NOVA IMS Information Management School

MAA

Mestrado em Métodos Analíticos Avançados
Master Program in Advanced Analytics

# A milestone in the Health governance of France

## The construction of a Health Information System

Cloé Dorali

Internship report presented as partial requirement for obtaining the Master's degree in Data science and advanced analytics, with a major in Data Science.

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

**NOVA Information Management School**
**Instituto Superior de Estatística e Gestão de Informação**
Universidade Nova de Lisboa

# A MILESTONE IN THE HEALTH GOVERNANCE OF FRANCE

by

Cloé Dorali

Internship report presented as partial requirement for obtaining the Master's degree in Data science and advanced analytics, with a major in Data Science.

**Advisor:** Vitor Duarte dos Santos

August 2019

# ACKNOWLEDGEMENTS

To **Vitor dos Santos**, **doctorated professor at the University of Lisbon – NOVA**, first of all for the transmitted knowledges, indeed very useful during this report redaction ; secondly for his help and guidance as a **tutor**.

To **Maryline Mouries, Project Manager & Consultant Manager at Keyrus**, for seeing in me the potential to integrate Keyrus and offering me this opportunity ; but also for her welcoming, her help and her time during this internship in her quality of **tutor**.

To **Mike Jardim, Project Manager and Tech Leader** on DIAMANT who was also a **great teacher** to me. Thank you for all the time you passed explaining the functionning of a project ; for all the time you passed correcting my development ; for all the time you saved me by finding the solution to my problem. I will be sure to walk in your footprints.

To **Delphine Charbonnel, Project Manager & Consultant Manager at Keyrus** (also Manager on GCS), for all the technical and managerial lessons you don't even know you gave me and for offering me more opportunities than anyone else. Thanks to your confidence, you led me to prove myself and my abilities. Some, numerous, that I owe you. Thank you for being an **amazing mentor.**

To **the DIAMANT's et GCS's teams**, and particularly to,

- **Adou Yapo-Ako, Developer** on DIAMANT;
- **Ludovic Romano, Tech Leader** on GCS ;
- **Imran Zamoum, Developer** on GCS and
- **Tom Simard, Intern** on GCS.

Each one of you and all the things you gave or showed me enriched my knowledges and capacities may them be professional, technical or personal skills. You all made me feel as a real player of the team, implicated and empowered.

To **Flavie Rochereuil, Project Engineer and Co-manager of Amadeus' project at Keyrus**, for her monitoring and her advices during this internship in her role of "**coach**".

Finally, I would like to say thank you to **all the collaborators of Keyrus** Lyon HQ for their integration and contribution – no matter the form taken by this contribution. Some of you became advisors or friends, some transmitted knowledges, and some transmitted joy. You all made this experience valuable to me.

# ABSTRACT

Although France is recognized as one of the countries with the best care support, it is also a country far behind on the integration of health data and the constitution of an HIS. Yet, in some aspects, France is not an entirely autonomous country in its government. Indeed, since the integration in the European Union, certain subjects - of which health - are subjects of common agreement, for a common application that can - at this scale - be qualified as a quasi-continental application. And in its goal of global HIS, the European Union is pressuring France to build its own HIS, which will then be absorbed by the HIS of the 27 countries. It is in this scheme that France's gave full authority since ten years to the Regional Health Agencies (and through them, to Keyrus, one of the leaders in business intelligence in France) to build this information system. This is not easy because the French administration is complex and has been solidly and strictly structured for several decades. Building this decisional model is long and will take many more years. But with projects as DIAMANT and GCS, the country is in the process of building a complete HIS taking into account the innovations of the practice of medicine today.



# KEYWORDS

Healthcare Organization, public environment, service provider, Business Intelligence, data integration, data transformation, data quality, data analysis, knowledge extraction and presentation, Decision Support System, Health Information System, project management.

# INDEX

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF APPENDIXES

# LIST OF ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| **AD** | Active Directory |
| **CDSS** | Clinical decision support system |
| **CNSA** | National Solidarity Fund for Autonomy |
| **DIAMANT** | Decisional Inter-ARS (=RHA in English) for the Mastery and the Anticipation |
| **DSS** | Decision support system |
| **DSV** | Data Source View |
| **DTA** | Delivery and Technology Accelerator |
| **EMR** | Electronic Medical Records |
| **EHR** | Electronic Health Record (another name of EMR) |
| **EHPAD** | Accommodation facility for dependent elderly people |
| **ESATs** | Workforce Assistance Facilities and Services |
| **ETL** | Extract, Transform, and Load |
| **EU** | European Union |
| **FTP** | File Transfer Protocol |
| **GCS** | Health Cooperation Group |
| **BFC** | Bourgogne-Franche-Comté |
| **GDP** | Gross Domestic Product |
| **GHT** | Territorial Hospital Trust |
| **GNUpg** | GNU 'Not Unix' Privacy Guard |
| **GUI** | Graphical user interface |
| **HAPI** | Harmonization and Information Sharing |
| **HAS** | National Health Authority |
| **HIS** | Health Information System |
| **IDS** | Intrusion Detection System |
| **KDC** | Keyrus Delivery Center |
| **MCO** | Medical, Chirurgical, Obstetrical |
| **MoH** | Ministry of Health |
| **PMSI** | Information System Medicalization Program |
| **RAD** | Rapid Application Development |
| **RHA** | Regional Health Agencies |
| **SDLC** | Systems-Development Life Cycle |
| **sFTP** | Secured File Transfer Protocol |
| **SHI** | Statutory health insurance |
| **SSO** | Single sign-on |
| **SSR** | Follow-up care and rehabilitation |
| **TMA** | Third party Application Maintenance |
| **VHI** | Voluntary Health Insurance |
| **VPN** | Virtual Private Network |

# 1. INTRODUCTION

Keyrus is a forerunner in the Business Intelligence and Digital fields. To the clients, the Keyrus' business model is attractive: it is low effort and investment from their side, for both the translation of the needs, the development and even the maintenance of the application; and this, at a competitive price as Keyrus - thanks to the many projects it won – can be really competitive. Which is important as, in France only, there is more than 150 SSII/ESN. Among them, Micropole (53$^{th}$) and Umanis (79$^{th}$) are huge rivals of Keyrus (56$^{th}$) (Dum, 2017). The competition in this market is indeed tough. Even more when adding all the BI companies of the world (especially the ones from India) which are taking a great part of the business mainly because of the low prices they can offer. This is the reason why Keyrus develops itself in all the areas of the BI, Data and Digital and on its presence around the world. By doing so the company seeks to be recognised in the industry and to concurrence the low-cost foreign companies and the top one in France. To do so, Keyrus acquires many companies among which Cyborg, Kadris and BIPB (Keyrus, 2018) and others in 2019.

As a service provider, Keyrus works on projects that can be inside ones (for Keyrus itself) or more often for other companies. The subject of this report is two projects called DIAMANT and GCS BFC. Both projects are proposed by the Regional Health Agencies of France (GCS BFC being a specific demand of the BFC RHA), under the aegis of the Health Ministry of France. The projects were developed till the beginning by Keyrus, respectively on 2009 for DIAMANT (on December 2018, Keyrus won for the 4$^{th}$ time the 4 years' DIAMANT's contract (until December 2022)) (Keyrus, 2016) and 2019 for GCS. Seeing and participating to all the process' phases is essential to get hands-on experience in the diverse areas of a project: from management, to development. The managerial aspect of this internship was a plus, but the development one goes in the way of the application of the knowledge dispensed in the master's degree in Data Science and Advanced analytics. The present report is a proof of the completed tasks and refers to the internship done at the Delivery Center of Keyrus between September 3, 2018 and February 28, 2019. This report also concerns the tasks completed after this internship and till the delivery of this report by the student in his quality of project engineer at Keyrus.

## 1.1. ACADEMIC CONTEXT

At Nova Information Management School, the master specialized in Data Science and Advanced Analytics aims at forming data specialists able to clean, transform and integrate data as a data architect, to look to the past and what's happened as a data analyst but also to look at the future to uncover patterns and make predictions as a data scientist. To do so, the program uses a practical approach on each of the following subject:

| 1$^{st}$ semester | 2$^{nd}$ semester | 4$^{th}$ semester |
|---|---|---|
| Computational Intelligence for Optimization | Predictive models | Text mining |
| Data Mining | Inferential analytics | Deep learning |
| Big Data | Business Intelligence | |
| Data Warehousing | Descriptive analytics | |

The projects combine Data Mining, Data Warehousing, Predictive models, Business Intelligence and Descriptive analytics axes. In this sense, the projects seen during this one year period are great topics of application of the classes followed last year and so to complete the master program.

## 1.2. ORGANIZATIONAL CONTEXT

Eric Cohen created Keyrus in 1996. The company is now, two decades later, present in 15 countries, counts 3000 employees all over the world and generated 273.2 million of euro on 2018,

representing a global annual growth rate of 12.9% (Keyrus, 2018). The growth rate of the company is in constant expansion with an average of 5% of growth by year since 2005. The company is directed by an entrusted board of representatives. Nevertheless, Eric Cohen has most of the shared parts of the company. Despite the stock exchange listing of Keyrus in 2000, the board of direction and the international presence, the culture of Keyrus is to try to keep a low hierarchical structure with essentially, technical skilled managers and directors of business units.

Keyrus counts 3 domains of activity and expertise: the "Management and Transformation" activity, the "Digital Experience" activity and the "Data Intelligence" activity. This last one is the core domain of the Delivery Center where the internship took place (Keyrus, 2018). This center, created in 2012, aims at designing hybrid BI models based on data exploration and visualization, the rationalizing and strengthening of the enterprise's architectures, the construction of transformation plans (social networks, Open Data and connected objects), the optimization of the performance of the company with the development of advanced and tailor-made algorithms and the implementation of highly upgradable predictive and prescriptive analytics solutions. Today, the DTA Center (previously known as KDC) represents more than one hundred collaborators in France, divided on 3 different centers (Lyon, Paris and Rennes). The center generates 8 million € in 2018 of which approximately 70% were generated by the Lyon's quarters.

## 1.3. GOALS OF THE INTERNSHIP

*The goals of the company regarding the intern*

By integrating the team working on these projects, the intern's role is to integrate each phase of the projects and to participate on each of them: to understand, formulate and answer to the client's needs and questions and to develop the applications needed. As an intern, his role is to integrate teams of highly efficient developers on MSBI and so on to have adequate competences on this technology, to respect and put in place "good practices" methodologies. Moreover, the projects need an "analytics profile" resource to answer to the newly added needs on Data Visualization and Predictive models.

*The intern's goals and expectation regarding the internship at Keyrus*

When given the choice to do a research or an internship, the internship always was the favourite option. At a young age, with no valuable experience in the domain nor any certitude of going further in the educational field, it seemed to be the best choice to make. Having decided in favour of the internship, there was a phase of research of a company expert in BI and Big Data to: firstly, put in application the obtained knowledge; secondly, to gain a valuable experience in the field concerned by the master degree; thirdly, to have an idea of the continuation (study or work) to settle on.
As for the chosen company, Keyrus has a position of interest on the market. It enables the interns to have an experience on a company renowned on the market and to make a reference regarding the position they can have on companies of this sector. By choosing Keyrus, one can be certain to be around experienced and highly trained developers and data analysts/scientists and so on to go further on the mastery of the technologies seen during classes.

# 2. THEORICAL FRAMEWORK

## 2.1. HEALTH INFORMATION SYSTEMS

Haux (Haux R. , 2002) gave in his analysis "Health care in the information society: what should be the role of medical informatics?" a definition of HIS. "As common information systems, a Health Information System process data and provides information and knowledge in a healthcare environment". This definition came belatedly compared to the emergence of (a) healthcare services (in France, 1945) (b) information systems (1961). Haux said that Healthcare went through a paradigm shift, moving from 'Industrial Age Medicine to Information Age Healthcare' (an expression of Smith R, 1997). The industrial health system compared to the information age system as Smith calls it, wasn't a centralized system nor in terms of hospitals, nor in terms of information. As Haux and Reinhold in their "Health Information Systems - Past, Present, Future" (Haux & Reinhold, 2006) explained, the shift from paper-based to computer-based processing and storage, as well as the increase of data in health care obliged or leaded to the standardization of HIS. Other changes are to be considered regarding the emergence of HIS as the shift from institution centered departmental and, later, hospital information systems towards regional and global HIS. In addition, as Kuhn and Giuse highlighted it (Kuhn & Giuse, 2001), new concepts became mandatory when speaking of HIS. An HIS has to be able to record and locate important information quickly. Based on this, a trend towards clinical computing and patient centered computer-based record appeared worldwide in the 1980s. And the hospital information system of earlier - mainly administrative - had become more focused on the clinical perspective and the patient record. Today, HIS seems to be essential for monitoring and evaluation, but also to provide alerts, to support patients and health facility management, to enable planning, to support and stimulate research and finally to permit health situation and trends analysis (Almunawar & Anshari, 2012).

To complete its duty, a HIS needs (a) individual level data, (b) health facility level data, (c) population level data and (d) public health surveillance level data (Almunawar & Anshari, 2012). To be clearer, it needs information regarding:

- Health determinants meaning socio-economic, environmental, geographic and genetic factors, as well as contextual environments within which the health system operates;
- Health System determinants related to processes, policy, organization, infrastructure, facilities, equipment, costs, human and financial resources;
- Performance determinants of the health system such as availability, accessibility, quality and use of health information and services, responsiveness of the system to user needs, and financial risk protection.

These different levels of data enable the HIS to answer to multiple purposes (a) clinical decision-making, (b) administrative decision-making (c) supplies acquisition (World Health Organization, 2008).

## 2.2. DATA SECURITY & DATA PRIVACY

Privacy is defined as having the ability to protect sensitive information about personally identifiable information. While security is the protection against unauthorized access. The first one focuses on the use and governance of individual's personal data to ensure that patients' personal information is being collected, shared and utilized in right ways. When the second focuses on protecting data from pernicious attacks and stealing data for profit. Althrough the two subjects may look similar, they are two different concepts as shown on the table below (Abouelmehdi, Beni-Hessane, & Khaloufi, 2018).

| Security | Privacy |
|---|---|
| Security is the "confidentiality, integrity and availability" of data. | Privacy is the appropriate use of user's information. |
| Various techniques like encryption or firewall, are used in order to prevent data breach from technology or vulnerabilities in the network of an organization. | The organization can't sell its patient/user's information to a third party without prior consent of the user. |
| It may provide for confidentiality or protect an enterprise or agency. | It concerns with patient's right to safeguard their information from any other parties. |
| Security offers the ability to be confident that decisions are respected. | Privacy is the ability to decide what information of an individual goes and where to. |

Table 1. Differentiation between security and privacy

## Security in Healthcare organizations

Kim and al. (Kim, Kim, & Chung, 2013) argued that security refers to (a) data security, (b) access control and (c) information security. When adapted to the healthcare system, this means that healthcare organizations must implement security measures and approaches to protect their data from internal and external risks. To be able to do so, one has to understand and follow the lifetime logic of an information security model. This kind of model ensures that appropriate decisions are made about retention, cost effectiveness, reuse and audit of historical or new data will be taken. Many researchers proposed such schema (e.g.. (Xu, Jian, Wang, Yuan, & Ren, 2014), (Yazan, Yong, & Raj Kumar, 2015)).

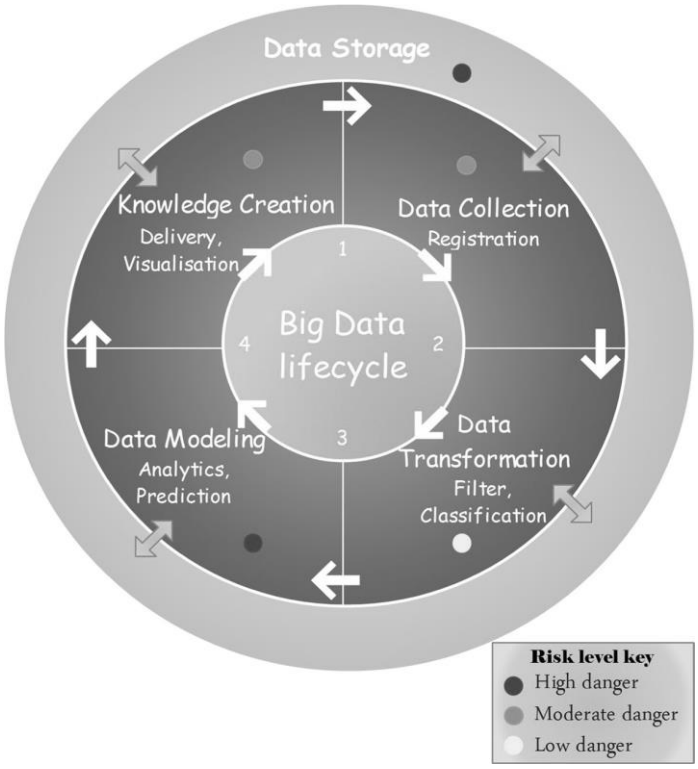



Figure 1. The big data lifecycle in healthcare

**The first phase: Data collection concepts**

During this phase, the collector has to take care to (a) gather data from trusted sources and (b) to preserve patient privacy by storing the data in specific environments.

**The second phase: Data transformation concepts**

The data has to be filtered, classified and transformed in order to perform meaningful analysis, while taking care to preserve sensitive information. The database where the data is stored must be isolated and protected by maintaining access level security and control. Furthermore, security measures should be defined.

**The third phase: Data modeling concepts**

This phase focuses on generating useful knowledge based on the use of supervised data mining techniques. Data mining techniques are useful to highlight features that can explicite sensitive data. This goes against the concept of privacy of the patients' data. So, the process of data mining must be configured and protected against data mining based attacks and security breach and only authorized staff must work on topics of this phase.

**The fourth phase: Knowledge creation**

The result of the third phase is the extraction of sensitive data useful for decision making. To ensure de security of the data, healthcare organization have to make sure that the data mining result is not to be publicly released.

5 ways to implement these lifetime phases of a secured system can be listed (Abouelmehdi, Beni-Hessane, & Khaloufi, 2018).

**Authentication**: establishing or confirming claims made by a real, authentic and confirmed actor. Il means to secure access to corporate networks, protect the identities of users, and ensure that the user is really who he is pretending to be. The information authentication can pose special problems, especially man-in-the-middle (MITM) attacks.

**Data encryption:** data is converted from a readable form to an encoded version that can only be decoded by another entity with access to a decryption key.

**Data masking:** replaces sensitive data elements with an unidentifiable value. It uses a strategy of de-identification.

**Access control:** grants permissions for users and is typically based on privileges and rights of each user having access to the information system. It gives access to the system but also ensure that users perform only the activities for which they have permissions.

**Monitoring and auditing:**

- monitoring is the act of investigating network events to catch the intrusions;
- auditing is the act of recording users activities in chronological order (maintaining logs and modification of data).

## Privacy in Healthcare organizations

The most used technic to ensure the privacy of data in healthcare organization is the "Identity based anonymization" also called de-identification (Abouelmehdi, Beni-Hessane, & Khaloufi, 2018). It is a traditional method to prohibit the disclosure of confidential information by rejecting any information that can identify the patient. But the method is not sufficient. Indeed, re-identification is a possibility on the phase of data mining for example. To help mitigate the risk of re-identification, privacy-preserving algorithms emerged as :

1) *k-anonymity* : a technic of P. Samarati and L. Sweeney

**Definition k-anonymity.** Let RT(A1,...,An) be a table and QIRT be the quasi-identifier associated with it. RT is said to satisfy k-anonymity if and only if each sequence of values in RT[QIRT] appears with at least k occurrences in RT[QIRT].

**Definition quasi-identifier.** Given a population of entities U, an entity-specific table T(A1,…,An), fc: U → T and fg: T → U', where U ⊆ U'. A quasi-identifier of T, written QT, is a set of attributes {Ai,…,Aj} ⊆ {A1,…,An} where: ∃pi∈U such that fg(fc(pi)[QT]) = pi.

**Lemma.** Let RT(A1,...,An) be a table and QIRT be the quasi-identifier associated with it. RT is said to satisfy k-anonymity if and only if each sequence of values in RT[QIRT] appears with at least k occurrences in RT[QIRT].

A k-anonymized data set has the property that each record is similar to at least another k-1 other records on the potentially identifying variables. The most common implementations of k-anonymity use transformation techniques such as generalization (example of variables ZIP code in table 3) and suppression (example of variable Name in table 3).

| Name | Birth | Sex | ZIP code | Religion | Disease |
|---|---|---|---|---|---|
| **Yasmine** | 12/03/1962 | Female | 20502 | Muslim | Heart-related |
| **Khalid** | 21/11/1962 | Male | 20042 | Muslim | Cancer |
| **John** | 01/08/1964 | Male | 20056 | Christian | Viral infection |
| **Aicha** | 30/01/1962 | Female | 29004 | Muslim | Diabetes mellitus |
| **Abraham** | 15/09/1964 | Male | 20303 | Jewish | Pneumonia |

Table 2. A non-anonymized database

| Name | Birth | Sex | ZIP code | Religion | Disease |
|---|---|---|---|---|---|
| * | 1962 | Female | 20000 | * | Heart-related |
| * | 1962 | Male | 20000 | * | Cancer |
| * | 1964 | Male | 20000 | * | Viral infection |
| * | 1962 | Female | 20000 | * | Diabetes mellitus |
| * | 1964 | Male | 20000 | * | Pneumonia |

Table 3. 2-anonymity database

K-anonymity can lead to distortions of data and hence greater information loss as excessive anonymization can make the data less useful and may produce analyses that are biased or erroneous.

2) *l-diversity* : a technic of A. Machanavajjhala, J. Gehrke, D. Kifer and M. Venkitasubramaniam

**Definition l-diversity.** An equivalence class is said to have l-diversity if there are at least l "well-represented" values for the sensitive attribute (attribute of a database that can't be disclosed).

***Definition Distinct l-diversity.*** Each sensitive attributes have at least l-distinct values in the dataset.

***Definition Entropy l-diversity.*** The entropy of an equivalence class E is defined to be:

$$Entropy(E) = -\sum_{s \in S} p(E,s) \log p(E,s)$$

, in which S is the domain of the sensitive attribute, and p(E,s) is the fraction of records in E that have sensitive value s. A table is said to have entropy l-diversity if for every equivalence class E, Entropy(E) ≥ log l. In order to have entropy l-diversity for each equivalence class, the entropy of the entire table must be at least log(l). But the entropy of the entire table may be low if a few values are very common.

***Definition Recursive (c, l)-diversity.*** This form of diversity makes sure that the most frequent value does not appear too frequently and that the less frequent values do not appear too rarely. Let m be the number of values in an equivalence class, and ri (with $1 \leq i \leq m$) be the number of times that i (the most frequent sensitive value) appears in an equivalence class E. Then E is said to have recursive (c, l)-diversity if: $r1 < c(rl + rl+1 + ... + rm)$

| Patient's ID | ZIP Code | Age | Salary | Disease |
|---|---|---|---|---|
| **1** | 47677 | 29 | 3K | gastric ulcer |
| **2** | 47602 | 22 | 4K | gastritis |
| **3** | 47678 | 27 | 5K | stomach cancer |
| **4** | 47905 | 43 | 6K | gastritis |
| **5** | 47909 | 52 | 11K | flu |
| **6** | 47906 | 47 | 8K | bronchitis |
| **7** | 47605 | 30 | 7K | bronchitis |
| **8** | 47673 | 36 | 9K | pneumonia |
| **9** | 47607 | 32 | 10K | stomach cancer |

Table 4. Original Salary/Disease Table

| Patient's ID | ZIP Code | Age | Salary | Disease |
|---|---|---|---|---|
| **1** | 476** | 2* | 3K | gastric ulcer |
| **2** | 476** | 2* | 4K | gastritis |
| **3** | 476** | 2* | 5K | stomach cancer |
| **4** | 479** | 4* | 6K | gastritis |
| **5** | 479** | 5* | 11K | flu |
| **6** | 479** | 4* | 8K | bronchitis |
| **7** | 476** | 3* | 7K | bronchitis |
| **8** | 476** | 3* | 9K | pneumonia |
| **9** | 476** | 3* | 10K | stomach cancer |

Table 5. A 3-diverse version of Table 4

However, L-diversity method is subject to skewness and similarity attack (ex. Table 5).

## 2.3. DATA INTEGRATION

Data Integration is the action of combining data residing in different sources and providing the user with a unified view of these data (Bergamaschi, et al., 2018). In order to discover compelling insights from data, it is often required to extract, transform, and load it into an integrated data (ETL).

### 2.3.1. Concepts[1]

| Concept | Definition |
|---|---|
| **Source system** | A system from which the data is extracted. In the context of healthcare organizations, the source system is the databases of all the Health institutions, plus the one of the organisms concerns by Health topics. |
| **Target system** | The system where the collected data will be load. The target system of Healthcare organization can be each organization or a central one. |
| **Metadata** | Data about the data |
| **Data Mapping** | To process the data between these two systems, relationships between the databases objects are created. It is called Data mapping. |
| **Staging area** | A database area where all processing of the data will be done. |
| **Scalable** | It means to be able to integrate large amounts of heterogeneous data and to manage any change in the system. |
| **Interoperable** | It means that the system must work with other needed applications already existant or not and this with no acces restriction. |

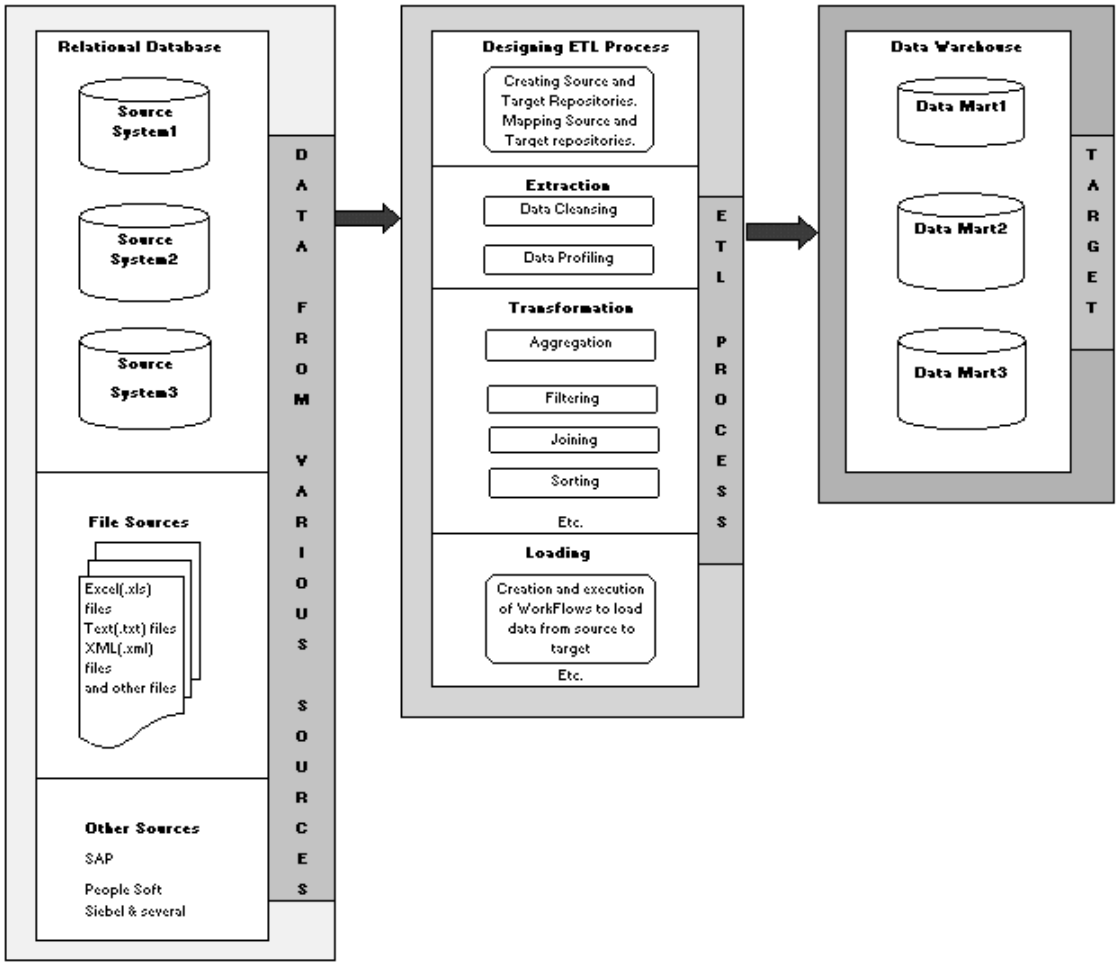
Table 6. Concepts of Data integration

[1] This table was built based on the information taken from the source Kathiravelu, Sharma, Galhardas, Van Roy, & Veiga, 2018.

### 2.3.2. ETL[2]

| Step | Process |
|---|---|
| **1. Extracting** | It covers data extraction from the source system and makes it accessible for further processing. The main objective of the extraction step is to retrieve all required data from source system with as little resources as possible. |
| **2. Transforming** | In this step, certain rules are applied on the extracted data. The main aim of this step is to load the data to the target database in a cleaned and general format because when the data is collected from different sources each source will have their own standards. |
| **3. Loading** | This is the final step in the ETL process. In this step, the data is loaded to the target database. |

Table 7. Descripted process of an ETL




(Source :(ETL concepts, 2018))

Figure 2. Schema of an ETL

[2] This table was built based on the information taken from the source Applied Informatics, 2018.

## 2.4. DECISIONS SUPPORT SYSTEMS

### 2.4.1. Concepts

The pioneers, Peter Keen and Charles Stabell, worked on the concept of DSS during the late 1950s and early 1960s. After that, many researchers - as Michael S. Scott Morton, Bonczek, Holsapple and Whinston, Turban and Aronson or even Steven Alter - worked on the topic. The last one published in 1980 an important book (Alter, 1980) giving three majors characteristics of DSS: they are designed specifically to facilitate decision processes, to support automate decision making and also to answer quickly to the changing needs of decision makers.

One can also retain a more comprehensive definition of DSS proposed by Turban and Aronson in 2004. To him, "a DSS assists decision makers in problems which cannot be solved by standard procedural methods or tools. It tries to improve the effectiveness of decision-making (quality) rather than its efficiency (the cost of decision-making) but doesn't tend to the remplacement of decision makers. The decision support system also facilitates several interdependent and/or sequential decisions that may be made once, several times, or repeatedly by carrying out all parts of the decision-making process and by covering a variety of decision analysis tools".

Taking into account the many researches and characteristics highlighted regarding DSS, Rashidi et al. in 2018 presented a diagram of DSS's features.

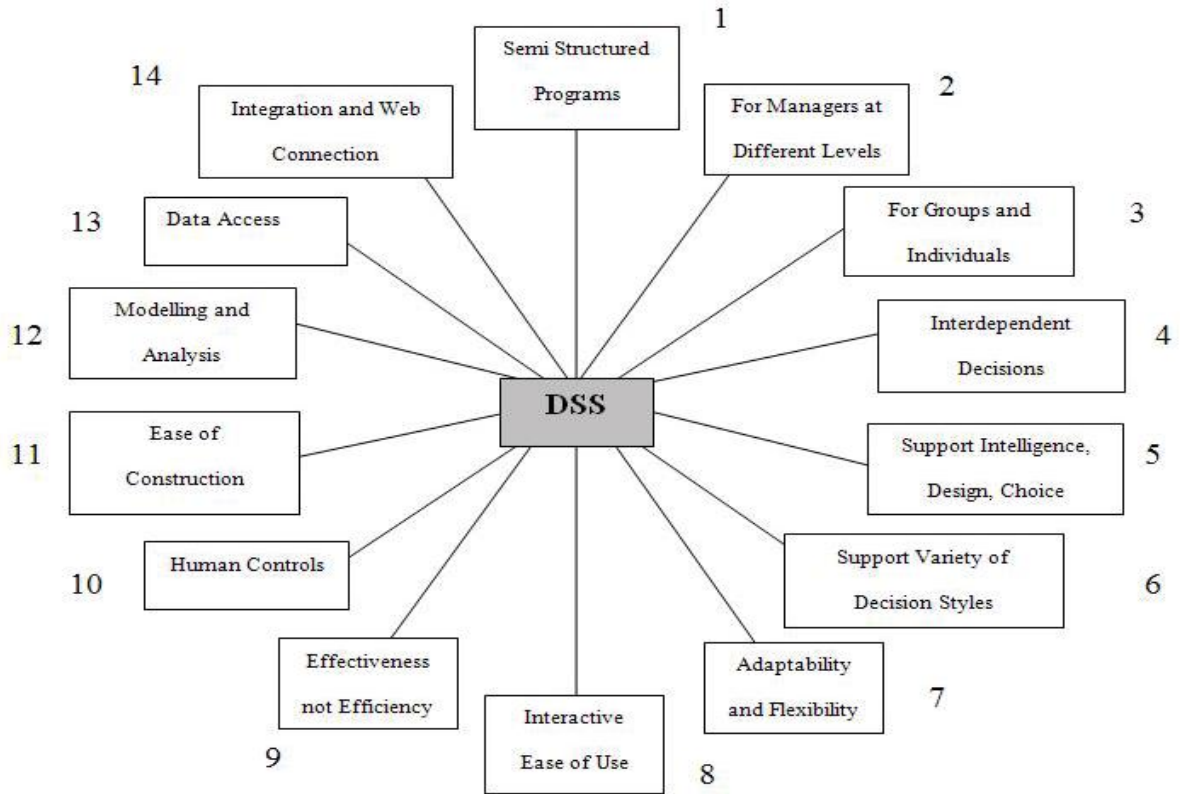



Figure 3. The desirable characteristics and capabilities of DSS

### 2.4.2. Model & Architecture

There exists multiples DSS' types. These classifications depend on the attributes/criterion on which the definition is build. Indeed, a definition of DSS can be made using two distinct criterions: the relation between the DSS and its users (Haettenchwiler) or the scope of utiity of the DSS (Power). If taking Haettenchwiler's theory (Holsapple, 2008) a DSS can be passive (only collecting and organizing data), active (is as passive's DSS but also showing specific solutions) or cooperative (is an

active DSS that allow the decision maker to refine the solutions suggestions). If taking Daniel Power's theory (Holsapple, 2008) a DSS can serves an entreprise-wide scope (serves many users in the company) or a desktop scope (a DSS that runs on an individual's PC) and take the form of a :

- Communication-driven DSS, integrating tools (e.g. audio and video conferencing) that enable the communication to support more than one person (usually internal team or partners) working on a shared task;
- Data-driven DSS, fitting data according to the decision makers' needs and deployed on a main frame system that enables access and manipulation of data through queries with the aim to solve specific purposes;
- Document-driven DSS, stocking unstructured information in a variety of format and ables to search specific data in its database based on a keywords;
- Knowledge-driven DSS can recommand specific actions to specific problems by proposing specialized problem-solving expertises stored as facts, rules and procedures;
- Model-driven DSS propose simulation models and help to choose between the different option regarding specific needs and wants.

Structurally, Bonczek et al. in 1981 described a DSS through the presence of four essential components:

(a) A Language System which includes all the recognizable messages;
(b) A Presentation System for all messages emitted by the DSS;
(c) A Knowledge System addressing all the imbedded knowledge in a DSS;
(d) And a Problem-Processing System that tries to diagnose and solve problems.

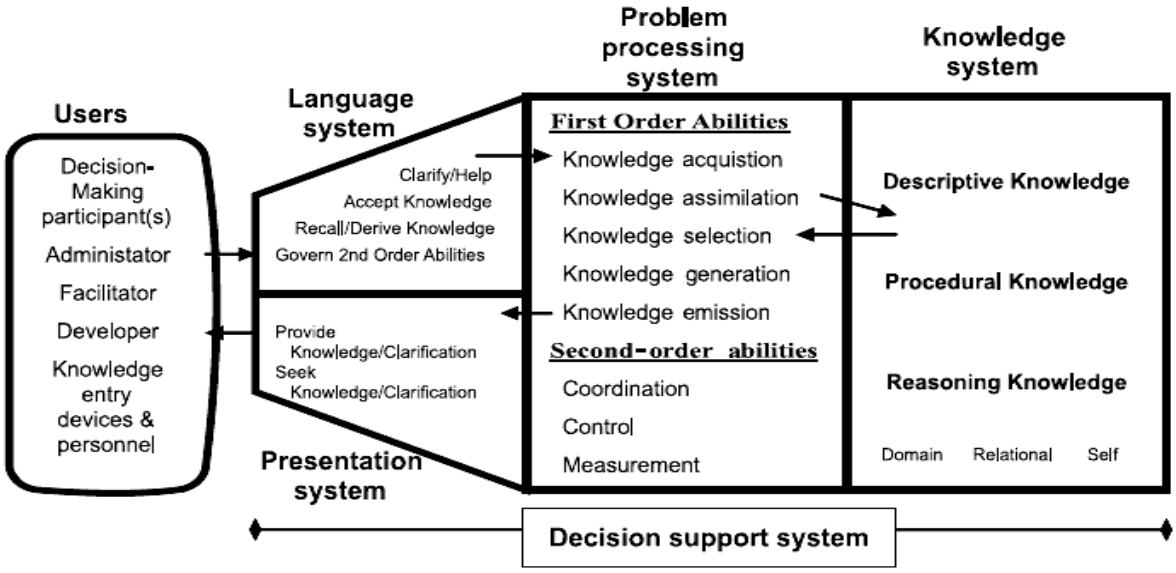



Figure 4. Basic architecture of a DSS with the 4 components highlighted

### 2.4.3. Predictive Analytics

Predictive analytics are specific studies of historical data aiming at developing an understanding of the phenomenon that impact the studied data with the goal to predict what will happen in the future. This type of analysis encompasses a variety of statistical and leads to an optimized process of decision making and represents a high achievement in terms of value and difficulty.

Companies now expect to know what the future will be made of so they can prepare themselves, decide what they have to do (a prescriptive decision making). As Halper highlighted it in 2014, the top five reasons to use predictive analytics are: (a) to predict trends, (b) to understand customers, (c) to improve business performance, (d) to drive strategic decision making and (e) to predict

behaviours. The method to achieve these five goals is based on the mapping of a mathematical function with a satisfactory condition. The complete methodology of the building of a predictive model starts with the identification and definition of an objective and a problem to solve. Then the data is explored so that the adequate data can be selected, then prepared (pre-processing) to create a model based on the data records from the training dataset (30% of the all selected data) with appropriate techniques. The model is iteratively tested then validated regarding the goals fixed by the definition of the problem to solve. The model is meeting a satisfactory condition only if (a) it enables the restitution of the expected outcomes also called target variables or if (b) it enables to understand the invisibles links between all the data of the input data. Once validated, the model is deployed, integrated and managed (automatized) (Guazzelli, 2012). A model that can map input to target variables is the outcome of a specific predictive analysis called supervised learning. As the name explains it, it means that when beginning the analysis, one dispose of the input and the output (Guazzelli, 2012). The best-known supervised learning algorithms are classification and regression algorithms such as neural networks, support vector machines and decision trees. On the contrary, the second type of model – understanding links in the input data – is called unsupervised learning for the simple reason that the target variables aren't known. The best-known supervised learning algorithms are clustering and association algorithms (Guazzelli, 2012).

### 2.4.4. Success factors

Surely, a sucessfull DSS is an used DSS as it is the only way for the DSS to have an impact. But there are others factors, too, more specific. Kawamoto et al. in 2005 examined and highlighted four of them. A sucessfull DSS should:

(a) provides alerts/reminders automatically as part of the workflow;
(b) provides the suggestions at a time and location where the decisions were being made;
(c) provides actionable recommendations;
(d) computerises the entire process.

# 3. APPLIED METHODOLOGIES

## 3.1. PROJECT METHODOLOGY

*« A system development methodology refers to the framework that is used to structure, plan, and control the process of developing an information system »* (Centers of Medicare and Medicaid services, 2008). It includes procedures, techniques, tools and documentation aids which help the system developers in their efforts to implement an IS.

Avison and Taylor (Avison & Taylor 1997) proposed a classification of the methodologies' types. To them, the main types of methodologies belong to one (or more) of the following classes:

- Class 1. Well-structured problem situations with a well-defined problem and clear requirements;
- Class 2. Well-structured problem situations with clear objectives but uncertain user requirements;
- Class 3. Unstructured problem situations with unclear objectives;
- Class 4. Situations where there is a high user interaction with the system.

DIAMANT and GCS share their methodology which is a complicated one to define based on what the literature states. To classify the projects, they are projects of the kind of the class 2. The situation is clear, the RHA and their collaborators need an efficient common database because they are all functioning with their own, which isn't efficient on a national level. Based on this, clear objectives are stated: a common database with cleaned data, modelled on cubes and reports, enabling the users to conduct the Healthcare strategy of France. But the requirements often change. This is because of the uncertainties in the requirements that the projects can't be part of the class 1. And this, despite the fact that this class 1 (related to projects using the SDLC methodology) is in many aspects similar to the strict methodology that Keyrus and the client want to apply on DIAMANT and GCS.

Keyrus and the clients of DIAMANT's and GCS's projects would have like to implement a SDLC methodology.

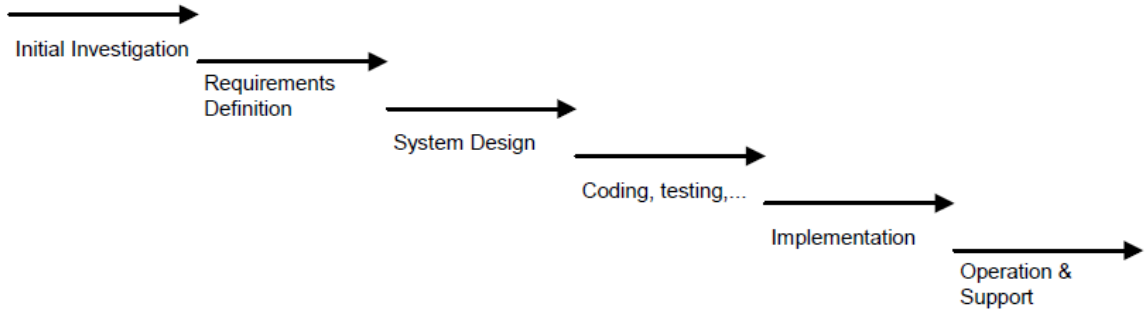



Figure 5. SDLC's steps

A project using this methodology is divided into phases, with minimum overlap. The planning and budgets are stated at the very beginning which implies a high level of understanding of the technology used and requirements of the client. The methodology is strict in its control as documentation must be updated, approval should be obtained for each development, which help to ensure the quality, reliability, and maintainability of the developed software. In all these aspects, each project at Keyrus hopes to apply this methodology. Indeed, the hierarchical structure of the company requires projects managers to be the makers of the needs' translation and to be fully-updated to the development, which goes in the way of a strict control. They also hope that each

implementation respects the chosen system design and is done with respect of time and cost agreements.

However, important components of this methodology aren't applicable in both cases. First of all, the managers at Keyrus aren't all technical people; it can be difficult in this situation to define clearly the requirements of the project and its architecture. This leads the managers to ask to the developers to be active parts of the project management. Secondly, the waterfall methodology is costly because of the tight control. Keyrus tries to avoid an excessive cost and being a service provider, the company can't afford to lose clients because of "useless" costs. Finally, this methodology isn't efficient to respond to changes because it only goes forward and so on isn't keen to use iteration, which is in practice not possible when developing for an external client.

DIAMANT and GCS are less "strict" or structured projects. But the process is the same as SDLC process:

- A preliminary analysis phase: the commercial team with the technical leaders review the request to know if financially, technically and legally, the society can participate to the project ;
- A system analysis phase: analysts, developers and final users determine the specific requirements, procedures are documented, key players are interviewed and data requirements are pronounced in order to get an overall picture of exactly what the system is supposed to do ;
- A system design phase: technical leaders develop the specific technical details required for the system;
- A programming phase : developers develop the solution;
- A test phase: the developers and final users look for errors or bugs. Any bugs, errors, or problems found during testing are addressed and then tested again.
- An implementation phase: the implementation/use in the organization and the training of the users;
- A maintenance phase: reported bugs are fixed and requests for new features are evaluated and implemented. Moreover, system updates and backups are performed on a regular basis.

Seeing the many similarities between the SDLC methodology and DIAMANT/GCS methodologies, it could have been a consensus to say that the incremental methodology is the best representation of the use method as it presents each step but with less rigidity. But, to be critical, choosing this methodology would be an excuse to explain why sometimes so many problems appended during the development : why requirements weren't goodly specified or why specification of the needs are to be defined again despite the advancement of the development. A more realistic consensus can be made on the RAD methodology. The method focuses on quickly building a working model of the software, getting feedback from users, and then using that feedback to update the working model. After several iterations of development, a final version is developed and implemented. The RAD methodology consists of four phases:

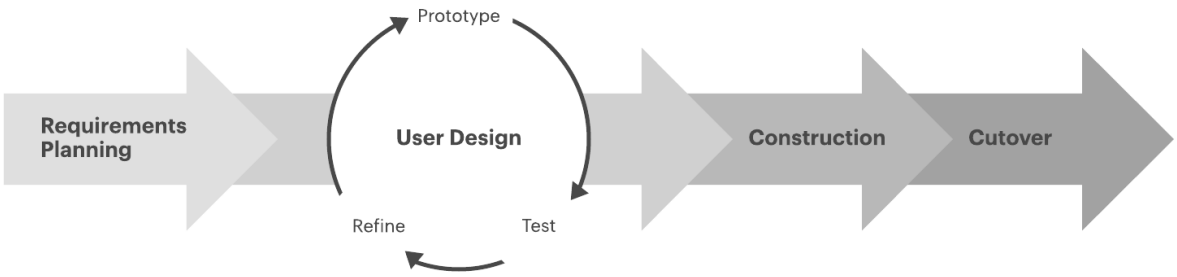



Figure 6. RAD's steps

- A requirements' planning phase: like the preliminary-analysis, system-analysis and design phases of the SDLC. In this phase, the overall requirements of the system are defined, a team is identified and feasibility is determined;
- An user design phase: technical leaders interactively create the design of the system while having discussions with the users about the design of the system and the expression of the requirements;
- A construction phase : developers de velop and proceed to changes in parallel with the User Design step;
- A cutover phase: similar to the implementation and maintenance phases of the SDLC.

As one can see, the RAD methodology is much more compressed than SDLC and focuses on user participation and iteration. This iteration gives the chance to improve the model as well as exploiting the gained knowledges on the subject. Nevertheless, in DIAMANT and GCS, this need of iterations often comes from the bad expression of clear needs and requirements. The iterations make the solution less rigid so that the developers can integrate corrections at each identification of a problem. But this methodology is a problem to Keyrus. Indeed, this methodology has a real weakness that can be easily explained by the quality triangle.

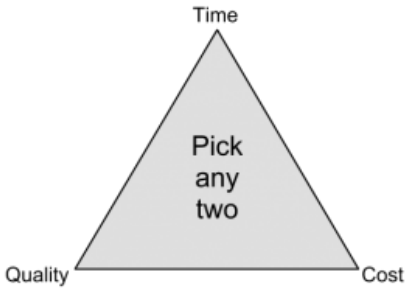



Figure 7. The quality triangle

This triangle stipulates that in a client- services provider's relationship, there always existed a tension relates "to how quickly the software can be developed (time), how much money will be spent (cost), and how well it will be built (quality). The quality triangle states that for any product or service one can only address two of the following: time, cost, and quality" (Bourgeois, 2014). This is the real and understanding explication that can be given to the many problems existing in DIAMANT and GCS. Unfortunately, this isn't always possible to explain it to clients, even more public ones.

## 3.2. DATA SECURITY AND PRIVACY METHODOLOGIES

Working in a sensitive topic as Health obliges the users / developers of an application to be aware and very strict on the security and privacy measures. The process elaborated by the identified users, managers and main developers concerning the securities measures are a) to extract, transmit and load data only coming from national institutions (INSEE, RHA, Ministries and others recognised sources) ; b) to store the data and database in specific and protected environments ; c) to control and monitor the access to these environments ; d) to ensure that data mining techniques are useful to highlight features useful for the users ; e) to forbid public released of data.

### 3.2.1. Data security

DIAMANT and GCS' teams can guarantee that the 5 security technics identified by Abouelmehdi, Beni-Hessane, & Khaloufi, 2018 are respected. First of all, an authentification is obligatory to have access to the data and global database. Secondly, on Keyrus side, the corporate network is secured and the team is using a virtual machine to develop. This machine is controlled, secured and maintain

by Wallix, a cybersecurity software vendor. The access to this machine is linked to a identified profile of user, at a specific IP adress and with a specific password. Each connexion to the machine produces a log that can identified the exact user and some actions on this machine. On the client side, the coporate network also is secured. The connexion to the database only is possible throught Excel, with a fixed password added to a changing password generates each 60 secondes before the connection by a token. The wrong use of the token blocks it and needs Keyrus' intervention to reactivate it. The access to Excel also is based on the specific rights attributed to each user. These measures limit the major problem of a lack of authentification measure : a man-in-te-middle attack. Thirdly, data are sent – encrypted - by the client to Keyrus throught a sFTP. To decrypt the files, Keyrus uses Kleopatra, a certificate manager and GUI for GnuPG. Moreover, data mining methods are not applied in both projects or not for the highlighting of patients' profiles by Keyrus. Only the clients do have data miners for analyses of this type. And finally, Keyrus does not have the right to disclose any data from its clients.

### 3.2.2. Data privacy

Privacy measures aren't managed by Keyrus' teams on DIAMANT and GCS projects. The reasons behind this, is that:

- Keyrus isn't collecting the data itself, the data are transmitted to Keyrus by the client ;
- The transmitted informations should be anonymized before any sending to Keyrus.

To be critical, the data is "name-anonymized" meaning that the name of the patients is transformed in numeric identifiants. But the data isn't de-identified in its globality. The database links each ID to a postal code, a pratician, even an hospital or clinic, exact birth or death dates are known as well as diseases or death cause. Indeed, re-identification is possible (or almost if looking for a vey specific cases or famous cases – for example a certain formule 1 pilot severely injured while skiing and who was transported to a known infrastructure for a known traumatism at dates that are publicly disclosed, can be easily identified) during a data mining phase. It should be kept in mind that a complete de-identification is tricky on projects as DIAMANT and GCS as the goal is here to develop a precise database that will be the main application at a national and european levels. Then, masking the specificities held by age, region and pathology will biased the final result. Moreover, the results achieved by both projects haven't vocation to be publicly disclose.

# 4. APPLICATIONS AND TECHNOLOGIES

This chapter serves the purpose to present all the applications and technologies used in the context of the projects. A list of website links regarding each application or technology can be found in appendix 1.

*Windows Server*

When the RHA proposed DIAMANT on the market, they asked for the use of WS 2008 R2. This solution was one of the best solutions in 2009. The capacities and possibilities of this solution enable the project to still be going on even now, then years later. It could continue like that, however, the use of MSBI 2017 isn't possible with WS 2008 R2. For now, the migration is on hold as the client doesn't want to pay for the migration as it has a huge impact on the on-going integrations. GCS, the project of "test" uses upgraded servers on the 2016 Essential version of WS which is also Azure connected.

*Microsoft SQL server (SSIS, SSAS and SSRS)*

DIAMANT uses SSIS as an ETL, SSAS to build cubes and SSRS to build reports using requests directly made on the SSAS' cubes. GCS used SSIS as an ETL and SSRS to build reports using requests directly made on SSIS' databases.

When DIAMANT was put on the market for the 1st time, the RHA asked for the use of MS 2008 R2 as it was an "up to date" software having the advantages of being free and rich-featured. 4 years later, for the 2nd version, an update was asked and so on the project was developed with MS 2012. The 3rd version was proposed in 2015 and then, MS 2012 still was the preferred version of the tool instead of MS 2014 because the 2014 only proposed few changes that weren't so relevant to explain a new investment. For the 4th version of DIAMANT, the client wants to keep using MS. The latest version is 2017 and this is what the client asked for, but the client also wants to implement it only after a period of test. Indeed, the 4th DIAMANT contract goes from 2019 to 2022 so DIAMANT's direction decided that the project will migrate on this latest version starting on 2020 after a 1-year period of test on a new project that Keyrus just won: GCS.

The reason behind this choice (Microsoft technologies) is the experience and well-known expertise of Microsoft. Nevertheless, the client is hesitant regarding the use of applications that aren't known or too young because it can lead to problems that will not be caused by the service provider's but by the company which developed the technology of storage and ETL.

There exist tons of other tools competing MS: SAP, BO. Technologies that Keyrus' workforce shows great skills on. Nevertheless, the migration on another solution is risky and includes lots of unknown elements which the RHA and Keyrus want to avoid.

*Excel*

The client has zero access to the database through Microsoft SQL Server, nor any knowledge on database development or SQL requests. The tool they are using is Excel, so the tool put at their disposition for the request of the database is Excel, the 2010 and 2013 versions (the 2010 version requires the installation of a component). Keyrus is paid to form the users on Excel. On DIAMANT, this also is the tool used to test the integration with no regression tests based on count of rows by period. The tests are historically kept on DIAMANT's servers; nevertheless, it is a poor method of tests as time shows. GCS on the contrary uses a more adequate method of test: Squash.

*Squash*

Squash is an open source project aiming to structure and industrialize functional testing activities.

It offers two functionalities Squash TA and Squash TM. The project's context here only presents the need for Squash TM. The tool enables the management of requirements, test cases and campaigns execution in a multi-project context.

On GCS, the tool is used to ensure each phase of the project development, meaning to ensure the integration and management of the source files during the phase of development, then the restitution during the phase of test and finally, the final presentation during the phase of production.
Squash is indeed very useful as it enables the administrator to construct repositories for each project, to reference test cases on an inter-project repository level. During the process each step of the tests obtains a status (success, failure, on-hold, not applicable) depending on the requirements applied by the administrator. They are useful as it allows the tester to evaluate the global project quality. Moreover, Squash TM integrates reporting functionalities based on the results and advancements process of the tests. These visuals can be exported and shared, in GCS context, shared with the client. The client is then kept up to date with the tests applied and the obtained results. The client can also have access to the tool without the installation of it and then can add tests or even test itself.

*SharePoint and Qlikview*

These 2 technologies are used:

- SharePoint: by Keyrus to develop and maintain the online platform created for DIAMANT. This platform is a way to stay close (the client and Keyrus). Keyrus is the administrator of the platform; it updates it with news of integration, with SSRS reports and Excel shits it formatted.
- Qlikview: by the client exclusively. Some of the RHA have developers with Qlikview appetences. They asked them to develop Qlikview applications in order to dispose of Data visualization as SSRS can't answer to this need and Excel isn't efficient enough on this topic. At Keyrus, the project's team publishes these applications on the DIAMANT portal and is administrator of them but does not develop them.

*PowerBI*

DIAMANT and GCS lack a real tool of Data Visualization. From now on, they were using SSRS from MSBI and sometimes graphs on Excel. This was decided by the client. Nevertheless, as a "connection analysis" (study of the number of connections by each user on the SSRS tool) showed it, the users do not find interest in this tool and mainly use their own Excel shits (that they share among them).
Their needs were reevaluated. GCS is a young project with only one built report for now on. This will change but the client is taking its time to evaluate its needs and decide with restitution tool to use. For DIAMANT, this is different as they ae convinced that they need a restitution tool because they feel the need to nowadays enter a new phase of the project: a predictive phase. Using the Microsoft solutions, they often ask for PowerBI as it is great Data Visualization software which also integrates R codes, so Keyrus' DIAMANT's team is working on propositions.

*R - RStudio*

As said above, the client wants to enter a second phase of the project. The first one was a phase of massive integration and development. The second is mainly correction, further analysis and the building of a solution more attractive with visuals and some propositions of predictions.
The predictive axe of the project is beginning slowly so instead of paying for software that could become useless, the client wants to test R capacities.

# 5. DIAMANT

## 5.1. PROJECT PRESENTATION

Being entrusted to guide the Health's strategy of the country, the 24 Regional Health Agencies need a common database and shared platform of information with the idea of building an efficient community of users. This information system should consider each IS of the 24 distinct RHA and a large variety of other sources, to build an ETL and program a restitution tool of the integrated data. Their request includes: resumption of the existing, initialization, detailed functional specifications, technical design and implementation, integration and data quality, scalable and corrective maintenance and user support. Precisely, their desire is a data-driven DSS on which the RHA are the one and only actors/decision makers. The automation of information in DIAMANT intended to give a vision of the performance of the institutions, to be able to identify risk situations, to pilot the common strategy of the health institutions under the aegis of the RHA of France (controlled by the Ministry of Health), to build the budgets, to make inter-regional benchmarking. In this sense, DIAMANT isn't an HIS. In many aspects is looks more like a classic IS as the data scope makes it more financially interesting.

### A public environment

As said, DIAMANT is a tool used for the control of the health strategy in France. Health is indeed a very important subject in France as the country spends around 10% of its national gross income on it. It is one of the highest rates in the world. Thanks to this, all resident in France have unrestricted access to doctors and specialists. This system is called "assurance maladie" (medical insurance) and is a dual system composed of a public agent (the government through a SHI) and a private one (private health insurances also called VHI) (Durand-Zaleski, 2016). In 2014, France spent in total, EUR257 billion or USD310 billion in its health expenditures which represent that year 12% of GDP: of which 76.6 percent was publicly financed and 13.5% were financed by VHI (Chevreul, Durand-Zaleski, Bahram, & al., 2014).

The SHI is half financed by employers and employees' payroll taxes, for 35% through a national earmarked income tax, for 13% through taxes levied on tobacco and alcohol, the pharmaceutical industry and voluntary health insurance companies and for the last 2% by the state subsidies. In France, this coverage is universal and compulsory, provided to all residents by noncompetitive SHI thanks to the "Protection universelle maladie" (universal health care coverage) law of 2016. SHI isn't reimbursing for the entire amount of the care. So even if using a VHI is not compulsory, it is a great complementary. VHI nonetheless are competitive (through limited as providers are supervised by the Mutual Insurance Funds Control Authority). Complementary insurance is provided mainly by not-for-profit, employment-based mutual associations or provident institutions (Chevreul, Durand-Zaleski, Bahram, & al., 2014).

This system is under severe financial pressure with a growing shortage of suitable staff in many regions, and a higher proportion of the running costs must be picked up by patients but also, an increased population, the integration of immigrant, the new territories (e.g. Mayotte) and poor gross increase). To fight this situation, the government has - over the past few years - introduced a series of reforms to increase efficiency and drive down costs. Social security health contributions have been increased, as have fees for medical consultations, whilst reimbursement levels have been reduced.

Seeing the country's situation, the Health Ministry (as each Ministry) had to apply a new strategy of control. This new strategy implies an increase of the budget and quality rigor in the appearance of

new processes and new institutions with new functions (an organigram of this system is available in appendix 2). Therefore, (born of the law concerning the « hospitals, patients and health policy in all territories » reform, promulgated the 21st of July 2009) the Regional Health Agencies became (the 1st January 2010), the unique authorities to decide of the health system at the regional level. It is their duty to put in place the regional health policy.

DIAMANT nowadays

Nowadays, DIAMANT, it is:

- 911 tables in the ODS and 923 in the DWH;
- 4 SSRS reports and one Excel report;
- 11 SSIS solutions;
- 13 cubes;
- And more than 1500 users and the number is growing.

It can be said that in 10 years the project became massive. Which is a difficulty when integrating it for an internship. The project is so advanced that not so many missions - valuable for the validation of the master - were proposed. DIAMANT isn't nowadays a succession of big development projects. It is mainly maintenance of the application (data loading, correction and users help), actions of evolution of the solutions and minor new projects development. The new projects that came during the internship are : social report (analysis of the patients population), Arianne (analysis of the slum areas of France), Ovalide (detailed report of the patients' trajectories), financial evolutions (major evolutions with many integration of new sources) and C3 (creation of a digital application).

## 5.2. MISSIONS

During the internship, the maintenance as well as OVALIDE and HAPI were the intern's tasks. Each of these projects go through the same process (5.2.1). Working on these projects took almost the 6 months of the internship. But as the company offered the intern a permanent contract, more ambitious tasks were also offered (5.2.2).

### 5.2.1. First task : Business Intelligence development as a developer

Developments at Keyrus are structured as follow, needs reception and expression, development, two sets of tests with possible rework and evolution and launch.

#### 5.2.1.1. Support and exploitation

Maintenance is indeed very important in the lifecycle of a project. It requires knowledge about the different integrated data, but also on their strategic meaning. It requires knowledge on the development tool and on the methodology. It is therefore a good way to integrate a project, before going on, on further development tasks. At Keyrus, the maintenance is called TMA.

The client wishes to benefit from Keyrus services for what is called in the company "support and exploitation" activities for the period from January to December 2018. These functions include:

- The manual execution of periodic loadings and recognition and correction of anomalies on the current perimeter (PMSI MCO, PMSI SSR, PMSI HAD, PMSI PSY, RPPS, Certification, Finess, DADS Medico-Social, DADS Sanitary, Social Balance Sheet and SAE);
- The monitoring and correction of the anomalies for the automatic periodic loadings on the current perimeter (CF, EPRD, DM, RIA, PMSI MCO, PPT);
- The changes related to edited files and solutions;

- The technical-functional support for the ARS;
- And the technical-functional support for the GHT.

These tasks can the can be made on the ETL solution, the cubes or on the presentation tools (SSRS and Excel). They represent the daily work of at least one resource for 1 year. This is an important and time-consuming task that can pose problems of stabilization of the solution. It is therefore a task that is expensive for the client.

| | | | Qté | Prix unitaire | Prix total HT | Prix total TTC |
|---|---|---|---|---|---|---|
| **DIAMANT Support Exploitation 2018** | RUN2 | Exécution de chargements périodiques | 1 | 41 524 € | 41 524,00 € | 49 828,80 € |
| | SUPT52 | Support technique pendant 1 an | 1 | 52 260 € | 52 260,00 € | 62 712,00 € |
| | SUPF52 | Support Fonctionnel pendant 1 an | 1 | 34 406 € | 34 406,00 € | 41 287,20 € |
| | | **TOTAL** | | | **128 190,00 €** | **153 828,00 €** |

Figure 8. Cost charge for support and exploitation for the RHA users on DIAMANT in 2018

This cost corresponds to the data loadings. Some of which are daily but automatic and some of which are trimestrial or even semestrial but not automatic. This cost also counts for RHA's users support.

| | | | Qté | Prix unitaire | Prix total HT | Prix total TTC |
|---|---|---|---|---|---|---|
| **DIAMANT Support Utilisateurs GHT 2018** | SUPF52 | Support Fonctionnel pendant 1 an | 1 | 34,406 € | 34,406.00 € | 41,287.20 € |
| | | **TOTAL** | | | **34,406.00 €** | **41,287.20 €** |

Figure 9. Cost charge for support and exploitation for GHT users on DIAMANT in 2018

The data are coming only from the RHA, that is why the GHT only have to support the cost represents by the support they need.

Not all the types of actions taken during this task are possible to present, so one specific case will be described. If taking the perimeter delimited by the automatic periodic loadings, each legal structure has to present and validate its financial status of the year to the state via validation organisms. Once validate these accounting documents are sent to Keyrus for integration. In general, the n+1 files have the same format than the n files. In this case, the loading is basic. But from year to year, the norms are charging and so the files' formats too. This has to be taken into account for the future loading of data as it changed many things on the ETL process and calculated measures made on SSAS. In DIAMANT case, the campaigns are at the number of 6 (in this order: EPRD, DM, RIA 1, RIA 2, RIA 3, CF). The EPRD (revenues and expenditures estimation) are results expectations. On these "expectation", modifications can be made, then an up-to-date document called DM (amending request) is produced. Each four months, the structures updated their current situation and report it (in the form of a document called RIA (sub-annual report) to their validation organism. Finally, at the end of the year, the global financial situation is reported in a report called CF (financial account).

Starting in September, the internship gave the intern the possibility to have to take care of the integration of the RIA 2 2018, RIA 3 2018, EPRD 2018 and CF 2018. Except the last one, they all changed of format. These changes require the parameters' evolution with the update of a parameters' file (csv), some changes in the SSIS loading process and in the SSAS measures' calculation and a test analysis. But no further developments on SSIS or SSAS are needed. Indeed, the developments are made so that the data integration process is not perturbed by small changes. It, nevertheless, requires work on presentation tools as new data have to be presented. Some excel shits are directly constructed and managed by the users as they are building their own specific queries based on their role in the organization. But one massive file is used by all users. This file, called OSEF (Tool for monitoring the financial balance of public institutions and ESPIC (Private health institution of collective interest)), serves to obtain a multitude of financial data in a macro or detailed way over several years. The intern role on this subject was to integrate each new financial campaign on OSEF: meaning building the parameters for the new data, but also the new tables and graphs.

As for the tests' phase on DIAMANT, it is simple a simple process. One has to look for data adequacy between the files and SQL tables then between those tables and the cubes' data, then between the cubes and OSEF. It is not a traced process. Then the integration is entering a second phase of test, this time made by the client. Some profiles are set up as testers on the cubes. It is their goals to verify the data adequacy in a functionary way. Finally, if the testers do not highlight problems in the changes and developments, the project is launched, meaning that the development is going fom the development and test environment to the production one and that all the users are able to use it. The final task when doing the integration on data is to update the portal so that the platform's users know what data they should expect or not in the application.

| Cube | Source | Données | Fréquence d'alimentation | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CUBE_COMMUN_PMSI_MCO | PMSI | PMSI Détaillé MCO | Mensuelle | - | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | extract. ATIH 20/03/2019 reçu le 21/03/2019 | - |
| CUBE_COMMUN_PMSI_HAD | PMSI | PMSI Détaillé HAD | Mensuelle | - | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | extract. ATIH 20/03/2019 reçu le 21/03/2019 | - |
| CUBE_COMMUN_PMSI_SSR | PMSI | PMSI Détaillé SSR | Bimestrielle | - | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | extract. ATIH 20/03/2019 reçu le 21/03/2019 | - |
| CUBE_PMSI_PSY | PMSI | PMSI Détaillé PSY | Bimestrielle | - | - | - | 2014 | 2015 | 2016 | 2017 | extract. ATIH 20/01/2019 | - |
| CUBE_RPU_ATU_ACE | ATIH | Résumé de passage aux urgences | Mensuelle | - | - | - | - | - | 2016 | 2017 | 07/2018 IDF | - |
| CUBE_DIAMANT_SAE | SAE | SAE Base administrative | Annuelle | - | - | 2013 | 2014 | 2015 | 2016 | 2017 | - | - |
| CUBE_COMMUN_FINANCE | ANCRE | EPRD | Quotidienne | - | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 2019 |
| | ANCRE | DM | Quotidienne | - | - | - | - | - | - | Données non exhaustives ! | 2018 | - |
| | ANCRE | RIA | Quotidienne | - | RIA3 | RIA3 | RIA3 | RIA3 | RIA2 | RIA2 | RIA1 | - |
| | ANCRE | CF | Quotidienne | - | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | - | - |
| | ANCRE | Plan de Trésorerie Prévisionnel | Quotidienne | - | - | - | - | 2015 | 2016 | 2017 | 2018 | - |
| | HAPI | HAPI | Mensuelle | - | - | - | - | - | - | 2017 | 2018 | - |
| DEMOGRAPHIE | ASIP | RPPS (Méd/ChirDen/SageF/Pharm/MKiné/PédicureP) ADELI (Autres professions paramédicales) | Mensuelle Semestrielle | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | 15/04/2019 18/03/2019 |
| CUBE_COMMUN_RTC | ANCRE | Retraitement comptable | Annuelle | 2011 | - | 2013 | 2014 | 2015 | - | - | - | - |
| CUBE_RH_MEDICO_SOCIAL | DREES | DADS/DSN Médico-Sociale | Annuelle | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | - | - |
| RH_PERFORMANCE | DREES | DADS Sanitaire | Annuelle | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | - | - |
| | ATIH | Bilan Social | Annuelle | - | 2012 | 2013 | 2014 | 2015 | - | - | - | - |
| CUBE_COMMUN_DETAIL | PMSI | PMSI Agrégé MCO | Hebdomadaire Clôture données samedi pour dispo mardi Extraction compémentaire le 16 de chaque mois | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | M12 | - |
| | IPAQSS | Qualité et sécurité des soins | Annuelle | - | - | - | - | 2015 | - | - | - | - |
| | QUALHAS | Certification | Annuelle | 2011 | 2012 | 2013 | 2014 | 2015 | - | - | - | - |
| | PMSI | PMSI Détaillé OVALIDE | Mensuelle | - | - | - | - | - | - | 2017 | 2018 | - |
| Tous | FINESS | Structure | Semestrielle | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 12/09/2018 | - |

Figure 10. DIAMANT's Data Integration state[3]

### 5.2.1.2. HAPI

The client wishes to benefit from Keyrus services for the restructuration of both SSIS and SSAS solutions regarding the financial scope of DIAMANT and the addition of new sources to these solutions.

The restructuration is composed of:

- sources' modifications;
- axis' creation ;
- indicators' creations.

The integration is composed of:

- project HAPI;
- project DGFIP.

At the beginning, the whole financial subject was supposed to be assigned to the intern as his profile (a bachelor in global management and financial strategy) is an advantage to understand functionally the data; which is supposed to lead to less errors. But, during the development, another project

[3] The colour code is as follow : red if there is no data, yellow if the data is still incomplete and green if the data is complete. Some box have comments in them : it refers to the date of the extraction of the last files integrated or to the period of time integrated.

(GCS) came and was proposed to the intern. So here, only the HAPI project will be presented as only this was the intern's task.

The HAPI application was created to harmonize and secure pricing practices between regions. Since 2012, the regional health agencies use it to rate the institutions under the jurisdiction of the CNSA. And thanks to the Qlikview restitution module, they can track the consumption of their regional envelope. In 2017, HAPI is evolving in order to integrate the impacts of the various tariff reforms in the sector (notably the integration of ESATs into the CNSA budget or the reform of EHPAD pricing) and to take into account the comments of users.

HAPI has two main objectives:

- At the regional level, for the RHA, to price the establishments and services until the publication of the tariff notifications and follow the consumption of the regional envelopes in real time. HAPI guarantees the use of the same documents and same harmonization of the tariff calculation methods based on the Social Action and Family Code.

- At the national level, for the CNSA and Ministries, to have visibility on the consumption of the envelopes delegated to the RHA to finance the operations of the establishments and services "in real time". This visibility on the consumption of the envelopes delegated to the RHA allows the CNSA to have reliable and precise information for the definition and monitoring of the overall objective of expenditure, one of the main missions set by the State to the CNSA in its objectives and management agreement.

This project represented a 2-weeks work for development, tests (by Keyrus) and documentation writing, 1 week of tests (by the client) and 1-week of corrections, evolutions and launch.

| | | | Qté | Prix unitaire | Prix total HT | Prix total TTC |
|---|---|---|---|---|---|---|
| **Projet ARS – DIAMANT v3 Devis Évolutions Finance 2018 Lot 2** | CUBE3 | Réalisation d'un Cube SSAS complexe | 1 | 5,151 € | 5,151.00 € | 6,181.20 € |
| | LIV | Livraison en environnement REC / PROD | 2 | 678 € | 1,356.00 € | 1,627.20 € |
| | MMCD2 | Maintenance de composants Mise à jour base de données Moyen | 5 | 351 € | 1,755.00 € | 2,106.00 € |
| | MCD2 | Modèle conceptuel des données moyen | 3 | 2,660 € | 7,980.00 € | 9,576.00 € |
| | REC | Recette | 2 | 1,716 € | 3,432.00 € | 4,118.40 € |
| | SFD2 | Spécifications fonctionnelles détaillées moyennes | 1 | 4,305 € | 4,305.00 € | 5,166.00 € |
| | MSCRIPT2 | Maintenance de composants Mise à jour d'un script SQL Moyen | 8 | 346 € | 2,768.00 € | 3,321.60 € |
| | MCUBE2 | Maintenance de composants Mise à jour de cubes SSAS moyen | 3 | 376 € | 1,128.00 € | 1,353.60 € |
| | SQL2 | Réalisation d'une requête SQL moyen | 3 | 467 € | 1,401.00 € | 1,681.20 € |
| | STD2 | STD ODS / DWH moyen | 5 | 678 € | 3,390.00 € | 4,068.00 € |
| | STD3 | STD ODS / DWH complexe | 2 | 1,476 € | 2,952.00 € | 3,542.40 € |
| | STD4 | STD Datamarts / Cubes simple | 5 | 654 € | 3,270.00 € | 3,924.00 € |
| | STD5 | STD Datamarts / Cubes moyen | 2 | 1,702 € | 3,404.00 € | 4,084.80 € |
| | **TOTAL** | | | | **42,292.00 €** | **50,750.40 €** |

Figure 11. Cost charge for the Financial restructuration on DIAMANT

This cost represents:
- The construction of the tables in ODS and DWH based on their complexity;
- The construction of the cubes based on their complexity;
- The different redaction to deliver to the client to explain the taken actions;
- The tests;
- The differents deliveries and
- The global maintenance.

| | | | Qté | Prix unitaire | Prix total HT | Prix total TTC |
|---|---|---|---|---|---|---|
| **Projet ARS – DIAMANT v3 Devis Évolutions Finance 2018 Lot 1** | CUBE2 | Réalisation d'un Cube SSAS moyen | 1 | 3,027 € | 3,027.00 € | 3,632.40 € |
| | LIV | Livraison en environnement REC / PROD | 1 | 678 € | 678.00 € | 813.60 € |
| | MCD2 | Modèle conceptuel des données moyen | 1 | 2,660 € | 2,660.00 € | 3,192.00 € |
| | REC | Recette | 1 | 1,716 € | 1,716.00 € | 2,059.20 € |
| | SFD2 | Spécifications fonctionnelles détaillées moyennes | 1 | 4,305 € | 4,305.00 € | 5,166.00 € |
| | STD2 | STD ODS / DWH moyen | 2 | 678 € | 1,356.00 € | 1,627.20 € |
| | STD5 | STD Datamarts / Cubes moyen | 2 | 1,702 € | 3,404.00 € | 4,084.80 € |
| | **TOTAL** | | | | **17,146.00 €** | **20,575.20 €** |

Figure 12. Cost charge for HAPI

This cost represents the exact same things than the previous costing but at the level of HAPI only.

| | |
|---|---|
| **COMPLEXITY** | **LOW**<br>1 new file must be integrated (AAAA_Engagement-MM.xls - with YYYY for the year and MM for the month) containing all the data to be added in the Finance cube. |
| **SOLUTION** | • Data integration: this project does not require the creation of a new SSIS project, but the integration of new packages in the solution: SSIS_DIAMANT_COMMUN. 3 new packages were created: PAK_ODS_HAPI_MENS.dtsx, PAK_DWH_D_HAPI.dtsx, PAK_DWH_F_ACT_HAPI_MENS.dtsx.<br>• Cube and indicators: this project does not require the creation of a new SSAS project, but the integration of new indicators in the solution: SSAS_DIAMANT_COMMUN. |
| **ODS** | • Creation of table on SQL Server: only one table (ATIH_HAPI_MENS)<br>• Truncate/insert if and only if a file is detected<br>• No rule to applicate |
| **DWH** | • Creation of 6 new tables in DWH (5 dimensions: D_HAPI_ARBRE_ANALYTIQUE, D_HAPI_ACTION, D_HAPI_ACTION_TEMP, D_HAPI_PAYEUR, D_HAPI_PAYEUR_TEMP and 1 fact table : F_ACT_HAPI_MENS)<br>• Integration into DWH is done with application of specific rules and the implementation of lookup and an insert / update distinction. |
| **CUBE** | It was necessary to :<br>• Add the new tables to the DSV of the cube;<br>• Then in the cube;<br>• Add a measure group and a sum function measure;<br>• In the "Use dimensions" tab, perform the different crossings on the new measurement group;<br>• In the "Calculations" tab, create 2 indicators. |
| **TECHNICAL TESTS** | It must be ensured that the integrity of the data is ensured and that the management rules have been applied. These tests are done on Excel by a cube connection and the creation of a pivot table.<br>With finite developments and satisfactory non-regression tests, deadlines were met with delivery a week ahead. |
| **FUNCTIONAL TESTS** | Started on November 20, 2018<br><br>This test showed no problem of integration but a problem of restitution. |
| **CORRECTION/ EVOLUTION** | To correct it, a new time axis had to be added to the cube.<br><br>The data was attached to the single time axis of DIAMANT, but the HAPI data required the creation of a new time axis specific to HAPI. As it is, each HAPI file corresponds to a HAPI campaign and therefore to a campaign year, but the campaigns run over 2 years (X and X-1). Thus, the time axis of DIAMANT does not make it possible to understand which data belong to which campaigns. |
| **LAUNCH DATE** | The realization of this axis and its proposal in test validated the project HAPI which could be put into production on December 5, 2018 (in the respect of the deadlines) with data for the 3 campaigns 2015, 2016 and 2017. |

This phase of tests and deliveries allows the building of important “criticisms” on the project methodology and client’s implication.

- Firstly, DIAMANT - which is a big project bringing to Keyrus 3 million euros in 4 years - should implement more advanced and traceable testing methods.

- Secondly, the position of a services provider is complicated. It would be easier to be a technical resource with a functional vision of the project but it isn't possible to know each sector and to be a specialist of BI with sanitary or financial knowledge. Indeed, from a purely technical point of view, the developments did not show any malfunctions. But from a functional point of view - here financial -, the solution did not meet the needs. But this problem is not easily recognizable for a developer.

#### 5.2.1.3. OVALIDE

The client wishes to benefit from Keyrus services for the addition of new data into an existing cube called DETAILS (DETAILED in English). This cube is in the scope of the PMSI cube but with more detailed information.

This represents:

- The creation of tables in ODS and DWH (with their relations and dependencies);
- The loading of the ODS;
- The loading of the DWH:
- Modification of the cube (142 indicators to create, several to modify or to delete);
- Verification and unit testing phase;
- Assistance during the tests;
- The different deliveries;
- The creation and update of documentations regarding the cube.

This project represents a 3-weeks work for development, documentation writing and tests (by Keyrus) and 1 week of tests (by the client), corrections, evolutions and launch.

| | | | Qté | Prix unitaire | Prix total HT | Prix total TTC |
|---|---|---|---|---|---|---|
| **Projet ARS – DIAMANT v3 Devis OVALIDE** | SFD2 | Spécifications fonctionnelles détaillées moyennes | 1 | 4 305 € | 4 305,00 € | 5 166,00 € |
| | MCD2 | Modèle conceptuel des données moyen | 1 | 2 660 € | 2 660,00 € | 3 192,00 € |
| | STD2 | STD ODS / DWH moyen | 6 | 678 € | 4 068,00 € | 4 881,60 € |
| | CUBE2 | Réalisation d'un Cube SSAS moyen | 1 | 3 027 € | 3 027,00 € | 3 632,40 € |
| | ETL1 | Réalisation Alimentation simple | 5 | 113 € | 565,00 € | 678,00 € |
| | ETL2 | Réalisation Alimentation moyen | 6 | 422 € | 2 532,00 € | 3 038,40 € |
| | MMCD2 | Maintenance de composants Mise à jour base de données Moyen | 1 | 351 € | 351,00 € | 421,20 € |
| | LIV | Livraison en environnement REC / PROD | 1 | 678 € | 678,00 € | 813,60 € |
| | REC | Recette | 1 | 1 716 € | 1 716,00 € | 2 059,20 € |
| | | **TOTAL** | | | **19 902,00 €** | **23 882,40 €** |

Figure 13. Cost charge for OVALIDE

This cost represents the exact same things than the previous costing but at the level of OVALIDE only.

| | |
|---|---|
| **COMPLEXITY** | **LOW**<br><br>It is necessary to integrate 1 new file (appli_09.csv) containing all the data to be added in the cube Finance.<br><br>This file has to be loaded through an Access Service by calling the URL : https://epmsi.atih.sante.fr/appli_09.csv?year=2017&period=12&action=7&statut=1&champPmsi=1&arbustPassword=AbR45-678UWXzzzftR3edZe<br><br>The parameters are as follows:<br>• year: corresponds to the year<br>• period: corresponds to the transmission period (for MCO / HAD / SSR the allowed values are from 1 to 12 ; for the PSY the allowed values are: 3, 6, 9, 12)<br>• action: must take the constant value 7<br>• status: can take values 1 for exDG or DGF (Public) or 2 for exON or OQN (Private) |

| | |
|---|---|
| | • PMSI field: can take the multiple values (1 for the MCO / 2 for the SSR / 3 for the PSY and 5 for the HAD)<br>• arbustPassword must not be modified.<br><br>The source file contains a multitude of indicators (old - to keep or delete and new - to create and integrate). It is then necessary to create a parameter file (PMSI_DETAILS_PARAM.csv). This file is used to resume the indicators to load). |
| **SOLUTION** | • Data integration: this project does not require the creation of a new SSIS project, but the integration of new packages into the solution: SSIS_DIAMANT_COMMUN_PMSI.<br>New packages must be created (PAK_DEP_URL_ATIH_EPMSI.dtsx, PAK_ODS_PMSI_DETAILS_MENS.dtsx, PAK_PMSI_DEAILS_PARAM.dtsx and PAK_DWH_F_ACT_PMSI_DETAILS_MENS.dtsx).<br><br>• Cube and indicators: this project does not require the creation of a new SSAS project, but the integration of new indicators in the solution: SSAS_DIAMANT_COMMUN. |
| **ODS** | • Creation of the tables on SQL Server: 3 tables to build (PMSI_DETAILS_PARAM, ATIH_PMSI_DETAILS_MENS, ATIH_PMSI_MENS)<br>• Truncate/insert if and only if a file is detected<br>• No rule to applicate except specific filters |
| **DWH** | • Creation of 2 tables (D_PMSI_DETAILS_INDICATEURS and F_ACT_PMSI_DETAILS_MENS)<br><br>• Integration into DWH is done with application of specific rules and the implementation of lookup and an insert / update distinction. |
| **CUBE** | It was necessary to :<br>• Add the new tables to the DSV of the cube;<br>• Then in the cube;<br>• In the "Use dimensions" tab, make the different crossings on the new measurement group;<br>• Add group and in this measure group, add a sum function measure;<br>• In the "Calculations" tab, create one by one the 142 indicators and modify the 30 other ones. |
| **TECHNICAL TESTS** | With finite developments and satisfactory non-regression tests, deadlines were met with early delivery. |
| **FUNCTIONAL TESTS** | Started on October 15, 2018.<br><br>The test for this project has been very interesting from a purely organizational and non-technical point of view. First, it showed the need to create more indicators than expected. Indeed, of the 142 created, some had correspondences for ex-DG and ex-OQN fields, while others did not. |
| **CORRECTION/ EVOLUTION** | To correct it, it was necessary to create about twenty new indicators. |
| **LAUNCH DATE** | Launch on November 11, 2018. |

OVALIDE was in many aspects an interesting project. Firstly, this project has required a longer period of pure development (especially since the creation of indicators is very time-consuming despite the simplicity of the action). This enabled the intern to work more on development and less on TMA. Moreover, the needs of OVALIDE have highlighted certain components or functioning still unknown as the integration of a web service.

Secondly, and as for HAPI, the complexity of being technical and not having a functional expertise caused a problem for OVALIDE. This is problematic, and even more when the client shows signs of

disorganization or fails to analyze its own needs. It is even more problematic if Keyrus has to manage by itself on situation like that. Indeed, the addition of these indicators had the effect of mobilizing a resource for several days instead of beginning another project. In case like that, the corrections are considered by Keyrus to constitute evolution of the project scope, which are subjects of new command and payment. This is not necessarily understandable for the customer. To avoid too much conflict between the parties (and because DIAMANT is an historic project for Keyrus), the actions were made without being considered as evolutions (meaning that Keyrus itself paid for the time the developer had to pass on the project).

Finally, on OVALIDE, it happened that DIAMANT experienced a deactivation of the new project. The many indicators of the cube have undergone a restructuring. This was planned and desired by the client to make the cube clearer and in line with the many new indicators of OVALIDE. Still, this new architecture has severely disrupted end users whom are not the decision-makers of the projects. It turned out that they were not even aware that an action like this was going to take place, which changed their habit. The client has therefore requested the deactivation of the OVALIDE project and the reactivation of the previous cube. While this has little impact on Keyrus' side, it does appear that the customer has paid a multi-thousand-euro solution to see it applied only several months after the end of the creation.

### 5.2.2. Second task : Projects management

In September, DIAMANT's team was made of 2 project engineers, 1 junior developer, 1 maintenance ressource and the intern. Then both project engineer and the maintenance ressource quit Keyrus in December and January. DIAMANT was only composed of a junior developer suddently becoming the major knowledge source on the project and the intern. And it was planned that the intern when hired quit DIAMANT to begin GCS. To remplace the 4 elements that quit or changed project, Keyrus hired, at mid-time a project engineer, a junior developer, 2 apprentices and 2 interns. On the 6 new ressources, 4 have Data Science profile. Nevertheless, Data Science isn't an usefull competence for DIAMANT as the client hasn't yet agree to pay for this kind of analysis.

During the internship, it was well discussed between the parties that the the internship should have integrated data science tasks as discussed during the interview. The managers thought it can be a great idea to give an executive role to the newly emloyee. They thought that the intern having competences in data science and being interested by doing some during his time at Keyrus, it could be interesting to offer him a team to manage, knowing that the apprentice and intern of the team do need to present to their school a data science project. This team is composed of the mid-time project engineer, one of the apprentice and one of the intern. The managers gave free credits to the team to come with projects propositions using data science on DIAMANT. The proposition should be serious and hold huge interest for the client. It should be applicable with as least investment as possible in term of tools licenses. The team is working in autonomy under the control of the manager (Cloé Dorali) and with the assistance of an expert project engineer (Diane Forin).

*It was decided till the very beginning that all the propositions the data science team will come with will proposed the use of onlt the two following tools: R Studio and PowerBI. R would allow data analysis in statistical and graphical forms, so that a further understanding of the data and the phenomena that generate it can be presented. The development being too technical, the use of PowerBI would make this very technical work more meaningful for a customer who is not technic oriented. The visual representation is the most telling and so on the most selling and so the work on R must be completed by visuals. Moreover, both technologies are free so usable without investment. it IS also a way to present PowerBi reports to the client which is thinking of changing of restitution tool as the project is for now using SSRS and that Keyrus also desires to see emerge in their developers' domain of competences. During the first year of the Master, PowerBi was a studied tool. So the first mission of the manager was to form the members of the team. Yet, inconvenients of using PowerBI on DIAMANT exist. The database is too large; as a consequence, any connection of data on PowerBI takes a lot of time. PowerBI is also refreshing the reports at each change which leads to a performance problem in this specific case. Nevertheless, the client isn't really understanding these problems and despite the possibility to use Qlik, prefers PowerBI.*

## Cross chirurgy - demography analysis

The PMSI MCO cube relates to medical, surgical and obstetric procedures. This cube is one of the most consequent, the one with the largest fields of application and the more assured quality of the data. Moreover, it is the most used cube of the solution and it has the advantage of allowing predictions. It would be interesting to consider crossing the analysis of the PMSI MCO cube with an analysis of the DEMOGRAPHY cube relating to the geographical data of the medical staff. DIAMANT is a vast information system by volume but quite incomplete. This can be explained by two reasons: the imperative of confidentiality of data which reduce the number of variables available, and also the fact that this cube for example, is filled by professionals themselves when they think about filling it. Analyzing only demographic can produce an uncomplete and then not-reliable prediction. This is the main reason why a cross-MCO analysis with DEMOGRAPHY could be more interesting. The desirable results that the team is hoping for is that if similar phenomena are observable on both patients and staff. With this, the elements of conclusions and predictions would be easier (and more interesting) to highlight ; having in mind that the expected results of these analyzes are (a) highlighting groups of patients, employees and geographic areas that are similar in terms of behavior or appearance of phenomena. (b) And to try to understand the reason for these phenomena.

The analysis goes through each phase of Data mining process, from cleaning to knowledge extraction by respecting the cross-validation principle. On the train set are tested different descriptive models such as classifications, associations, principal components as the analysis looks for similarities between the studied populations. The Descriptive part being done, the Predictive part began with the application of predictive models such as linear regression and decision trees. Unsupervised methods will be used as the data is only the input data and no outputs are known. These outputs will be assumed and generalized to make conclusions and reports. The report on this cross analysis will serves as a Proof of Concept for the client (something that the client doesn't have express a need for but could be interested to implement and pay Keyrus for).

## Trajectories analysis

Parcours (pathways in english) is a cube whose order was made in May 2017. Since then the cube has been subject to evolutions. These lasts were launched in December 2018. The interest of this cube is that it makes it possible to identify the inter-establishments' trajectories of the patients.

And this, without limit of fields, establishments and regions where the care is provided and on a history of 8 years. In this, this cube is a central element of the challenge of building an HIS that the Ministry of Health wants to put in place. Indeed, unlike other cubes, focused on the elements of a stay (with an "activity" vision with pricing and valuation data). Parcours or "inter-institutions" focuses on the patient and his trajectories. A pathway is unique and generally includes at least one passage in an MCO Unit, to an SSR Unit of a different institution.

Because of the modern way of life, human health is degraded. Because of the ever-improving medicine, the longevity of man continues to grow. Both of these phenomena mean that care needs are growing as well. The Parcours cube can help to highlight these phenomena with numbers. In this context, it was chosen to conduct a study on this cube because it can be interesting to determine typical trajectories, action plans to apply to improve the existing (time taken, support, transfer, scheduling of stays, rehabilitation time, etc.). Thus, the Data Science team also meets the imperative of interest for the client as an analysis of this cube responds to social issues. Indeed, it is a health issue and goal to be able :

- To map the actual stages of the patient's trajectories throughout his or her health journey. This mapping must make it possible to identify the place and the way in which the care is carried out. Then, it must allow to note the level of effectiveness of the interventions but also to note the first assumption of responsibility and the conformity of the trajectories for the chronic pathologies;
- To identify the successive rehabilitation of a patient;
- To follow the progression of chronic diseases which implies a more complete care of the patients;
- And to note the possible optimizations on the support of the isolated events.

To do this, the team would:

- Analyze the orientation of the patients: it is an analysis at a departmental level taking into account the professionals' of health's behaviour concerning their decision of orientation for the patients regarding the patient's city of residence.
  - ⇨ With the assumption that geographic proximity is a factor that greatly influences the orientation of patients;
- To analyze the orientation of the patients by main diagnosis: if specific diseases emerge of the analysis this could mean that the institutions having identified the pathologies and that transferred the patients are not able to take care of subsequent care. And maybe need of ressources.
  - ⇨ Some institutions lack skills or space (saturation of services) to take care of patients suffering from certain pathologies (especially cancers) and for patients in need of palliative care or extreme recovery;
- To analyze the number of course breaks : a rupture is taking place when a patient has benefited from an act of surgery but not rehabilitation.
  - ⇨ The basic assumption is that breaks occur on minor surgical procedures. The interest is to find patients who do not fulfill this condition and who come back from this act for a re-hospitalization. This could highlight some procedure's lack.
- And finally, to analyze the number of re-hospitalizations for a given institution.
  - ⇨ With the expectation to highlight key establishments :
    - With many out-patients that come back ;
    - With many in-patients being send to them for care.

## Reporting evolution

As seen above, the client is interested by PowerBI. And as previously said, the SSRS reports aren't used by the clients. The file they use the most is OSEF which is an Excel file. It is known that the SSRS visualizations aren't really attractive. SSRS and Excel are both Microsoft tools and the visuals are really similar, so the users tends to use Excel to make their own. It is also true that SSRS isn't an user-friendly technology which means that Keyrus is building the reports but the users do not have the possibility to load one and modify it. It is this situation that leads the client to think about using another data visualization tool. As Microsoft is offering a more intuitive/user-frienly tool for data visualization with PowerBI, the goal is here is to take the reports and OSEF and to transform them on this tool. So the first phase is here to understand the data, to analyze the current visuals and to come with a proposition. The order gave about this proposition is that the proposition should come with:

- rethink/improved visuals (if a shape is more understandable or impactant then the present shape should be changed) ;
- visuals should be created (if some interesting visuals are missing, then they should be added, as well KPI are important for financial analysis, so KPI should be created).

Fullfilling this data visualization "rebuilding" task is time-consuming. If only taking OSEF, the rebuilding means to build more than 50 queries and 20 visuals. Moreover the resuming table is really complex.

The second phase of this report "rebuilding" project is to come with a predictive model of the trends that the financial situation of the Health organisation will tend to. This is think as the game-changing attribute of this project. Indeed, the model can be really interesting for the client so the model should be really controlled and nothing should be left aside. The determination of the applied process is still on-going. As said the visualizations are time-consumming. Moreover, as a master student, the person assigned to this task should have control on the process to put in place.

## Show me the tools

This project is a little bit particuliar. When working on GCS - as it will be present afterwards - I had to build a SSRS report. It was a first approach of the technology. Until then, the only known visualization tool was PowerBI. At this time, a client (Elegia : an urban and regional development company) that Keyrus works for, was looking for a presentation tool. On cases like that, the process is to present to the client all the technologies spectrum Keyrus proposes : Qlik, PowerBI, SSRS, and others. GCS is one of the only projects using Microsoft technologies and moreover, the only one to use the last version of SSRS. As the main developer of GCS, the Keyrus' managers approached me to ask if I could present the technology to the client. On the presentation date, the SSRS presentation was ready. But the commercial responsible of Elegia had a impromptu, the PowerBI presentator will not make it. It ended up with me doing the PowerBi presentation with a report found during the SSRS presentation. So I discovered the report at the same time as I had to explain it to Elegia's directors.

This situation leads to a reflexion. Why can't developers with appetences on the tools that Keyrus proposes, develop applications that will be given to the commercials so that situations like that didn't appear again. Indeed, it can be a domaging event to organize a presentation without presentators, presentation supports or badly prepared presentators. The commercial told that some "Show me <the tool>" are created when employees have time on some tool but none exists on SSRS and PowerBI. It is why it was decided that Show me PowerBI and Show me SSRS 2017 will be created as part of the data science team.

DIAMANT, which represent the majority of the internship, is a health project for the Ministry of Health in France. The project is today too vast to allow an intern to express himself and show real competences. The creativity of the intern isn't appreciable on this project and even when seeing bad methodologies in application, no corrections can be applicate as one small correction on a system of this size can have negative impacts on multiple dimensions et facts tables. The realized tasks allows one to understand that the intern is able to integrate a team and a project and take in charge missions in complete autonomy and with little corrections need after presentation to the client. As a plus, the intern shows managerial capacities which corresponds with the role of executive a master graduate can ask for. On the other side, this project isn't enough to show that the student is an an expert in its domain. So DIAMANT was indeed a good preparation project for a more complex one : GCS.

# 6. GCS

DIAMANT isn't an HIS. Nevertheless, its aspirations make it possible in the future to assimilate it with an HIS. On the contrary, GCS is thought in its construction as a real HIS with individual level data, health facility level data, population level data and public health surveillance level data as Almunawar and Anshari described a DSS should integrate (Almunawar & Anshari, 2012). Which is the reason why it is interesting to be able to see both projects, their differences and similarities. The client is the same (DIAMANT : all the RHA of France, GCS : one unique RHA in collaboration with a hospitals trust). The technologies used are also the same, except for the versions (DIAMANT : MS2012, GCS : MS2017). But the projects' philosophy are completely different as well as the goals, management and control measures.
DIAMANT is a project for health strategy management based on financial situation of the infrastructure when GCS is way more social-oriented. GCS' goal is to integrate EMR and to form a real HIS, which Regional Agencies were asked to develop by both France's administration and the European Union. It is indeed true that the other european countries are ahead of France in the integration of EMR. But *in fine* both projects have the aim to be merged together. So in fact, GCS serves as a test using only one RHA data in order to, in the future, be applicated to all the RHA.

## 6.1. PHASE 1. PROJECT INITIATION

In project management, the first phase is initiation. During this first phase, actors, goals and requirements are defined, needs are documented as well as the development strategy. Last but not least, a cost proposition is made.

### 6.1.1. Needs presentation

The GCS BFC wishes to set up a decision-making information system in order to benefit from a common reporting solution between the different regional e-health projects. The major challenge is to improve the management of the regional health policy and to be able to adapt the supply of care (by analyzing the patients' pathways, the ambulatory shift among others). Moreover, the project's administrators desire to change the weekly report solution. The objective is to greatly improve the quality of the monitoring and to offer to professional actors and structures a personalized reporting at the level of their geographical perimeter.

The main objectives of the project are:

- The automatic integration of the already available data sources and the progressive integration of new ones. The solution must provide a holistic view of the performance of institutions, departments, regions and so on, of the country, by cross-referencing data that comes from different sources.
- The implementation of a reporting solution using an adapted tool.

This centralized system will therefore serve both the members of the GCS, the RHA and their partners.

### 6.1.2. Keyrus' proposition

Keyrus proposed an architecture and its principles of functioning and also alimentation's principles. The proposition concerned the different points listed in the table below.

| Step | Description | Requirements |
|---|---|---|
| 1 | Selection and installation of the application material system | Sizing validation, order form issue. |
| 2 | Creation and management of reception and archive folders. sFTP and firewall configuration. | Protocol validation and signature of accords. |
| 3 | Applications' installation, server configuration, connections creation and scheduling. | |
| 4 | ETL's development on SQL Server SSIS | |
| 5 | Reports' development on SQL Server SSRS | |
| 6 | Test | |
| 7 | Solutions' launches | |

#### 6.1.2.1. System architecture and functioning

Regarding the technical identification of the application, the project has 3 machines with 3 bodies or environments (development, test and production). A description of the machines can be found below:

| Environment | DEV/TEST |
|---|---|
| OS | Windows 2016 Datacenter |
| CPU | 4 |
| RAM | 16Go |
| Disks | C:\100Go<br>D:\100Go<br>E:\100Go |

| Environment | PROD SSIS + DB | Portal + SSRS |
|---|---|---|
| OS | Windows 2016 Datacenter | Windows 2016 Datacenter |
| CPU | 4 | 4 |
| RAM | 16Go | 16Go |
| Disks | C:\100Go<br>D:\100Go<br>E:\100Go | C:\50Go<br>D:\50Go<br>E:\100Go |

In order to respect the rules of privacy and security, the project uses data anonymization and an IDS, a device that monitors a system for malicious activity violations. This IDS had recommendations that oblige the users to use a VPN with at most a connection lasting 4 hours. This VPN only enables users to access the database and the environments. This VPN when connected gives access to a secured portal that traced the users' activity on the environments. Regarding the reporting solution, the access is also restricted by the use of a SSO which needs a customization and an extraction of the client's AD. In order to ensure the project proper functioning, different backups are to be put in place. The backup politic is to weekly save a copy of the D disk. This disk is composed of a folder DATA (archive), a folder PROJECT where are all the deployed developments, a folder DATA_EXCHANGE where the FTP is putting the files to be loaded, a folder SOURCES with all the executables and a folder BACKUP. Only the backups of the 3 lasts months are kept.

| Creation of backups | Delete of oldest backups |
|---|---|
| **dbo.SP_BackupDatabase**<br>**@ DatabaseName = 'DWH',**<br>**@ FolderPath = 'D:\PRODBDD\BACKUP'**<br>**GO** | $Now = Get-Date<br>$TargerLocalFolder = "D:\PRODBDD\BACKUP\"<br><br>$Days = "95"<br>$LastWrite = $Now.AddDays(-$days)<br>$Files = get-childitem $TargetLocalFolder –include *.bak –recurse \| Where {$_.LastWriteTime –le"$LastWrite"} |

```
Foreach ($File in $Files) {if ($File) {Remove-item $File}}
```

Regarding the whole virtual machine, the backups are weekly realized by the hosting company.

### 6.1.2.2. Data integration

The project is divided in three parts. Each of them will be treated separately. The first one is called eTICSS. The data is coming from an organism called eTICSS and concerns patients trajectories, medical records and personal information. Keyrus charged around 85K for this development.

| Référence | Quantité | Description | P.U. HT | Prix total HT |
|---|---|---|---|---|
| | 1 | Spécifications générales/cadrage | 3 765 | 3 765 € |
| | 1 | Modèle conceptuel complexe | 4 600 | 4 600 € |
| | 3 | Spécification technique flux simple | 668 | 2 004 € |
| | 3 | Spécification technique flux moyen | 1 002 | 3 006 € |
| | 3 | Spécification technique flux complexe | 1 723 | 5 169 € |
| | 2 | Spécification technique rapport moyen | 1 032 | 2 064 € |
| | 1 | Spécification technique rapport complexe | 1 439 | 1 439 € |
| | 22 | Alimentation flux simple | 167 | 3 674 € |
| | 25 | Alimentation flux moyen | 598 | 14 950 € |
| | 18 | Alimentation flux complexe | 946 | 17 028 € |
| | 2 | Réalisation d'un rapport moyen | 900 | 1 800 € |
| | 1 | Réalisation d'un rapport complexe | 1 295 | 1 295 € |
| | 4 | Assistance à la recette | 1 255 | 5 020 € |
| | 3 | Livraison des composants | 1 669 | 5 007 € |
| | | | Total HT € | 70 821 € |
| | | | TVA | 14 164 € |
| | | | **Total TTC €** | **84 985 €** |

Figure 14. Cost charge for initiation and eTICSS development on GCS

This cost correspond to a 4 months involvment.

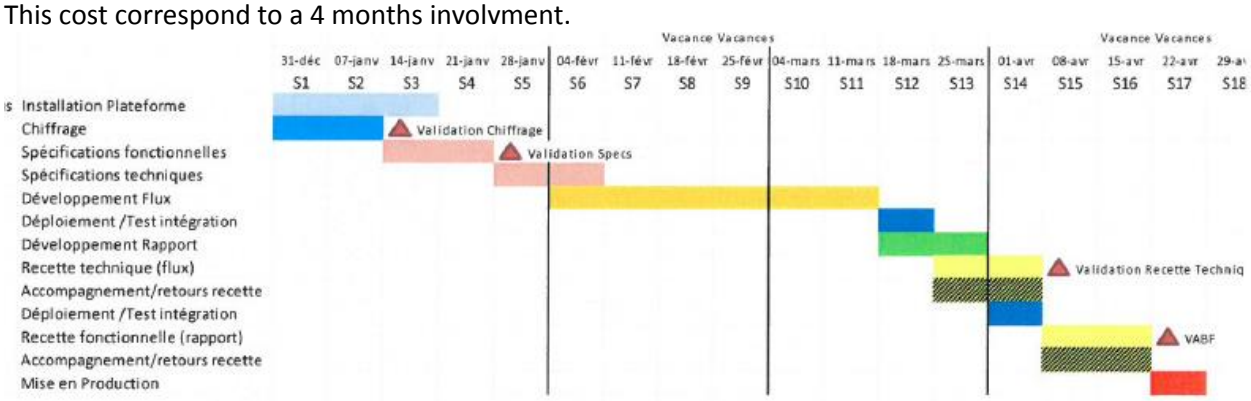



Figure 15. Charge planning for GCS' first batch

As the previous images showed it, the initiation of the project is beginning with platform installation, cost declaration and documentation redaction. As a service provider, Keyrus is obliged to document everything and to transmit it to the client. On each project, it is the developers' mission to write some of these documents:

- SQL, SSIS, SSRS norms: explain the norms to use on the development as naming and functionalities;
- SSIS, SSRS functional and technical specifications: explain in a functional or a technical way what the development puts in place.

One last document that is useful before the development. It is the data mapping. The client sent the input data and Keyrus build the model which is documented. The mapping shows a version of each table with notes, description, primary and foreign keys, look up and formula to apply for both the ODS and the DWH.

## 6.2. PHASE 2. DEVELOPMENT AND DATA INTEGRATION

### 6.2.1. Data warehouse construction

GCS also is a data-driven DSS. Only taking the first batch (eTICSS), the data are coming from three sources:

- DIAMANT (the project seen above) for geographical data. A SSIS solution for data extraction was created. This extraction produces 4 files;
- ROR (an organism) for organizational data (5 files);
- eTICSS (an organism) for data regarding the patients (19 files).

Based on these files was built a database which model (DWH) can be found on appendix 3. This database is not directly reachable by the final users. Meaning that users can't directly query the database with SQL queries nor can they see the base. Their only access to the information is through the report developed using SSRS. This report is for each RHA employees of the BFC region to use.

The loading process has been thought as follows:

- a solution, for the dimensions and facts tables common to the 3 batches of the project;
- a solution for each specific batch;
- each solution respects (by axis) this structure:
    - an ODS package: with a truncate / insert load;
    - a DWH package: an insert / update load with management rules;
    - a GLOBAL package: successively calls the execution of the ODS package and the DWH package.
- Even if load in the production environment is configured to use only the GLOBAL package, manual launches can be made, in this case they are only made by using the ODS and/or DWH packages;
- each package launch generates an execution ID (example for an ODS package):
  @[User::ID_EXECUTION] = (@[User::ID_EXECUTION] == "" ? "ODS_ROR_"+ REPLACE(REPLACE(REPLACE((DT_WSTR,4)DATEPART( "yyyy",GETDATE()) + "_" +(DT_WSTR,2)DATEPART( "mm",GETDATE() ) + "_" +(DT_WSTR,2)DATEPART( "dd",GETDATE() ) + "_" + (DT_WSTR,2)DATEPART( "hh",GETDATE()) + "_" +(DT_WSTR,2)DATEPART( "mi",GETDATE() ) + "_" +(DT_WSTR,2)DATEPART( "ss",GETDATE() ), "_", "_" ),"-", "" )," ","_") : @[User::ID_EXECUTION])
  This ID is written in an execution tracking table and is assigned to each rejection happening during the launch;
- the files are imported from the sFTP to the Keyrus files, then unzipped;
- before any truncation of tables, there is a check of presence of the files. All files must be present for the integration to begin;
- then a second series of tests is added to the process with the verification of the headers of each file. Wen agreeing on the format, the parties agreed that the files should not be modified without a first period of announcement and modification from Keyrus' part. Based on this agreements, a repository of the headers was saved in a table;
- at the end of the integration, the files are zipped, an archive is created with the name of the source and a timestamp and is moved to an archive area;
- if successful, the loading line entered at the beginning of the execution is updated with a status S (success) and an end date of execution; in case of failure, this same line is updated

with a status E (error). In both cases, an email is sent to the database manager with a message specifying the success or failure of the integration and in case of failure, the failure's reason.

### 6.2.2. Report construction

Report principles

| | |
|---|---|
| **Public** | Direction |
| **Access** | SSRS Portal |
| **Export** | Optimization for PDF export |
| **Data refresh** | Instant |
| **Data sources** | 2 : ODS and DWH (all the requests can be found on appendix 4) |
| **Logos** | |
| **Graphical charter** | |

| | | | |
|---|---|---|---|
| CMJN<br>C=10 M=60<br>J=100 N=0 | CMJN<br>C=66 M=44<br>J=32 N=0 | CMJN<br>C=30 M=100<br>J=62 N=26 | CMJN<br>C=0 M=0<br>J=0 N=70 |
| RVB<br>R=223 V=124<br>B=9 | RVB<br>R=83 V=181<br>B=180 | RVB<br>R=147 V=17<br>B=57 | RVB<br>R=112 V=113<br>B=115 |
| Hexadécimal<br>#DF7C09 | Hexadécimal<br>#53B5B4 | Hexadécimal<br>#931139 | Hexadécimal<br>#707173 |
| Pantone<br>145 PC | Pantone<br>3262 C | Pantone<br>216 C | Pantone<br>7540 C |

Report parameters

| Parameters | Definitions | Default values |
|---|---|---|
| **Beginning of the inclusion period** | Patient's consent collection date | 01/01/1900 |
| **End of the inclusion period** | | Date of the day |
| **Patient's region** | Region of attachment of the patient's postal code | All regions |
| **Patient's department** | Department of attachment of the patient's postal code | All departments |
| **Inclusion's region** | Region of attachment of the postal code of the institution where the patient was received | All regions |
| **Inclusion's department** | Department of attachment of the postal code of the institution where the patient was received | All departments |

Report presentation

| Characteristics | Value |
|---|---|
| **Distribution** | Confidential |
| **Number of pages** | 5 |
| **Number of graphical display** | 13 |
| **Time of development** | 15 days (with formation and corrections included) |
| **Formula** | Appendix 4 |

### 6.3. PHASE 3. TESTS AND LAUNCH

After the developments come the tests. For GCS, these tests are done on Squash, which allows a more efficient tracing and especially a more reassuring presentation for the customer. The tests were complicated to trace because of the late reception of the right files. Because of these delays, not all tests could be traced because other developments (including reports) were waiting. Nevertheless, the tests gave good results when they were done.
We have created specific ODS and DWH test campaigns (appendix 5). The former mainly checked the similarity of the source files with the database, while the latter mainly checked the correct application of the management rules. Another subtlety for the DWH tests, there was in this campaign, both basic integration tests (for the verification of mechanisms: writing of the execution ID, rejections, etc.) and specific tests on each management rules (including reject line accounts).

These developments took longer than expected for many reasons.
First, the source files have never been sent in the correct format. A good practice of GCS concerns interface contracts. Keyrus and the customer agreed on an extraction format. These formats are the basis of any integration, their change can cause problems, and so by doing interface contracts, Keyrus is preserving itself from any failure caused by wrong format. Despite the agreement, the extractions were not good and so it was impossible to test a real integration. The compliant files were received 3 days before the planned date of test by the client. As a first step, it was necessary to develop taking into account bad file formats and then update all file connections and correspondence tables, do the integration tests and corrections if necessary.
Secondly, the machines on which we operate are unstable and simultaneous development on two competing solutions has not been effective. Because of this lack of machine performance, many optimization problems were observed and required fixes recognized as optimizations knowing that at Keyrus, integration, data quality and optimization are different subjects and comply with different purchase orders.
Finally, many changes have been made especially in the sense of data quality (appendix 6). This was not planned as well, but it happens that during the test phase, the client realized that his own database was not correctly filled and thus the files were in an incomplete state and so was the integration. In terms of restitution, this gave very unsatisfactory results with a lot of data labelled as “not applicable” or regional filters that did not work. It was therefore necessary to evolve solutions by taking data quality measures by creating default values in all the join tables in order to link data to fictitious accounts. The solution was finally launched on May, 9$^{th}$. It is an efficient solution, both in terms of loading time and restitution of the data. It is true that mechanisms of improvement were set up before the hour but they allow one to highlight the little-perfectible character of the development. Yet, it might be interesting, however, to add a mechanism for dealing with rejects. This would go in the direction of a “Data Quality” logic.

As for the report, its complexity was not high, but it must be assumed that the technology was never studied or used before these reports were made. This has therefore generated a need of

training and longer development time. Moreover, the multiple changes to put in place for the integration happened at the same time that the report construction. The estimated charge was 6 working days. Actual development times are 15 working days. This led to a delivery date in receipts on April 18, 2019.
The report as presented does not have the same form as the report defines at first. Graphs proved to be preferred over others and many changes occurred. It was necessary to create the graphs 6, 7, 8 and 11 to modify the graphs 10 and 14 and to delete 4 graphs not presented. It was also necessary to go back to the SSIS developments to add a dimension serving as a repository for user authorizations (for graphs 7 and 8). Between the various modifications, waiting for approval and different phases of tests, the final delivery of the report was made on May 28, 2019.

Despite a complicated deadline and many necessary changes, the customer was satisfied with the deliveries. The GCS team has always tried to respond as quickly as possible and to manage problems with the greatest amplitude and efficiency. This earned us a very satisfactory performance rating since it ranks us among the top 5 customer satisfaction projects. Indeed, Keyrus at the signing of the contract proposes to the client a monthly project monitoring meetings during which the client will rate 5 performance indicators that are essential to him. These indicators concern the accompaniment of the client, adequacy of the implemented system, the functional and technical mastery of the subject, the project management, the involvement of the team, the quality of the service and the customer relationship. Keyrus has built a list of 29 indicators from which the customer must choose 5 to score from 0 to 5 and a global rating to rate from 0 to 10. In the last session, the GCS project was rated 9/10 (appendix 7).

Despite the launch of the solutions, this project isn't finished. Indeed, GCS is made of 3 major developments : eTICSS (that was presented above), Via Trajectoires and Télémédecine.

- "Via trajectoires" includes 3 main axes: old age, disability and rehabilitation. As well as a transverse axis of structures' habilitation and regional directories. It was decided that only the transversal axis as well as the disability scope would be first delivered by the end of July 2019. The 2 others will be developed afterwards.

- "Télémédecine" is both a migration and construction project. Indeed, the "Télémédecine" database is currently housed at INOVELAN for the Bourgogne region and QIMED for the Franche-Comté region. Since the two regions merged during the restructuring of the French regions in 2016, the databases have to merge. Therefore, it was decided that INOVELAN will delete the database they are hosting by 2021. This database is in MySQL, it should be replicated on Keyrus servers in SQL. At this, will be added all the other sources of Franche-Comté. This batch gathers telemedicine information, that is to say, data regarding medical consultation made remotely and practiced by certified doctors (an act to fight against medical desertification).

Integrating the GCS project was indeed a huge opportunity. It was really interesting to see closely the project management and even more to implement processes that now, beginning the second batch, I can see the great impact in efficiency terms. Indeed, during the first batch, it was necessary to install several dimensions common to all the batches and to familiarize with the data but more importantly, it was necessary to set up several data management, rejections and loads processes. In this, the future batches will be simpler because the main lines are all already created. Nevertheless, in terms of deadline and volume, they may be more complex.

# 7. DISCUSSION

All the work done, the people met, and the difficulties surmounted were great opportunities to learn a lot about myself: my capacities, limits and career projects but also the norms and methodologies that should be applicate in order to have the most industrialized project possible.

### 7.1.1. Enrichment

Working on these projects is highly interesting as the presented similarities and differences enable one to apply the courses studied while still learning.

During classes, MSBI was the technology used to build ETL. Having now 1 year of experience to add to the knowledge acquired in class, I feel more confident with my abilities on the subject. ETL wasn't my favorite subject in class and I was hoping to do something more Data Science oriented. Yet, today, ETL seems so much simpler than it was at school. An every-day basic work on the technology allows huge improvement. As a result, the company proposed to pay for a Microsoft Certification.

Social and professional skills: being the employee/tutor that they need

Being a student and being an employee are similar yet so different. Being a good student doesn't mean one is going to be a good employee and *vice versa*. As a true first work experience, the student learned to become a collaborator. The student should go out of the schema of a professor that will teach you something "without" waiting something from you as a proof of capacities. The student in a company is a workforce that should bring value to the company. It taught one: independence in the application, stubbornness in the resolution of problems, humility in the realization because there are still so many things to learn but also pride in the completion of projects and meeting of deadlines. As a real collaborator, one learns how to be a piece of a whole and to work sometimes alone and sometimes with other people, to receive instructions and advices but also to give some. This is truly different of the student between themselves despite the collaboration during projects because at school, students all have the same status and usually same experience and they tend to work with people they share affinities with. At work, no one has the choice of his collaborators, one must do efforts to sometimes maintain cordiality and develop respect, well-being and results.

Indeed, cordiality and respect are important between collaborators as a bad communication can lead to huge problems. It is even more important when a hierarchy exists. When beginning to work on GCS, I was an intern and I had to supervise and form on MSBI an experienced developer. He was tense at the beginning because he felt like an intern was giving order to him despite his experience. Regarding my feelings, I didn't felt at ease to be put above him. We had to take the best from each other: he had to understand that on this technology I was at my advantage compared to him and so he had to take my knowledge on the subject; when I had to take from him his experience and the reflection leading to a good development. It was easier when he felt like he wasn't underestimated and that his advice was appreciated. For weeks we developed our own respect and friendship. Once formed, we became "equals" on the subject. It was easier after hand to form an intern with no experience and who didn't meet me when I was myself an intern.

For GCS, I had to be the comprehensive supervisor, not fighting for recognition and respect. On the contrary, for DIAMANT, I had to be the strong tutor that will represent the interns in front of the managers. Being an intern isn't easy, especially when the company isn't offering you subjects that can be presented for the completion of your degree. Having been in this situation, the interns found in me someone to talk to, someone to help them in their redaction and in their defense, which led us to develop so many data science projects after all.

### Managerial skills: mastering your subjects

In a development, it is the developer who does it all. Yet, when the developer is manager or work with technical leaders or project managers that are highly efficient and have at heart to applicate only the best process, the development is nothing compared.
On DIAMANT, we were managed by successively 2 managers with no technical competences or knowledges. Regarding the process application, we were free. If our solutions are working then it is done, despite the fact that the solutions can be improved. This situation is really problematic as it leads to malfunction.
On GCS, the manager is completely aware of the data model, the existing junctures and rules we applicate. The model is clear in her. When present the result of the integration, she always was able to highlight problems in the integration. As for the tech leader, he was always capable to resolve problems or to critic our development by proposing better alternatives that take into account optimization, loading time, a possible future migration, etc.

The collaboration with people with these positions and knowledge taught me a lot not only regarding technics to apply in an ETL construction but also a lot regarding project management (which isn't a class of my master).

## 7.1.2. Difficulties

As for each experience/project, difficulties arose. It was the case for DIAMANT, GCS and even the internship.

### Keyrus' vision of an internship

Being a service provider, Keyrus, often is in a complex situation. To win contract, the company should be able to propose adequate profiles to the clients. To do so, it should have unstaffed resources available. Because of this, the company has a vision of internship and apprenticeship (but it is also the case for employees) that isn't meeting the needs of the interns or apprentices. Indeed, the company seeks to acquire competences on the long term and so, mainly focused on the multi-technologies' knowledges acquired by the candidate. Keyrus saw the internship as a way to test the intern before offering him a long-term contract on the company. This vision of the internship was a difficulty it itself as the intern has to focus on a specific mission to write a report on, but because the company sees beyond the internship, the missions weren't defined nor even did the company really had the need to find someone on DIAMANT at that time. During the internship, the hierarchy understood the need to implicate more the intern. So they decided to place the intern on another project as the lead developer enabling him to participate in the project management and to write a more substantial report. Nevertheless, I was willing to develop my abilities on Data sciences and not on Business Intelligence. I accepted this internship because at the interview, I was offered Data sciences tasks. Yet, I chose to finish the internship at Keyrus and prolonged my career within the company because doing Data science while staying at Lyon is almost not possible in France where only the capital does offer Data science projects.

### Being a service provider

With the tasks' diversity came problems specific to a service provider working with a public client: planning problems, cost problems, client problems. It is surprising to experience a situation in which the clients don't fully understand their own needs. It obliged the service provider (Keyrus) which isn't an expert in the functional domain of the project (here health) to determine by itself the needs in results and data. This situation leads to many problems as wrong interpretations and needs

expressions, long period of waiting for the client to accept the contract terms and deliver the right data in the right format. The waiting generates time gaps; gaps that have to be taken into account in the development of the project, its implementation and its cost. This is a real problem because Keyrus tends to keep the minimum latitude possible to the development of the project, in order to win contracts. And this strategy isn't always an easy one to hold when such gaps are to be taken into account but it is efficient to gain client. Yet wining and keeping clients are different matters. To keep them, Keyrus wants to build proximity with its clients through the RAD methodology. This methodology offers to the client an impression of involvement and possession over the built-solution. This possessiveness merged with the time gaps often leads to dissatisfaction and this even if the problems came from the client side.

Another problem existing in a client – service provider relationship is the misunderstanding of the impact caused by a change in the project scope or parameters (it was the case for OVALIDE). The projects evolve and often need new additions. And despite the fact that everything is possible; it can necessitate rework of existing and approved parts of the projects. The client not always understands the impacts because technical notions aren't so clear to them.

## Having multiple roles to play

Being a full-time employee and student was difficult. As an executive at Keyrus, the newly employee must fulfill the needs of the company on DIAMANT and GCS, having also to complete tasks of management on DIAMANT and collaborator training on Microsoft technology on GCS. The hierarchy saw a human side and human interaction need in its new employee and offer her to integrate a team of 3 people with the role of "workplace wellness" managers (which consist in the integration of the new collaborators/interns/apprentices and events referents). As a student, NOVA was waiting for its student the completion of a report but also in my case the completion of two new topics in the master program: text mining and deep learning. Wearing several hats when so many deadlines are present and so many things are in line is stressful and even confusing even more when these caps aren't compatible.

# 8. CONCLUSION

There are many topics that I have been able to apply and deepen during this last year at Keyrus and also many topics that are not part of the core learning of the Master in Data Science and that I could see from close and even put myself into application. After 1 year in business, I have a clear idea of the role I want to play in a company.

Today, the technologies and methodologies of building BI tools are what I master the best and I think that this will remain my specialization, but in the long term not as a developer but as a business engineer. So, I would now like to emancipate myself from MSBI and go to other ETLs. In the same way, I would like to go towards less obsolete restitution technologies than SSRS, like PowerBI which is a tool that customers are asking more nowadays and Qlik which is a very used tool. Thanks to this knowledge of several technologies and a personality that is firmly focused on customer needs and creation, the role of commercial / negotiator is more adapted to my profile.

I could see that by helping the commercial on ELEGIA. This is why after this master and a longer experience as a developer, I have as project to continue my studies on a degree in business engineering or if offered the confidence of my company, to access this role as a junior and to be formed on the field. The business engineer has a double commercial and project manager cap. It is a technical profile that makes the technological watch. It is also a creative profile that understands and tries to imagine the solution to put in place for a customer. But unlike the project manager, he negotiates the delivery and must win the contracts, so it is a strategist and a negotiator who must know how to assert his positions without forgetting the human behind the negotiations because he is also in the end the manager of a team and a project.

To conclude, despite my preference for data science, this business intelligence experience has been highly rewarding. It allowed me to implement some of my courses and in this, this internship is consistent with my degree. It allowed me to grow professionally speaking because it was my first real professional experience and it was not only done as a trainee but also as an employee. In addition, Keyrus is recognized in the sector as an excellent training center, which is a very interesting gateway to the sector. This experience also taught me a lot technically speaking and I will keep from this experience all the good practices and the need to go deeper into the theoretical knowledge of them. Finally this experience allowed me to discover that my interests are multiple and therefore to direct myself more precisely on my career project. All in all, Keyrus will have fulfilled my expectations and my needs. I hope to have fulfilled their own as an intern, as an employee, as a trainer and as a social manager. Moreover, Keyrus gave me some valuable inputs to fulfill a desire of mine by letting me work on a very interesting sector (Health) to me. Indeed by doing this Master at NOVA, I was expecting to one day work on Smart City subjects. Health is an important component of the Smart City/State program so it was really rewarding to work on both projects, to experience this proximity with the administration system of France and all the requirements to respect in terms of security and privacy. But also to understand what the state desires to put in place. And finally, to be able to say that the work done during this last year is helping the State to achieve a milestone in the construction of the French HIS and the application of a more global health management.

## BIBLIOGRAPHY


[1] Abouelmehdi, K., Beni-Hessane, A., & Khaloufi, H. (2018, January 9). *Big healthcare data: preserving security and privacy.* Retrieved from Journal of Big Data: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0110-7

[2] Almunawar, M., & Anshari, M. (2012). Health Information Systems (HIS): Concept and Technology. *CoRR*.

[3] Alter, S. L. (1980). Decision support systems: Current practice and continuing challenges. In S. L. Alter, *Decision support systems: Current practice and continuing challenges.* (p. 316 ). Addison-Wesley Publishing Co.

[4] Applied Informatics. (2018). *ETL, Extract, Transform and Load process concept*. Retrieved from Applied Informatics: http://blog.appliedinformaticsinc.com/etl-extract-transform-and-load-process-concept/

[5] Avison, D., & Taylor, V. (1997). Information systems development methodologies: a classification according to problem situation. *Journal of Information Technology*, 73 - 81.

[6] Bergamaschi, S., Beneventano, D., Mandreoli, F., Martoglia, R., Guerra, F., Orsini, M., Magnotta, L. (2018). *From Data Integration to Big Data Integration.*

[7] Berner, E., & La Lande, T. (n.d.). *Overview of Clinical Decision Support Systems.* Retrieved from Semantic scholar: https://pdfs.semanticscholar.org/8534/879f72139a6e78c1be749ff5f02cfa25bb2a.pdf?_ga=2.265111439.1666811625.1547635652-1808244430.1547217308

[8] Bonczek, R. H., Holsapple, C. W., & Whinston, A. (1981). *Foundations of Decision Support Systems.* Academic Press.

[9] Bourgeois, D. (2014). Information Systems for Business and Beyond. Saylor Foundation.

[10] Centers of Medicare and Medicaid services. (2008). *Selecting a development approach .*

[11] Chevreul, K., Durand-Zaleski, I., Bahram, S. B., & al., i. a. (2014). France: Health System Review. In *vol. 12, no. 6.*

[12] Coiera, E. (2003). *The Guide to Health Informatics (2nd Edition).* London: Arnold.

[13] Dum, S. (2017, August 08). *La bonne santé des SSII en France*. Retrieved from LesEchos.fr: https://www.lesechos.fr/08/08/2017/LesEchos/22503-071-ECH_la-bonne-sante-des-ssii-en-france.htm

[14] Durand-Zaleski, I. (2016). *French Health Care System*. Retrieved from French Property: https://www.french-property.com/guides/france/public-services/health/system-overview#1.2.

[15] *ETL concepts*. (2018). Retrieved from Learndatamodeling: https://learndatamodeling.com/blog/etl-concepts/

[16] Guazzelli, A. (2012). *Predicting the future , part 1 : what is predictive analytics ?*

[17] Hagerty, J. (2016, October 13). *2017 Planning Guide for Data and Analytics.* Retrieved from Gartner: HTTPS://WWW.GARTNER.COM/BINARIES/CONTENT/ASSETS/EVENTS/KEYWORDS/CATALYST/CATUS8/2017_PLANNING_GUIDE_FOR_DATA_ANALYTICS.PDF

[18] Halper, F. (2014). *Predictive analytics for business advantage.* Tdwi best practice report.

[19] Haux, & Reinhold. (2006). Health Information Systems - Past, Present, Future. *International journal of medical informatics*, 268-81.

[20] Haux, R. (2002). Health care in the information society: what should be the role of medical informatics? *Methods Inf Med*.

[21] Holsapple, C. W. (2008). *DSS Architecture and Types.* Retrieved from ResearchGate: https://www.researchgate.net/publication/226283106_DSS_Architecture_and_Types

[22] IBM. (2013). *Data-driven healthcare organizations use big data analytics for big gains.* IBM.

[23] Kathiravelu, P., Sharma, A., Galhardas, H., Van Roy, P., & Veiga, L. (2018). *On-demand big data integration. Distributed and Parallel Databases.*

[24]Kawamoto, K., Houlihan, C., Balas, E., & Lobach, D. (2005). *Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success.* BMJ.
[25]Keyrus. (2016). *Projet ARS DIAMANT - Livret d'accueil.*
[26]Keyrus. (2018). *About Keyrus*. Retrieved from Keyrus: http://www.keyrus.com/en/about-keyrus/
[27]Kim, S.-H., Kim, N.-U., & Chung, T.-M. (2013). Attribute relationship evaluation methodology for big data security. *Iinternational conference on IT convergence and security (ICITCS)* (pp. 1 - 4). IEEE : https://doi.org/10.1109/icitcs.2013.6717808.
[28]Kuhn, K., & Giuse, D. (2001). From Hospital Information Systems to Health Information Systems. *Methods Inf Med*, 275-287.
[29]Marin, G. (2008). Decision support systems. Bucharest, Romania: Romanian American University - Faculty of Computer Science for Business Management.
[30]Nay, O., Béjean, S., Benamouzig, D., & al., e. (2016, May 28). Achieving Universal Health Coverage in France: Policy Reforms and the Challenge of Inequalities. *Lancet*, pp. 2236–49.
[31]Ninghui, L., Tiancheng, L., & Suresh, V. (2007). t-Closeness: privacy beyond k-anonymity and L-diversity. *Data engineering (ICDE) IEEE 23rd international.*
[32]Protti, D., Johansen, I., & Perez-Torres, F. (2009). *Comparing the application of Health Information Technology in primary care in Denmark and Andalucía, Spain.* Elsevier.
[33]Rashidi, M., Ghodrat, M., Samali, B., & Mohammadi, M. (2018). *Decision Support Systems.* IntechOpen.
[34]Smith R. (1997). *The future of healthcare systems.* BMJ (Clinical research ed.).
[35]Sweeney, L. (2002). K-anonymity: a model for protecting privacy. . Int J Uncertain Fuzziness.
[36]Turban, E., & Aronson, J. E. (2004). *Decision Support Systems and Intelligent Systems.* Prentice-Hall, Inc. .
[37]World Health Organization. (2008). *Toolkit on monitoring health systems strengthening - Health Information System.*
[38]Xu, L., Jian, g. C., Wang, J., Yuan, J., & Ren, Y. (2014). *Information security in big data: privacy and data mining.* Retrieved from Doi: https://doi.org/10.1109/ACCESS.2014.2362522.
[39]Yazan, A., Yong, W., & Raj Kumar, N. (2015). Big data life cycle: threats and security model. *21st Americas conference on information systems.*

## APPENDIX

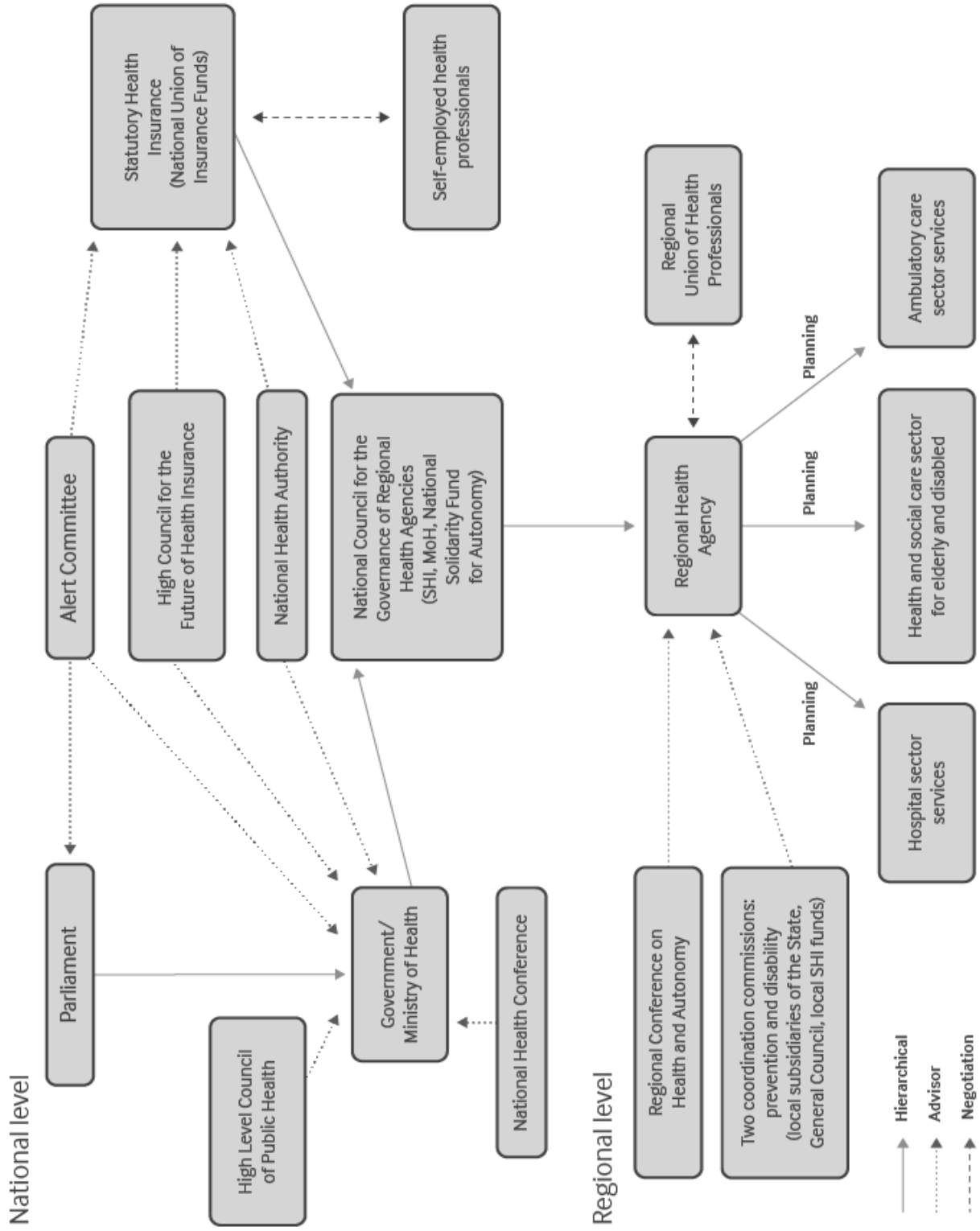
### *Appendix 1 – Technologies and applications list*

Last update: 18/02/2019

| Technologies or applications used | References |
|---|---|
| **Excel 2010 / 2013** | 2010 : https://products.office.com/en-us/previous-versions/microsoft-excel-2010<br>2013 : https://products.office.com/en-us/previous-versions/microsoft-excel-2013 |
| **MS 2008 R2 / 2012** | 2008 R2 : https://www.microsoft.com/en-us/download/details.aspx?id=26113<br>2012 : https://www.microsoft.com/en-us/download/details.aspx?id=29062 |
| **MS 2017** | https://www.microsoft.com/en-us/sql-server/sql-server-2017 |
| **PowerBI** | https://powerbi.microsoft.com/en-us/ |
| **Qlikview** | https://www.qlik.com/lp/ppc/qlik-sense-desktop/QV?CampaignID=701D0000001qoPV&ppc_id=&kw=qlikview&utm_content=s_pcrid_318516269645_pmt_e_pkw_qlikview_pdv_c_mslid__pgrid_65579563612_ptaid_kwd-1400175604&utm_source=google&utm_medium=cpc&utm_campaign=QlikView+Ad+Copy+Test&utm_term=qlikview&gclid=Cj0KCQiApbzhBRDKARIsAIvZue-yNfHjHtjbXBhK6YMdWnrzQQoVEBbn7hKMiMCedcC-OU8pAExbp7oaApW8EALw_wcB |
| **R – RStudio** | https://www.rstudio.com/ |
| **SharePoint** | https://sharegate.com/products/sharegate-desktop?gclid=Cj0KCQiApbzhBRDKARIsAIvZue9SjtuRbQZpDXbcWjWTaWD7twXPkQfDUgkfSWaW_-Us8XJZI-VUUcUaAncKEALw_wcB |
| **Squash (TM)** | https://www.squashtest.org/ |
| **WS 2008 R2 / 2016 Essential** | 2008 R2 : https://www.microsoft.com/en-us/download/details.aspx?id=11093<br>2016 Essential : https://docs.microsoft.com/en-us/windows-server-essentials/get-started/what-s-new |

Appendix 1. List of the used technologies and applications on the projects

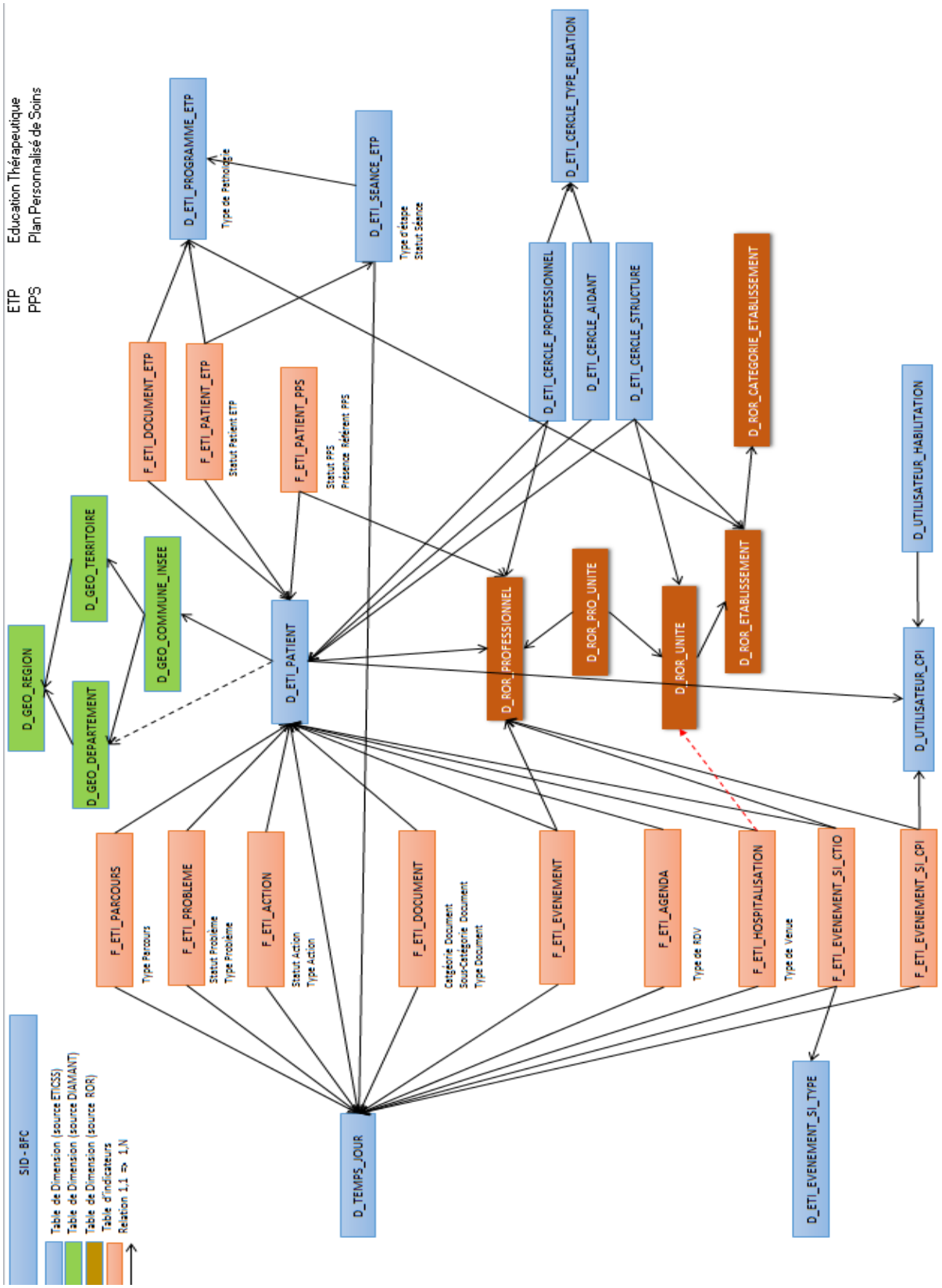
## *Appendix 2 [4] – French Health Administration's organization*




Appendix 2. Organization of the french Health Administration

[4] Adapted from K. Chevreul, I. Durand-Zaleski, S. B. Bahrami and al., "France: Health System Review " Health System in Transition, vol. 12, no. 6, 2010.

## *Appendix 3 – Data model of eTICSS*




Appendix 3. ETICSS' data model

## *Appendix 4 – eTICSS report's requests*

```
-- Dts_PatientGeneral --
SELECT
pat.ID_PATIENT,
pat.STATUT_DOSSIER,
pat.STATUT_FILE_ACTIVE,
pat.STATUT_PPS,
pat.STATUT_SITUATION_SOCIALE
FROM  [dbo].[D_ETI_PATIENT] pat
left join [dbo].[D_ROR_PRO_UNITE] pu on pu.ID_ROR_PROFESSIONNEL =
pat.ID_ROR_PROFESSIONNEL and pu.FLAG_PRINCIPAL = 1
left join [dbo].[D_ROR_UNITE] u on u.ID_ROR_UNITE = pu.ID_ROR_UNITE
left join [dbo].[D_ROR_ETABLISSEMENT] e on e.ID_ROR_ETABLISSEMENT =
u.ID_ROR_ETABLISSEMENT
left join [dbo].[D_GEO_DEPARTEMENT] d on d.ID_DEPARTEMENT = pat.ID_DEPARTEMENT
left join [dbo].[D_GEO_DEPARTEMENT] dept on dept.ID_DEPARTEMENT = e.ID_DEPARTEMENT
left join [dbo].[D_GEO_REGION] r on r.ID_REGION = d.ID_REGION
left join [dbo].[D_GEO_REGION] reg on reg.ID_REGION = dept.ID_REGION
WHERE pat.DATE_PREMIER_CONSENTEMENT between
CONVERT(DATE,@ParamPeriodeInclusionDeb,103) and
CONVERT(DATE,@ParamPeriodeInclusionFin,103) and pat.ID_DEPARTEMENT in
(@ParamDeptPatient) and e.ID_DEPARTEMENT in (@ParamDeptPro)

-- Dts_PatientCumul --
SELECT
ID_MOIS,
MOIS_LIB,
NB_PATIENT,
SUM(NB_PATIENT) OVER(ORDER BY ID_MOIS ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT
ROW) AS NB_PATIENT_CUMUL
FROM (
  SELECT
         CASE WHEN T.ID_MOIS <= 201612 THEN 201612 WHEN T.ID_MOIS <= 201712 THEN
201712 ELSE T.ID_MOIS END AS ID_MOIS,
         CASE WHEN T.ID_MOIS <= 201612 THEN 'Déc' WHEN T.ID_MOIS <= 201712 THEN
'Déc'ELSE T.MOIS_LIBELLE_COURT END + '-' + CAST(T.ANNEE AS NVARCHAR(4)) AS MOIS_LIB,
         COUNT(ID_PATIENT) AS NB_PATIENT
  FROM [dbo].[D_ETI_PATIENT] P
inner join [dbo].[D_TEMPS_JOUR] T ON T.ID_JOUR = P.ID_JOUR
inner JOIN [dbo].[D_ROR_PRO_UNITE] pu on pu.ID_ROR_PROFESSIONNEL =
P.ID_ROR_PROFESSIONNEL and pu.FLAG_PRINCIPAL = 1
inner join [dbo].[D_ROR_UNITE] u on u.ID_ROR_UNITE = pu.ID_ROR_UNITE
inner join [dbo].[D_ROR_ETABLISSEMENT] e on e.ID_ROR_ETABLISSEMENT =
u.ID_ROR_ETABLISSEMENT
inner join [dbo].[D_GEO_DEPARTEMENT] d on d.ID_DEPARTEMENT = P.ID_DEPARTEMENT
inner join [dbo].[D_GEO_DEPARTEMENT] dept on dept.ID_DEPARTEMENT = e.ID_DEPARTEMENT
inner join [dbo].[D_GEO_REGION] r on r.ID_REGION = d.ID_REGION
inner join [dbo].[D_GEO_REGION] reg on reg.ID_REGION = dept.ID_REGION
WHERE [DATE_PREMIER_CONSENTEMENT] BETWEEN CONVERT(DATE,@ParamPeriodeInclusionDeb,103)
AND CONVERT(DATE,@ParamPeriodeInclusionFin,103) AND P.ID_DEPARTEMENT in
(@ParamDeptPatient) and e.ID_DEPARTEMENT in (@ParamDeptPro)
GROUP BY
       CASE WHEN T.ID_MOIS <= 201612 THEN 201612 WHEN T.ID_MOIS <= 201712 THEN 201712
ELSE T.ID_MOIS END,
       CASE WHEN T.ID_MOIS <= 201612 THEN 'Déc' WHEN T.ID_MOIS <= 201712 THEN 'Déc'ELSE
T.MOIS_LIBELLE_COURT END + '-' + CAST(T.ANNEE AS NVARCHAR(4))
) F
ORDER BY ID_MOIS

-- Dts_TOP_Structures --
SELECT TOP (10)
```

```
etab.ID_ROR_ETABLISSEMENT,
etab.NOM_ETABLISSEMENT,
count(*) as nb_patient
FROM D_ETI_PATIENT f, D_ROR_PROFESSIONNEL d, d_ror_pro_unite unite, D_ROR_UNITE U,
D_ROR_ETABLISSEMENT ETAB
WHERE f.ID_ROR_PROFESSIONNEL=d.ID_ROR_PROFESSIONNEL
and d.ID_ROR_PROFESSIONNEL= unite.ID_ROR_PROFESSIONNEL
and unite.ID_ROR_UNITE = U.ID_ROR_UNITE
and U.ID_ROR_ETABLISSEMENT= ETAB.ID_ROR_ETABLISSEMENT
and etab.ID_DEPARTEMENT in (@ParamDeptProfessionnel) and FLAG_PRINCIPAL = 1 and
f.ID_DEPARTEMENT in (@ParamDeptPatient) and f.DATE_PREMIER_CONSENTEMENT BETWEEN
CONVERT (DATE,@ParamDebInclusion, 103)  and CONVERT (DATE,@ParamFinInclusion, 103) and
NOM_ETABLISSEMENT!= 'ETABLISSEMENT FICTIF INTEGRATION'
GROUP BY etab.ID_ROR_ETABLISSEMENT, etab.NOM_ETABLISSEMENT
ORDER BY count(*) desc

-- Dts_Nb_Utilisateurs --
SELECT [ID_UTILISATEUR]
  FROM [dbo].[D_UTILISATEUR_CPI]

-- Dts_TOP_CONNEXION_PROFIL --
-- Dts_TOP_CONNEXION_DISTINCT_UTILS --
-- See line with comments to compare the two datasets
select
HAB_CNX,
       ID_MOIS,
       MOIS,
       NB_CONNEXIONS,
       NB_CONNEXIONS_M1
from
 (
SELECT
ref_hab.HAB_CNX,
j.ID_MOIS,
j.MOIS_LIBELLE_COURT + '-' + CAST(j.ANNEE AS nvarchar(4)) AS MOIS,
CASE
       WHEN j.ID_MOIS = substring(convert(varchar, getdate(), 112), 1, 6) - 1 THEN 'M'
       ELSE 'M-1'
END AS FLAG,
count(*) as NB_CONNEXIONS,
row_number() over (partition by j.ID_MOIS order by count(*) desc) as rank,
LAG(count(*)) over (partition by ref_hab.HAB_CNX  order by j.ID_MOIS) as
NB_CONNEXIONS_M1 -- (Dts_TOP_CONNEXION_PROFIL)
LAG(count(distinct e.ID_UTILISATEUR)) over (partition by ref_hab.HAB_CNX  order by
j.ID_MOIS) as NB_CONNEXIONS_M1 -- (Dts_TOP_CONNEXION_DISTINCT_UTILS)
  FROM F_ETI_EVENEMENTS_SI_CPI e
  inner join [dbo].[D_UTILISATEUR_CPI] util on util.ID_ROR_PROFESSIONNEL =
e.ID_ROR_PROFESSIONNEL
  inner join [dbo].[D_UTILISATEUR_HABILITATION] ut_hab on ut_hab.ID_UTILISATEUR =
util.ID_UTILISATEUR and HABILITATION not in ('ET_SupportProxival', 'Administrateurs',
'ET_Stats', 'ET_AdministrateurETICSS')
  inner join [dbo].[D_ETI_REFERENTIEL_HABILITATION_UTILISATEUR] ref_hab on
ref_hab.HAB_TYPE = ut_hab.HABILITATION
  inner join D_TEMPS_JOUR j on j.ID_JOUR = e.ID_JOUR
  WHERE TYPE_EVENEMENT_SI = 'Login' and (substring(convert(varchar, getdate(), 112),
1, 6) - 2 = j.ID_MOIS or substring(convert(varchar, getdate(), 112), 1, 6) - 1 =
j.ID_MOIS)
GROUP BY ref_hab.HAB_CNX, j.ID_MOIS, j.MOIS_LIBELLE_COURT + '-' + CAST(j.ANNEE AS
nvarchar(4)),
CASE
       WHEN j.ID_MOIS = substring(convert(varchar, getdate(), 112), 1, 6) - 1 THEN 'M'
       ELSE'M-1'
```

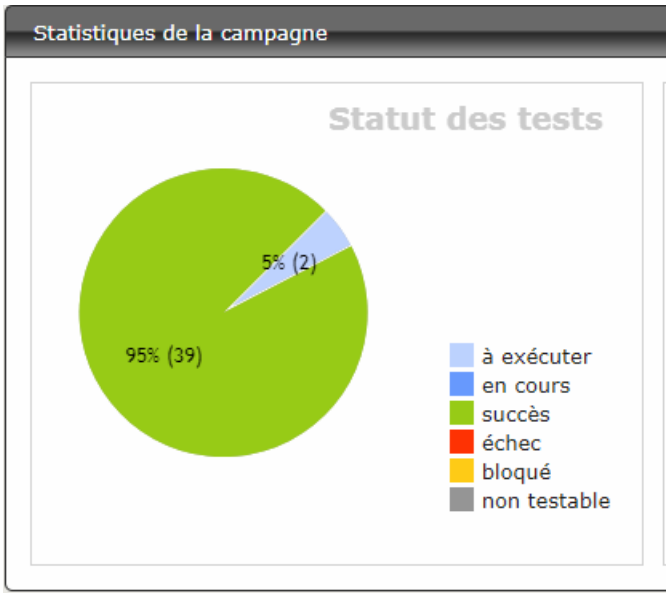
```
END
) F
WHERE rank <= 6 and FLAG = 'M'

-- Dts_Demarches --
SELECT
count(*) as 'nb_patients',
nb_parcours_patient,
count(*) * nb_parcours_patient AS AVG_NB_DEMARCHE
FROM (
SELECT
par.ID_PATIENT,
count(par.TYPE_DE_PARCOURS) as 'nb_parcours_patient'
FROM [dbo].[F_ETI_PARCOURS] par
left join [dbo].[D_ETI_PATIENT] pat on pat.ID_PATIENT = par.ID_PATIENT
left JOIN [dbo].[D_ROR_PRO_UNITE] pu on pu.ID_ROR_PROFESSIONNEL =
pat.ID_ROR_PROFESSIONNEL and pu.FLAG_PRINCIPAL = 1
left join [dbo].[D_ROR_UNITE] u on u.ID_ROR_UNITE = pu.ID_ROR_UNITE
left join [dbo].[D_ROR_ETABLISSEMENT] e on e.ID_ROR_ETABLISSEMENT =
u.ID_ROR_ETABLISSEMENT
left join [dbo].[D_GEO_DEPARTEMENT] d on d.ID_DEPARTEMENT = pat.ID_DEPARTEMENT
left join [dbo].[D_GEO_DEPARTEMENT] dept on dept.ID_DEPARTEMENT = e.ID_DEPARTEMENT
left join [dbo].[D_GEO_REGION] r on r.ID_REGION = d.ID_REGION
left join [dbo].[D_GEO_REGION] reg on reg.ID_REGION = dept.ID_REGION
left join [dbo].[D_TEMPS_JOUR] tps on  tps.ID_JOUR = pat.ID_JOUR
WHERE tps.JOUR between CONVERT(DATE,@ParamPeriodeInclusionDeb,103)  and
CONVERT(DATE,@ParamPeriodeInclusionFin,103)  and pat.ID_DEPARTEMENT in
(@ParamDeptPatient) and e.ID_DEPARTEMENT in (@ParamDeptProfessionnel)
GROUP BY par.ID_PATIENT
) a
GROUP BY nb_parcours_patient
ORDER BY nb_parcours_patient

-- Dts_VoletSocial_Renseignement --
SELECT
'GIR' AS CATEG,
SUM(CASE WHEN GIR IS NULL THEN 0 ELSE 1 END) AS COUNT_VALUE,
Count(1) AS COUNT_TOTAL,
100.00* sum(CASE WHEN GIR is null THEN 0 else 1 end) / count(*) AS PROP_VALUE
FROM D_ETI_PATIENT pat
left JOIN [dbo].[D_ROR_PRO_UNITE] pu on pu.ID_ROR_PROFESSIONNEL =
pat.ID_ROR_PROFESSIONNEL and pu.FLAG_PRINCIPAL = 1
left join [dbo].[D_ROR_UNITE] u on u.ID_ROR_UNITE = pu.ID_ROR_UNITE
left join [dbo].[D_ROR_ETABLISSEMENT] e on e.ID_ROR_ETABLISSEMENT =
u.ID_ROR_ETABLISSEMENT
left join [dbo].[D_GEO_DEPARTEMENT] d on d.ID_DEPARTEMENT = pat.ID_DEPARTEMENT
left join [dbo].[D_GEO_DEPARTEMENT] dept on dept.ID_DEPARTEMENT = e.ID_DEPARTEMENT
left join [dbo].[D_GEO_REGION] r on r.ID_REGION = d.ID_REGION
left join [dbo].[D_GEO_REGION] reg on reg.ID_REGION = dept.ID_REGION
WHERE pat.DATE_PREMIER_CONSENTEMENT between
CONVERT(DATE,@ParamPeriodeInclusionDeb,103) and
CONVERT(DATE,@ParamPeriodeInclusionFin,103)  and pat.ID_DEPARTEMENT in
(@ParamDeptPatient) and e.ID_DEPARTEMENT in (@ParamDeptProfessionnel)

UNION ALL

SELECT
'Assurances sociales' AS CATEG,
SUM(CASE WHEN ASSURANCES_SOCIALES IS NULL THEN 0 ELSE 1 END) AS COUNT_VALUE,
Count(1) AS COUNT_TOTAL,
100.00* sum(CASE WHEN ASSURANCES_SOCIALES is null THEN 0 else 1 end) / count(*) AS
PROP_VALUE
```

```
FROM D_ETI_PATIENT pat
left JOIN [dbo].[D_ROR_PRO_UNITE] pu on pu.ID_ROR_PROFESSIONNEL =
pat.ID_ROR_PROFESSIONNEL and pu.FLAG_PRINCIPAL = 1
left join [dbo].[D_ROR_UNITE] u on u.ID_ROR_UNITE = pu.ID_ROR_UNITE
left join [dbo].[D_ROR_ETABLISSEMENT] e on e.ID_ROR_ETABLISSEMENT =
u.ID_ROR_ETABLISSEMENT
left join [dbo].[D_GEO_DEPARTEMENT] d on d.ID_DEPARTEMENT = pat.ID_DEPARTEMENT
left join [dbo].[D_GEO_DEPARTEMENT] dept on dept.ID_DEPARTEMENT = e.ID_DEPARTEMENT
left join [dbo].[D_GEO_REGION] r on r.ID_REGION = d.ID_REGION
left join [dbo].[D_GEO_REGION] reg on reg.ID_REGION = dept.ID_REGION
WHERE pat.DATE_PREMIER_CONSENTEMENT between
CONVERT(DATE,@ParamPeriodeInclusionDeb,103) and
CONVERT(DATE,@ParamPeriodeInclusionFin,103)  and pat.ID_DEPARTEMENT in
(@ParamDeptPatient) and e.ID_DEPARTEMENT in (@ParamDeptProfessionnel)

----- UNION ALL 5 others categories

-- Dts_Patient_Age --
SELECT
f.ID_PATIENT
,f.ANNEE_NAISSANCE
,substring (convert (nvarchar(10), GETDATE(), 102), 1, 4) as GETDATE_SUBSTRING_OF_YEAR
,substring (convert (nvarchar(10), GETDATE(), 102), 1, 4) - ANNEE_NAISSANCE as AGE
,CASE
       WHEN substring (convert (nvarchar(10), GETDATE(), 102), 1, 4) - ANNEE_NAISSANCE
< 18 THEN '1 - Moins de 18 ans'
       WHEN substring (convert (nvarchar(10), GETDATE(), 102), 1, 4) - ANNEE_NAISSANCE
>= 18 AND substring (convert (nvarchar(10), GETDATE(), 102), 1, 4) - ANNEE_NAISSANCE
<= 50 THEN '2 - Entre 18 et 50 ans'
       WHEN substring (convert (nvarchar(10), GETDATE(), 102), 1, 4) - ANNEE_NAISSANCE
>= 51 AND substring (convert (nvarchar(10), GETDATE(), 102), 1, 4) - ANNEE_NAISSANCE
<= 75 THEN '3 - Entre 51 et 75 ans'
       WHEN substring (convert (nvarchar(10), GETDATE(), 102), 1, 4) - ANNEE_NAISSANCE
>= 76 AND substring (convert (nvarchar(10), GETDATE(), 102), 1, 4) - ANNEE_NAISSANCE
<= 85 THEN '4 - Entre 76 et 85 ans'
       WHEN substring (convert (nvarchar(10), GETDATE(), 102), 1, 4) - ANNEE_NAISSANCE
>= 86 AND substring (convert (nvarchar(10), GETDATE(), 102), 1, 4) - ANNEE_NAISSANCE
<= 95 THEN '5 - Entre 86 et 95 ans'
       ELSE '6 - Plus de 95'
END AS AGE_CATEGORIE
FROM
D_ETI_PATIENT f, D_ROR_PROFESSIONNEL d, d_ror_pro_unite unite, D_ROR_UNITE U,
D_ROR_ETABLISSEMENT ETAB, D_GEO_DEPARTEMENT DEPTpro, D_GEO_DEPARTEMENT DEPTpat,
D_GEO_REGION r, D_GEO_REGION reg
WHERE f.ID_ROR_PROFESSIONNEL=d.ID_ROR_PROFESSIONNEL
and d.ID_ROR_PROFESSIONNEL= unite.ID_ROR_PROFESSIONNEL and FLAG_PRINCIPAL=1
and unite.ID_ROR_UNITE = U.ID_ROR_UNITE
and U.ID_ROR_ETABLISSEMENT= ETAB.ID_ROR_ETABLISSEMENT
and ETAB.ID_DEPARTEMENT = DEPTpro.ID_DEPARTEMENT
and f.ID_DEPARTEMENT = DEPTpat.ID_DEPARTEMENT
and r.ID_REGION = DEPTpat.ID_REGION
and reg.ID_REGION = DEPTpro.ID_REGION
and DEPTpro.ID_DEPARTEMENT in (@ParamDeptProfessionnel)
and DEPTpat.ID_DEPARTEMENT in (@ParamDeptPatient)
and f.DATE_PREMIER_CONSENTEMENT between CONVERT(DATE,@ParamPeriodeInclusionDeb,103)
and CONVERT(DATE,@ParamPeriodeInclusionFin,103)

-- Dts_Provenance_Patient --
SELECT
r.LIB_REGION as lib_region_patient,
DEPTpat.ID_DEPARTEMENT as dept_pat,
DEPTpat.LIB_DEPARTEMENT as lib_dept_pat,
```

```
count(*) as nb_patient
FROM
D_ETI_PATIENT f, D_ROR_PROFESSIONNEL d, d_ror_pro_unite unite, D_ROR_UNITE U,
D_ROR_ETABLISSEMENT ETAB, D_GEO_DEPARTEMENT DEPTpro, D_GEO_DEPARTEMENT DEPTpat,
D_GEO_REGION r, D_GEO_REGION reg
WHERE f.ID_ROR_PROFESSIONNEL=d.ID_ROR_PROFESSIONNEL
and d.ID_ROR_PROFESSIONNEL= unite.ID_ROR_PROFESSIONNEL and FLAG_PRINCIPAL=1
and unite.ID_ROR_UNITE = U.ID_ROR_UNITE
and U.ID_ROR_ETABLISSEMENT= ETAB.ID_ROR_ETABLISSEMENT
and ETAB.ID_DEPARTEMENT = DEPTpro.ID_DEPARTEMENT
and f.ID_DEPARTEMENT = DEPTpat.ID_DEPARTEMENT
and r.ID_REGION = DEPTpat.ID_REGION
and reg.ID_REGION = DEPTpro.ID_REGION
and DEPTpro.ID_DEPARTEMENT in (@ParamDeptProfessionnel)
and DEPTpat.ID_DEPARTEMENT in (@ParamDeptPatient)
and f.DATE_PREMIER_CONSENTEMENT between CONVERT(DATE,@ParamPeriodeInclusionDeb,103)
and CONVERT(DATE,@ParamPeriodeInclusionFin,103)
GROUP BY r.LIB_REGION, DEPTpat.ID_DEPARTEMENT, DEPTpat.LIB_DEPARTEMENT
```

Appendix 4. eTICSS report's requests

## *Appendix 5 – Squash representation of tests executions*




Appendix 5. Squash representation of tests executions

As said in the previous parts, Squash enables one to see advancement of the testing process. This graph is only one example of the numerous visuals generated automatically by Squash. The graph shows that 41 tests where created on Squash for GCS. 39 were executed and successful and 2 were not executed.

## *Appendix 6 – GDI's list of evolutions and corrections to do*

Liste des bogues( 1 - 15 )

| | P | Identifiant | | Catégorie | Impact | Statut | Mis à jour | Résumé |
|---|---|---|---|---|---|---|---|---|
| | _ | 0016134 | 7 | Evolution | majeur | fermé (cloe.dorali) | 2019-05-29 | Valeur par défaut ProxyRor_Pro_ID |
| | _ | 0016113 | 6 | Anomalie | mineur | fermé (cloe.dorali) | 2019-05-29 | SSRS : Indicateurs Nombre de patient inclus |
| | _ | 0016101 | 4 | Anomalie | majeur | fermé (cloe.dorali) | 2019-05-29 | Rapport ETICSS - Indicateur patient inclus |
| | _ | 0016087 | 4 | Evolution | mineur | fermé (cloe.dorali) | 2019-05-29 | ODS - Longueur longueur champs "Structure_EVENEMENT" dans table CPI_PATIENT_EVENEMENT |
| | _ | 0016064 | 5 | Evolution | mineur | fermé (cloe.dorali) | 2019-05-29 | Rejet Patient "CPI_PATIENT_PROFIL" - Cause ID_PROXYROR_CONSENTEMENT = NULL |
| | _ | 0016063 | 6 | Evolution | mineur | fermé (cloe.dorali) | 2019-05-29 | ODS - Longueur longueur champs "type de problème" dans table CPI_PROBLEMES |
| | _ | 0016062 | 4 | Anomalie | mineur | fermé (cloe.dorali) | 2019-05-29 | Rejet utilisateurs existant dans table D_UTILISATEUR_CPI |
| | _ | 0016206 | 3 | Anomalie | mineur | fermé (cloe.dorali) | 2019-05-29 | Rejet patient null - F_ETI_EVENEMENTS_SI_CTIO |
| | _ | 0016136 | 4 | Evolution | mineur | fermé (cloe.dorali) | 2019-05-29 | [SSIS] Rejet ROR_ETABLISSEMENT - Code postal Nul |
| | _ | 0016135 | 3 | Anomalie | majeur | fermé (cloe.dorali) | 2019-05-29 | [SSIS] Suppression zip des fichier ROR |
| | _ | 0016131 | 4 | Anomalie | mineur | fermé (cloe.dorali) | 2019-05-29 | [SSRS] Doublon sur le tableau page 3 |
| | _ | 0016112 | 3 | Evolution | mineur | fermé (cloe.dorali) | 2019-05-29 | SSRS : Valeurs de filtre par défaut |
| | _ | 0016334 | 9 | Evolution | mineur | A recetter (cloe.dorali) | 2019-05-28 | [SSRS]Amélioration rapport ETICSS |
| | _ | 0016156 | 1 | Evolution | **majeur** | A recetter (cloe.dorali) | 2019-05-20 | SSRS - Evolution rapport ETICSS |
| | _ | 0016071 | 3 | Anomalie | mineur | prise en charge (cloe.dorali) | 2019-05-03 | Rejet Patient sur commune non existante dans la table commune |

Appendix 6. GDI's list of evolutions and corrections

The image shows that 15 demands were made by the client on the proposed development. So 15 subjects were to be improved (field "catégorie" at "Anomalie") or to be created (field "catégorie" at "Evolution"). From all these demands, the majority were set as not major one, so the impact were minor.

Today, all the tickets are closed. Yet on the picture 3 tickets aren't. When a ticket is closed it means that the application is in place an approved by the client. If a ticket is in blue, it means that the problem or evolution is in test on the client's side. If in yellow, the label is "taken in charge", so an action is in implementation or currently tested by Keyrus' teams.

To give examples of the subjects of those tickets :

- The yellow ticket concerns the geographical data that we extract directly from DIAMANT. The tables aren't up-to-date (with missions postal codes, not well regrouped postal code and old regions names). An action from their side is needed to solve this ticket.
- The tickets in green concern solutions to apply for the files that didn't respect the interface contract that Keyrus and the client agreed on. So this problem finds its source in the client.
- The tickets in red concerns the creation of fictive lines of regroupment : for example we do not have the exhaustivity of the establishments and an establishment can be composed of several units; yet we can have on the unit file units that are attached to establisments we do not have. Those units are directly going to the rejection tables after the lookup. The domino effect causes the rejection of all the patients attached to those units after the lookup. This is not possible here as the client doesn't send us the exhaustivity of its database each time it sends us files. So fictives lines ou regroupment were created and all the lines are updated if we receive more completed information afterwards.

## *Appendix 7 – GCS obtained grades*

**SATISFACTION CLIENT**

| | Indicateur | Note Avr-19 | Note Juil-19 | Tendance | Commentaire client |
|---|---|---|---|---|---|
| Q00 | Quel est votre niveau de satisfaction global sur la prestation réalisée par Keyrus (sur une échelle de 0 à 10) ? | 9 | **9** | → | Entièrement satisfait de Keyrus. Réactivité par rapports aux nouvelles demandes (Modification rapport, VT), propositions pertinentes |
| Q02 | Qualité du transfert de compétences | *N/A* | **4** | ↗ | Transfert de compétence sur l'ETL, le lancement des chargements et le processus de chargement (gestion des rejets..) très bon. |
| Q18 | Qualité de la documentation | 4 | **4** | → | Documentations claires et précises et mise à jour régulièrement |
| Q19 | Qualité des livraisons logicielles | *N/A* | **4** | ↗ | Le développements des flux et des rapports correspondaient parfaitement à nos attentes. |
| Q21 | Respect des délais/niveaux de service | 5 | **5** | → | Malgré toutes les difficultés de mise à disposition de la donnée, les délais sont respectés. |
| Q28 | Réactivité et réponse face aux nouveaux besoins | 4 | **4** | → | |

Appendix 7. GCS obtained grades at the COPIL of May and July 2019

Depending on the project and its advancement, the global team (composed of Keyrus' team and client's team) decided to meet more or less each 2 months. In May, GCS knew its first COPIL (after 4 months, the first developments were on test on the client's side). At this meeting, the client is giving its grades to Keyrus based on 6 indicators.

In fact, at the signature of the contract, Keyrus proposes a list of indicators to all of its clients. Those indicators are numerous. It can concern pure technical skills or more social skills as communication. The client has to choose 5 of them (and to grade up to 5). And a general indicator is obliged for the projects as it refers to the global satisfaction of the client regarding the whole project (graded up to 10).

GCS chose:

- The quality of the competences' transfer,
- The quality of the functional deliverables,
- The quality of technical deliverables,
- The compliance with required project' deadlines and
- The responsiveness to users' needs.

Based on the grades, the client's satisfaction is really good. This makes GCS one of the top graded projects at Keyrus. The client commented that they are satisfied of the development, our reactivity to their questioning and the quality of the deliverables. They highlighted that despite the difficulties faced by the client to give us files to load in time and with the good format, we came with good propositions, availability and developments that weren't late.