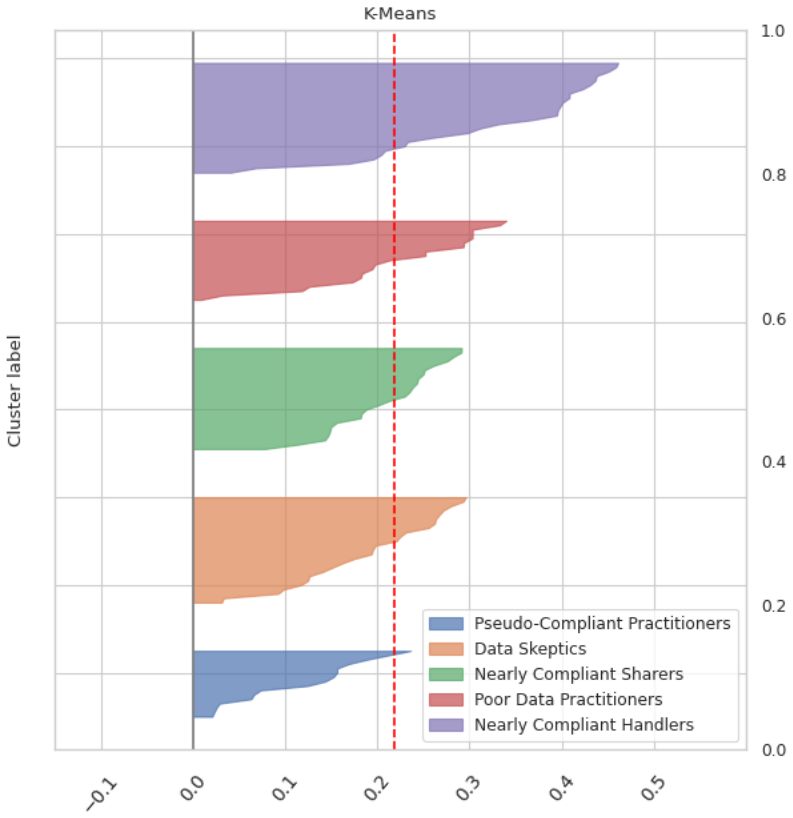


(A) A implementação do Código de Conduta requer mais esforços do que traz benefícios.
(B) A implementação do Código de Conduta conduz a burocracia desnecessária.
(C) O Código de Conduta deve ser adaptado às necessidades e valores específicos da organização para ser verdadeiramente eficaz.
(D) Os códigos/regulamentos já existentes são suficientes para orientar a organização.



(E) O Código de Conduta é uma ferramenta valiosa para promover a ética e a responsabilidade no tratamento de dados.
(F) A implementação do Código de Conduta é um processo complexo e dispendioso.
(G) A implementação do Código de Conduta é um investimento a longo prazo na segurança e integridade dos dados.
(H) A criação de um Código de Conduta seria útil para a implementação de boas práticas de gestão de dados na minha organização.

Appendix 14: Individual Silhouette Scores



Appendix 15: Elements per Cluster

Clusters	Pseudo-Compliant Practitioners	Data Skeptics	Nearly Compliant Sharers	Poor Data Practitioners	Nearly Compliant Handlers
Element Count	16	25	24	19	26

Appendix 16: Centroids coordinates for each cluster

Clusters	Pseudo-Compliant Practitioners	Data Skeptics	Nearly Compliant Sharers	Poor Data Practitioners	Nearly Compliant Handlers
Dimensions					
9.dpo	0.69	0.4	0.46	**0.21**	0.54
12.informed_withdraw_consent	**1**	0.64	0.79	0.37	**0.81**
12.informed_rectification	0.5	0.4	0.79	0.21	0.77
12.informed_erasure	0.62	0.44	0.79	0.32	0.77
12.informed_none	0	0.32	0	0.42	0.04
15.1.priority_protection	0.8	0.76	0.81	0.7	0.82
15.2.clear_policies	0.72	0.66	0.73	0.42	0.81
15.3.regular_training	0.5	0.44	0.58	**0.25**	0.6
15.4.contingency_plans	0.36	0.36	0.56	**0.2**	0.58
15.5.pseudonymization	0.45	0.33	0.54	0.33	0.56
15.6.test_efficiency	0.31	0.35	0.54	**0.2**	0.5
16.share_data	**1**	**0**	**1**	**1**	**0**
17.procedures_breach_none	0.31	0	0	**1**	0
17.procedures_breach_notify_immediately	0	0	**1**	0	0
17.procedures_breach_notify_CNPD	0.25	0	0	0	0
17.procedures_breach_notify_in_time_limit	0.44	0	0	0	0
17.procedures_breach_not_applicable	0	1	0	0	1
20.own_CC_no	0.06	**0.92**	0.25	0.68	0
20.own_CC_dont_know	0.06	0.08	0.12	0.16	0.04
20.own_CC_yes	**0.88**	0	0.62	0.16	**0.96**

Appendix 17: Decision Tree