A framework for fair and responsible data market ecosystems

Nwatchock A Koul, Aman Sabrina

A framework for fair and responsible data market ecosystems

THÈSE

présentée à la Faculté des sciences de la société
de l'Université de Genève

par

Aman Sabrina Nwatchock A Koul

sous la direction de

prof. Jean-Henry Morin

pour l'obtention du grade de

Docteur ès sciences de la société
mention systèmes d'information

Membres du jury de thèse:

Mme. Giovanna DI MARZO SERUGENDO, Professeure, présidente du jury, Université de Genève
Mme. Solange GHERNAOUTI, Professeure, Université de Lausanne
M. Dimitri KONSTANTAS, Professeur, Université de Genève
M. Jean-Henry MORIN, Professeur, Directeur de thèse, Université de Genève
M. Jean-Philippe WALTER, Docteur, Expert indépendant et commissaire à la protection des données du Conseil de l'Europe

Thèse no 125
Genève, 02 juillet 2019

La Faculté des sciences de la société, sur préavis du jury, a autorisé l'impression de la présente thèse, sans entendre, par-là, émettre aucune opinion sur les propositions qui s'y trouvent énoncées et qui n'engagent que la responsabilité de leur auteur.

Genève, le 02 Juillet 2019

Le doyen
Bernard DEBARBIEUX

Impression d'après le manuscrit de l'auteur

Table of Contents

List of figures

List of tables

Résumé

Alors que l'accès à l'information revêt une importance critique pour notre société, nous assistons à une "course à l'information" où de nombreuses initiatives d'accès aux données prolifèrent pour permettre à cette information d'être facilement disponible et utilisable. Récemment, une nouvelle économie des données a émergé avec un nombre croissant de marchés de données.

Le terme marché des données couvre tout un ensemble d'activités qui tirent de la valeur des données, offrant ainsi des avantages à de nombreux intervenants comme les producteurs de données, les courtiers en données, les consommateurs de données, etc. Aujourd'hui, différent problèmes menacent le développement futur de ces marchés. Parmi les plus urgents figurent le commerce de données sur de multiples canaux qui ne sont pas nécessairement rendus publics ou transparents, le manque de services et d'outils permettant aux acteurs du marché d'échanger en toute sécurité dans ce secteur en pleine expansion. Par conséquent, la valeur de l'échange de données n'est pas partagée de manière équitable entre les participants au marché. Les écosystèmes de marchés des données sont pour l'essentiel incontrôlés et les actions visant à créer un espace sécurisé sont très fragmentées. Comme tout autre marché, les marchés des données exigent un niveau commun de confiance et de transparence afin de garantir leurs durabilités. Très peu de travaux ont été faits jusqu'à présent pour établir les bases d'un échange de données sûr et digne de confiance dans notre société.

Cette thèse aborde donc la question de la conception d'écosystèmes de marché de données équitables et responsables. Premièrement, nous analysons les catégories de données en fonction de critères spécifiques en vue d'établir des règles d'échange entre les parties prenantes. Ensuite, sur la base de la méthodologie de recherche de design science nous étudions les composantes des marchés des données et proposons une approche globale pour la conception d'un cadre pour des écosystèmes de marché des données équitables et responsables permettant la transparence, la confiance, l'équité et la responsabilité des acteurs. La conception couvre également la gestion des accords contractuels sous-jacents qui sont nécessaires entre les parties prenantes.

Pour appuyer la proposition, nous démontrons la faisabilité du cadre au moyen d'un prototype de mise en œuvre basé sur un scénario. Le prototype couvre l'échange de données entre les participants du marché, ce qui permet de vérifier ses propriétés de sécurité, de protection des données, de respect de la vie privée et, dans une certaine mesure, la conformité avec le GDPR.

Ce travail représente un pas en avant vers un échange et une utilisation responsables des données. En particulier, il fournit une base pour discuter de la souveraineté des données et de l'autodétermination dans des écosystèmes de marché équitables et responsables.

Abstract

As access to information has become critically important in our society, we are witnessing an “information race” where many initiatives for data access are proliferating to allow for this information to be readily available and usable. Recently, a new economy around data has emerged with a growing number of data markets.

The term data market covers a whole range of activities where value is derived from data, thus providing benefits to many stakeholders like data producers, data brokers, data consumers, etc. Nowadays, some issues are threatening the future development of such markets. Among the most pressing problems are the trade of data on multiple channels that are not necessarily made public nor transparent, the lack of services and tools allowing market actors to exchange safely in this growing area of data marketplaces. Therefore, the value of data exchange is not shared in a fair way between the market participants. Thus, the data market ecosystem is for its most part uncontrolled, and the actions for creating a secure space are highly fragmented. As with any market, in order for it to work properly, data markets require a common level of trust and transparency. Very little work has been done so far to address the foundations for the safe and trustworthy exchange of data in our society. These main issues undermine the emergence and the development of this critically important ecosystem for the future.

Hence, this dissertation addresses the question of the design of fair and responsible data market ecosystems. We first analyze the data categories according to specific criteria towards, establishing exchange rules. Then, based on a design science research methodology, we study the constituents of data markets and propose a global approach towards the design of a framework for a fair and responsible data market ecosystems enabling transparency, trust, fairness, and accountability. The design also covers the necessary underlying agreement management among the stakeholders.

To support the proposition, we demonstrate the feasibility of the framework through a prototype implementation based on a scenario. The prototype covers data exchange among the marketplace participants allowing to verify its properties of security, data protection, privacy and to a certain extent GDPR compliance.

This work represents a step forward towards enabling responsible data exchange and usage. In particular, it provides a basis for discussing data sovereignty and self-determination in fair and responsible market ecosystems.

Acknowledgement

First, I would like to thank God Almighty for giving me the strength and the opportunity to carry out this research study and to complete it with satisfaction.

In my research journey, I have found a teacher and an inspiration model, Prof Jean-Henry Morin who provide me support and guidance at all times. He has given me invaluable guidance, suggestions in my quest for knowledge and all the freedom to pursue my work while ensuring its accomplishment. I shall eternally be grateful to him for his encouragement.

I would like to thank the thesis committee: prof. Giovanna Di Marzo Serungendo, Prof. Solange Ghernaouti, prof. Dimitri Konstantas, Doc. Jean-Philippe Walter for their insightful comments, support, and trust towards my work.

I express my gratitude to my colleagues in CUI, especially I take pride in appreciating the stimulating discussions with my dear colleague Arbër Sahili to whom I am indebted for sharing his valuable time and joyful moments, as well as Camille Tardy and Alan.

I certainly do not want to forget my dear friends Jessica Gomez and Nico Hillah who have assisted me, in whatever manner possible and enable good times to continue.

Ultimately, I thank the biggest source of my strength, David, Eden, and Edna that made me cheer up in difficult times. I thank also my dear Sharon, Anna, Nelly, Mimi, Muriel, Therese, Jack, Christian, Judith, and Yoan, who have all made a tremendous contribution in helping me reach this stage.

I dedicate this work to my mother and my father whose dreams for me have led this achievement.

Chapitre 1. Introduction

When the first electronic computers development started in 1941, the foundation was laid for the future of innovative technologies. The digitization process was underway, spurring change in the overall functioning of our societies. Peattie and Peters (1997) summarized the main transformations during this process into three phases. The first one was "*the computer age*" characterized by large machines, mainframes and the beginning of small computers production. At that time started a growing consciousness about the strategic importance of Information and Communication Technologies (ICT) in information collection, storage, and management. The next period of technological advancement in the middle of the 1980s, "*the PC age*", was marked by the adoption of computers in the business arena, and to a lesser extent, by the individuals. And finally, "*the communication age"* which started from the beginning of the 1990s, was related to the improvement of internal and external communication powered by the increasing use of ICT in all areas.

A major step during this age was definitely the advent of the Web, which undoubtedly has triggered great transformations by connecting people and organizations in a vast and distributed network of computers. On one hand, the new web users empowered by the availability of information through the web and the ease of access to new internet services began to search for information, to buy online, to work, to meet people and progressively produce web content. Thus, they moved from being simple internet users to web actors, offering their own services and content, thus participating in the development and the wide adoption of the ICT. This is also known as the transition from the Read-Only internet to the Read-Write internet. On the other hand, a new category of tech organizations, e-business, e-commerce, and many others had emerged, using the Web for service delivery and collaboration in a worldwide context. E-business refers to the use of digital technology and the internet to execute the major organization business process, while e-commerce deals with the buying and selling of goods and services over the internet (Laudon and Laudon 2007).

Hence, over the years, these new organizations as well as the traditional ones have adopted ICT and made massive investments for software and hardware acquisitions in order to digitally manage their business. The digitization process has also influenced the public sector, mainly with the creation of electronic government services called e-government for services delivery to citizens, employees, and businesses. The goal was "*to empower citizens, by giving them easier access to information and the ability to network electronically with other citizens*," (Laudon and Laudon 2007).

During these phases, numbers of digital technologies have been created and improved to respond to the growing needs of organizations and individuals. From the 2000s onwards, technological advancements, such as the improvement in internet bandwidth, the falling cost of massive storage capacity and the democratization of the computing power have changed the way organizations operate and develop their activities. The tipping point was the availability of cloud computing which, although introduced in 1961 by John McCarthy (1992), took off in 2006 with Amazon through the Amazon Web Services (AWS). The massive

adoption of this technology on a commercial basis by organizations occurred around 2009-2010. Amazon offered a variety of cloud-based services including storage, computation, and other online services which allow organizations to rent virtual computers and use their own applications. Soon after, other large tech organizations such as Microsoft and IBM followed this trend. This advent of cloud computing was a huge opportunity that enabled more productivity for organizations and individuals as it became more economical to outsource their services and data storage, rather than owning them (Carr 2003, 2005).

New digital gadgets based on hardware and software technologies such as mobiles phones and connected devices have also been created and have become over time pervasive in our society, thereby contributing to the increase of digital content. They are now routinely used for professional tasks as well as personal tasks, and to live without them is now unthinkable. All of these technological advances led to the next major technological change: "big data" which is defined as "*a high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making*" (Gartner, 2019). Big data originates from many sources, one of the most important being social networks which have become an extremely valuable tool for people to network with each other. These services, which were simply allowing people to gather and connect, have enabled the creation and collection of information on individuals which have been, until now, difficult to harvest.

Today, social networks represent an ideal medium for brands, content providers, and advertisers to reach millions of people worldwide. While many industries have been disrupted by the rise of social networks, none has been impacted more than the advertising industry which relies now on social network providers to collect and use data from their users for profiling, marketing, and managing customer intimacy to better understand customer needs (Spiekermann and Novotny 2015). Ultimately, the technological advancement and digital services have led to an ever-increasing growth in the amount of digital data, which are harnessed by organizations.

Concurrently with these major advances and the associated opportunities, major security, ethical and social issues closely related to IT and its use has also emerged. Accordingly, different data protection frameworks have been created or are currently being revisited to regulate this environment, the first national data protection law being the *Data Act* (Sw. Datalagen) enacted in Sweden on 11 May 1973 in order to prevent the disclosure or misuse of personal information. As new advances in the IT domains come with new challenges, regulatory frameworks have evolved to respond with more or less adequate legislation to address these challenges. Figure 1 shows an overview of major advances of IT along with the main data protection regulatory frameworks and breaking date in the European context.

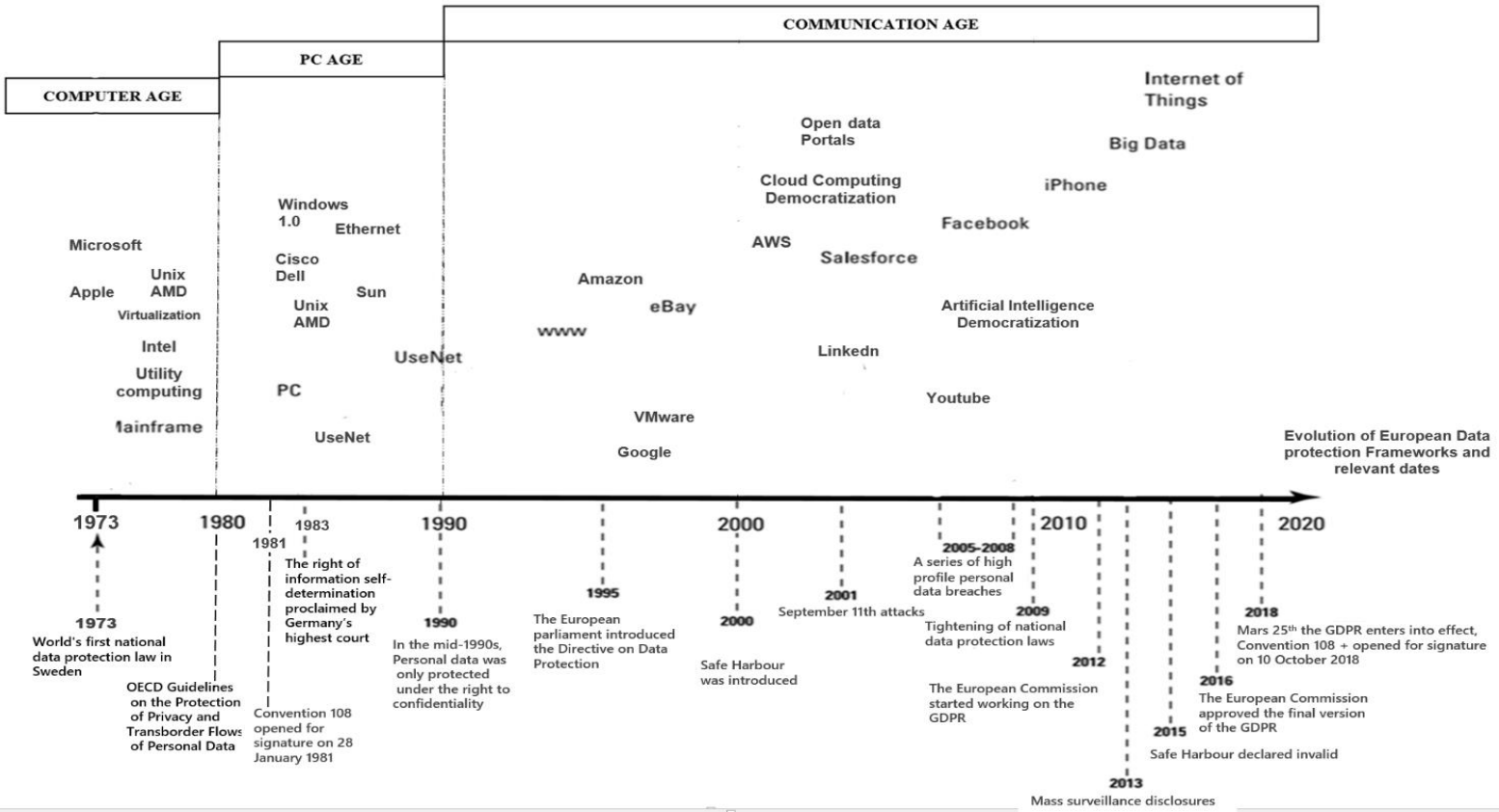

Figure 1: Information technology and European data protection evolution

Nonetheless, the collection of massive personal data which has led to many issues is more than ever at the forefront of legal reforms. In the 2000s, frequent data breaches have been observed, exposing millions of personal information records (De Groot 2019).

Personal data have also become a tool for mass surveillance by governments for political and industrial espionage purposes as well as counter-terrorism (Oscar H. Gandy 1989). This mass surveillance issue pointed out long before by Duncan Campbell in his report about the Echelon program (Campbell 1988) is exacerbated today by the magnitude offered by digital services combined with ever-increasing algorithmic performance. Thus, on June 6th, 2013, the "PRISM program" (James 2013), a clandestine electronic surveillance program handled by the National Security Agency (NSA) has been revealed. This program gave the NSA the power to gain direct access to the data of millions of customers of various internet services, including those provided by Google, Facebook and Apple. The mass surveillance by the NSA and its partners along with the multiple data breach denunciations have triggered a new conversation about the personal data privacy and security issues and even more importantly the control of the transparent flow of personal data.

Thereafter, the European Commission was engaging in the reform of personal data protection with the goal to enhance the rights of data subjects and to create an environment of fair competition for information technology actors. Two main reforms occurred in the European arena of data protection. The first was the modernization of the Convention 108, called Convention 108 +, on 18 May 2018 which has been, for the past decades, the only international legally binding instrument on privacy and personal data protection (Council of Europe 2018). It deals with the innovations in the area of information and communication technology since its original adoption in 1981 and aims at strengthening the convention's effective implementation. The inclusion of state parties across the globe in this convention makes it a strong international standard for privacy and data protection.

The second great reform was the General Data Protection Reform of the European Union (GDPR 2018) which entered into force on 25 May 2018. A significant element about the modernized convention 108 and the GDPR are their compatibility, as both instruments tend to tackle similar data protection and privacy challenges posed by ICT and the massive data collection. For instance, both reforms provide new rights for data subjects in response to technologies such as profiling, automated data processing, algorithms, etc., and introduce new obligations for controllers relating to transparency, accountability, privacy by design and by default, risk assessment, and data breach notification. However, both initiatives are different in several ways such as the difference of scope and focus as described by Greenleaf (2017). In the Convention 108 +, each party undertakes to apply to data processing subject to its jurisdiction in the public and private sectors, while the GDPR is applicable to the processing of data of people in the EU. Moreover, the GDPR has some requirements which are not explicitly required by the Convention 108 + such as local representation required of foreign controllers or processors, right to portability of data subject, mandatory data protection officers for sensitive data processing, etc. Nonetheless, the Convention 108 + provided a framework of fundamental principles around which nations can

build regulatory data protection frameworks which are, at a low level, in adequacy with the GDPR. It constitutes a universal and common basis for data protection, on which it is possible to build and differentiate according to legal tradition and regional specificities, without having to reduce the level of protection.

1.1 Information race

Access to information has become critically important in our society. It is an essential part of our environment, whether for purpose of innovation, research or different business and social good activities. It has the potential to transform business models for more economic profit and competitive advantage, and even to set up whole new industries. For the scientific community whose research is heavily dependent on data availability, data re-use is as important as data gathering. According to Borgman et al. (2007), the advance in science will increasingly rely on the existence of a common information infrastructure enabling the life cycle management of available data effectively and efficiently. During the last decade, some projects for the development of such information infrastructure emerged like Kaggle and DataMarket (2019), allowing companies to supply data to scientific communities and organizations in general. Currently, we are witnessing an “information race” where many initiatives in data sharing, data trade, and services for data access are flourishing to allow for this information to be available and usable. Hence, the issue of data provision and reuse that is spreading in various domains where valuable data are abundant represents a key challenge in the digital landscape. Noteworthy initiatives like Open Data and Data Liberation that will be discussed in the next chapter have emerged to disseminate as much data as possible to sustain the economic growth of the entire community.

However, personal information are arguably the most sought-after as they become a real commodity that are collected and sold for advertisement or other related services. The commoditization of these data is a growing business which involved many actors such as social network providers and entities called data brokers. Data brokers gather information about individuals from a variety of sources. Then, they create the profile of each individual for marketing and other purposes and finally sell them to business. In 2014, the global industry of data broker in the United States was estimated to comprise thousands of companies generating some US$200 billion in annual revenue (Mott 2014). Today, a new opportunity is given to organizations to get income through the monetization and commoditization of the personal information they capture. According to a survey performed on 476 executives (The economist 2015), an increasing proportion of their respective companies are preparing to monetize their data. This trend is ongoing as estimated by the Transparency Market Research (TMR 2018) report on the data broker industry development in the United States, Europe, and China. According to them, the global data brokers market is expected to grow at a compound annual growth rate (CAGR) of 11.5% for the forecast period between 2017 and 2026. The demand for consumer data is estimated to contribute more than one-third income share of the overall market by the end of 2026 and North America will probably control the data broker market over the 2017-2026 forecast period. The information race has led to the emergence of a “data market” where a wide range of data is sold or shared for public and private consumption.

1.2 Emergence of data market

The term data market was coined recently (Dimitru and Gatti 2016, Elbaz 2012) to designate a structure where data are exchanged for money, or for free. Data markets do not only refer to traditional market models where providers sell goods to earn money as argued by Gil Elbaz (2012), they also cover a whole range of activities where value is derived from data, thus providing value to many stakeholders. This is a wonderful opportunity for the society to acquire data, especially if we consider them as part of an ecosystem that could serve different purposes.

Many parties are involved in the data market such as data providers, data consumers, individuals, data brokers, etc. Data markets are expanding rapidly, and this trend is likely to further increase for the foreseeable future to support the large demand and need for data. As with any market, in order for it to work properly, data markets require a common level of trust and transparency. Nowadays, data are traded on multiple channels that are not necessarily made public or transparent. Therefore, the value of data exchange is not shared in a fair way between the market participants. Some recent studies (FTC 2014) show that transactions on personal data between data brokers and organizations are obscure, if not ethically debatable.

The development of data marketplaces now enables the trade of data on platforms with a certain level of visibility of data products and transaction transparency. However, participation in such markets is still hard in terms of accessibility and roles (e.g., data brokers, individuals, etc.). Among the main reasons is the fact that individuals often do not even own both legally and physically their information. As a result, there is often no interest to interact directly with them to negotiate their information through informed consent, which is often not even “informed.” Moreover, the data marketplaces are mainly centralized, which implies the influence of a third party in the management of the data market. Current regulatory frameworks do not provide answers on the issues of data commoditization in a way that really protects individuals. Consequently, service providers exploit this situation through their business models forcing users to release their rights and to cooperate under their conditions, often in a one size fits all way.

The data market is for its most part uncontrolled, and the actions for creating a secure space are highly fragmented. This research is an opportunity to explore the issues undermining the emergence and sustainable development of data market ecosystems in a fair and responsible way by the involved parties. Addressing data provisioning from a market perspective must be done in a holistic way, thus including the study of data that are part of the transaction between parties. These data have some characteristics which required specific conditions of exchange in a market. Depending on the nature of data, their collection and usage may raise considerable concerns such as privacy, ethics, usage rights, etc. Accordingly, these issues must be addressed in the light of principles that will govern data collection and sharing. Therefore, this thesis attempts to contribute to the development of fair and responsible data market ecosystems.

1.3 Thesis structure

The dissertation is structured as follow. Chapter 2 presents the research background and context behind this research. We discuss the main obstacles currently hindering the development of fair and responsible data market ecosystems. Chapter 3 defines the term "data market ecosystem" and summarizes related issues. In addition, it addresses the research question and research methodology. In Chapter 4, we provide a state of the art and related works on the topic of data processing. We provide an overview of the addressed issues on academic, industry, and regulation perspectives. Chapter 5 identifies the requirements to address the research question. Chapter 6 proposes a categorization of the data in the context of our study with the goals to further derive the principles to guide their exchange. Chapter 7 introduces the main contribution of the thesis which is a framework for fair and responsible data market ecosystems. This part describes the main components of the framework. Chapter 8 elaborates on agreement management in data market ecosystems and describes the transaction flow protocols between the parties. Chapter 9 describes an instantiation of the proposed framework in order to validate its core functions and feasibility. In Chapter 10, we evaluate the framework instantiation by analyzing the design elements and then, we discuss the contributions of the thesis as well as its limitations. Finally, Chapter 11 concludes the study by summarizing the contributions and then gives an outlook on possible future research.

The Appendix section provides a glossary of terms, the high-level description of the implementation code and a template of a GDPR compliant data processing agreement.

Chapitre 2. Research background and context

The ongoing digitization process has profoundly transformed our society. In this chapter, we identify the key aspects of these transformations that will constitute the foundation of our research on data collection and sharing.

The first part of the chapter elaborates on the increasing value and the democratization of data. This aspect is at the heart of the scientific and industry and regulation efforts towards better information management, in particular in the domain of personal information. The second part outlines the data protection reforms that contribute to the effort of transparent data processing. Especially we address the evolution of the European Data Protection, and we discuss its implication for the parties involves in the data collection and sharing process. The third part describes the current business models of the internet organizations, their issues and the initiatives around personal information management, and we analyze how these new approaches fit into the global effort of user empowerment.

2.1 The increasing value and the democratization of data

Data are increasingly compared to raw material as they have become an essential part of most socio-economic activities. Meglena Kunewa (Kuneva 2009) was the first to coin this idea stating that: “Personal data is the new oil of the Internet.” Likewise, this analogy has been extended to other data categories to express the idea that data access and use is essential (GOV.UK 2012). Marta Teperek (Teperek 2016) goes even further when she argues that: “data is the new water – it is renewable and it is crucial to the ecosystem”; this idea stresses that data is vitally needed today and sustainable access to data is critical for the future. Any data type has the potential to leverage new business opportunities or be useful for social progress. As a strategic asset, their most salient characteristics are their diversity, heterogeneity and the fact that they are in silos. While this could be an advantage for local control and governance in a way that complies with some legal and regulatory considerations, it prevents a connection of things which is highly desirable in this century. Silos represent an obstacle for innovation as they restrain the discovery of valuable data. For decades, powerful institutions have invested in tools and methods to capture and retain data. They have acquired powerful analytic techniques to extract useful knowledge from this body of data. Multidimensional analysis is such a technique allowing companies to turn their corporate raw data into valuable knowledge (IMB 2009) to be used for many purposes like market research, performance improvement, and more recently predictive analytics. Many examples of innovative services are emerging based on such technique and new trends in data usage continue to emerge.

As transparency and openness lie at the heart of the concept of democracy and is vital to enabling trust and accountability in our society, new initiatives towards these key ingredients are increasingly making data available not only to organizations and the public in general. Since 1966, the US Freedom of Information Act (FOIA), one of the first tools for greater transparency in government activities, has given people the right to request information held by federal government agencies (Rehnquist 1969). The rationale behind this law was that governments hold information not for themselves but on behalf of the public.

Several decades later, in 1995, the term open data appeared in the scientific community to promote the disclosure of geophysical and environmental data in a complete and open exchange of scientific information (NRC 1995). Subsequently, the open data movement sprung up for enabling the unrestricted access of data produced by public and private players involved in the context of a public service mission. This data type is a public good. Unlike the FOIA which provides data access when requested, the open data movement requires data to be published online by default. Open data is collected using taxpayer money and should always be available at no cost, including business. The goal of this movement is to improve transparency in public governance as well as to make public data available in a reusable format so that value could be generated through the re-use of these data.

Open data has emerged from the open movement supporting the idea of a free, open and collaborative society where everyone can access the work done by others, use it and build on top of it. From this perspective, a huge amount of public government data has been made available for public consumption. Open data initiatives have increasingly impacted other sectors such as scientific communities and private organizations. Some studies based on citizen science which enables the collaboration of large groups of individuals, often without financial compensation have generated datasets published in open format (Heipke 2010, Bradley et al. 2009). Private sectors are more and more publishing open data. The company Uber, through a private open data portal, provides anonymized data from over two billion transportation movements for non-commercial re-use (Uber, 2019).

More than ever, an increasing volume of data is available and ready for consumption provided these data become more accessible. Technologies around big data contribute also to the democratization of data. Big data comes from public data and private data sources. Interestingly, services around big data are increasingly available. Data cloud services are examples of services that have emerged to address the big data challenge by offering on-demand and scalable storage solutions using cloud computing technology and infrastructure. The cloud computing shifts information technology costs to a pay-as-you-go model, hence small companies can then use this technology without investing in costly infrastructure for data storage and analytics. Xignite (2019) and Azure Marketplace (2019) are examples of platform supplying data on demand. Another example is a data exchange project, where businesses can share their data with others (Data-XC 2015). As the democratization of data is growing exponentially, it becomes essential to rethink the legal aspects, the business and service design around data consumption in our society.

2.2 Evolution of European data protection reform

Privacy and trust are important issues in the digital age. According to Mason (1986), the growth of information technology and the increasing value of information in decision-making are the main forces that threaten privacy. The significant growth of computer power, the ubiquitous availability of computing and storage resources and the increase of digital content production have adversely affected the privacy world. Documents leaked by Edward Snowden (Ewen et al.

2013) in the PRISM scandal where evidence of extensive internet and phone surveillance by the NSA. This issue confirmed again the generalized infringement of individual privacy. In Europe, these revelations have raised strong concerns for European politicians who called for more stringent measures to ensure privacy, preservation and data protection.

In January 2012, the European Commission has unveiled its plans for improving the data protection directive (DPD) dating back to 1995, as it failed to keep up with the ongoing change of digital technologies. The core element of this reform is the re-appropriation of personal data by individuals. On 25 May 2018, the General Data Protection Reform (GDPR 2018) entered into force harmonizing the regulations for the processing of personal data by companies and public authorities within the EU and also any organization processing data from EU business and residents. The major changes that come under the GDPR are:

- **Enlargement of personal data**: the DPD apply to personal data that are defined as “any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity”. The GDPR expanded the definition of 'personal data' by including online identifiers such as IP addresses, mobile device identifiers, geo-location and biometric data like fingerprints, retina scans, etc.
- **Consent**: the GDPR provides detailed provisions for valid consent that were not given in DPD. An explicit opt-in is required for personal data processing, a short and straight description for data use, consent must be in plain language, informed, specific, and unambiguous, and with the requirement that the Data Subject is able to opt-out of profiling and object to the results of profiling. The age of the data subject must be appropriate, and he may withdraw consent at any time.
- **Obligations for data controllers and data processors**: in the DPD, data controllers were held accountable for any mishandling of personal data and for all actions of the data processors Art.17(1) DPD, Art.23 DPD. Both data controllers and processors are required to abide by the GDPR and are liable for violations. Data processors are required to have a contract with data controllers to process personal data. The controller or processor must appoint a data protection officer in any case where (Art. 37 GDPR):
    - The core activities involve “regular and systematic monitoring of data subjects on a large scale.
    - The processing is carried out by a public authority, except for courts acting in their judicial capacity.
    - The core activities consist of processing on a large scale of special categories of data or personal data relating to criminal convictions and offences.
- **Additional measures for information governance and security**: GDPR requires that organizations consider compliance with the regulation from the inception of systems and processes. They must implement “privacy by design” features. It also implies that controllers

discard personal data when they are no longer using it. Organizations must conduct impact assessments for automated data processing activities, large-scale processing of certain kinds of data, and systematic monitoring of publicly accessible areas on a large scale.

- **Rights for data subjects**: the GDPR (Chapter 3) strengthens certain rights granted under the DPD and adds additional rights summarized in Table 1. These new rights enable data subjects to have more control and responsibility over their personal data.

Rights	Definitions
The right to be informed: New right	**GDPR**: organizations must be completely transparent in how they are using personal data (personal data may include data such as a work email and work mobile if they are specific to an individual).
The right of access: Improved right	**DPD**: individuals will have the right to know exactly what information is held about them and how it is processed.
	GDPR: the data subject can also know about retention period, the existence of certain rights, the data source and the consequences of processing.
The right of rectification: Unchanged right	**DPD & GDPR:** individuals will be entitled to have personal data rectified if it is inaccurate or incomplete.
The right to restrict processing: Improved right	**DPD:** the data subject can block processing of data on the grounds of data inaccuracy or incomplete nature of data.
	GDPR: this law is more elaborate and defined in this respect more grounds with consequences of enforcement of this right and obligations on the controller.
The right to object : New right	**GDPR:** in certain circumstances, individuals are entitled to object to their personal data being used. This includes if a company uses personal data for the purpose of direct marketing, scientific and historical research, or for the performance of a task in the public interest.
The right to be forgotten: Improved right	**DPD**: it merely mentions that the data subject has the right to request erasure of data on grounds of data inaccuracy or incomplete nature of data or in case of unlawful processing.
	GDPR: it has strengthened this right by laying out 7 conditions for enforcing this right including five grounds on which the request for erasure shall not be processed.
The right to data portability: new right	**GDPR:** it allows individuals to retain and reuse their personal data for their own purpose.
Rights of automated decision making and profiling: Improved right	**DPD:** the intent is that data subjects should have the right to obtain human intervention into their personal data.
	GDPR: it has put in place safeguards to protect individuals against the risk that a potentially damaging decision is made without human intervention and the decision-making excludes data concerning a child.

Tableau 1: Differences of data subject rights under the GDPR (2018) and DPD (1995)

- **Measures for data breach notification and penalties**: In DPD, EU member states were free to adopt different data breach notification laws and the DPD does not specifically mention or require administrative fines for Data Protection violations. Organizations must report data breaches to the individuals whose data was compromised and to their supervisory authority within 72 hours. The authority will evaluate the data compromised and the preventative security measures in place at the time of the breach to assess repercussions and ensure future compliance. Organizations in breach of GDPR can be fined up to 4% of annual global turnover or €20 Million whichever is the highest.
- **Extraterritorial applicability of the laws**: Article 3 of GDPR introduces the "lex loci solutionis" which means that GDPR it is not mandatory that a data processor has a physical establishment with the EU. In fact, an organization that provides goods or services to people in EU must comply with the GDPR, regardless of its physical location, and of where the processing takes place.

Despite the improved and new rights for data subjects, existing tools and services are designed around a "one size fit all" model which compel them to be locked-in, without any possibility to take decisions about their personal data.

Current applications of the GDPR in confined to obtaining digitally informed consent to terms of service of organizations. While this practice provides a low transparency level to the potentials uses the company has for personal data (Jones et al. 2018), more is needed to enable data subjects to not only exercise control over their data but to use their personal data in different context. A radical shift from opt-out to opt-in culture has been leveraged by GDPR in the EU, in contrast with the opt-out culture still maintained by the service providers in the United States context. In order to build trust, organizations will have to adhere to an opt-in culture.

Data portability from one service to another is currently not operational as few service providers enable personal data collection, which limits the possibility of flexible data sharing. One priority of the EU commission is to create a connected digital single market for the EU (EPC 2010) and this will be achieved by ensuring trust and confidence in digital services as well as interoperability between them. The implications of these new rules are several as they will affect the way businesses are dealing with individual's data and in parallel, will give more responsibility and right to individuals about their data. Moreover, it is important to consider a paradigm shift for new business models, along with new tools and services for personal data management.

2.3 Personal information management and current business models around data processing

Different definitions of personally identifiable information (PII) have been proposed over the years by scientific communities and by regulatory frameworks. One definition provided in the scientific communities is those of Jones and Teevan (2007) which described PII as digital information held by an individual and remaining under his direct control and responsibility. However, as the current context shows that PII are for now held by companies and managed on behalf of

individuals with or without their knowledge, this initial definition is not valid anymore.

In analyzing the definition provided in the context of data protection frameworks regulations in the United States and the European Union, we made the following observations: in the United States, where there is a diversity of amendments addressing data protection, personal information have different meanings according to a particular amendment. The personal data definition in the US varies across the states and the regulations. Some types may be considered to be personal data in one context but not in another. For example, in the California Online Privacy Protection Act (CalOPPA 2004), personal data refers to “details collected on the Internet about an individual consumer, including an individual’s first and last name, a physical street address, an email address, a telephone number, a Social Security number, or any other information that permits a specific individual to be contacted physically or online.” While in the California data breach notification law (California S.B. 1386 2003) personal data means: “an individual’s first name or first initial and last name in combination with any one or more of the following data elements, when either the name or the data elements are not encrypted:(1) Social security number, (2) Driver’s license number or California Identification Card number, (3) Account number, credit or debit card number, in combination with any required security code, access code, or password that would permit access to an individual’s financial account.” In the European Union context, a unique definition of personal information provided by the General Data Protection Regulation (GDPR 2018) states that: “personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person”. For this reason, we will retain that definition in this work.

Personal data are at the center of the business models upon which the world’s digital companies are built. The whole development of digital services has been based for the past 20 years on business models relying on advertisement subsidized by users’ data. Thus, individuals access freely online services mostly in return to their PI that become the property of their services providers. The statements of Meglena Kunewa (Kuneva 2009) “personal data is the new currency of the digital world” perfectly illustrates the fact that user’s data have become the main currency in an environment where using a service involves allowing their exploitation. The collection and sale of personal data are normal practices among companies. The problem lies in the facts that the users have no choice to select another model than the “one size fits all” model. People have poor awareness of user consent when they're subscribed to service because most of the Terms of Services (ToS) and Service Level Agreement (SLA) are too complex and generally unreadable. According to Macnish (2014), signing the form for terms and conditions when registering to use a service is often not an act of informed consent because users agree on these obscure documents without reading them in order to use the services. Even with adequate consent, there is no granularity of the data collection element and the service providers remain equally unclear on the final destination of the collected personal data.

Domains such as commercial surveillance and data trade have highly exacerbated the growth of data collection and sharing practice. Data are provisioned by entities called data brokers who are, for the most part, not consumer-facing (Crain 2016) thus posing the particular challenge of data collection transparency (FTC 2014). This practice is illegal from the view point of the convention 108+ as well as the GDPR, which among other provisions, requires the data processors to provide data subjects not only personal data access but a way to exercise their right. One transparency compliance effort taken by some data brokers consist in providing information portal which enables individuals (data subjects) to review some substantial information about them, edit and add more information and sometimes opt out (Hicken, 2013). The limitation of this effort stem from the fact that only a subset of data are disclosed with limited data collection element, the information about the source boils down to how the data are gathered from a range of anonymous data collectors and there are no information concerning the destinations of the data. According to Crain (2016), the data brokers and other surveillance entities unilaterally control the conditions under which personal data are sold. All of these issues threaten the effective implementation of the data subjects' self-determination and empowerment.

Others issues on the intensive data collection by data brokers are: the indefinite data storage, the collection of more data than needed as data can serve purposes other than the original one, the lack of guarantee about the data accuracy when the updates are not operated on a frequent basis and the lack of transparency of multiples data sources. This problem may affect the final data consumer who needs guarantees on the data origin, its accuracy, and its usability in a timely way. While individuals are increasingly aware of this state of affairs (Turow et al. 2015), their implication in the process is still hard. First, they are not enough educated about the challenges of personal data empowerment and also they lack resources to get involved. As the development of data marketplaces now stimulates the sharing of data on platforms, one could imagine this as an opportunity for them to participate in the process of data sharing. However, participation in such markets is still hard because individuals do not even own physically their personal data. Possibility to transfer all of their personal data or part of their data collection element in a structured ontology comprehensible by the parties from one service to another without being engaged in a cumbersome process is rare. As a result, there is often no interest to interact directly with them to negotiate PI access. Even if organizations were complying with the regulations, what could be the motivation of individuals to be completely engaged in provisioning their data when needed in order to not deprive the industries and the entire community of valuable data? Hence the need to define the interest of the data subjects in the outcome of data collection and sharing.

Recently, many services in the area of personal information management have been proposed to assist individuals in the collection, the re-use and the organization of their PI. These solutions focus on storing individual's information and providing features for their sole use. The supporting systems are personal information management system (PIMS) (Rustom Al Nasar et al. 2011) and personal clouds". These systems solve the problem of the personal data collection and involve only the individual side. Notable examples in the domains are summarized in the Table 2.

Personal clouds and PIMS	Features
Phlat (Cutrell et al. 2006)	Optimize personal information search engine
Memomail (Elsweiler et al. 2006)	Email management
Photomemory (Elsweiler et al. 2005)	Management of personal photographs
Stuff I've Seen (Dumais et al. 2003)	Information retrieval
My Cloud (2009)	Personal information Collection and Management
Nextcloud (2009)	Personal information Collection and Management
Cosy (2009)	Personal information Collection and Management
Freedombox(2009)	Personal information Collection and Management
Things (Things 2015)	Personal Task management

Tableau 2: Examples of personal clouds and PIMS

However, none of these approaches address personal information management in a holistic way (centered on users). In addition, a tremendous amount of PI and other data are scattered and stored in proprietary servers of organizations we know nothing about, thus hindering the freedom of individuals to keep control over their digital information. Also, such dependencies hamper the autonomy of people by creating lock-in effects and poor commercial practices. In our democratic societies, laws govern such issues as privacy, data protection, etc. This issue, taking for example the EU, is addressed by the article 18 of GDPR which introduces a right to data portability which entitles individuals to move their data from one service to another (Hertab et al. 2018). This right aims at ensuring that people regain control over their PI. If we assume a legal context, then the issue of the technical implementation needs to be addressed. Questions relating to data formats and system interoperability are important. Some initiatives such as the Data Portability Workgroup (DPWG, 2008) have attempted to address these issues through for example a Data Portability Reference Design to specify the process of developing data portability technologies.

Recently, new efforts in the domain of customer relationship management (CRM) has emerged basically aiming at improving the value of this relationship by making the customer a fully empowered actor in the marketplace. This represents a major paradigm shift. Doc Searls (Mitchell et al. 2008) brought forward this paradigm through the idea of "Intention Economy" where customer demand will drive supply efficiently and vendors will respond to the actual intentions of customers instead of aggressively looking for customer attention. He argues for Vendor Relationship Management (VRM) approaches. In this proposition, he claims that customers will more likely engage with suppliers if customer independence and privacy are better preserved. He further argues this will take place through a set of tools allowing for this reversal of relationship management. The VRM tools should allow to:

- Control the management, flow and, use of PI
- Build personal loyalty programs
- Negotiate personal terms of service
- Express the needs in terms of how, where, how much and when they want to be serviced

Following the VRM initiative, projects like Midata (Midata 2014) and Mesinfos (La Fing 2012) propose specific prototypes and trials to explore and implement such scenarios to re-empower consumers with respect to their PI and data. The purpose of these initiatives was to give back to consumers their information previously held in machine-readable format in order to allow new customer-centric services and experimentation to take place. The results were interesting as they allowed to generate new knowledge and understanding of the usage of PI by empowered customers. These initiatives have shown promising opportunities for companies to build new services leveraging trust and collaborative environments.

Similarly, tools like Personal data lockers and personal data store (PDS) (Narayanan et al. 2012) have also emerged as new business and service opportunities allowing for the control shift of PI from companies to customers. Table 3 provides some example of PDS along with their offers and associated business models. These solutions focus on giving back control to users, enabling them to 'own' their data and control access. Some collect copies of personal data from organizations in addition to storing self-generated data. And then the user can allow third parties to access or indirectly use their personal data in exchange of services. This approach has two important sides which are the individuals and the organizations interest on personal data access.

Personal Data Store	Offers	Business Models
Mydex **(Mitchell 2017) a hyper-secure storage and service to manage personal data.**	**Individuals**: Store personal data attributes. Individual's data is kept "safe", private and users can use their data to get useful insights. Individuals can provide data or "proofs" to others, for a limited time and purpose. **Organizations**: can access the personal data source.	Free for individuals. For organizations, there's a £10k set-up, and a cost of 15p per individual connected +25% after the first year, + 4% of fees paid by individuals for access.
Digi.me **(2019) enables the importation and use of personal data scattered around apps and websites by connecting various online services to the individual cloud storage app of choice.**	**Individuals**: Store copy of personal data across many accounts including social media, health, finance etc. Search and browse data. **Organizations**: Create data-driven apps with access to thousands of fields of accurate, normalized data provided directly by users. Comply with GDPR consent requirements for data processing.	No cost for start-ups that have raised less than $10M or have less than $1M revenue, no cost. Data transfer fees: between $0.10 per data transfer, max $3.00 per user/app/year; or 7.5% revenue share on fees charged for applications or app-related service.
Hub of all Things **(2019) enables a person can legally own the rights to their personal data, and benefit from all the many apps and personal AI tools that draws from all of their personal information, no matter where they are created".**	**Individuals**: Collect data once, use it everywhere. Ability to revoke access when no longer desired. **Organizations:** Avoid the risk of protecting/sharing personal data. Reduce development time by using auth, account creation and API instead of a database. Get access to users who are into the HAT ecosystem.	£4.99/month or £50/year after for individuals.

Personal Data Store	Offers	Business Models
SOLID **(2016) is a proposed set of conventions and tools for building decentralized Web applications based on linked Data principles. Enable interoperability between apps.**	**Individuals**: Choose where individual data is stored and who can access it. Solve the issue of account proliferation and avoid lock-in by being able to change providers and retain access to data. **Organizations**: Application developer can benefit from existing data created by individuals.	No Business model and pricing have not yet been announced.
Meeco **(2018) provides access, control, delegation and consent from the perspective of the individual user.**	**Individuals**: Securely manage all personal data and exchange it on individual terms. Assistance for selling data and getting value out of it. **Organizations**: Remove the burden of regulation compliance about data collection. Provide Live API for data access.	Free for consumers and Paid access to API for business.

Tableau 3: Examples of personal data store

While there are many features claimed by PDS the main focus is on the guarantee of data ownership and privacy of individuals by enabling personal data storage and access control to other organizations. One issue is that PDS does not integrate with existing services used by individuals and their social networks. PDS enables individuals to act as their own data broker where they do not possess expertise. Currently, the adoption of the PDS services by individuals are very low and organizations would not adopt it as a channel for data access if there are not enough individuals using these services. GDPR could be a potential enabler since it places major burdens and restrictions on personal data processing.

2.4 Summary

This chapter summarizes the background research that enables us to analyze the data market environment and the issues that undermine its sustainable development. We discussed the influence of technologies and the open data movement in the process of data democratization. Additionally, we presented the GDPR and the new rights of the data subjects in Europe through this reform of data protection and how it could enable new alternatives for data collection and sharing. Finally, we describe the current business models of the internet-based companies, the paradigm shift in the domain of personal information management, and how the GDPR could enable new alternatives for data collection and sharing. We summarized the issues as follow:

- The lack of services and tools allowing market actors to engage safely in this growing area of data marketplaces and the enforcement of cooperative partnership and mutual gains for parties.
- The lack of transparency around PI transaction and the incentives that individuals should get in return to their PI if they no longer want to subscribe to the current “take-it-or-leave-it” choice.
- The lack of alternative business models that take individual needs into account in order to re-create trust and user empowerment when using digital services. Businesses need to understand the new implications of

the GDPR directives and design new business models in order to comply with the new data landscape.

- Individuals' copy of personal data are not recognized as the authoritative source, and they have no control over how every other copy of it is used.

In the next chapter, we present the research problem and the methodology used to address it.

Chapitre 3. Research problem description

The research background reviewed data processing practices, in particular, data collection, sharing, and usage according to regulatory, economic, and technological perspectives. A significant number of issues in this area have been brought to light by the GDPR and the scientific literature.

In this chapter, we introduce the concept of data market ecosystems to support data exchange. We also discuss how it should enable responsible interaction among stakeholders of an ecosystem. The design of such ecosystems which represents our research challenges must be based on the effective implementation of regulatory frameworks and provide fair value to every party in such ecosystems. Next, we present the research question around which this research is built and finally, we describe our methodology.

3.1 Data market ecosystems

The term of data market ecosystem represents a system on which data are exchanged, namely they are offered and consumed by a set of parties whose collaboration is supported by some interrelated components. The analysis of the current state of data market in chapter two clearly emphasizes ethical, social and regulatory compliance issues like:

- The power imbalances at all levels between the stakeholders such as the capacity of data collection and sharing by the data subjects, the lack of tools to operate in the data offering and consumption process, the one-side terms for service consumption
- The advertising business model and the societal concerns raised by big data processing which relies on continuous tracking of online activities,
- The unfair distribution of the value to the stakeholders involved in the process of data sharing and consumption, and service accessibility
- The opacity of the data collection and sharing process for the involved market parties
- The lack of compliance to the GDPR
- The lack of adoption of current technology such as PIMS, etc.

In light of these issues, we endeavor to question the mechanisms needed to support and control these market participants while operating in such ecosystems. In fact, any market, in order for it to work properly, requires a common level of trust and transparency among the parties, which should be leveraged by some constraints and regulations that guarantee the fairness of their collaboration. Therefore, we need to lay the foundation for cooperation among the parties in a way that supports ethical and social values, comply with regulatory requirements such as GDPR and enable the design of a set of empowering mechanism and tools for the market participants.

Laudon and Laudon (2014) define ethical choices as decisions made by individuals who are responsible for the consequences of their actions. As such, data collection, and data sharing in a socially responsible way means that one should be held accountable for all the consequences of its actions. It means also that mechanisms must be established to determine the actors' responsibility, and

to provide the elements of guarantee and liabilities for operating in a data market ecosystem. Building such ecosystems requires establishing rules governing the way they work. It is only under such conditions that they can become trustworthy and consequently develop as the basis for transactions around data collection and sharing. Hence, the general problem that needs to be addressed is how to enable fair and responsible data exchange. The emergence of such ecosystems adds value to data by enhancing a trustworthy collaboration among parties which in turn will contribute to its sustainability. Different parties may participate and offer a set of value-added services for the development of such ecosystems.

3.2 Research question

Our overall research attempts to address the design of data market ecosystems which allow for social responsibility and ethics of the actors and the respect of market principles based on data collection and sharing best practices, data protection frameworks in such a way that trust and fairness are leveraged. Moreover, the design must support the traceability of the collaboration between the actors. Finally, the design will address particularly the empowerment of data subjects in these ecosystems, the cooperation in this multi-stakeholders environment and the added-values services which can enforce cooperative behaviour. Therefore, the research question addressed in this dissertation is: *How might we design a framework for fair and responsible data exchange in order to bring transparency and sustainability in data market ecosystems?*

Addressing this question requires that we consider a holistic design approach, that can be applied to any data market ecosystem, and that can mediate their internal functions, in addition to their inputs and their outputs. In fact, a data market ecosystem is an interconnected whole that is part of a larger world data market. In order to give an adequate, sufficient and comprehensive answer to the central research question the following sub-questions will be discussed:

- What are the main characteristics of data and how do they influence the exchange conditions in data market ecosystems?
- How can information retrieval be enhanced by the parties of the ecosystems?
- What mechanisms must support fair and responsible collaboration between the market participants in the data exchange process?
- What are the elements of guarantee provided at varying levels and responsibilities of the parties, in such a way that their rights be adequately protected?

To address these specific concerns, we select a research methodology described in the following section.

3.3 Research methodology

Research studies build on research methodology to carry the entire research process and reach a set of objectives. In the information systems field, different research methodologies can be applied with respect to their compatibility with the addressed problems in order to get accurate results. As the main research goal of this study is the design of a framework for fair and responsible data market

ecosystems, it means designing and building an artifact that is usable by ecosystems' parties, and that responds to the current challenges in this area. The appropriate methodology in the information system fields which satisfy this problem resolution is the Design Science Research (DSR) which has been defined as "an attempt to create things that serve human purposes" (March and Smith 1995) and "that builds and evaluates new artifacts for problems solving or improvement" (Alturki et al. 2013).

This research methodology follows the general design cycle that Vaishnavi et al. (2013) describe as a process of five steps illustrated in Figure 2. Each step is associated with its corresponding elements to develop during the research process.

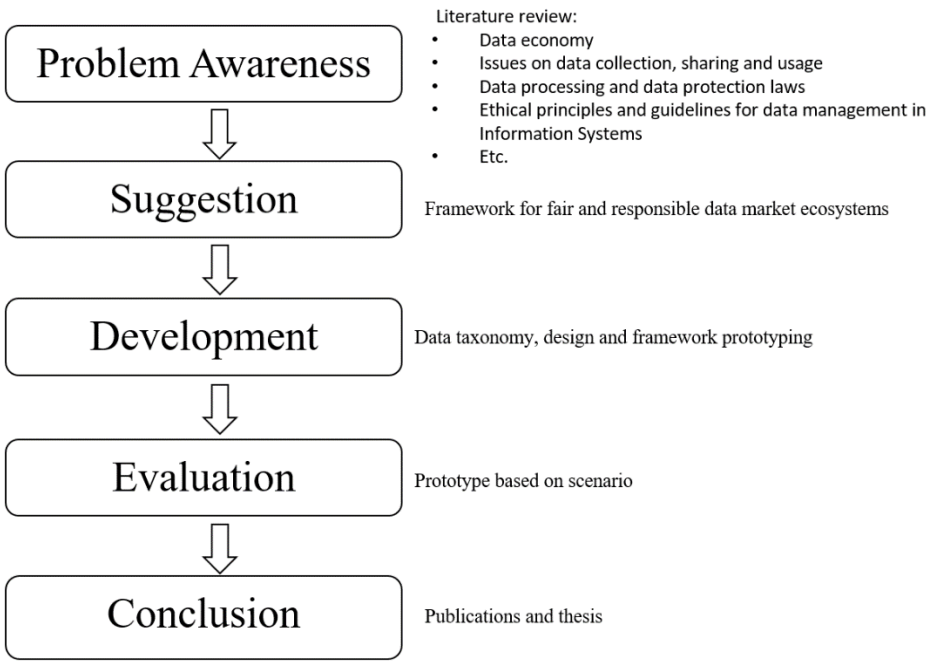

Figure 2: Design science research methodology (Vaishnavi et al. 2013)

March and Smith (1995) outline the design science framework with two axes which are the research activities and the research outputs (see Table 4). The research outputs cover constructs, models, methods and instantiations and the research activities comprise: building, evaluating, theorizing on and justifying the artifacts. In this dissertation, the framework design covers the build and the evaluation of the research artefact, as a research activity does not necessarily cover all the cells. Table 4 illustrates the cells at the intersection of research activities and research outputs of March and Smith's framework that are covered by this thesis.

RESEARCH ACTIVITIES					
RESEARCH OUTPUTS		BUILD	EVALUATE	THEORIZE	JUSTIFY
	CONSTRUCTS	Define a taxonomy for data in the context of data market ecosystems (Chapter 6)	Investigate completeness and usability (Chapter 6)		
	MODEL	Define a framework for fair and responsible data market ecosystems (chapter 7 and 8)	Framework prototyping (Chapter 10)		
	METHOD				
	INSTANTIATION	Instantiate the prototype based on a scenario (chapter 10)	Describe, demonstrate the prototype (Chapter 10)		

Tableau 4: Design Science Research Framework (March and Smith 1995)

Each cell contains a research objective addressed in a specific chapter of the dissertation. The “build” column covers the data categorization in the data market ecosystem (construct), the definition of a framework for a data market ecosystem and the prototyping of the framework. The evaluate column includes evaluating the completeness of the data categorization, and the application of the prototype based on the reference model. This research does not cover the “theorize” and “justify” columns

3.4 Summary

This chapter introduced the research problem and the research questions around this work. First, we defined the term “data market ecosystem” and summarized the current issues that undermine the sustainability of such an environment. Further, we formulated our research question which consists of designing a framework for fair and responsible data market ecosystems. Answering this question requires to handle some requirements addressed as sub-questions of this study. Finally, we ended this chapter by providing an overview of the research process followed in this work and the main artifacts delivered throughout the process. The next chapter describes the state of the art and works related to this research.

Chapitre 4. State of the art and related works

Research and initiatives on data processing are addressed in different domains with specific goals, yet sometimes complementary.

We highlighted the efforts in these domains as well as some approaches and technologies developed over the years, starting with the regulatory environment for data processing and the difference between two key players in this domain, the United States and the European Union, in their strategy for enabling a competitive environment for data processing. Then, we concluded with the viewpoints leveraged by the GDPR adoption. Second, we analyzed some proposed principles, outside the regulatory sphere, for enabling trust and fairness in data processing and globally in the internet ecosystem. The business models under which internet organizations are built strongly affect data processing activities. In this respect, we provide an analysis of current and emerging business models in the internet ecosystem, and later on, we summarize the services and solutions for enabling efficient data flow and processing. Finally, we provide an overview of the emerging privacy preserving solution for data processing and discuss how they can benefit from the future of data market ecosystems.

4.1 Regulatory perspectives for data flow enhancement and data processing

Data are precious assets that organizations are more careful to capture and retain for enhancing their competitiveness. In this respect, the European Union, for purposes of digital single market creation, has established rules for high standard protection of personal data and provided a comprehensive and coherent approach to the free movement of all data in the EU. The ultimate goal is to fully unleash the data economy benefits allowing companies and public administrations to have access to valuable data and to process them wherever they choose in the EU. Formally signed by the European Parliament and the Council on 14 November 2018, the regulation on the free flow of non-personal data ensures: "*the free movement of non-personal data across EU borders, the availability of data for regulatory control, the easier switching of data service providers for users and the EU cloud cyber-security framework, and finally the transition to a sustainable green cloud."*

4.1.1 Comparison between EU and US regulation approach on personal data protection

Regarding personal data processing regulations, we focus our attention on the unified approach to data protection across the EU and those of the United States where are located the digital tech pioneers, impacted by the EU's regulations. Although having the same root, both regulation strategies diverge in several areas.

The United State (US) and EU data protection approaches lie at the root of the Fair Information Practices (FIP) which consists of a set of principles governing the collection and use of personal information, to support a mutuality of interest between the data processor and the individual (FTC 2010). Over the years, the

US data protection laws have evolved regarding the number of amendments addressing specifics of personal data types. Table 5 presents some U.S. federal Laws that establish the conditions for handling individuals' personal information in different contexts. According to Cobb and CISSP (2016), the interests other than those of the individual have tended to prevail in US data privacy legislation, notably the interests of commerce, as well as those of state security agencies.

GENERAL FEDERAL PRIVACY LAWS	PRIVACY LAWS AFFECTING PRIVATE INSTITUTIONS
Freedom of Information Act of 1966 as Amended (5 USC 552)	Fair Credit Reporting Act of 1970
Privacy Act of 1974 as Amended (5 USC 552a)	Family Educational Rights and Privacy Act of 1974
Electronic Communications Privacy Act of 1986	Right to Financial Privacy Act of 1978
Computer Matching and Privacy Protection Act of 1988	Privacy Protection Act of 1980
Computer Security Act of 1987	Cable Communications Policy Act of 1984
Federal Managers Financial Integrity Act of 1982	Electronic Communications Privacy Act of 1986
Driver's Privacy Protection Act of 1994	Video Privacy Protection Act of 1988
E-Government Act of 2002	The Health Insurance Portability and Accountability Act of 1996 (HIPAA)
	Children's Online Privacy Protection Act (COPPA) of 1998
	Financial Modernization Act (Gramm-Leach-Bliley Act) of 1999

Tableau 5: Example of US data protection Laws (Laudon and Laudon, 2014).

While most of the U.S. federal privacy laws apply only to the federal government, they regulate very few areas of the private sector (Laudon and Laudon, 2014). As stated by Crain (2016) there is no comprehensive federal law governing the commercial collection of personal information, and only a few privacy protections regarding certain types of data are covered by their disparate statutes. Table 6 compares the US approach to the EU's, in terms of principles and individual rights.

In contrast to the US approach, The EU data protection framework addresses all personal data types and covers also the industry data processing activities. It provides more general and stringent rules which bring new rights to European citizens and binds some design requirements that will impact future services for data processing and service delivery.

Data Protection and Security Frameworks	Key principles						Individual rights								
	Transparency	Lawful basis for processing	Purpose limitation	Data minimization	Proportionality	Retention	Right of access to data or copies of data	Right to rectification of errors	Right to be forgotten	Right to object to processing	Right to restrict processing	Right to data portability	Right to withdraw consent	Right to object to marketing	Right to comply to the relevant data protection authority
General Data Protection Framework (GDPR)	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Health Insurance Portability and Accountability Act of 1996 (HIPAA)							X					X			
CAN-SPAM Act										X				X	
Fair Credit Reporting Act							X								
California Children's Online Privacy Protection Act Rule (COPPA Rule)									X						
Telephone Consumer Protection Act										X			X	X	
California's Shine the Light Act														X	

Tableau 6: Comparison of USA data protection principles and individual rights and EU

4.1.2 Diverging debate on a property right for personal data

Proposals to property rights in personal data which has emerged in 1970 in the United States have been accentuated by the recent improvement of data subjects' rights. In particular, those provided by the GDPR which are perceived by some as the missed opportunity to introduce a property right on personal data. Arguments put forward to justify property rights on personal data are the privacy-protective potential of property rights, the economic interest, and investment in data and finally the fair sharing of data generated wealth and data access (Purtova 2017).

A strong proponent of this view, the British jurist Christopher Rees (2013) argues for the extension of ownership right over personal data in such a way that enables data subjects to use them, whether by transferring them or not, temporarily or definitively, in whole or in part. Likewise, the "Generation Libre" movement (Landreau, 2018) argues that the GDPR failed to really empower individuals who cannot negotiate their data usage by organizations. Moreover, they argue that data markets for personal data could re-balance the power between platforms and individuals, by endowing each party with real capital. According to (Belleil 2009), this could help to fight direct marketing as selling personal data would increase the cost of producing these campaigns, which would, in turn, discourage businesses from collecting such data. However, a property right implies to waive personal data protection guarantees, hence relinquishing total control over personal data once sell to another party.

Some arguments have also been made against property rights on personal data. As discussed by the (AEDH 2017), a property right on personal data would not serve individuals interests nor resolve the difficulties associated with accessing digital services, because individuals are forced, under the appearance of consent, to renounce either services access or their private life.

Article 8 of the Charter of Fundamental Rights of the European Union

1. Everyone has the right to the protection of personal data concerning him or her.

2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.

3. Compliance with these rules shall be subject to control by an independent authority.

In addition, personal data protection is a fundamental right provided for in Article 8 of the Charter of Fundamental Rights of the European Union (europarl 2010). On this basis, a complete alienation of personal data for waived or reward would be unfeasible (Peres 2015). Another viewpoint, held notably by the essayist Evgeny Morozov (Calimaq, 2017), advocates for the availability of personal data in the public domain and unalienable, which guarantee the right to access and use to the whole community. Organizations will use them by paying for license fees, and such mechanism will ensure that companies do not impose their access conditions to data contributed by the community.

Regardless of what might be the viewpoints, whether the GDPR truly enforces or not the right of individuals, it will have wide implications for the digital economy, and maybe, enable the emergence of new business models in the internet ecosystem.

4.2 Emerging business models

Among the most dominant business models in the digital area is the advertising model, which appears to be systematically integrated by internet organizations which deliver free services and, in return, collect and share their user's data with advertisers. Thus, part of their revenue is generated by selling or monetizing their user's data. Not only High-Tech organizations having an advertising model are involved in this practice of data monetization. Many organizations such as telecom service providers (Deloitte 2014) have got the opportunity to extend their revenue from selling their customers' data. To share the value of these data with their service users, Telefonica (Cryptonomis 2016), a telecommunication organization has launched a decentralized platform that enables their customer to trade their personal data and certificates in exchange for remuneration.

Hence, emerging business models are integrating users as a beneficiary of their information. For instance, Connexions Asia (CXA) provides solutions to analyze companies' insurance spending and the problem affecting their workforce health and welfare with the goal to improve the health of the workforce. The employees of the insurance companies participate by providing information about their claims and update their coverage information according to their need. In return, they may get some benefits and also get rewarded if they are getting healthier.

These business models are attractive because they integrate the participation of data subjects in the collection and usage of their data. Moreover, data subjects get a reward for their participation according to some conditions. While these business models enable individuals to participate in the data collection and sharing process, those individuals cannot take control of their data as the service provider retain data and can continue to profit from them. Furthermore, the service providers unilaterally defined the conditions for service usage.

The GDPR gives an opportunity to explore new business models better aligned with the new regulations principles and data subjects' rights. Alternative business models, described in Table 7, are for the moment the subscription model which deliver superior service for loyal customer base and micropayment which enables users to pay small amounts for product or service access.

Business models	Description	Examples of Organizations
Free or Advertisement Model	It involves selling personal data harvested from free product or service users.	Facebook, Google
Freemium Model	Users pay for a basic product or service with their data and can get charge when upgrading to full service or product usage.	LinkedIn, Vimeo, Flickr, Spotify
Subscription Model	A customer pay subscription fee to a vendor for continued access to a product or service.	Netflix, Apple Music
Micropayment Model	It enables a user to pay fractions of penny for product or service usage	Tsu
Community-based model	In this model creator of content get directly rewarded through micropayment.	Steem

Tableau 7: Business models descriptions and examples

The crypto-currencies has enforced micropayment models which allow paying lower cost service. Another business model is about investment by participation and consumption. Steem (2018) is such a blockchain-based social media platform which reward their users for content creation and viewers' attraction. In the "Steem" model, all data are public while in the advertisement model, all data are the private property of the platform. More steps are needed to empower individuals with tools and services that will allow them to act proactively, making their own choice in terms of services and data processing. Furthermore, the internet ecosystem must become a fair environment, the enabler of responsible service delivery and data processing.

4.3 Some guidelines for fairness and compliance in the internet and data ecosystems

With the aim to provide global approaches to enhance trust and fairness in the internet ecosystem regarding service providers, platforms and data usage, the CNNum (CNNum, 2015) has provided an activity report which describes a principle of faithfulness of the services and suggests some guidance on its application. The main elements of this principle are:

- Transparency of service behavior in order to ensure compliance between the stated service promises and actual practices.
- Compliance with a general principle of non-discrimination of users regarding the faithfulness of algorithms for customization, indexing, and ranking, the legibility, and disambiguation of TOS.
- Faithfulness among economic actors regarding economic conditions of access to platforms and the conditions for opening services to third parties.

Research conducted by Zook et al. (2017) for enhancing fairness and responsibility in data processing has suggested “ten simple rules” for addressing the complex ethical issues in research-based big data summarized in Table 8.

Ten principles for big data research
Principle 1: Acknowledge that data are people and can do harm. Start with the assumption that data are people (until proven otherwise), and use it to guide your analysis.
Principle 2: **Recognize that privacy is more than a binary value**. Situate and contextualize data to anticipate privacy breaches and minimize harm. The availability or perceived public-ness of data does not guarantee lack of harm, nor does it mean that data creators consent to researchers using their data.
Principle 3: Guard against the re-identification of your data. Identify possible vectors of re-identification in your data. Work to minimize them in your published results to the greatest extent possible.
Principle 4: Practice ethical data sharing. Researchers should consider the best interests of the human participant, proactively considering the likelihood of privacy breaches and re-identification issues.
Principle 5: Consider the strengths and limitations of your data; big does not automatically mean better. In order to do both accurate and responsible big data research, it is important to ground datasets in their proper context including conflicts of interest. Document the provenance and evolution of your data. Do not overstate clarity; acknowledge messiness and multiple meanings.
Principle 6: Debate the tough, ethical choices. Engage your colleagues and students about ethical practice for big data research.
Principle 7: Develop a code of conduct for your organization, research community, or industry. Establish appropriate codes of ethical conduct within your community. Make industry researchers and representatives of affected communities’ active contributors to this process.
Principle 8: Design your data and systems for auditability. Responsible internal auditing processes flow easily into audit systems and also keep track of factors that might contribute to problematic outcomes. Systems of auditability clarify how different datasets (and the subsequent analysis) differ from each other, aiding understanding and creating better research.
Principle 9: Engage with the broader consequences of data and analysis practices. Recognize that doing big data research has societal-wide effects.
Principle 10: Know when to break these rules. It is important to recognize when it is appropriate to stray from these rules, especially, in case of emergency. Nonetheless, the review of regulatory expectations and legal demands associated with dataset privacy protection must be carried.

Tableau 8: Ten rules for responsible big data research (Zook et al. 2017)

They argue that “*Statements to the effect that “Data is already public” are unjustified simplifications of much more complex data ecosystems embedded in even more complex and contingent social practices*” (Zook et al. 2017). These principles aim to direct researchers by recognizing the human participants and

complex systems contained within their data and to deal with ethical questions of their big data management workflow. We believe that some of these principles can be generalized to other data ecosystems where social and ethical issues are noticeable.

4.4 Consent and agreements for data processing and service usage

4.4.1 Consent and agreement for personal data processing

The GDPR provides six legal grounds required for the lawful processing of personal data (GDPR 2018).

Lawful bases for personal data processing Article 6 of the GDPR (2018)

(a) **Consent**: the individual has given clear consent for you to process their personal data for a specific purpose.

(b) **Contract**: the processing is necessary for a contract you have with the individual, or because they have asked you to take specific steps before entering into a contract.

(c) **Legal obligation**: the processing is necessary for you to comply with the law (not including contractual obligations).

(d) **Vital interests**: the processing is necessary to protect someone's life.

(e) **Public task**: the processing is necessary for you to perform a task in the public interest or for your official functions, and the task or function has a clear basis in law.

(f) **Legitimate interests**: the processing is necessary for your legitimate interests or the legitimate interests of a third party, unless there is a good reason to protect the individual's personal data which overrides those legitimate interests. (This cannot apply if you are a public authority processing data to perform your official tasks.)

The new requirements under the GDPR for consent, described in the chapter 2 section (2.2), entail organizations to invest in consent management mechanisms which cover the whole consent lifecycle, prove the validity of the collected consent, and enable data subjects to exercise their rights. Some commercial solutions are TrustArc GDPR compliance (2019), OneTrust GDPR consent management platform (2019) and the Evidon GDPR consent solution (2016). The main functionalities offer by these solutions are:

- Data flow mapping to standardize and operationalize the mapping process of customers and organizations data flows
- Operationalize Data Protection Impact Assessment (DPIA) and Privacy by Design

- Meet EU Privacy Cookie Compliance Requirements
- Provide a Data Subject Rights Request Portal
- Provide a Framework for Consent Management
- Prepare an Incident Reporting & Breach Management Workflow
- Review and Remediate Vendor Risks
- 1st-Party and 3rd-Party consent controls

Data processing agreement (DPA) is the legally binding document to be considered between the controller and the processor in writing or in electronic form for personal data processing. It demonstrates the compliance of data processors with GDPR by providing sufficient guarantees for the protection of the data transferred to them. The elements of the DPA are (see the checklist in Appendix C):

- **The object of the agreement** concerns all activities related to the contractual relationship between partners.
- **The scope, nature, and duration of data processing** describe the usage of personal data and the party responsible for ensuring that the processed data meets the requirements of GDPR.
- **The subjects of data processing** define the category of data subjects.
- **Data category** that will be handled by a data processor. A special data category should be differentiate to the regular types of personal data as they should be processed in a more restricted fashion.
- **Data storage restrictions** related to the transfer of data beyond EU borders must be satisfied. In that case, data processors must describe the steps undertaken to ensure a level of security equivalent to that provided within the EU.
- **Terms and conditions of contract termination** include information regarding the controller's clients data that should be removed from the processor's databases and enumerate cases in which each party has a right to terminate the agreement.

Both consent and DPA govern the contractual relationship within the GDPR bounds and focus on the granularity of both processing activities and the data elements. Data processing clauses are usually specified in legal documents such as the DPA and are not yet subject to automated processing for enforcement.

4.4.2 Other agreements for data processing

Different studies have proposed solutions to automate the processing of data sharing agreements. Hence, Egea et al. (2015) addressed this issue by defining a machine process-able multilateral contract based on the Italian data protection Law. The proposed solution is composed of three elements: the predefined legal background information which encodes the law for the sharing of personal data, the rules specific to the domain related to data collection, the sharing preferences of the data subjects along with some other adjustment to control their data disclosure. In the context of data as a service (DaaS), Truong et al. (2011) have defined models for encapsulating data processing agreements and for exchanging data agreements among DaaS service providers, data providers, and data consumers. Furthermore, they proposed a data agreement exchanging service (DAES) for enabling the composition, analysis, and management of these

agreements. The agreement is composed of: the agreement types (such as data licensing, data privacy compliance, quality of data), the agreement identifier, the data asset, the source, the data asset provider, the data asset consumer, the creation date, the DAES, and the agreement status. The DAES is used to check the compatibility of the data privacy policies and the service that consumes the data asset. Troung et al. (2012) have introduced an abstract model for data contracts that can be used to build different types of data contracts for specific types of data. Based on the abstract model, they also propose several techniques for evaluating data contracts that can be integrated into data service selection and composition frameworks.

4.4.3 Agreement for service delivery

Different service agreements exist for defining the relationship between an end-user client and their service providers. The cloud standards customer council (CSCC, 2015) defines the cloud service agreements as a set of three artefacts which are: TOS which carries the explicit definitions of the roles, responsibilities, and execution of processes, the acceptable use policy (AUP) which defines illegal use of service and the service level agreement (SLA) which includes the metrics for the levels of service.

Ludwig et al. (2015) defined an SLA specification in an XML Schema composed of the following elements: the stakeholders, the services definition where the SLA parameters and metrics are described and the obligations which specify the service level objectives (SLO) and corresponding actions.

Limited efforts have been made to address the processing of agreements on data asset delivery along with the supporting service in the DaaS. Vu et al. (2012) propose to solve the separation of information about provided services for data provisioning and supplied data assets. They define a cloud data service which describes data provisioning agreement at data asset level and service level. The service level provides a general description of DaaS and the data asset level includes information specific to particular data assets. Each agreement explicitly states the service that the end-user expects to receive from a service provider and clarify the performance metrics used to measure the service quality. In case of any disagreements around the delivered service, all involved parties must turn to the service agreement to resolve the dispute.

4.5 Data services and marketplaces

In recent years, many solutions for data exchange have emerged. The first ones were centralized cloud-based data services and data marketplaces such as Amazon Data Sets (2019), Factual (2019), Gnip (2019), Azure Marketplace (2017), and Xignite (2016), Marsa (2017), which allowed the exchange of different data type from public data, finance data, IOT, social network data etc, delivered on batch, near real-time and real-time. Using these platforms, a data provider can upload his or her data manually or automatically using APIs.

However, new categories of data marketplaces have emerged adopting a decentralized model leveraged by Distributed Ledger Technology (DLT). One argument is about removing silos created by centralized data marketplaces which have limited offering of datasets and constitute the central authority for pricing

data or get fees in all data transactions. Another argument is that these marketplaces are not suitable for personal data exchange as they do not empower data subject in term of data access authorization and usage control.

4.5.1 Consent and agreement for personal data processing

A distributed ledger is an asset database shared across a network of multiple nodes. Participants within a network can retain an identical copy of the ledger and each decentralized copies reflect any changes to the ledger. The security and accuracy of the assets stored in the ledger are maintained cryptographically through the use of 'keys' and signatures (UK Government 2016).

A sub-category of this technology is the Blockchain, a time-stamped series of immutable records of data called blocks, linked and cryptographically secured. Blockchains can be classified into two categories depending on the access permission they offered: permissionless blockchain which is completely open and where anyone can join and operate in the network without restriction other than economic ones and permissioned blockchain which is a closed and monitored ecosystem where the access of each participant is well defined and differentiated based on role. One feature is its distributed nature and management through consensus algorithm which guarantees an unambiguous ordering of transactions and blocks (Androulaki et al.2018). Moreover, some Blockchains support the execution of smart contracts which are: "*computerized transaction protocols that execute terms of a contract*" (Szabo 1997). The blockchain design bound intrinsic trust which enables its use as a trusted layer in the design of solutions for data processing. These solutions, having different properties depending on their specific use, allow for:

- Decentralization of data source: moving away from single-source data to data source decentralization.
- Data providers' empowerment through self-data sharing: data owners can make their data available to others and to directly benefit from the incentives.
- Removal of central authority for data pricing, transactions management, and data storage. Data providers can set their own prices. The smart contract enables automated, secure and fast transactions with no representatives and no fees. Moreover, the traceability feature of the blockchain allows complete data transaction traceability.
- Micropayment infrastructure allows buyers to pay small amounts for data consumption.

Because of the diversity of use cases in this area, we categorized data marketplaces by the predominant type of data exchange and their main features in Table 9.

Features	Personal data	Business data	IOT data
Value proposition	Personal data monetization	Business data exchange	Device monetization
Transaction type	Business-to-consumer	Business-to-business	Machine-to-machine
Data type	Personal data	proprietary data	IOT data Stream
Interface	App (data subjects) &API (Buyers)	API	API
Pricing	Pay-per-user	Pay-per-datapoint	Pay-per-hour

Tableau 9: Type of data marketplaces and features

Personal data marketplaces aim at empowering data subjects by allowing them to monetize their data directly and on their own terms. Examples include Datum (Haenni 2017), DataWallet (2014) and physical which enable individuals to trade any range of their data such as social media streams, their location, etc.

Business data marketplaces are designed to enable efficient business-to-business data exchange. Ocean Protocol is an example which allows organizations to trade computerized artificial intelligence (AI) data. Another example is DX Network which provides platforms to trade enterprise knowledge such as industry-specific data or scientific experiment results. Data is aggregated within the marketplace and is thus served ready for immediate use.

Sensor data marketplaces allow for the purchase of real-time data feeds from remote devices. For instance, IOTA Data Market (2015), DataBroker DAO (2018), Datapace (Draskovic and Saleh 2017) and Streamr (2017) offer pollution, power grid and vehicle telematics data feeds. The characteristic property of sensor data marketplaces is the real-time nature of the data for sale.

While proponents of the blockchain advocate for the end of existing centralized models, which concentrate the power in the hand of the marketplace providers, we should admit that the decentralized model prone of the blockchain suffers as centralization of power prevail at some level. One reason is the private ownership of Initial Coin Offering organizations (ICO), which are most of the time centrally ruled. Therefore, there is a concentration of power which enables the centralization of decision over the future of their services such as the adding of new rules. Hence, they constitute implicitly trusted parties. The Constantinople hard fork execution and cancellation of Ethereum (Buck, 2019) is an example of decision largely influenced by Ethereum core developers and security community. Another example is the governance model of Tezos (Goodman 2014) which utilizes a method called, on-chain governance. In this model, developers can submit upgrades proposal accepted based on the vote of token holders. Thus, the more a participant has token, the more he influences the direction of the network.

Moreover, the functionality that DLT provides such as validation and verification mechanisms, has traditionally been implemented with a trusted third party. Locher et al. (2018) addressed the issue of fully replacing a trusted party. They provide two essential criteria that must be met for adopting a ledger-based approach

without relying on any particular party in the system. These criteria show that the DLT only solve the trust issue when it is recognized as the supreme authority in that its consensus protocol controls internally the process of object creation and the predicate verification process. In the blockchain-based solutions, data represent object which are created outside of the consensus protocol. Therefore, to ensure data authenticity, we need to rely on third-parties, for data marketplace enables only to store the hash of data in blockchain to prove data source origin and that data have been untampered from the day of their registration in the blockchain. This is an important requirement for ensuring trust in data exchange.

4.5.2 Responsible use of blockchain for personal data processing

The very characteristics the Blockchain, namely transparency, immutability, and decentralization raises concerns about GDPR compliance when dealing with personal data processing. While the GDPR were designed in a centralized model of data management model where the responsible entities are strictly defined, the Blockchain has a decentralized data management model and the liability of the multiplicity of actors involved in data processing is hardly established. In this respect, a number of requirements have been proposed (Bundesblock 2018, CNIL 2018) to respond to the main issues such as the identification of personal data written in a blockchain, the responsibility of the blockchain' stakeholders and the exercise of the data subjects right in the blockchain ecosystem. Table 10 summarizes the issues and the key considerations to deal with them.

Issue: Personal data identification	
Description	**Key considerations**
Public Keys that can be associated with a natural person. Others personal data can be stored in the blockchain. Depending on use case, hashed personal data and encrypted personal data can be considered pseudonymous not anonymous.	Personal data should be truly anonymized before stored on a blockchain when it can be associated with a data subject. Zero proof knowledge could be applied in blockchain to prove data possession without revealing the content of personal data.
Issue: Legal status of the stakeholders	
Description	**Key considerations**
Determination of the actor responsible of data processing.	Participants who have a writing right on the chain and who decide to submit data for validation by minors may be considered as data controllers for these data. The users are in control of the processing of their personal data, which may be operated by a data processor. Service and applications providers that determine the purpose and means of personal data processing are likely data controllers.
Nodes and miners: difficulty to conclude data processing agreements with them as they are allowed to participate without permission from a central party.	Nodes and miners do not decide what data is written to the blockchain, the means, and purposes of the processing of data. They should be considered as infrastructure. Miners when executing the instructions of the controller, they have to check that the transaction meets technical criteria.
Developer of the blockchain are not considered controllers or data processors.	The developers of "smart contracts", who process personal data on behalf of the controller may endorse the role of data processors or data controller depending on their role in determining the purposes of the processing.

Issue: Data subject rights	
Description	**Key considerations**
Right to erasure	Anonymization procedure should be considered to be an alternative way of erasing data
Right to restriction of processing. Fully automated decision from a smart contract is necessary for its execution	Restriction on data processing may be required for application developers who are controllers but not on the part of nodes on a public blockchain network. The controller should therefore provide for the possibility of human intervention to challenge the decision by allowing the data subject to challenge the decision, even if the contract has already been performed, regardless of what is written in the Blockchain (Bundesblock 2018)
Right to data portability	Data written to a public blockchain should be deemed to comply with data portability requirements.

Tableau 10: Issues and the key considerations for blockchain use

The issue of personal data transfers outside the EU is problematic in the context of a public blockchain where control over the location of participants is difficult to exercise.

In general, personal data marketplaces deal with some of the above issues such as storing the hash of personal data in the blockchain while keeping the original dataset store off-chain. However, they do not establish the responsibility of each participant and the way for a data subjects to exercise their rights in terms of smart contracts executions, the portability of data across services and erasure of personal data, whether in public or permissioned blockchain. In order to comply with personal data processing, different approaches are being considered over the years through the design of trust architectures.

4.6 Trust systems and architectures

One of the most challenging research in this decade concerns the design of systems for data subject privacy protection and control during data processing activities. Different perspectives have been adopted for solving these issues.

One approach proposed by Mont at al. (2013) is to stick machine-readable policies to user's data in order to define the usage and obligations as it travels across multiple parties and to protect an unauthorized data sharing and usage. The proposed model including the obfuscation of personal information before it leaves users' premises, association of "tamper-resistant" sticky policies defined by users to the obfuscated data, the disclosure of data subject to the fulfillment of the sticky policies' constraints, enforced tracing and auditing of disclosures of confidential data, to increase data receivers' accountability. However, this system does not guarantee that data consumers cannot re-share the decrypted data and there is no means to penalize the data consumer for non-compliance behavior.

Another approach (Iyilade and Vassileva 2013) propose a framework that allows the data owner to log any access to his data and any illegal copies of his data together with an auditing mechanism. However, the framework provides no means of penalizing data consumers when data misuse is detected. Noorian et al. (2014) propose a trust mechanism that detects contract breaches of privacy in a provider-consumer marketplace for sharing user data after interactions have taken place,

based on incoming evidence of bad behavior or poor quality of service (complaints). However, the framework relies on the assumption that there is evidence of privacy violation that will trigger a user complaint.

Recently, another paradigm on trust architectures has emerged due to the need for collaborative data sharing. The first approach is to provide privacy-preserving mechanisms and data analysis algorithms and solve the issue of trust in data sharing, by enabling computation on encrypted data rather than making data accessible by all the parties involved in the computation. In the data mining domain, Zhan (2015) studied a methodology for performing data mining without data disclosure between the parties. This methodology relies on homomorphic encryption and digital envelope techniques. Similarly, Thuraisingham (2015), Malik et al. (2012) and Ashok et al. (2015) focused on the concept of privacy preserving data mining. Martinelli et al. (2016) propose a generic framework for privacy preserving mechanisms and data analysis algorithms which aims at being applicable in different contexts. The framework which implements a set of privacy rules associated with privacy mechanisms defines indexes to measure the compatibility between privacy requirements and includes a novel method to compute the trade-off between privacy and the analysis accuracy.

Egorov and Wilkison (2016) have designed an end-to-end encrypted database called zeroDB that enables clients to operate on encrypted data without exposing encryption keys or clear text data to the database server.

Zyskind et al. (2015) designed the Enigma protocol that is a decentralized network for running computations on encrypted data. Enigma uses the MPC structure to enable distributed data queries without the need for a trusted third-party. Data are split among differing nodes, which then compute functions at the same time without sharing information with other nodes. Enigma integrates blockchain technology to store public data and reference to private data. Use cases for Enigma live within the subset of solutions where datasets are needed to be consumed as data input for private computations. Dimutri and Gatti (2016) proposed a reference architecture of trusted data marketplaces for credit scoring data. They also rely on homomorphic encryption and MPC for data encryption and a blockchain technology to remove the need of trusted party for handling the transaction in the data market.

4.7 Summary

Many efforts have been achieved for enabling security and control while processing data. Recently, the GDPR, more than any data protection framework, has set the course for a paradigm change concerning services design for the foreseeable future.

This chapter highlighted the effort in regulating data processing activities, particularly, the enhancement of data subjects' right, the free flow of data in the EU market and the legally binding element for establishing relationships within the GDPR bounds. We provide an overview of the current business models, which are partly responsible for the lack of individuals' empowerment in the internet ecosystem. Following, we give some example of business models which cooperate with individuals at a certain level. Later, we discuss the need for individuals to engage differently with service providers by being able to negotiate

service agreements or better set the terms of agreements according to their need and those of the services. Moreover, we analyzed current solutions for enabling data exchange by data marketplaces in a centralized and decentralized model. The blockchain technology, highly adopted in the decentralized data marketplaces, is considered to solve the issues of trust, control privacy and disintermediation. However, an analysis of the blockchain governance and the criteria for total disintermediation show that the area of data exchange is not free from trusted third parties. Moreover, the compliance with the GDPR is hardly achievable in a public blockchain where roles and control are difficult to handle. To the question of the real benefit of personal data trade, Searls (2018) argues that “individuals do not get scale with organization by selling their data.” He call for personal agency that will enable individual to deal equaly with organization. Finally, we presented some research on privacy-preserving and trust in data processing. The proposed solutions addressed specific use- cases which are useful for controlling data access and usage.

All of these solutions and approaches are focused on privacy and control of data processing and are domain specific. They do not consider elements of guarantees for a fair collaboration between the participants of data market ecosystems, nor individuals' needs regarding service offering. The lack of these services adoption in responding to the participants’ interests may justify their low level of adoption and that more is needed to create fair and responsible data market ecosystems. Our approach aims at being freed from centralization and decentralization constraints, but focus on the need of the main ecosystem actors and their collaboration to create sustainable ecosystems. In the Next chapter, we define the requirements for designing a framework for fair and responsible data market ecosystems.

Chapitre 5. Design requirements for responsible and responsible data market ecosystems

The research question around this work focuses on designing a framework for fair and sustainable data market ecosystems. This requires new cooperation models among ecosystem participant in a way that:

- Supports ethical and social values shared by the participants,
- Enforce the equitable sharing of benefits resulting from data processing,
- Comply with regulatory requirements in the data protection area,
- Promote the design empowering mechanism for the participants,
- Encourage the emergence of new business models for service delivery.

Following these goals, we discusses the design requirements in this present chapter. First, we analyze ethical considerations for designing such ecosystems. Second, we identify the main actors, their functions, and their interactions with each other. Third, we consider generally accepted ethical principles applied to the use of information technology that may also be applied to these ecosystems. Following, we identified those relevant to our context and used these principles as a code of responsibility of parties operating in these environments. Finally, we analyze data characteristics and the requirement for their exchange data market ecosystems. Ultimately, we propose key requirements that lead to the design of the framework.

5.1 Ethical considerations for designing data market ecosystems

The domain of data exchange and services access supports a number of actors whose decisions and actions greatly impact the digital environment. Few corporations preempt digital markets with their own rules in closed ecosystems, limiting access to data and services, and pervasive commercial surveillance remains the standard. In this complex environment, paying attention to the ethical responsibilities of each actor is essential to consider sound and transparent principles for data market ecosystems regulation and provide better opportunities for each party. Accordingly, we use the five-step process describes in figure 3 (Laudon and Laudon 2014) to perform an ethical analysis in our context.

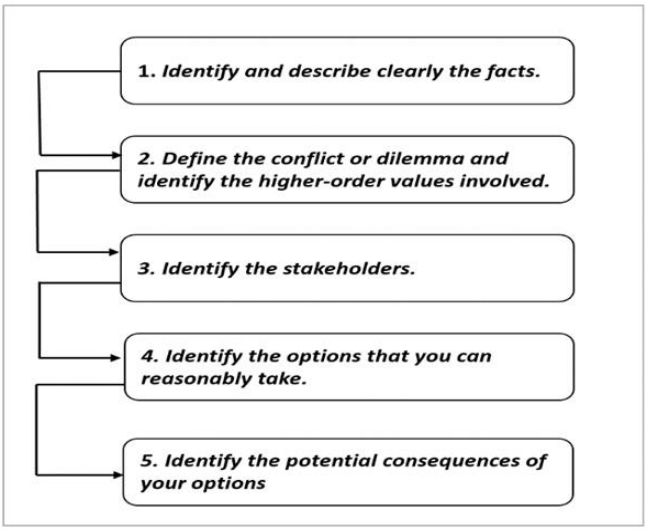

Figure 3: Five-step process for an ethical analysis in our context (Laudon and Laudon 2014)

The first step involves identifying and describing clearly the facts. This step has been done in the research background where the analysis of the current environment of data exchange and service access shed light on identified social and ethical issues. Parties have several competing values which lead to different conflicts of interest. In order to point out these competing values, we proceed with the actor's identification and their function. The literature review in Chapter 4 enables the identification of the following parties clustered into fives main roles described below.

- **Individual (Service Users or Data Subject).** Data are generated as service users mediate their lives with information technologies, more specifically, with the intensive use of digital devices and services. These devices and services maintain an ongoing relationship with their users and therefore enable the extensible collection, revision, and extension of their data. Users represent individuals using devices and services to perform some tasks including consuming the information provided by service providers. This result in the production of digital traces containing a rich corpus of data including personal data.
- **Service providers (Device and Service Providers).** Not only do they rule how Individual interact and operate their services but they also dictate which personal data must be disclosed for service access and how Individual can capture these data. In this context, we consider two moral values for enhancing individuals' self-determination on service terms and conditions and also for data processing. The first is individual privacy which enables an individual to express selectively the boundaries and content of data disclosed to services providers. These data boundaries and contents must be evaluated according to the legitimate use of personal data by service providers. The second concerns the freedom of service access according to various options that satisfy individual needs and offer equitable opportunities for their participation in the digital age.

- **Data providers.** They are responsible for supplying data for consumption. By their very nature service providers collect tremendous amounts of data, which enable them to assume the role of data providers, thus creating conflicts of interests with Individuals. This issue hampers the process of individual's empowerment when it comes to personal data exchange, in an environment where the authoritative sources of personal data are unestablished. Also, it is difficult for individuals to satisfy the demand for data supply as they lack experience and appropriate tools.
- **Data consumers.** They consume data with the help of applications and services or through web API. The compliance with data usage terms is hardly assessable once data are gathered. Moreover, Data Consumers rely on data brokers or services providers for data supply.
- **Data broker.** They are non-consumer facing entity that captures data from diverse sources and then sell them to different parties. Data provenance, data usage (transparency and traceability), as well as data quality are not open to inspection, hence hardly enforceable.
- **Data Marketplaces.** They provide services for data localization and for matching data demand and supply. In some cases, Data Marketplace assumes the role of data aggregator, impose their own rule in terms of participation, or set out the fees for data access. As they play the role of trusted third parties, they inherited the risk associated with these functions such as: the blindness of their activities regarding other parties which may lead to fraud and abuse, the entry barriers set by trusted third parties as they become prominent within a given industry.
- **Regulators.** These entities oversees the enforcement of personal data and information legislation. In the context of GDPR, we identify two entities involved in the enforcement process:
  Data Protection Authorities or (Information and data commissioner) are responsible for supervising, through investigative and corrective powers, the application of the data protection law. They provide expert advice on data protection issues and handle complaints lodged against violations of the GDPR and the relevant national laws.
  Data protection officers (DPO) are responsible for "working towards the compliance with all relevant data protection laws, monitoring specific processes, such as data protection impact assessments, increasing employee awareness for data protection and training them accordingly, as well as collaborating with the supervisory authorities." (*Art. 39 GDPR Tasks of the data protection officer)*. An organization should appoint an internal or external DPO if: the processing is carried out by a public authority or body; the core activities of the controller or the processor consist of processing operations, which require regular and systematic monitoring of data subjects on a large scale or the core activities of the controller or the processor consist of processing on a large scale of special categories of data or personal data relating to criminal convictions and offenses. These entities play a significant role of trusted parties which attest the compliance or not of organizations to data protection laws.

5.2 General requirements for a fair and responsible data market ecosystem

A data market ecosystem should be designed as a vector for value creation, service access, and data exchange. It should be composed of dependable and globally responsible parties for its sustainability. A data market ecosystem should support the empowerment of parties, the development of fair cooperation between them, and guarantee the flow of data and access to services. As the substantive law contains specific rules aiming to protect contracting parties usually considered to be the 'weaker party' (Van Bochove 2014), we identify individuals the weaker party whose capability should be enhanced by default.

The key considerations for defining the requirement stems from the previous analysis of key actors and the ethical analysis of data exchange. We use candidate ethical principles in information systems proposed by Laudon and Laudon (2014) that constitute our foundation to work towards the highest possible standards of integrity in data market ecosystems. These principles are the following:

- **The golden rule**: Do unto others as you would have them do unto you.
- **The Immanuel Kant's categorical imperative**: If an action is not right for everyone to take, it is not right for anyone.
- **Descartes' rule of change:** If an action cannot be taken repeatedly, it is not right to take at all.
- **The utilitarian principle:** Take the action that achieves the higher or greater value.
- **The risk aversion principle:** Take the action that produces the least harm or the least potential cost.
- **The no free lunch rule:** Assuming that data or services are useful to any party, they have value, and it should be assumed that the provider requires compensation for them.

A starting point is to use these general principles to address the key ethical issues inherent to all market environment. For addressing fairness and responsibility of the participants in data market ecosystems, one must consider their diverging interests and conflicting desires, for ethical assessments of ecosystems. The categorical imperative of Kant defines an action as ethical if it could become a general law, which would allow all other parties to behave the same. One instance of this principle application may be in the opaque exchange of data by any party which does not guarantee the traceability of data, hence hampering the exchange of valuable data.

For actor empowerment in data market ecosystems, we consider the utilitarian principle which goal is to satisfy the need of empowerment of the actors in such a way that maximizes the most considerable number of positive repercussions for the greatest number of actors while at the same time minimizes negative repercussions to the lesser number. One instance of the application of this principle may be in the empowerment of individuals in data market ecosystems that achieve broader value for an ecosystem than parties without the power to act. Furthermore, to protect weaker parties, each market participant must consider the golden rule, the descartes' rule of change and the risk aversion principles before

taking an action influencing another. The last principle applied in this research is the no free lunch rule. By assuming that all data asset belongs to someone or are the property of an organization unless there is a specific declaration otherwise, and have value for which compensation should be given.

Table 11 provides an overview of the opportunities offers by data market ecosystems to actors. It presents the risks, the ethical issues, the requirements for handling issues and the guarantee that is needed by each party.

	Opportunities	**Requirements**	**Guarantee**
Service Providers	•New market of Personalized services for Individuals •Improved competitiveness •Increased revenue •New collaboration with individual	•Mitigate Compliance risks •Loyalty and fairness of service access and usage •Responsibility by design and by default	•Service Level •Neutrality of services •Fair competition with data provider
Data Providers	•Better Data offerings •Increase data flow and usage •Individuals empowerment •Define terms and condition for data access •Guarantee flow of data in the ecosystem •Individual agency •Get personalized services	•Control data creation, collection and Sharing •Measure data usage	•Terms and condition for data usage •Service level •Authoritative data source management •Data authenticity and Quality
Data Consumers	•Access to new sources of data •Access only data needed •Saving storage cost •Comply with higher-level data protection laws	•Mitigate Compliance risks •Responsibility by design and by default •Responsibility by design and by default	•Data usage metrics •Consumer service definition
Data Brokers	•Change Business model •Comply with higher-level data protection laws	Mitigate Compliance risks	Fair competition with data provider
Marketplaces	•Open and regulated channel for data and services discoverability •Manage data transactions and services access	•Mitigate Compliance risks •Responsibility by design and by default	•Terms and condition for market access •Fair competition •Neutrality of services
Regulators	Asynchronous Audit of data processing	•Asynchronous audit •Full access to digital traces of data market ecosystems •Service evaluation	

Tableau 11: Opportunities, risks, requirements and guarantee for market participants

A data market ecosystem must enable ethical assessment by requiring the traceability of issues and responsibility of parties and transparent data flow for parties collaborating together. It must enhance trust between parties. This involve on the one hand considering:

- The conditions of lawfulness of data processing from the point of view of a data protection framework
- The respect of the condition of data usage which depends heavily on the guarantee provided by each party to ensure the right data are delivered to the right party and consume under the settle conditions
- New business models enabling fair cooperation.

On the second hand is the empowerment of the data subject in order to enable their active participation. This empowerment will only be achieved through three aspects: the guarantee of the data subject privacy which concerns the provisioning of their data to other parties, the access of their personal data and the availability of value-added services which will enable them to collaborate with the other parties and use their data in these services. The third element is the fair data value sharing among the ecosystem actors. It is important all the actors of the ecosystem get rewards in their collaboration to stimulate data exchange. And finally, the compliance to data protection and security framework. Based on the ethical principles and the key considerations of the issues summarized above, we propose some requirements for the data market ecosystem.

5.2.1 Common values goal for actors in data market ecosystems based on the utilitarian principles

To limit the variation of goals and priorities independently of jurisdictions, all actors of the ecosystem must abide by commons higher-level privacy and security goals. For example, a data market ecosystem can decide to comply with the GDPR while not being in the European Environment. As ecosystems are open environments, which communicate with each other, this will considerably reduce compliance efforts. Hence, ecosystems must define appropriate and transparent criteria for managing data and services assets they bring to parties, in a fair and equitable way. In addition, ecosystems must ensure all parties have an equitable capacity to participate, including ensuring equitable access to information, the ability to operate in, or if necessary represented adequately by another party.

5.2.2 Roles free of conflicts of interest

Service Provider. We believe that they have the responsibility to enhance individual agency towards their interactions and their service design.

- *Fair and transparent Services*. To act with fairness and responsibility in data market ecosystems, service providers must first guarantee the fairness of their services and produce predictable results for any user. The services must behave according to their promises without any discrimination such as user profiles or pricing.
- *Mitigation of Data Protection Risk Compliance.* Another requirement for Service Providers is the mitigation of data protection compliance risk. They should limit the need for individuals to rely on their right under data protection laws by enabling data processing over personal data without

their access. This could be achieved in self-tracking services where the result of data processing represents the salient point, and only useful for the service user. Technology such as homomorphic encryption and Secure Multi-Party Computing enable the processing of encrypted data. Data processing on the user side offers another solution where the services are pushed on the user side and the data generated is never accessed by the providers. However, some data processing does not allow such use cases. Particularly when for some audit and legitimate reason or for an agreement to be fulfilled the collection of personal data are inevitable. In such case, the Service Providers must handle proactive data exchange with individuals they collected data from with timely repackaging data in a usable format through a ready-made channel. The Service Providers must define a granular data schema for the data that are collected and provide a guarantee that data collected and shared with the user match the data schema.

Individuals. To exercise their agency in data market ecosystems, individuals need first to manage the incoming and out-going flow of their data in the ecosystem. Moreover, they should be able to formulate decisions over the use of their data, negotiate terms and conditions for services access in a range of business models and ask for added-value services that can fulfill their specific needs. Ultimately, they should be able to monitor their data usages in such a way that enable the transparency of actions.

- *Individual Agency.* True empowerment is about enabling individual agency over their data and services consumption. It starts by controlling the source of data production, the acquisition of these data any time they are generated and the delivery of data wherever it is demanded according to the rules imposed by individuals, and providing guarantees of data delivery. The second level is to control data disclose while using a service, and a third level is to have access to value-added services which fill the need of individuals.
- *Usage Right.* Personal data must be exchange on a Usage Right granted to Data Consumers which automatically excluded any future possibility of usage by Data Consumer without the approval of Individual as well as any appropriation of these data by them. In order for this right to be effective, individuals' data source must be the authoritative data source.

Data providers. Key considerations for data providers, whether Individuals or Organizations, are to ensure the sustainable delivery of data according to satisfy the need of data access in an ecosystem (Utilitarian Principles).

- *Timely delivery of Data Asset*. Data Provider must guarantee the timely delivery of data asset so as not to jeopardize the future sustainability of data access.
- *Usability of data.* Data provider must guarantee the usability of data by using adequate format.
- *Non-Hindrance of data exchange*. As data are valuable goods for any party, it is of necessity to enable as much as possible the accessibility of data without discrimination to parties satisfying the usage conditions.

Data provider must not hamper the accessibility to data asset on grounds other than non-compliance considerations.

- *Transparent valuation of data*. Data provider must define the real value of data in a way open to scrutiny.

Data consumers. They must ensure the use of data within the specified Usage Right by providing auditable proof of their data usage and the non-disclosure to other parties. Like the Service Providers, Data Consumers must apply the Mitigation of Data Protection Risk Compliance.

Data brokers. They can hardly be adapted without a drastic change in their business model. Nevertheless, one noteworthy feature of data brokers concerns the location of suitable data for data consumers. This role is significant for data market ecosystems and data brokers are already involved in this practice. They can perform data composition according to data schema and provide assistance for choosing the right data asset. Data brokers may also help Individual in defining customized service demand for their personal use.

Auditors. Each market participant should monitor their transactions. Therefore, we suggest that the monitoring and auditing of activities and transactions should be available to any concerned party. However, regulatory bodies like Auditors fulfill a critical role in the ongoing regulatory auditing of data market ecosystem. In particular, for the definition usage right and data protection compliance assessment.

Marketplaces. They play the role of trusted third parties for ensuring data transactions integrity and safety.

- *Informed trusted third party*. Given the risks implied by their nature, we introduce the concept of Informed Trusted Third Party which contrasts with the blind model through the integration of trust-enhancing technology such as DLT. Thus, whether in a centralized or decentralized form, Marketplaces must enable the transparent observation of their actions in an ecosystem.
- *Equality of access*. Moreover, marketplace must meet the requirements of equality of access by all the parties as well as the continuity and the neutrality of the platform.
- *Mitigation of data protection risk compliance*. Moreover, it must also apply the Mitigation of Data Protection Risk Compliance by handling a minimal amount of personal data for transactions processing.

5.2.3 Parties Compensation

It is essential to compensate or reward parties for their efforts in data market ecosystems. This requirement is based on the ‘No free lunch principles’ for data exchange or for providing value-added services. Compensation may take multiple forms depending on the type of data asset.

5.2.4 Responsibility by Design and by Default

By analogy to the concepts ‘Privacy by Design’ and ‘Privacy by Default (GDPR 2018) which enable considering privacy at the earliest stage of product or service development, we propose the concept of ‘Responsibility by Design’ and

‘Responsibility by Default to introduce upfront responsibility of market participants about designing services and applications that:

- *Give control for personal data exchange to Individuals by default.* For instance, an individual does not need to ask for data access or portability. Data generated must be repackaged in a reusable format and send automatically to the concerned party. This requirement enables the automation of data portability by default as actors do not need to ask for any data before getting data. These rights must be carried out at the service definition and initiate at the earliest moment an Individual subscribes to a service.
- *Engage the responsibility of market participant.* For instance, data usage should be monitored by default by auditors, data providers and data consumers.

5.2.5 Requirements for data exchange

We address the question of the rules under which data must be exchange and use in a data market ecosystem with respect to these categories.

Some studies proposed a variety of policies and practices regarding data access, sharing, and management. These can be used to address the aforementioned issues. In this part, we investigated these studies and selected the guidelines addressing them. These guidelines could provide guidance to share a category of data and set the boundary requirements for the intended uses of data.

We collected a number of established principles and guidelines that we transposed in the context of data market ecosystems to govern the sharing and use of a certain data category. We focus on existing principles on fair data sharing practices, data governance, and regulation apply to data processing. We outline specific areas that are closely linked to our issues in Table 12:

- Principles on “the responsibility of market participant” outline best practices on a data market ecosystem in general to enhance the confidence in a data market ecosystem. We considered the ten rules provided by Zook et al. (2017) in the Big Data ecosystem to address the general distrust of the data market economy. These principles could enlighten the condition under which data are exchanged in a data market. Each market participant might conform to all or part of these principles in order to build trust in the market ecosystem.
- Principles on “privacy” govern the sharing of sensitive data and build trust among the participant. In the context of privacy, the European data protection reform (GDPR 2018) set the basis of privacy principles associated with personal data. We selected those that are suitable for our context.
- Principles on “data quality” ensure the quality of data necessary to achieve desired outcomes. Some of the substantial principles for data quality in biological diversity domains are appropriate for our context and are detailed in Table 12 (Chapman 2005).
- Principles on “data sharing” propose good practices on data sharing. These principles based on the open government data principles provides

principles for sharing data (Opengovdata 2014). We selected the principles with respect to data sharing and adapted them to our context.

Principles	Descriptions
Privacy Principles (GDPR 2018)	Lawfulness, fairness and transparency: Data sharing and usage must be in accordance with the law and pursues a legitimate purpose. Purpose specification and limitation: the purpose of data usage must be visibly, explicitly defined. Data minimization: data shall be adequate, relevant and limited to predefined purposes. Storage limitation: data must be erased after serving the purposes for which the data were collected. Integrity and confidentiality: appropriate technical and organizational measures must be implemented to protect data. Accountability: parties must implement measures to promote and safeguard data.
Data Quality Principles (Chapman 2005)	Quality of Data Sources: accessibility, sustainability, license, trustworthiness, verifiability, primary. Quality of data: accuracy, referential, correspondence, cleanness, consistency, comprehensibility, completeness, typing provenance, versatility, traceability, correctness, granularity.
Data Sharing principle (Opengovdata 2014)	Timely: Data is made available quickly to preserve the value of the data. Machine processable and schema agnostic: Data is reasonably structured to allow automated processing and data format and meaning are sufficiently documented. Non-proprietary: Data is available in a non-proprietary format to reach a wide audience. Licensing: Data provider clear about what data is available and what licensing, terms of service, and legal restrictions apply.
Ten principles on Big data Research(Zook et al. 2017)	Acknowledge that data are people and can do harm. Practice ethical data sharing. Design data for auditability. Engage with the broader consequences of data and analysis practices.

Tableau 12: Principles Applicable to Data Exchange and Processing

A taxonomy of data must be defined for specific data market ecosystem: we discuss the utility of categorizing data and the necessity of establishing principles that will govern the data exchange among the stakeholders. Data categorization will provide a clear picture of data characteristics to further derive the conditions of their collection, access, and usage. Depending on their nature, data collection and use may raise considerable concerns only addressable through valuable principles.

All data must be auditable: Robust traceability and audit are essential in a regulated environment. This requirement guarantees all data exchange and processing leave an auditable trace in an ecosystem that must be accessible by

the involved parties. This requirement concerns the creation of data, the modification of data, the exchange of data and the use of data. In the data market ecosystem, it is noteworthy to guarantee data source and quality in such a way that prevents unauthorized data alteration for falsifying their value. Also, it is essential to ensure the authenticity of data asset in order to verify what it claims to be. All data use must leave an auditable trace. This requirement encompasses every possible action over data. The goal is to bring transparency over data transactions and processing in the ecosystem by keeping a record of every action of the actors triggered by transactions and data processing within the ecosystem.

Data formats shall be designed for data usability and data quality. Open formats are widely encouraged.

Agnostic data schema for data storage and retrieval. This requirement mitigates the problem of data portability and interoperability between services. Hence, a data market ecosystem must guarantee that any personal data collected while using a service is automatically sent by default to individuals in an agnostic data schema, with the semantic model and contextual information of data collection in such a way that enable the integration of data in another service. A semantic model or ontology is a set of consensus knowledge about a domain. The use of common language based on ontologies is necessary to remove ambiguity and confusion when describing data asset.

5.2.6 Requirement for service access and usage

Service Evaluation. Services, applications, and platforms in an ecosystem must be evaluated for assessing their transparent and fair behavior. This might be carried out by adopting open source evaluation or by a regulatory body by relying on a set of criteria including marketplace audit log. We advocate for an ethical marketplace which adopts an open source model, in such a way that enables the assessment of the design objective. In fact, the marketplace is at the core of data exchange and services access among parties.

Contractual agreements must support data exchange and service usage in an enforceable way. The contractual clauses must take into account the rights and obligations of the actors related to data protection frameworks and service levels. Hence it must be legally binding and also have an automated dimension that helps parties in executing the contractual clauses. Agreements must be made persistent to allow their observation and auditability by the contracting parties as well as auditors.

5.3 Summary

In this chapter, we defined the requirements for designing fair and responsible data market ecosystems. These requirements drew on values and ethical principles in the information systems domains.

Based on the analysis of current parties involved in the data exchange activities, we identified the market participants, their current role and their new functions based on ethical principles in such a way that resolve conflicts of interest in this environment. We also emphasized how these roles are significant in the data market ecosystems and how they contribute to their sustainability. Following, we

analyze some general requirement addressing data exchange and services access for minimizing the risks of compliance with data protection framework. We proposed a set of principles aiming at guiding the data sharing. These principles are derived from existing guidelines and practices in data processing. Data transactions aren't currently regulated or controlled like other common markets. Initiatives like the European data protection framework (GDPR 2018) or open data sharing principles (Opengovdata 2014) cover some elements about regulating the sharing of various data categories. However, none of the existing works address the regulation and the organization of a data market. The proposed principles in focus on data categories and represent an additional step in helping better understand the control of data transactions in a market. More work taking into account multi-stakeholder requirements is still needed to further refine these principles but they represent a good starting point.

Next chapter provides a high-level data taxonomy and elaborates on principles that should govern data exchange depending on data category.

Chapitre 6. A data taxonomy for data market ecosystem

Drawing upon literature from data collection studies and regulatory frameworks, it is noticeable that the issue of data collection and sharing encompasses a wide range of data types. These data types which appear scattered throughout scientific literature, government reports, and across the web are potential, if not yet the case, valuable data asset. They are classified differently depending on the authors and the domains, even if there are the same. We argue that this classification could be addressed in a broader perspective rallying these diverging ecosystems around some common data characteristics and rules to consider when exchanging data from the perspective of data market ecosystems (Nwatchock A Koul and Morin 2016).

The objective of this chapter is twofold. First, it offers analysis and categorization of data in the context of data market ecosystems, by analogy with other commodity markets. This classification should improve the understanding of the characteristics of data. Depending on the nature of data, their collection and usage raise considerable concerns from the point of view of issues such as privacy, ethics, usage rights, etc. Consequently, we need to address these issues with valuable rules that should govern the data exchange. Therefore, our second goal is to propose a set of rules for data exchange depending on the data category. These rules are based on literature and regulatory frameworks.

6.1 Methodology for the data taxonomy development

Taxonomy development is a search and iterative process changing overtime that is useful to describe objects in a particular field. Many approaches exist for this complex process. To perform our work properly, we turned our attention to two very practical and highly-cited taxonomy development process. The first one is the study carried out by Carl Von Linne (Hoquet 2005) in biology. He classified the living organisms in a systematic way, based on their natural characteristics, through a hierarchical classification scheme describing the groupings of kingdom, phylum, classes and then orders, families, genus, and species. This classification scheme results in a significantly low level of categories collectively exhaustive and mutually exclusive. Applying this classification scheme in our case may result in a large and redundant number of data categories and the mutually exclusive principle could not be achieved. The second is in Social Science and has been achieved by Bailey (1994) who provided three classification technics. The conceptual classification (typology) that is based on deduction, the empirical classification (taxonomy) where the categories are derived from empirical data clusters and the operational classification which combines both conceptual and the empirical classification approaches. In the operational approach, it is possible to start with the deductive approach and then examine the empirical data cluster or start with the empirical cluster and then formulate the conceptual categories.

In order to perform our data classification, we found a third option which is the taxonomy development method of Nickerson et al. (2009), a straightforward process specially designed for the field of Information Systems. They define a

taxonomy as a set of dimensions, each consisting of characteristics that are mutually exclusive, collectively exhaustive and sufficient to describe the objects. This approach combines both empirical-to-deductive and deductive-to-empirical approaches of Bailey (1994) to identify the dimensions and their related characteristics according to a predefined meta-characteristic (Figure 4).

6.1.1 First step - determine meta-characteristics

The first step is the identification of the meta-characteristics based on the purpose of the taxonomy and in turn based on the users and taxonomy usage. The users of this taxonomy are market participants. We elaborate the categorization of data for high-level characteristics of data according to the user needs and the data protection requirements. The purpose of this taxonomy is to provide a broad view of valuable data by finding a reasonable balance between user needs and data protection requirements. Therefore the meta-characteristic is the user needs with particular attention to data protection requirements. This takes into account the lawfulness, fairness, and the sustainability of data exchange.

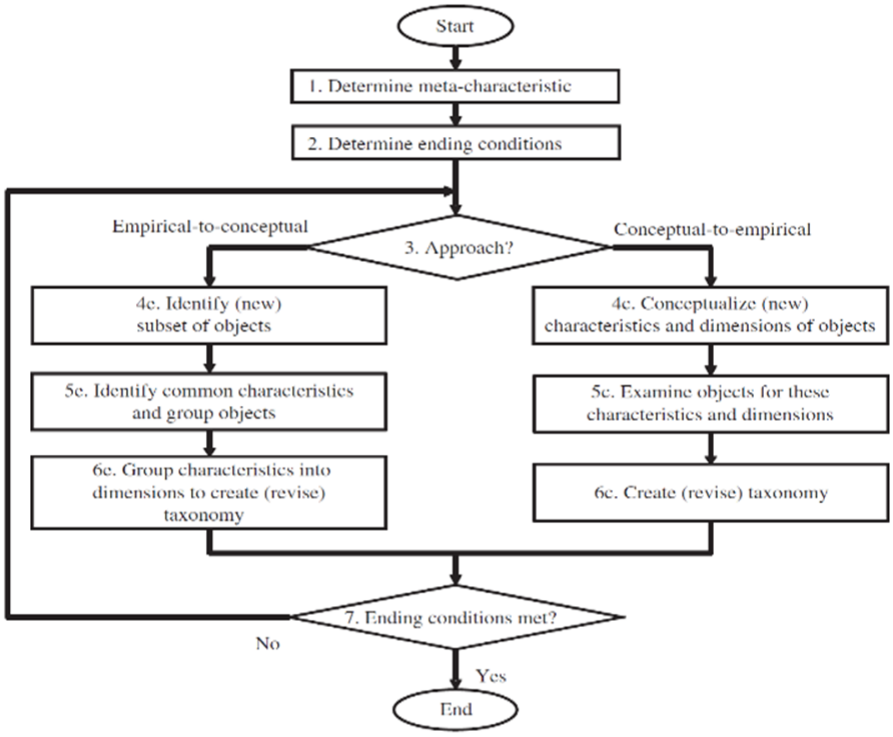

Figure 4: Classification development process of Nickerson et al. (2009)

6.1.2 Second step - determine ending conditions

The second step is to determine the ending conditions. The methodology requires that both objective and subjective ending conditions be met. We use six of the eight objective ending conditions that are:

- A representative sample of data has been examined; no new dimensions or characteristics are added in the last iteration
- At least one data category is classified under every characteristic of every dimension
- Every dimension is unique and not repeated; every characteristic is unique and not repeated
- Each cell is unique and is not repeated

We also use all of the subjective ending conditions that are met when the taxonomy is concise, robust, comprehensive, extendible and explanatory.

6.1.3 Third to seven step

From the third step to the seven step, we follow the iterative approach of the taxonomy development method by starting with the empirical-to-conceptual approach. We gathered substantial elements of data samples from scientific literature, governmental report, and periodicals of all kind addressing data trade and sharing and their related issues. We explored the IEEEXplore Digital Lib, google.com, the Web of Science Conference Proceedings Citation Index, Google scholar, some leading global management consultancies like Accenture and some economic organizations like the OECD, the World Bank. We use the keywords related to data ecosystems, data sharing and data trade and we get some data samples that are referenced in Table 13.

We found out that there are different characteristics of data according to specific data ecosystems and domains. Then, we analyzed these characteristics and selected those in connection with our meta-characteristic. The encountered obstacles in analyzing and comparing data classes are that different denominations for the same data type or the reverse and also the literature showed different perspectives of classifications of data according to the domain of application.

The analysis also reveals different granularity levels mostly because the literature addressed various issues. Therefore, it was not necessary to accurately specify a low granularity of data classes. We consider literature that includes many data samples. We combined similar characteristics and deleted those with very low granularity level. After identifying the differentiating characteristics of data classes, we grouped them into three dimensions that we called "*data content*", "*data staticity*" and "*data sensitivity*". We have not considered the volume of data, the distribution option, the sources of data, as they do not influence the purpose of this classification.

Finally, we applied the conceptual-to-empirical approach in our taxonomy development. Considering some data are more sensitive than others, we included further characteristics forming the sensitivity dimension following the privacy requirements of the meta-characteristic. We used these dimensions for the classification of our list of data in Table 13. By classifying these data, we identified

further relevant differences between them, which needed to be reflected in the dimensions and their characteristics. In some cases, we removed the irrelevant or redundant characteristics, and in others, we refined remaining ones. Based on the analysis, we derived high granularity classification elements offering a global outline of data.

Keywords	Literatures Reference	Data Samples
Economic data	OECD (2013)	Identifying data, Activity or Behavioral data, demographic data, Social data, Locational data, De-identified data, personal identifiable information data, non-personal identifiable information, personal health record,
	World Bank (2014)	Geospatial reference data , environment data, economic data, transport data, energy data, resources data, demographic data, weather data, road data, transport data, official registers data, company registers data and cadastres data
Data broker data type	FTC (2014)	Identifying data, Social and technology data, Home and neighborhood data , Court and public record data, Sensitive identifying data, purchase behavior data Health data, Financial data, travel data , vehicle data, General interest data
Access Scientific Data	OECD (2014)	Research data, publicly funded research data
Research data type	Burnham (2012)	Observational data, experimental data, simulation data, derived or compiled data, reference or canonical data
Open data value	Peter Murray-Rust (2008)	Scientific data, factual data, experimental data, public domain compliant data business data
Information reuse market	Vickery (2014)	Economic & business data, social data, legal data, geographic data, meteorological data, transport data, farming, forestry, agricultural and fisheries data, cultural data, political data, environmental and natural resource data, scientific and research data, tourism and leisure data, educational data, infrastructure and urban development data
Organization data	Liebig (2009)	Organizational data, enterprise data, firm data
Corporate data sharing	Verhulst (2014)	Corporate data, private data
Data value	Hjalmarsso et al.(2015)	Open static data, open dynamic data, open statistical

Tableau 13: Data categorization from literatures review

6.2 A taxonomy of data in the context of data market

We identified three dimensions that are: *data content, data sensitivity* and *data staticity.* Each dimension includes specific characteristics as detailed below.

6.2.1 Dimension 1: data content

Data content reflected the nature of the data in terms of property. We proposed three characteristics that have implications on the data exchange terms and conditions:

Personal data

The meaning of personal data has progressively evolved over the decades. Accordingly, personal data are "*any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person"* (GDPR 2018). Personal data are probably the most sought after data asset when analyzing the current data sharing environment. A number of studies suggest some personal data elements from the viewpoint of privacy and trading (OECD 2013; Federal Trade Commission 2014, GDPR 2018).

The issues raised by their usages like privacy, unlawful processing, and discrimination make it difficult for now to leverage the full access to them. With the GDPR aiming at empowering the individuals with their new rights, we anticipated the involvement of individuals in the data market ecosystem as data providers in a secure and fair environment.

Non personal data

We refer to any non-personal data which are whether under copyright, patent, trade secret laws or no intellectual property law. One example is the corporate data generated by private organizations to safeguard their competitive advantage and to manage their daily organizational tasks. Corporations collect and manage a variety of data about their customers, their products, their transactions and their organization (Liebig 2009). Currently, there is a thin line between corporate data and personal data related to the ownership issue. We argue that in our context, corporate data that can be exchanged in a data market ecosystem, should not contain any identifying customer data or non-anonymous customer data. The report on the business of data (The economist 2015) is an example of the assimilation of both corporate data and personal data partly because many companies monetize their customer's data.

Also by public administrations as part of its public service (Vickery 2014) are providers of this category. During the last decade, some public sector data held by government and public institutions have been made open and usable for the entire community. However, some remain private, under sharing restriction or accessible for a fee. However, public data are not just a matter for public organizations but also for any entity who generate valuable data, distinct to personal data, for whom the intellectual property laws are expired after a certain amount of time or under license usage.

6.2.2 Dimension 2: data sensitivity

The purpose of considering data sensitivity as a dimension of our taxonomy comes from the privacy requirement of data. According to the GDPR, some

personal data are classified as highly sensitive data requiring more controls when processed. The public sector has opened some public sector data while some remain closed for many reasons like high added-value data or confidential data. We deduced that, in the context of data market ecosystems, it is necessary to estimate the sensitivity of data to ensure the sustainable trade of any data category according to appropriate security measures. Although the literature review widely acknowledges three levels of sensibility, we focus on two categories of data sensitivity as there is no metric to evaluate the accurate level.

Sensitive data
It refers not only to special categories of data of the GDPR Art.9 but to any data whose privacy, protection and security need to be guaranteed during their collection and usage because:

- The disclosure of these data in potentially harmful for a party (personal data)
- Data have a high added-value (patents)
- Data are considered highly confidential

Security and fair use of these data should be guaranteed by the parties in charge of the processing.

Non sensitive data
It refers to data that required certain usage conditions that should be defined to ensure the security of data and fair use of these data. Even if these data are not sensitive, still, they require adequate protection and security because of their value in the data market ecosystem and the production effort.

6.2.3 Dimension 3: data staticity

Among the data exchange challenges are the sustainable delivery of data. To this end, we propose a dimension called "*staticity*" that addresses the modification of the data state. This dimension is crucial to determine whether some data assets need particular treatment to ensure their accuracy and integrity. We define two characteristics of the "*staticity"* dimension:

Static data
Static data is the property of fixed datasets. It requires less effort to publish since data is not updated and changes rarely. Consequently, it is easier to maintain data integrity, accuracy and provision in a data market ecosystem. The static data could be data that do not change by its very nature like a birthdate or a place. This data could also be some past-periods data or historical data like historical weather data or archive data.

Dynamic data
Dynamic data is the property of data state that is in flux and the use value is closely tied to the age of the data (Hjalmarsso et al. 2015). This category includes data that are updated frequently. The publishing of these data requires different mechanisms like scalable IT infrastructure, data versioning to ensure the sustainability of the delivering process.

6.3 Data taxonomy cube and usage

6.3.1 Basic rules associated to data dimension

Basic clauses associated to personal data. The GDPR provides terms of data processing agreements between data controllers and data processors that can be used as the basis for defining rules associated with personal data. Therefore, the key elements of this agreement, defined in Chapter 8 will be associated with personal data agreement definition. Another consideration is the compensation offered by a Data Consumer give for personal data consumption.

Basic clauses associated to non-personal data. The terms must consider by a data provider are the license under which data will be exchanged and fees for exchanging data.

Basic clauses associated to static data. Data provider must define the quality of data in order to deliver accurate data asset.

Basic clauses associated to dynamic data. Data provider must define the quality and the version of data for delivering accurate data asset.

Basic clauses associated to sensitive data. The marketplace must enforce secure data transport by supporting data encryption during the exchange process. Data consumer must provide a guarantee for security measures and privacy for data processing.

Basic clauses associated to non-sensitive data. Encrypted data during the exchange process is also required.

6.3.2 Data taxonomy cube

Using the dimensions of the proposed data categorization, we built a data taxonomy cube which is a synthetic representation of data categories in this market context. Figure 5 represents the data taxonomy cube which dimensions are: *data staticity (X), data content (Y), and data sensitivity (Z).*

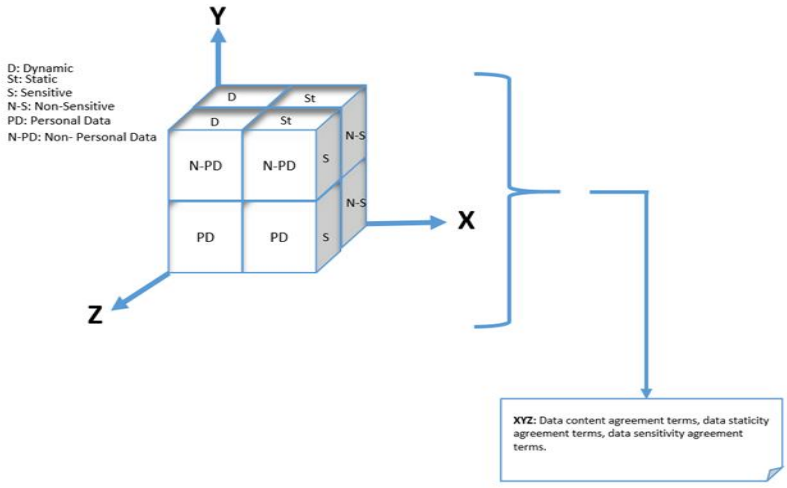

Figure 5: Data taxonomy cube

This cube enables to explore the main characteristics of data assets that are necessary to generate the clause of exchange and processing. To each characteristic, we associate a clause that will be validated or not by a data provider as a decision tool.

6.4 Summary

This chapter has first proposed a taxonomy of data in the context of data market ecosystems. The methodology is based on the work of Nickerson et al. (2009) who provided a taxonomy development method in information systems. The proposed taxonomy is likely to evolve, considering the number and diversity of existing data and the design of a specific data market ecosystem.

Second, we propose a data taxonomy cube that can be used as a decision tool when operation a data market ecosystem.

In the next, we propose a design of a framework for a sustainable and responsible data market ecosystem based on the predefined requirement.

Chapitre 7. A framework for fair and sustainable data market ecosystems

As described in several previous sections of this dissertation, the ultimate goal of this research is to provide a framework for a fair and sustainable data market ecosystem. In order to achieve this, we identified the main actors involved in data exchange and the main areas that constitute the essential issues in the domain.

Influenced by the representation of the NIST reference model of cloud computing liu et al. 2011), we adopt a framework which emphasizes on the actors and areas that data market ecosystems have to address: data supply by data providers, individual service usage, third-party service consumption, data usage by data consumer and transaction handling by the marketplace (Nwatchock A Koul and Morin 2017).

Further, we describe the main functions of each area and represent how data and transactions flow in the ecosystem.

Finally, we summarized the chapter and introduced the agreement management process.

7.1 Framework architecture

The high-level framework architecture is composed of six functional components as depicted in Figure 6.

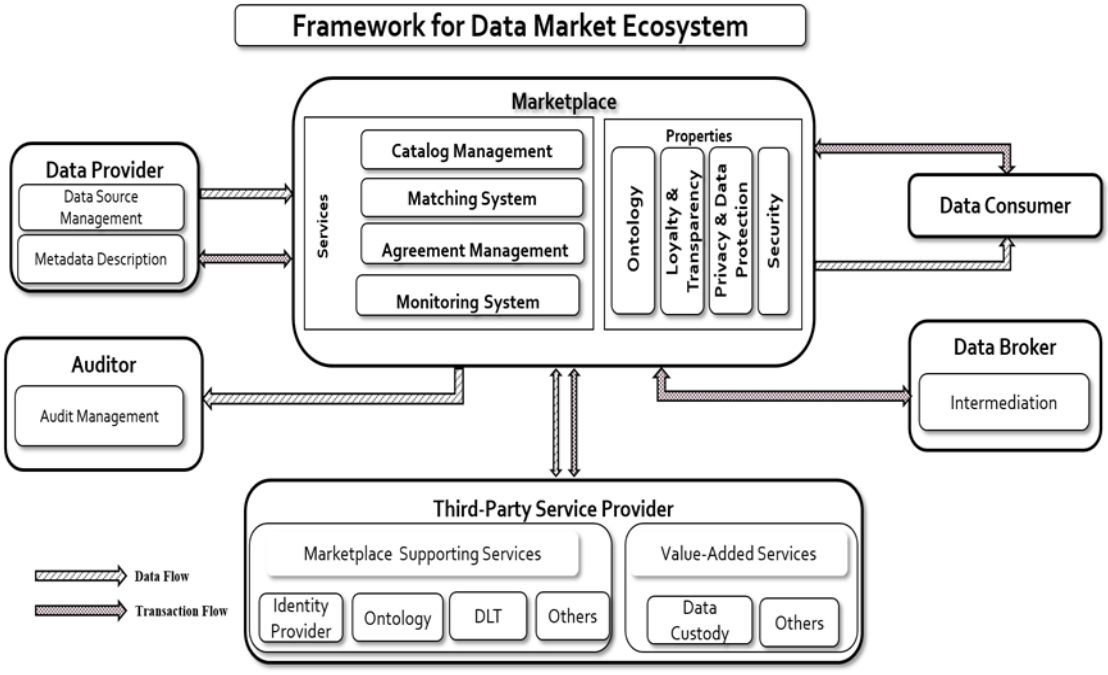

Figure 6: Framework architecture

- *Marketplace component*: The marketplace mediates the exchange between market parties.
- *Data provider component*: data provider are responsible for data provision in the ecosystem. Individuals acting as data provider must remain at the heart of the exchange of any personal data in order to maintain control over them.
- *Data consumer component*: data consumer subscribes to data supply by a data provider.
- *Third-party service component*: third-party service provider are responsible for service provisioning to any actors of the ecosystem.
- *Data broker component*: Data broker handles data and service retrieval for the ecosystem actors.
- *Auditor component*: auditors are independent, impartial and public supervisory authorities. They are responsible for analyzing the agreements and transactions audit trails of the ecosystem activities. They must be able to detect breaches of contract terms and apply corresponding sanctions.

7.1.1 Marketplace component

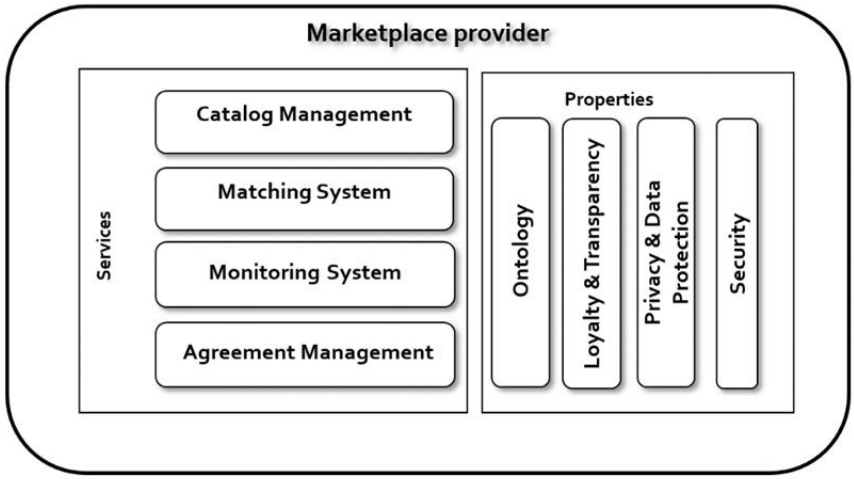

Figure 7: Marketplace component

A data market ecosystem must work within the ethical and legal boundaries defined and validated at the marketplace level which constitutes the medium for parties' collaboration. A marketplace culture must serve to enhance trustworthiness between parties. It should be established on core values like transparency, honesty, openness, equality of access and usage, empowerment of parties, trust in transactions, mutual respect and responsibility. Marketplaces should be designed out of ownership interest and opportunism and be involved only in the protection of the integrity of transactions process.

The proposed marketplace constituents are arranged in vertical and horizontal dimensions as shown in Figure 7. The vertical dimension describes the features that span across all the mechanisms of the marketplace and which must be effective at every level of the transaction. Each dimension possesses some supported functions that satisfy the requirements. Starting by the vertical dimension, we define the features that should impact the activities in the ecosystem.

Marketplace properties

- *Loyalty and transparency.* Loyalty and transparency are strongly linked concept. The French Digital Council (CNNum 2015) advocate for a principle of loyalty that any platform can adapt for fairness compliance. This principle aims at compelling platform providers to adopt transparent practices and ensure the compliance between platforms commitments and their real actions. In the data market ecosystem, this loyalty principle and transparency should apply to the marketplace and all the integrated services. The openness of the marketplace implementation will allow the evaluation of its behavior by experts to, in turn, provide a more trustworthy environment. Moreover, all the actions of the marketplace will leave an auditable trace handle by the audit management service. These traces provide information about data collected, data exchange, transactions, and agreements.
- *Privacy and data protection.* Data and transactions should be protected at the marketplace level. It essentially means encrypting all sensitive data collected or in transit on this platform, and ensuring that only the necessary information are disclosed to the marketplace. For that, we need to define any sensitive data that must be processed for limiting potential harm.
- *Security.* Security concerns are extremely important in the data market ecosystem, especially when considering the sensitive nature of data, the geographical location of service providers which are under different data protection laws and the difficulty of the evaluation of services security. For the marketplace, it primarily means providing a secure environment that addresses common information system security requirements like confidentiality, integrity, availability, identification and authentication, communication security, accountability, access and usage control, authorization, auditing and data protection (Krutz and Vines 2010).
- *Ontology*. The marketplace provides some ontologies besides the third-party ontology services plug-in usable by the market actors for data requests, data and service search, and any activities requiring a predefined vocabulary.

Marketplace mechanisms and functions

The horizontal dimension describes mechanisms and functions strictly necessary for parties' collaboration.

- *Authentication and authorization management*. This mechanism enables the authentication of parties and services for operating in a marketplace. The marketplace handles this task by providing an internal authentication

system or relying on a TIdP to authenticate each party. The marketplace authentication assertion from internal authentication service or a third-party identity provider proves that a party or a service has been authenticated and is known by its internal identity for service access. This mechanism is also responsible for providing corresponding resources to parties and handles agreement execution.

- *Catalog management*. A marketplace provides a catalog to register and search for data assets and third-party services offerings. In the next section, we described a generic data model and service model for registering data asset and third-party services. These models are used to describe the ontological vocabulary and are used to build an ontology-based catalog service with the functions describes in Table 14.

Functions	Description
`Verify(descriptions)`	Verify that data asset or service descriptions are filled with appropriate elements.
`Register(descriptions)`	Register a data asset or third-party service description in the ontological-based catalog.
`Get_Ontology(domain)`	Get an ontology for describing data asset or service. The selected ontology may come from some ontology service plug-in.

Tableau 14: Functions of marketplace catalog

- *Matching system*: The matching system assists data and service consumer on finding service or data asset offering. Table 15 describes the function of the matching system. The matching system handles the request for data assets and third-party services. It takes as input consumer's preferences for data asset or third-party service offerings. The matching system queries the ontological-based catalog storing data asset and service offering. Based on the queries, the system returns to a requester an offering that matches his preferences. A party can also perform a search by navigating in the ontology-based data catalog. The requester can accept an offering right away, or negotiate for a better offering, or cancel the session. If the offering is accepted, then an order is created and the agreement process is initiated as described in Chapter 8. If no offering satisfies the request, this is stored in the demands list carried by the marketplace and can be used by a data broker for searching appropriate data asset or services inside or outside an ecosystem. For example, an individual can publish a service request with its requirement in terms of service usage that will be handled by a data broker for finding the suitable service provider.

Functions	Description
`Publish_request(requirements)`	Publish a request for data assets or third-party services in the marketplace demand list that will be handled by a data broker
`Request_dataAsset(requirements)` `Request_service(requirements)`	Request a data asset or third-party service
`Filter(requirements)`	Query the ontological-based catalog for data assets or third-party services
`Get_offering()`	Get data assets or third-party services offering

Tableau 15: Functions of Matching System

- *Agreement management and data provisioning*: The authorization service of the marketplace handles the agreement's execution in the ecosystem. Chapter 8 provides a full description of the agreement management and data provisioning mechanisms.
- *Asynchronous monitoring and audit management*: Monitoring is used in real time to oversee and review the effectiveness of transactions in such a way that ultimately allows to detect any transgression from a party and risks as soon as possible. Auditing enables a systematic and independent examination of ecosystem parties and auditors to continually gather audit evidence to support data transactions activities by collecting data on transactions and accounts to establish compliance with data exchange agreements and service access. Monitoring and auditing apply to all activities involving trust and transparency for parties in ecosystems. Figure 8 provides a general description of the monitoring and auditing service.

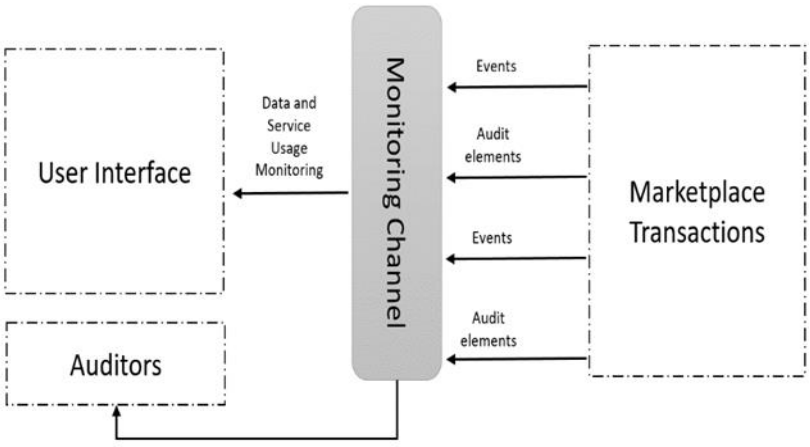

Figure 8: Monitoring and auditing service design

The monitoring mechanism addresses three areas: data generation through service use, data provision to third-party service and data usage. Table 16 describes the monitoring objectives, the element for performing monitoring and the parties receiving audit elements.

	Areas	Objectives	Audit Elements	Actors
Monitoring	Personal data generated during third-party services use	Proof personal data origin, verification third-party data collection	Data elements, data collection events, service usage events	Individuals, third-party services providers, auditors
	Data provisioning and usage	Verification of data provisioning match the agreement terms, get a proof of data origin, monitoring timely delivery of data, verification of data usage match the agreement terms	Data provisioning events, data elements, SLA metrics, service use events	Individuals, data consumers auditors
Auditing	Platform and third-party services loyalty	Verify the compliance of the platforms and services with their predefined goal	Auditors metrics	Auditors
	Data protection law compliance	Verify compliance of the platform and services with the privacy by design and defaults, responsibility by design and by default security, compliance with security, data protection effort, Perform data audit	Auditors metrics (audit elements and other metrics)	Auditors

Tableau 16: Monitoring and auditing objectives in data market ecosystems

The marketplace filters the audit elements corresponding to each party. These elements are collected in real-time as soon as a transaction is performed. The marketplace also stores the hash of these element as proofs in the integrated DLT.

7.1.2 Data provider component

Data provisioning is achieved by organizations, communities or individuals. We assume that an organization and communities hold non-personal data and an individual holds personal data that can be exchanged in an ecosystem. For an individual, becoming a data provider is preceded by services usage which generates digital trails. The main consideration for acting as a data provider is the description of the data asset to be exchanged in a data market ecosystem. An organization may define data description as well as the delivering service. For individuals lacking experience, service may be carried out by a data custodian which will handle the data provisioning on behalf of individuals.

7.1.3 Data consumer component

Data consumers search data assets according to their needs and subscribe to it (if available). Accordingly, they need a way to express accurately the selection criteria for enhancing the opportunities to find appropriate data assets. The use of an ontology can effectively direct the search and finding of a data consumer. Another constraint is the quality assessment of data that should be described fairly in such a way that data consumer pays data asset usage according to data quality and service delivery. Finally, the data access and usage monitoring may be a constraint of the data exchange, particularly for personal data where individuals offer a usage right. In this context, it is critical to dissociate data consumers as an entity from the services operating data processing on behalf of data consumers. Hence the role of data consumer can be restricted to find appropriate data and take an agreement for data use while the data access and usage is authorized and measured at the level of processing services.

7.1.4 Third party services component

The third party services component provides external service plug-ins and connectors. The goal is to support parties operating in an ecosystem and to extend the marketplace functions.

Marketplace supporting services

Services supporting the marketplace provide additional functionalities to complement the core marketplace, thus fostering the integration of a comprehensive range of solutions. Key services that can significantly enhance the core marketplace are:

- *Trusted identity provider services* (TIdP) are identity providers recognized and trusted by the ecosystem actors. They enable the authentication of markets participants through external credential storage and authentication standards support. A TIdP manages actors' identities outside the marketplace and gives these external actors identities permissions to access marketplace resources.
- *Ontology services*: One design requirement is the use of ontologies to maintain a consistent vocabulary understandable by each party. The goal is to bring correctness and clearness where necessary in every activity performed in the ecosystem. This service plug-in will take advantage of existing domain-specific ontologies, validated as sound and correct.
- *Dispute arbitration services*. This mechanism provides services to resolve disputes between parties through conciliation, mediation, and arbitration. Online dispute resolution (ODR) platform provided by the European Commission is an example of a service that helps dispute resolution with online customers without going to court (Cortes 2010). This service is employed to mediate contractual disputes arising from online purchases of goods and services, where the trader and consumer are both based in EU or other European territories. This kind of service can help to support regulators in capturing all the conflicts requests of parties and find evidence in the monitoring and audit services.
- *Distributed ledger technology connector*. Because of its immutable property, a DLT is suitable for continuously storing transaction and

agreement proofs for market participants. The DLT would then facilitate the auditing process in confirming the accuracy of agreements and transactions in such a way that enhances transparency and trust in an ecosystem.
- *Others*. It represents the other set of services that can extend the marketplace functionalities.

Value-added services
These services provide support to market participant operating in an ecosystem. We consider the range of service use by individuals for their routine tasks or to perform some specific tasks. We equally consider the data processors acting on behalf of data consumers.

- *Data custodian service*. Value-added services such as data custodian services are critical for individuals. They enable personal data gathering while respecting their privacy. These services should provide a range of storage option that will guarantee the use of encryption schema applied to data. The service should support the migration of data from multiple sources and also the provision of data to any party if required by an Individual. Broadly, data custodians must provide services that satisfies an individual in collecting his data and data consumer in delivering Individual's data.
- *Others*. They represent other value-added services individuals and other parties may get for their specific use.

7.1.5 Data broker component
A data broker provides an intermediation service for data localization and guidance. An individual can request an intermediation service from a data broker for expressing service needs that can be arranged or retrieved by a data broker. The service may be unavailable in the ecosystem. In that case, it makes sense to rely on a data broker to make it available. A data consumer can also use a data broker for retrieving data assets in data market ecosystems to reduce information overload by simplifying search processes for consumers.

7.1.6 Auditor component
An auditor is responsible for verifying the trustworthiness and validity of data exchange and storage transactions in a marketplace. An auditor can be a legal regulator. He has access to the monitoring service, which provides the main elements for handling an auditing process.

7.2 Considerations for data asset and service description
Describing data asset should take into consideration the utility and relevance of data asset for their proper use. One requirement around data exchange and processing is to design data for auditability. This involves providing data origin and authenticity, assessing data quality and capturing the full data life-cycle. Data asset description must encapsulate the metadata of data auditability. These elements should be informed by entities generating and maintaining data assets. While this task is trivial for organizations, it is more complicated for individuals

who rely on third-party services for data generation. Hence, we put this constraint at the third-party services which must associate metadata to each personal data that are generated. Another requirement is to consider data categories defined in Chapter 6 for establishing appropriate access and provisioning constraints. As data exchange is handled by a service (ex: data custodian), it is crucial to consider the service level of this service associated with a particular data asset. The service level should be adapted to the data category. For example, for dynamic data provisioning, the service must consider the appropriate delivery constraints for maintaining a high quality of service. Vu et al. (2012) proposed a description of data asset and service represented in Figure 9, in the context of data as service (DaaS).

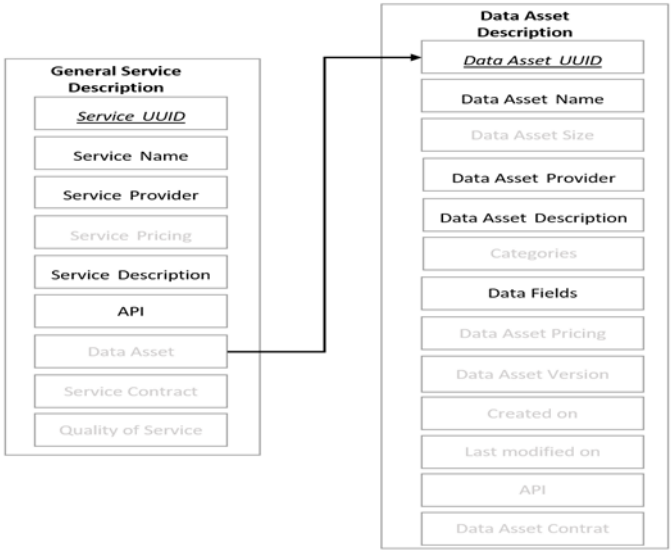

Figure 9: DEMODOS model for data and service description (Vu et al. 2012)

We use this model as a basis for defining a data and service description according to our requirements. However, this model does not integrate the element for data auditability as well as the guarantee for data delivery. To adapt this model to our requirements, the gray elements in Figure 9 have been redefined or removed from the proposed model. For example in data asset description, we remove the data asset size which is highly dependent on the dynamicity of the data.

7.2.1 Data asset description model

Therefore, data asset description should encapsulate metadata about:

- General information on data asset adapted from DEMODOS (Vu et al. 2012)
- Third-party Services which generate data asset
- Data asset version based to the update time
- The context of data generation which indicate the field corresponding to a data assets
- Fine-grained data elements

- Service level

We define a set of generic parameters for data asset description model.

- General_Description. This element corresponds to the data asset description of DEMODS model without the gray element. It enables the general description of data assets.
- Data_Category. This element corresponds to the category of data asset described in Chapter 6.
- Data_Version: This element provides an overview of the change operating in data asset, by whom and the goal of the change. It is an element of data traceability and auditability.
- Data_Elements: This element provide the fine-grained data element that will be exchanged.
- Data_Fields: The domain to which a data asset belongs such as health, transportation, etc.
- Sharing_Constraints: This element describes the rules for data access and usage. There are many strategies like subscription, pay as you go, pay what you want, pay per unit, etc., for organizations to exchange data. However, for an individual who makes an informed and voluntary choice of sharing their personal data, pricing them could be morally not permissible as personal data are inalienable goods. In such a context, the motivation of individuals could be through some non-monetary compensations offer by data consumer.
- Service_Level_Description: This element enables the description of service handling data provisioning. It is composed of several elements such as general service description as DEMODS model without the gray elements and with SLA that enables the description of guarantees provided by the custodian service.
- Optional: This element includes any other descriptive element that could inform data asset provisioning and usage.

The model represented in Figure 10 will be used as the basis for building an ontological description of the data asset in Chapter 9.

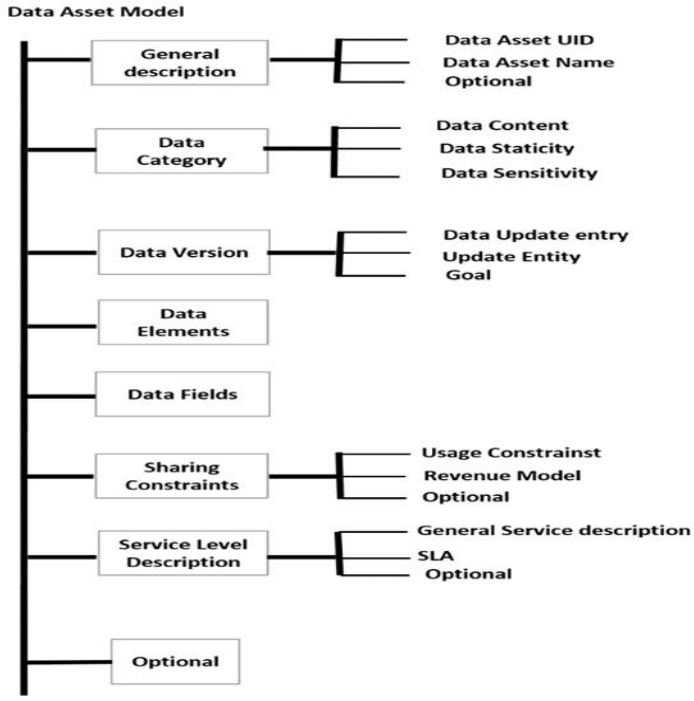

Figure 10: Data asset model

7.2.2 Service description model

The service description model provides a representation of the properties of all service metadata elements for describing third-party and data consumer services in data market ecosystems. Service description model must encapsulate all relevant elements for a fair description and responsible data processing. We propose a description based on semantics concepts used in industry (Jackson et al. 2014), which are suitable for both human-readable and machine-processable representations. Figure 11 provides the main elements for service description model (Jackson et al. 2014) that should be accessible to any market participants that consume third-party and data consumer services. It is composed of service profile, service interface and service implementation.

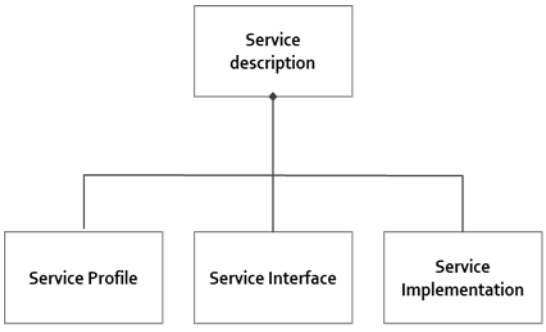

Figure 11: Service description Model (Jackson et al. 2014)

- *Service interface* specifies how a service consumer invokes a service.
- *Service implementation* describes the means by which the service is invoked, the underlying protocols and the endpoints for service invocation.
- *Service profile* represents the main component for advertising the service to potential consumers. It describes the parties responsible for service provision, the purpose of the service and limitations on service applicability. As such, it must provide a fair description of the service to enable a consumer to be informed of data collection and processing through service usage. The main properties of service profile are described in Table 17.

Name	Definition	Properties
Service Profile	The general description of a service.	Name, URI, description, version, service category
Organization	The organization responsible for service provisioning. For a data consumer service, the organization section will encapsulate the identity of data controller and those of data processor.	Name, URI, description, role, point of contact
Service Function	The activity that describes the functionality of a service.	Description and purpose
Service Level	The Service level that a service provider is obligated to deliver to a service consumer. In this thesis we limit the Service level to Quality of Service (QoS).	Name, value, definition, calculation method, unit of measure
Service Policy	The constraints that govern a service.	Policy attributes
Security mechanism	A protocol that describes and governs the implementation of service mechanism.	Document
Operation	It provides all the processing activities handle by a single service.	Operation Name, description
Processing	An action that is taken at service request	Description, data input, data output. (Data input and output are data taken as input for service use and data generated through service use that should be returned to service users.)
Data element	A unit of data that are collected or generated as part of service usage for which the definition, identification are specified.	Name, definition, data category, format, etc.
Data model	A lexical representation of data properties, structure and inter-relationships which specify the data collected or generated as part of service usage.	Data model schema, format

Tableau 17: Service Profile Model

7.3 Marketplace conceptual data model

Figure 12 defines the marketplace data model and the interactions between them. The core elements are data asset, service, orders, agreement, compensation, transactions, and audit trails. These elements enable the storage and the management of offering, agreements and all the transactions via the marketplace.

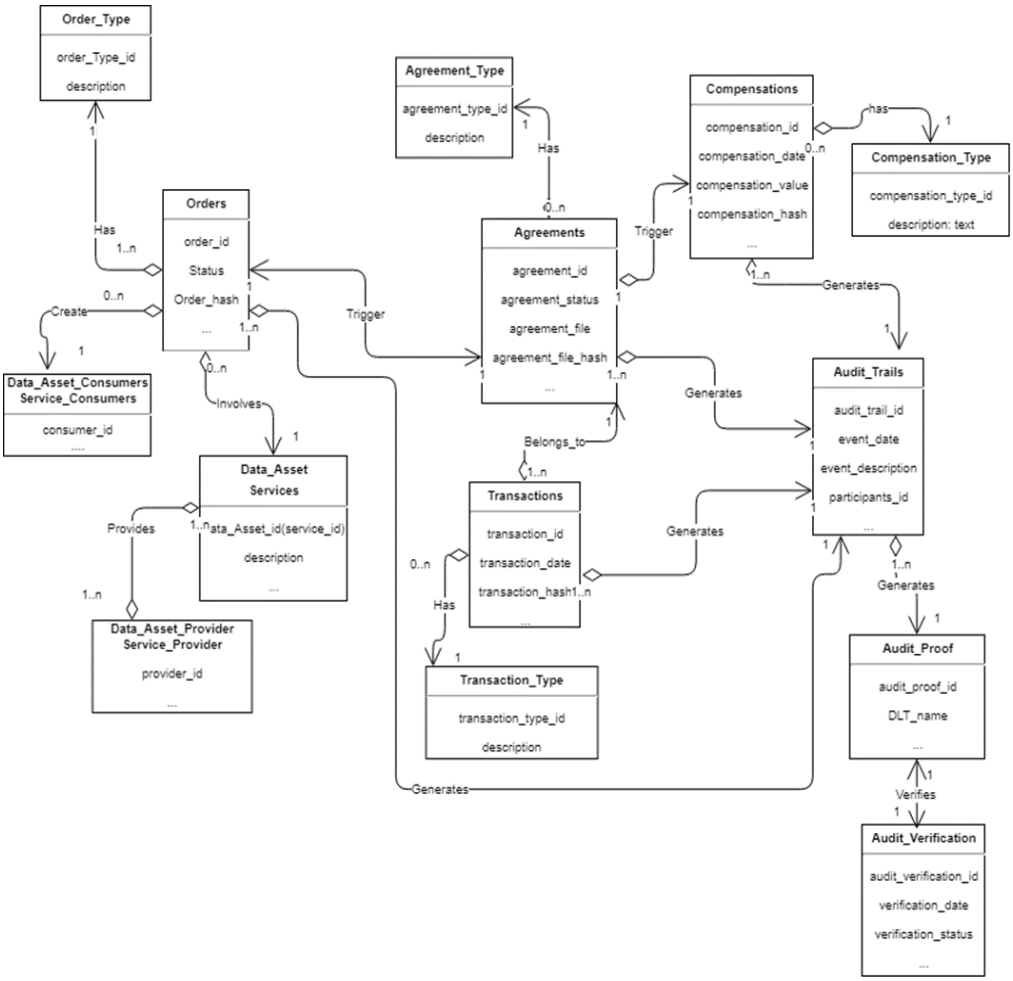

Figure 12: Marketplace data model

7.4 Summary

In this chapter, we proposed a framework as an architecture for addressing the issue of fair and responsible data market ecosystems. The framework design adopts the NIST reference model of cloud computing representation. It is based on the role of the main actors involved in data exchange and service usage and associated with tools and mechanisms allowing them to engage in an ecosystem. The main goal of this model is to outline the interaction between the participants of the ecosystem and to describe the required building blocks that will leverage

trust and sustainability in the ecosystem. Design elements defined in this chapter are independent of underlying technologies. They are intended to serve as guidelines for implementing a data market ecosystem. The next chapter will describe the agreement lifecycle for data exchange between parties.

Chapitre 8. Agreement management in data market ecosystem

Data exchange and service access in data market ecosystems require contractual agreements that provide a set of guarantee for market participants. In this regard, we analyzed the type and elements of agreements required for establishing fair cooperation.

Following, we propose an agreement manager for handling agreement lifecycle in an ecosystem and the associated monitoring activities that enable each party to verify the validity of their transactions (Nwatchock A Koul and Morin 2018). The monitoring process relies on an asynchronous collection of transaction event and data systematically processed and packaged for audit purposes. To provide a level of trust and transparency in this multi-stakeholder ecosystem, we rely on a distributed ledger to log the transactions and agreement proof to enable their traceability.

8.1 Agreements definition

Agreements govern the conditions under which parties exchange data assets and offer services in an ecosystem. Agreements are various as regards their objects, parties, binding obligations, and other constraints. As such, we define the type of agreements and their key components.

8.1.1 Agreement for marketplace usage

A marketplace is designed to protect the integrity of transaction for data exchange and service access. Access to marketplace' services is achieved by agreeing on the terms of the marketplace (TOS). TOS design must support the transparency and the fairness of the marketplace as well as the rights and obligations of parties operating in it. The common standard of terms of services includes the following elements:

- *User agreement* dictates and defines the general scope of rights and responsibilities between both parties. It integrates the service warrantee, the payment model, the dispute resolution service, etc.
- *Privacy statement* explains how a service may collect, retain, process, share and transfer personal data.
- *Acceptable use policy* is a set of rules for service restriction and sets guidelines for service use.

As an informed trusted party, the common standard will be enhanced by the ability to monitor marketplace activities, in such a way to inform the market participants about the fairness of the marketplace service execution. One relevant criteria of fairness is transparency over parties' data collection and exchange between the marketplace and third-party services supporting the marketplace activities:

- The data collected by the marketplace and supporting third-party services while delivering service support to each party
- The processing activities of the marketplace and supporting third-party services while delivering service support to each party.

8.1.2 Third-party services agreement

IT services and cloud computing domain widely use SLA as a contractual agreement between service providers and customers. Service Level Agreements (SLA) are service agreements between a service provider and an end user who are expecting the service usage in a given time frame. The main elements of an SLA as defined by Keller et al. (2003) are:

- *Involved parties*: these parties consist of one service provider and one service customer. Supporting parties represent third parties that operate on behalf of either or both signatories.
- *Service description*: Services are described and encapsulate SLA parameters, which in turn contain properties and indicate quantitative as well as qualitative metrics.
- *Obligations*: A service provider defines guarantees in the form of obligations either as service level objectives (SLO) or as action guarantee. SLOs represent measurable targets that service providers promise to fulfill during service execution.

In this research, we use SLA to express Third-Party Services agreement. In particular, we focus on SLO values that should be made available to any contracting party for Quality of Service, fine-grained data collection and fine-grained data processing activities, as described for each service type in Table 18.

Metrics measurements and monitoring	Marketplace service support	Value-Added services	Data consumer services
Quality of service attributes	Y	Y	Y
Fine-grained data collection	Y	Y	Y
Fine-grained data processing activities	Y	Y	Y

Tableau 18: Metrics measurements and monitoring

One example of value-added services is the custodian service which requires both parties, the data custodian and individual to agree on the precise bounds of service level for data coming from other services for storage. Accordingly, a data custodian must offer the following guarantees:

- Data custodian service should support data collection and storage
- Data custodian service should support data exchange between an individual and a data consumer
- Access to data for the custodian should be prohibited
- Data exchange and usage must leave an auditable trace for individuals, data consumers, and auditors

8.1.3 Data exchange agreement

Data exchange agreement is designed to support data exchange which involves data providers and data consumers and a data custodian if required. As a basis

for the agreement design, we refer to the data processing agreement template under the GDPR, required between data controllers and data processors for individual data processing. In this regard, we defined the parameters of the agreement, as follow:

- *Main parties* represent data provider and data consumer which have the authorization for monitoring data exchange activities.
- *Supporting services* represent data custodian and data consumer services along with (processor and sub-processors services) which handle the agreement for Data provider and Data consumer.
- *Fine-grained data elements* represent an atomic unit of data involved in the data exchange transaction.
- *Fine-grained processing activities and purposes* represent any atomic processing activity for data exchange with each associated purpose.
- *Technical measures and organizational measures of* Security means those measures aimed at protecting personal data against accidental or unlawful destruction or accidental loss, alteration, unauthorized disclosure or access and against all other unlawful forms of processing.
- *Report data breach* concerns the measures aimed at notifying without undue delay upon data consumer service becoming aware of a Data Breach in order to report or inform Data provider.
- *Monitoring and control metrics* are about making available all information necessary to demonstrate compliance with this agreement.
- *Governing Laws and jurisdictions* concern the legislation protecting the fundamental rights and freedoms of individuals or data provider, in particular, right to privacy for an individual with respect to the processing of personal data applicable to a data controller.
- *Dispute arbitration services* concern the services in charge of handling dispute between Parties by the jurisdiction of the courts or a third-party mediator.

8.2 Agreement instantiation, execution and monitoring

An agreement is created and used for two purposes. First, it legally binds the market parties. Second, it enables the enforcement of agreement execution and monitoring. Figure 13 provides an overview of an agreement instantiation, execution, and control.

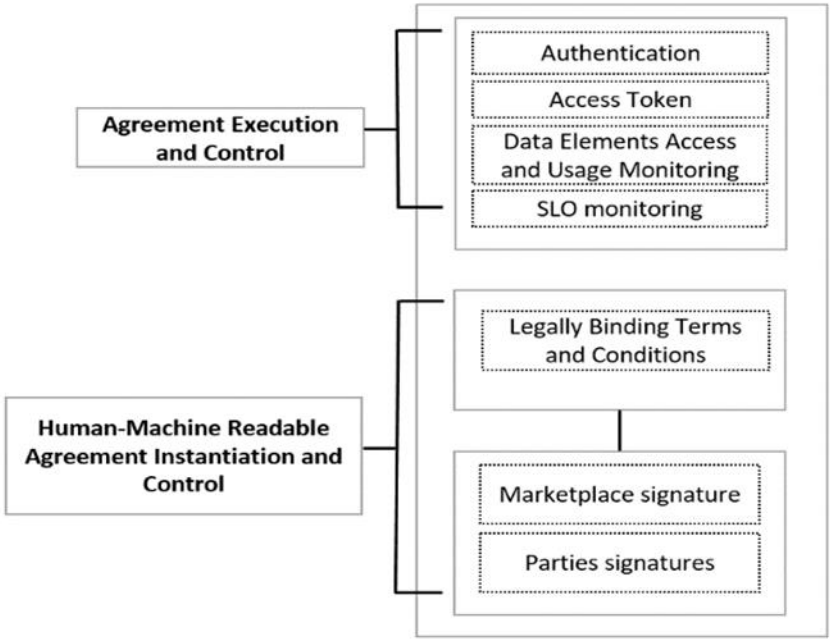

Figure 13: Agreement instantiation, execution and monitoring

Human-machine readable agreement
It constitutes a machine-processable agreement, available also in a human-readable format, and finally cryptographically signed and verified by the marketplace and parties. It is composed of the predefined agreement parameters depending on the agreement type whether for services access or data exchange. It is a legally binding agreement used for audit purpose. As such, this agreement that expresses the commitment of each party should be enforced to establish the concordance of parties' intention. The main elements of this agreement are:

- The legally-binding terms and condition which encapsulate all the elements of an agreement
- The marketplace signature which enables to verify the service which carries the agreement generation.
- The parties' signatures which establish binding obligation between parties.

Agreement execution and control
First, it enables the automation of the agreement execution process based on the enforceable parameters of an agreement. Second, it enables to assess the compliance with an agreement by comparing transaction and SLO metrics for audit purpose by any party involved. The main enforceable parameters are extracted from an agreement for enabling the automated execution. These elements are composed of:

- Parties identities verification them during the agreement execution process.
- Third-party services identities and API
- Fine-grained data elements and the agreement start data and end data for generating the corresponding access token for Third-party services
- SLA metrics for collecting corresponding metrics information.

8.3 Functions for agreement management

The main functions for agreement management are summarized in Table 29.

Functions	Descriptions
`Generate_Agreement()`	This function generates human-machine readable agreement
`Extract(information)`	This function extracts the information from an agreement
`Send(element)`	This function sends agreement or agreement hash to involved parties for signature or for analysis.
`Get(agreement)`	This function retrieves an agreement for verification of the terms and signature.
`Get_Event()`	This function collects the data pipeline transaction events
`Hash(element)`	This function hashes an element (eg. agreement)
`Sign(element)`	This function signs an agreement or an agreement's hash with its private key
`Verify_Signature (H(agreement))`	This function verifies a party signature for authenticity verification
`Match(H(agreement1, H(agreement2))`	This function compares two hash agreements for authenticity verification
`Store(element)`	This function stores an element (eg. agreement or an agreement hash) in the marketplace database
`Request(action)`	This function enables to request data storage transactions or data exchange transactions
`Notify(event)`	This function notifies an event to a party
`Verify(element)`	This function verifies that the agreement authenticity proofs against the DLT.
`Authenticate()`	This function authenticates a party identity
`Generate(token)`	This function creates a token for data collection.
`Send_Token()`	This function sends a token to a party
`Publish_Data()`	This function publishes data asset to the data pipeline
`Subscribe(token)`	This function subscribes to the data pipeline by passing the token to the marketplace for data collection.
`Submit_To_DLT(H(element))`	This function submits the final agreement hash or transactions to the DLT for permanent storage

Tableau 19: Agreements management functions

8.4 Agreement flow steps

8.4.1 Agreement creation

A *human-machine readable agreement* is generated automatically after the matching of an offering of data asset or service. Figure 14 describes the process for agreement creation.

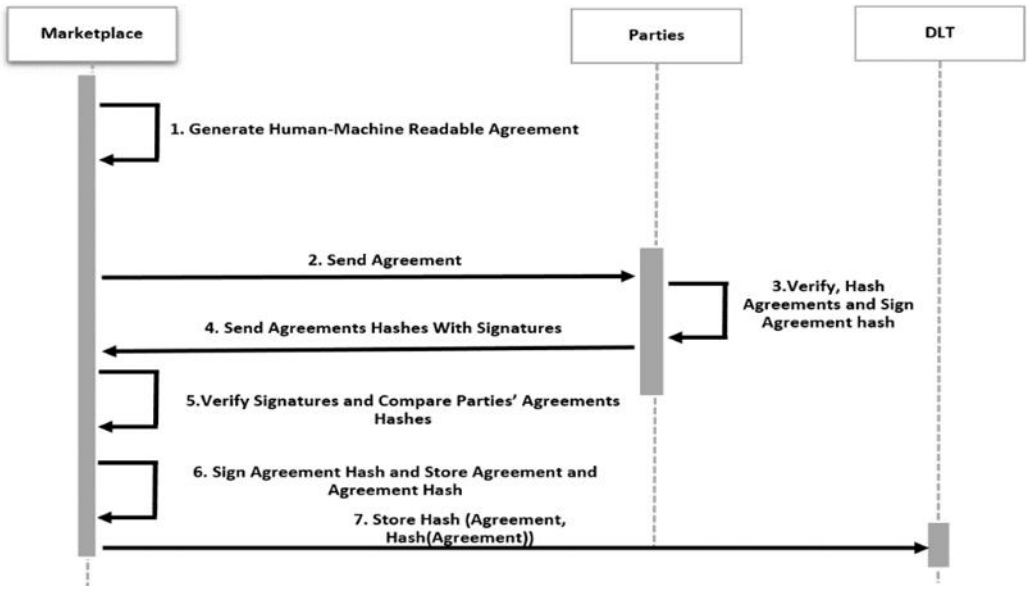

Figure 14: Creation of a human-machine readable agreement

An agreement is generated based on the agreement template, which encapsulates offering and negotiation terms.

STEP 1: In the initial step, the marketplace generates Human-machine Readable Agreement Object through the **`Generate_Agreement()`** function

STEP 2: The marketplace send the agreement to involved parties through the **`Send(agreement)`** function.

STEP 3: Each party retrieves and verifies the agreement, hash and signs it with his private key with the following functions: **`Get(agreement)`**, **`Hash(agreement`**) and **`Sign(H(agreement)).`**

STEP 4: Each agreement is sent back to the marketplace for registering in the with the function **`Send(H(agreement)).`**

STEP 5: The Marketplace verifies the signatures with the parties' public keys with the function **`Verify_Signature(H(agreement))`** and compares them with the function **`Match(hashagreement1, hashagreement2)`** and.

STEP 6: If successfully match, the marketplace select one agreement hash, signs the hash and finally hash the signed hash with the functions: **`Sign(H(agreement))`** and **`Hash(finalHash).`**

STEP 7: Finally, the marketplace stored the final hash and agreement in its database with the function **`Store(agreement, finalHash)`** and submit the final hash object in a DLT by calling the function **`Submit_To_DLT(H(finalHash)).`**

8.4.2 Agreement execution flow

Data storage execution flow
We assume an individual has selected an offering of a data custodian. Later, an agreement for data storage is set according to the agreement creation process. An individual subscribes to a value-added service which generates personal data for the individual throughout service use. By default, this value-added service requests data storage to the marketplace for handling the storage individual's data. We also assume the third-party service and the individual's data are repackaged in an open format and are usable by any data consumer. The authorization server of the data marketplace mediates the agreement execution between these parties.

As described in figure 15, the agreement execution flow for data storage between an individual and a data custodian service has the following process:

STEP 1: The subscription to a third-party service triggered a request for data storage by the third-party service by calling a **`Request(dataStorage)`** function. The goal is to meet the "responsibility by design" requirement which stipulates that an individual data is available for collection for his own use by default. Hence, the request is submitted to the authorization server, which carries all data exchange authorization.

STEP 2: The Authorization Server get the corresponding agreement proof stored by the marketplace, and that can be verified against the DLT by calling a **`Verify(proof)`** function. The function is performed to verify the authenticity of the agreement and locate the custodian service in charge of data collection. The value-added service ignores the data custodian identity.

STEP 3: If available, the agreement for data storage is returned to the Authorization Server which extracts the information about the data custodian responsible for collecting individual's data and other agreement information with the function **`Extract(information).`**

STEP 4: The marketplace creates a token for data collection with the function **`Generate(token).`**

STEP 5: The marketplace send a notification for storage request to the data custodian along with the token to the data custodian by calling these functions: **`Notify(storageRequest)`** and **`Send_Token().`**

STEP 6: The Third-Party Service publishes data asset into the data pipeline by calling the **`Publish_Data()`** function .

STEP 7: Upon receiving data, Authorization Server notifies the data custodian about the data availability with the function **`Notify(DataAvailability).`**

STEP 8: The data custodian calls the function **`Subscribe(token)`** by passing the token to the marketplace data pipeline.

STEP 9: The marketplace collects the data pipeline transaction events along with the metadata from the third-party service with the function **`Get_Event().`** These events are sent to the monitoring service of the marketplace, hashed and save in the DLT with the two following functions: **`Hash(transactionEvents`**) and **`Submit_to_DLT(H(transactionEvents)).`**These events will be published to each involved party's dashboard.

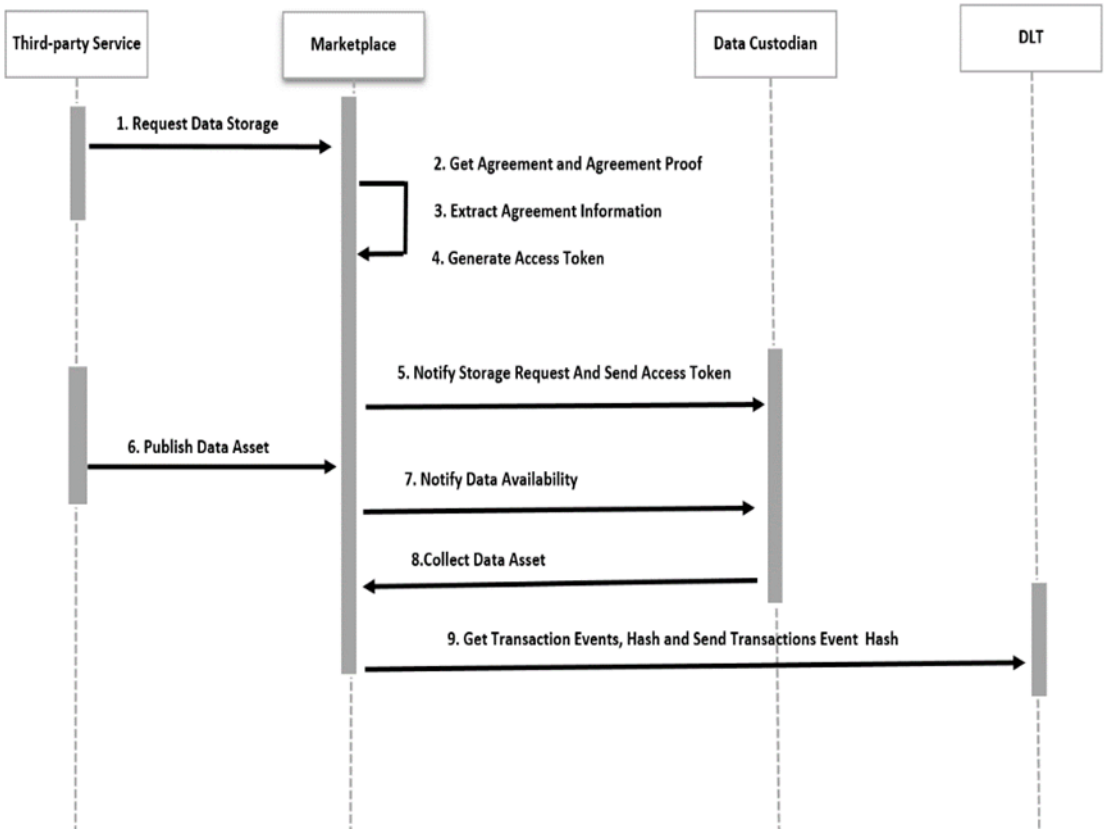

Figure 15: Diagram flow for data storage agreement

Data exchange agreement execution flow

Data exchange agreement occurs between a data consumer, a data provider and also a data custodian when the data provider is an individual. In any case, data custodian corresponds to service carrying data exchange on behalf of an organization. Figure 16 summarized the flow step for this agreement fulfillment.

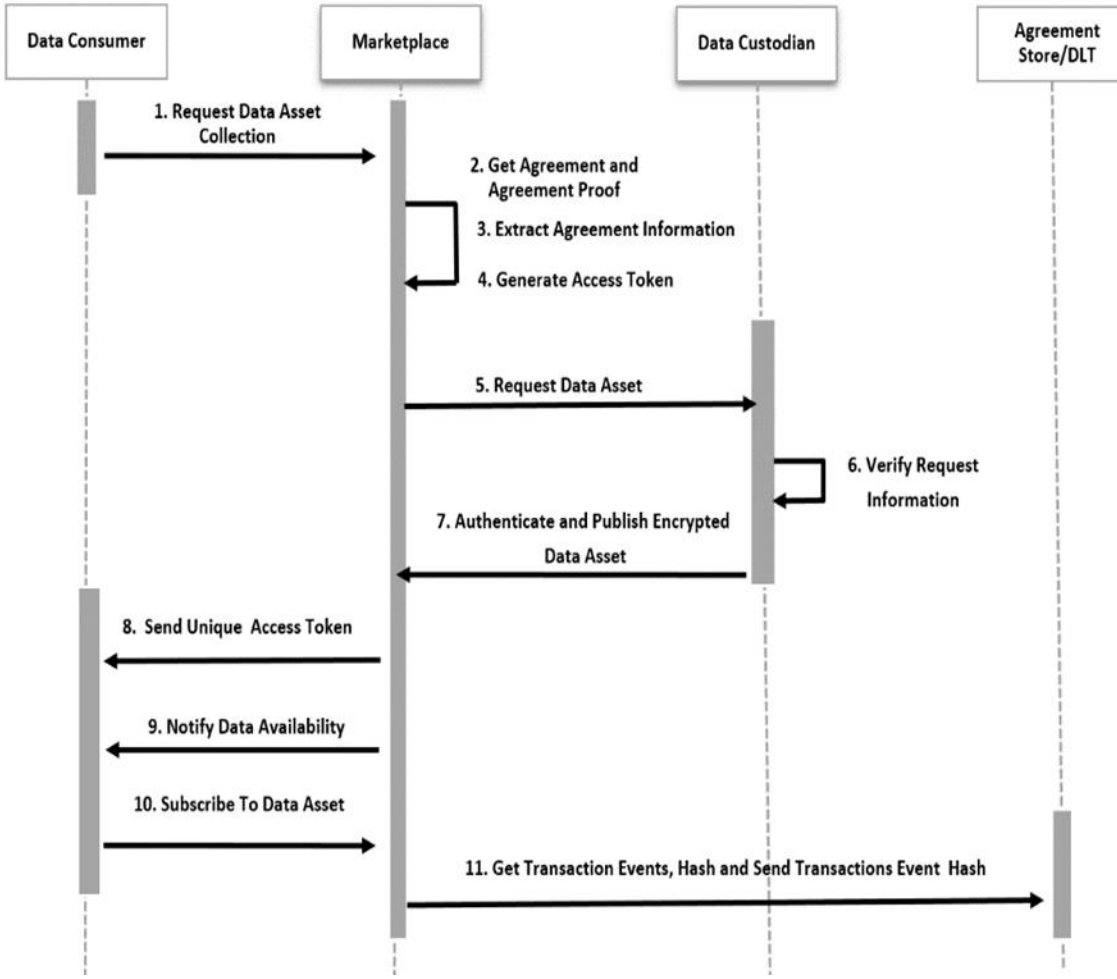

Figure 16: Data exchange agreement execution flow

STEP 1: After the creation of a data exchange agreement between a data consumer and a data provider, a data consumer can request data asset to the authorization server of the marketplace by calling the **`Request(dataExchange)`** function.

STEP 2: The authorization server get the data exchange agreement proof stored in the agreement database and that can be verified against the DLT by calling a **`Verify(proof)`** function. This operation is performed to verify the authenticity of the agreement and locate the reference of data storage agreement along with the data custodian service in charge of data provisioning.

STEP 3: If available, the data exchange agreement along with the proof is returned to the authorization server which extracts the reference of the custodian service and the granular data asset elements with the function **`Extract(information).`**

STEP 4: The authorization server creates a One-time Access token for data collection with the function **`Generate(oneTimeAccessToken).`** The goal is to request a token for each data usage.

STEP 5: The authorization server sends a request for data asset collection to the data custodian by calling these functions: **`Request(dataExchange).`** The request is sent to the identified data custodian with the following information: data storage agreement reference, marketplace signature, and granular data asset elements.

STEP 6: The data storage agreement and the marketplace signature enable to verify the request origin and the condition of data exchange fulfillment. The Data Custodian Service verifies the Marketplace Signature, the storage agreement, and the agreement proof by calling the **`Verify(signature, agreement, proof)`** function.

STEP 7: If successful, the custodian service authenticates to the Marketplace with the **`Authenticate()`** function and publish the data asset with the **`Publish_Data()`** function.

STEP 8: The Authorization Server sends a One-time Access Token to data consumer with the **`Send_Token()`** function.

STEP 9: Upon receiving data, the Authorization Server notifies the Data Consumer about Data asset availability with **`Notify(DataAvailability`**).

STEP 10: The data consumer calls the function **`Subscribe(token)`** by passing the token to the Marketplace marketplace's data pipeline.

STEP 11: The marketplace collects the data pipeline transaction events along with the metadata from the Data Custodian with the function **`Get_Event().`** These events are sent to the monitoring service of the marketplace, hashed and save in the DLT with the two following functions: **`Hash(transactionEvents)`** and **`Submit_to_DLT(H(transactionEvents)).`** These events will be published to each involved party's dashboard.

Agreement termination

Agreement termination involves the end of market participant relationship under a particular agreement. An agreement specifies the terms and conditions of contract termination and situations under which the relationship between market parties should be legally ended. The termination of agreement initiates a termination process. All data related to each party should be removed or transferred to them. Only essential information must be retained by the marketplace and other Third-party services for legal compliance.

8.5 Individual data alteration

An alteration of individual data in the data custodian service must be handled through the marketplace in order to provide proof of data source authenticity. Any modification that occurs on the individual data must be handled by the marketplace even when initiated by an individual itself. Like third-party services, an individual must request authorization for updating or delete his data without the need to rely on an agreement. By verifying the identity of the individual, the

marketplace will provide a token for data update, creates the corresponding transaction and submits the corresponding event to the DLT. We also need to consider the calculation of the data asset hash each time that a modification occurred by the marketplace and logged in the DLT will enable to keep a proof of origin of the data asset and the version of the data asset.

8.6 Access token generation and verification

With valid agreement, a data consumer or a third-party service can get an access token from the marketplace authorization server. The authorization server sets the token claims according to the data usage permission. Figure 17 provides an overview of the access token design.

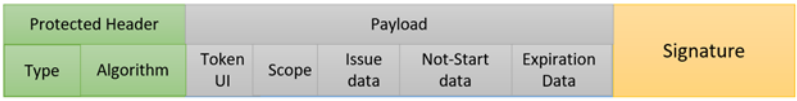

Figure 17: Token model for data exchange

The token claims use the same structure as the JSON Web Token (JWT) specification (JWT, RFC 7519) consisting of three parts such as:

- Protected Header: The protected header is a JSON object that includes the header elements that has to be integrity protected by the signing or MAC algorithm. JWS Header declares that the encoded object is a JWT and the JWS Header and the JWS Payload are signed.
- Payload: The payload carries the token claims. This includes information extract from an agreement such as the issuer of the token, the scope of the token, the expiration date value, the time before which the token cannot be accepted for processing, the time the JWT was issued, the unique identifier for the JWT. Can be used to prevent the JWT from being replayed.
- Marketplace Signature: This signature is composed of a hash of the protected header, the payload, and a marketplace secret.

Token payload information

The unique token identifier fields contain a subject field and an audience field. The first identifies Third-party service API or data consumer API that is the target of this token. The latter stores the Third-party service public key for which the token is created. The scope of the token describes the different data element a third-party service or data consumer is allowed to collect once the token gets validated. These actions are application dependent. The expiration, not-before and issued-at fields store timing information on when the token can be used and when it was created.

8.7 Distributed ledger integration

A DLT integration provides an option to prevent tampering and ensuring integrity and auditability of data and transactions (Xu et al, 2016). In a data market ecosystem, issues such as data provenance, agreement integrity, and auditability of transactions and data are mandatory to build trust. DLT integration as an

immutable registry of transactions and data exchange can radically enhance the handling of these issues in a data market ecosystem. Key considerations for DLT integration must consider the transactions that should be traced, the data processing that should be monitored and the privacy regarding personal data disclosure.

A DLT is used for logging all activities around data processing. The content of a DLT reflects historical and current states of information recorded in the ledger maintained by its network. A data market ecosystem must define the information and associated data model to be stored in the ledger in such a way that respects privacy and enable the search in the DLT. Only information required to be tamper-resistant, transparent and traceable should be recorded in the Distributed ledger. Therefore, a party may rely on a DLT as trusted service. The agreement stored in the marketplace database and its hash stored in a DLT will be used as a proof to check an agreement authenticity. By storing an agreement hash in a DLT at its creation, a marketplace insures that the terms of the agreement are immutable. The marketplace will get the references of each block created for a specific agreement hash and send it to the parties involved in order to facilitate the search of an agreement.

Moreover, any transaction or activity should be logged in a Distributed ledger. The events should contain information about:

- The party responsible of the activity
- The activity purpose
- The date and Time of the activity
- The activity's object

The Hash of personal data transaction can be recorded in a distributed ledger for data provenance, authenticity and integrity checking.

8.8 Summary

In this chapter, we covered the definition of agreements between market participants for service usage and data exchange.

We proposed some agreement elements for enhancing transparency on service consumption and data processing. Furthermore, we define agreement flows for service consumption and data exchange. Finally, we discuss the need to constrain data alteration from an Individual and the integration of a DLT to the marketplace for providing a higher level of trust and fairness in this multi-party ecosystem.

In the next chapter, we describe the prototype implementation of the proposed framework as a reference implementation.

Chapitre 9. Framework implementation

In this chapter, we present the implementation of the framework architecture as a reference implementation. We start by describing a scenario that serve in designing the prototype.

We implement the core elements of our general framework, which is the marketplace to demonstrate the feasibility of our approach. The marketplace is composed of seven services: an authentication and authorization service, an ontology-based data catalog service, a matching service, an agreement manager service, a monitoring audit service, and an API gateway connector. We also implement some REST services that plays the role of third-party services. To support our implementation model, we made a number of assumptions regarding the ideal environment for operating a fair and responsible data market ecosystem. We assume that data are exchanged in a structured format.

9.1 Requirements for system prototype

The system requirements covers the technical components necessary to implement our framework. We define the requirements in such a way that gives the possibility to parties to interact following the predefined design concept. The system requirements are divided into four main parts:

- *User interface* allows market participants to interact with the marketplace' services and third-party services. For instance, data providers and service providers can compose and publish their agreement templates, obtain their transactions audit trails, etc. Therefore, the user interface enables any party to interface with the marketplace for performing specific tasks like proposing data asset for exchange, manage transactions, monitor transaction, etc.
- *Marketplace's services and persistent storage* handle the marketplace's activities and enable the storage of the information concerning those activities and transactions.
- *Marketplace's connectors and plugins*. They work as a bridge between the Marketplace and the third-party service API.
- *DLT* is used to verify agreements integrity and enable data assets and transactions events auditability. With the DLT integration, it is possible to prove the existence of an element (example: agreement, data asset, transactions events) at a specific time as well as its authenticity. The use of hashing algorithms enables to keep the original element publicly unavailable in the DLT. Key considerations for choosing a DLT are separate in primary properties which are mandatory and secondary properties which are desirable.
  Primary properties. The DLT must accept different format of data. As an example, the storage and retrieval of hash information must be supported. The DLT network must not be controlled by a single organization, in such a way that prevents a single authority to validate and controls the hashed elements stored in it. Furthermore, the DLT should be widely available and sustainable as the operational stability of this network is absolutely critical to enable the long-term preservation of the hashed elements.

Secondary properties. The DLT does not require owning a currency for use. We should prioritize the use of public DLT over private DLT to allow everyone to become a node and verify transactions. In fact, private DLT has a select group of entities that can become nodes. In such a context, we must define the governance body for the DLT.

Figure 18 provides a high-level overview of the key components of the technical architecture.

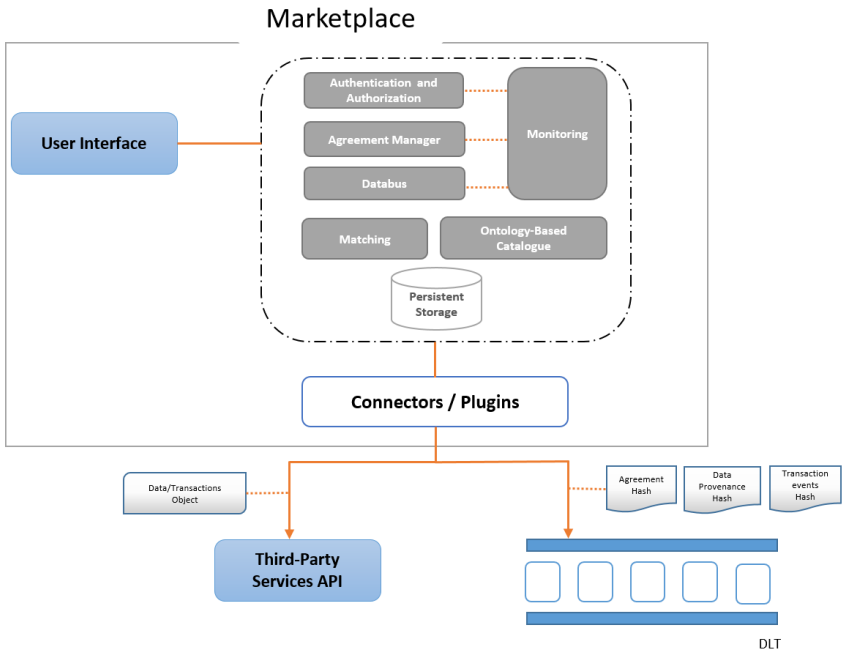

Figure 18: Prototype architectural component

9.2 Implementation

The main goal is to demonstrate the framework instantiation in the context of real-world applications. For that purpose, a scenario borrowed from the quantified-self domain is used to exhibit the relationship between the framework concepts and the phases followed for the design of a data market ecosystem. The prototype is developed with JavaScript frameworks and based on REST architecture. Parties process and receive data through the calling of RESTful APIs. Besides the data asset and service offering description stored in a knowledge graph, all marketplace's related information is stored in MongoDB documents.

9.2.1 Application to a simulated scenario

The quantified-self movement, defined as self-knowledge through self-tracking, aims at tracking every moment and aspects of lives via wearable technologies, fitness apps, monitors, etc. It is also concerned with the collection and analysis of data related to our daily lives. In this research, we chose to study the requirements

for collecting and exchange personal data from a quantified-self application as an example for four reasons:

- The categories of data processed by the quantified-self applications and services are a useful predictor for handling fair and responsible data exchange.
- The availability of real personal data from quantified-self applications;
- The self-involvement of individuals in the collection of different facets of their life to take action for their lifestyle improvement;
- The focus of quantified-self application on straightforward monitoring of individual habits enables a mere accumulation of data and the processing of personal data whose result is only profitable to individuals and organizations delivering services for healthy lifestyles. Much of these services fail to provide their users with broad access to their own data without burdens. Moreover, they did not acquire granular consent for their use (Hutton et al. 2018). In this context, we are able to separate the roles and concerns of each party.

Therefore, our scenario is about the exchange of quantified-self personal data with the following market participants:

- **Value-added Service**: the quantified-self service for tracking sleeping, calories, and sports activities data of an individual. We simulate a self-tracking service that generates personal data for Individual. It is a REST API supported by a MongoDB database and which has two functions:
  - **`Verify(token)`** which verify the access token delivered by the marketplace
  - **`Get_Data()`** which delivers data to a data custodian API according to the token parameters. This function returns fine-grained data in a JSON schema
- **Individual**. The data provider whose data are generated from the use of the quantified-self service;
- **Data consumer**: A practitioner service which collects specific data asset like calories and sport activities data from multiple individuals for analyzing the relationship between the burnt calories and the sports activities of an individual in order to derive daily efficient sports practices and recommends healthier habits such as daily aerobic exercise, accurate diet, etc. Data consumer service is a REST API exposing one function:
  - **`Process(data)`** which consumes the calories and sport activity data of an individual.
- Data custodian: A custody service which collects personal data from the quantified-self service on behalf of the individual. We design a REST API service, which stores data, accepts and verify token for data provisioning. The API is supported by a MongoDB database and has three functions implemented:
  - **`Store(data)`** function which stores personal data coming from a given value-added service as a MongoDB document,
  - **`Verify(token)`** function which verify the access token delivered by the marketplace,

- **`Get_Data()`** function which sends data to a data consumer API according to the token parameters. It returns fine-grained data in a JSON schema.
- **Marketplace**: A set of services (described in section 9.2.2), which enables the interaction with the main parties. The marketplace will enable to exchange these quantified-self data assets and the integration of all the parties.

In the beginning, each party (quantified-self service, data custodian service, the practitioner service, and the individual) registers and authenticate to the marketplace. The individual subscribes to both custody service and quantified-self service via the marketplace. It is assumed that the parties reciprocally have agreed on the terms, conditions, and policies. The individual gets in return the data model and data element description from the quantify-self services and the SLA of the data custodian as a JSON file. Next, he uploads these descriptions in the marketplace catalog which concatenated them as a data asset offering. Therefore, the marketplace generates the corresponding agreement template, which is fulfilled partially by the individual and registered in the agreement template store.

Each usage of the quantify-self service generates data. In order to send the quantified-self data to a data custodian, the quantify-self service API connects to the marketplace, authenticates and initiates the data storage process flow. Likewise, the exchange of the quantified self data between the practitioner and the custody service follows the data exchange process. The marketplace APIs integrates a data bus which manages data transfer between parties. Each transaction is logged and counted by the monitoring service and regularly updates the information on the marketplace. The events are hashed and then stored in a DLT.

9.2.2 Core marketplace prototype

The marketplace is designed as a set of micro-services interacting through an API Gateway (Figure 19). Each micro-service is supported by a MongoDB database except the ontology-based catalog service which is backed by a knowledge graph.

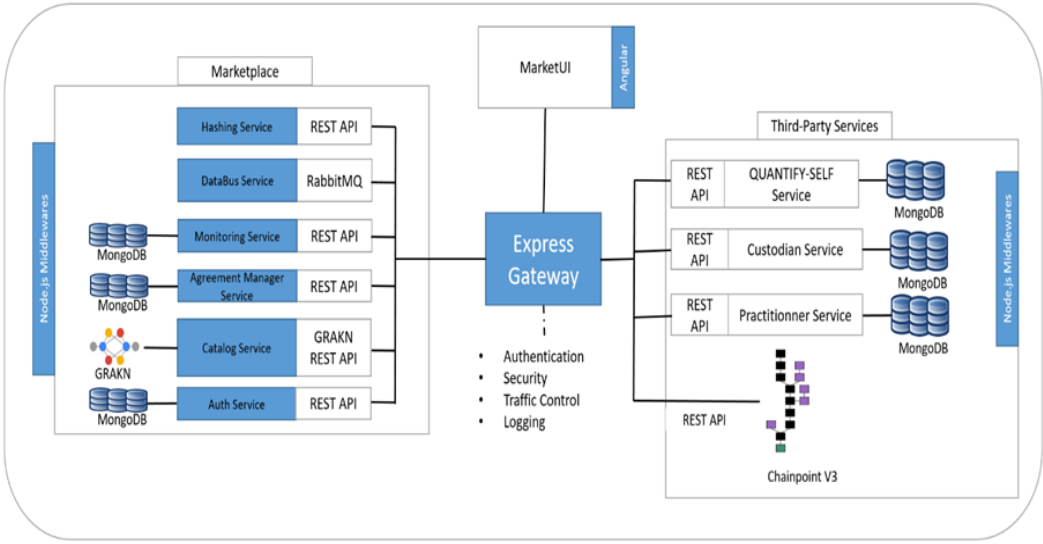

Figure 19: Prototype technical architecture available in Github (Appendix B: https://github.com/sabrina-ossey/MarketFramework)

The data marketplace exposes an Express Gateway (2018) as the API connector. The Express Gateway is a microservices API gateway that sits at the heart of any micro-services architecture, secures micro-services and expose them through APIs using Node.js, Express and Express middleware. As it also supports third-party services integration, all the parties and marketplace components interact through the Express Gateway.

We use Chainpoint (2017), an open standard for creating a timestamp proof of any data, file, or process to link our monitored information to the Chainpoint Calender blockchain (CAL) for storing the monitored elements hashes. For the integration, we use then Chainpoint REST API. For our purpose, we simulate the storage of hashed elements on 2 Chainpoint V3 nodes. Each Chainpoint node receives the monitored elements hashes which are aggregated together using a Merkle tree. A Merkle tree, or binary hash tree, is a data structure used in the DLT for efficiently summarizing and verifying the integrity of transactions.

User interface

The user interface is a web-based graphical interface developed with Angular framework enabling a party to connect to the marketplace and operate on it. It provides an authentication interface, a data and service catalog interface, an agreement manager interface, and a dashboard where a party can observe current marketplace activities as well as transactions history.

Authentication and Authorization service

Each party and third-party service API must authenticate before operating in the marketplace in such a way that uniquely identifies them and provides adequate services. For the authentication process, we use self-signed certificates to authenticate the parties. This is carried out by OpenSSL that generates the certificates for enrolling and operating in the marketplace. The authorization

server assigns an identifier to the party and API that is bound to them each time they request a service. The parties also use the private key of their certificate for signing agreement hashes.

We use the JWT of OAuth 2.0 Bearer for generating the access token for data exchange. For designing a one-time access token, we use an HMAC-based JWT signature. The authorization server generates a unique secret that is used to sign the token and is securely shared with the data consumer, hence guaranteeing a single-usage of every issued token.

Ontology-based catalog service

This service enables a data provider and a service provider to describe their data asset or service offering. Data assets and services are described following the data model and service model description in Chapter 7 as the basis for ontology modeling. The ontology-based catalog service is built with GRAKN.AI (Grakn, 2019), a knowledge graph which provides an integrated and intelligent database for semantic data search. We choose Grakn as it exposes a high-level knowledge model, allowing to represent an application domain as an ontology, specifying it in terms of:

- *Entities*. An entity is anything with a distinct existence in the domain such as “Organization”, “Individual”, and “Data Asset”.
- *Relations*. A relation describes how two or more entities are in some way connected to each other.
- *Roles*. Describes the participation of entities in a relation. For example, in a data processing "Relation", there are roles of data controllers and data processors respectively.
- *Resources*. Represents the properties associated with an entity or a relation, for example, a name or date. Resources consist of primitive types and values, such as strings or integers.
- *Attributes*. An attribute is a piece of information that determines the property of an element in the domain.

We define both ontologies for data assets and services according to the ontology formalism of GRAKN. We use the GRAKN Loader Client API for uploading data asset and service description in GRAKN. This allows objects and relationships to be categorized into distinct types, enabling automatic reasoning over the represented knowledge, such as inference (extraction of implicit information from explicit data) and validation (discovery of inconsistencies in the description). The step for building and populating one GRAKN ontology are:

STEP 1: we define the elements of GRAKN ontology in an “ontology.gql” file. Figure 19 shows the definition of data asset ontology.

STEP 2: we load and test the ontology in the GRAKN Keyspace of the GRAKN server.

STEP 3: we design an API for integrating the data asset description or service description into the GRAKN server. This API enables to load a description in JSON format into the GRAKN server. The description file is parsed and then loaded in the Grakn Keyspace. A data provider or service provider can load his description file in JSON format via the user interface.

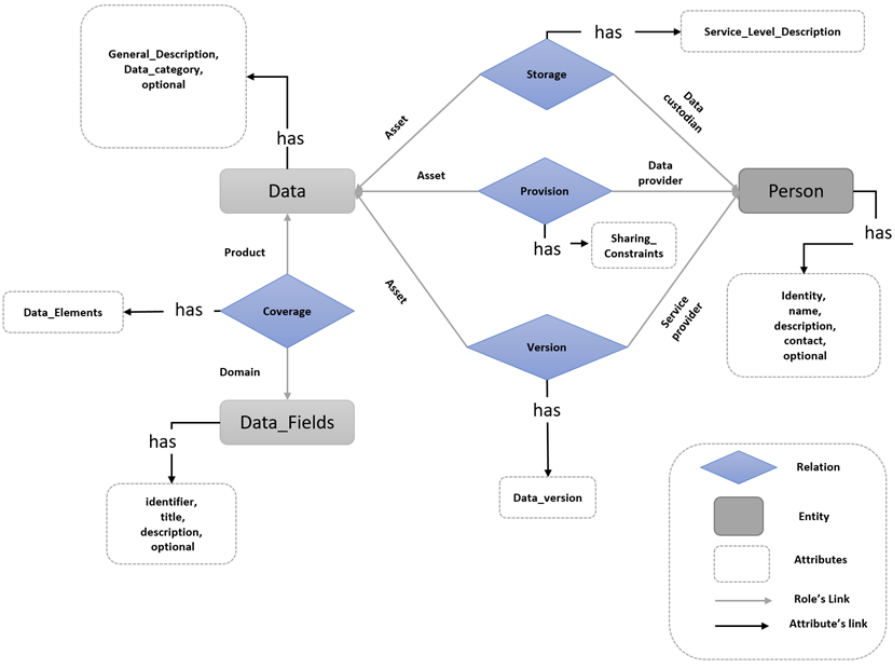

Figure 20: Ontological data asset model schema

Matching service

We use the Graql language, the query language of GRAKN that uses machine reasoning to retrieve data assets or service offerings. The matching service exposes an API which enables a party to query the knowledge graph by entering some keywords in a simple search bar or by selecting an offering from the list of available offerings.

Agreement manager services

The agreement manager, represented in Figure 21, handles the agreement creation and storage. The agreement manager consists of two micro-services with their functionalities described below:

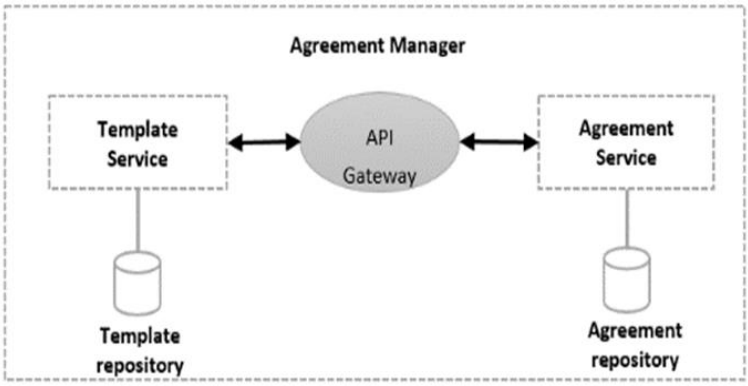

Figure 21: Agreement Manager Service

- *Agreement template service API*: Based on a data asset category, an agreement template for data exchange is generated following the loading of data asset ontology in the ontology-based data catalog. Then, a data provider fills the agreement template and submits it to the template service for recording. This template is saved in the template storage and contains prefilled fields along those to be filled in during the agreement process. The SLA template is uploaded by a third-party service provider at the load of the service ontology in the ontology-based data. Both agreement templates contain a set of constraints on the fields that express the data exchange guarantees and the quality of service. We set up one MongoDB template database with two documents. One for the storage of data exchange agreement template and another for SLA template. The generation of agreement template for data exchange is based on the data taxonomy cube developed using dynamic components generation of the Angular framework. Each data property of the data taxonomy cube is associated with a list of corresponding agreements clauses.
- *Agreement service API*: At the creation of a data asset or service order, the agreement manager retrieves the agreement template offered by a provider. This results in the creation of a human-machine readable agreement file in JSON format which is sent to each party for filling and final validation. Each party must hash the agreement and sign the agreement hash. Three options are possible to handle this task:
    - The marketplace can propose hashing and signing operations that execute locally.
    - Each party can handle the hashing by using a third-party service
    - Each party can handle the hashing and signing operations with his own service.

We use the first solution for our implementation. Each party calls SHA512 hashing algorithm of the crypto module of angular which hashes the copies of the agreement of involved parties and enables each party to sign the agreement hash with their private keys. The hash is then uploaded to the agreement service API with the following information: the signed hash, and the agreement reference. The agreement service retrieves the agreement and compares the submitted hash and signatures of the involved parties. Next, the agreement service submits a copy of the agreement hash to both chainpoint nodes through the chainpoint API. When successfully submitted to and validated by the chainpoint network, an event is returned with the proof about the hash submission. Next, the agreement manager stores the agreement hash and the proof in the agreement database and send them to the monitoring service.

Real-time monitoring service

For designing our monitoring service, we use a Node.js framework Socket.io for bi-directional, event-based communication between the marketplace services and clients. It allows us to receive and emit events in real time whenever a data exchange or storage transactions happen in the marketplace. The monitoring service listens for creation and change events within the marketplace. It registers

these events into a MongoDB database according to the involved parties (who), the purpose (why), the date (when), the elements (what).

When an event is registered by the monitoring service, it initiates the process of creating an immutable audit trail following these steps:

STEP 1. An element hash is submitted to 2 Chainpoint nodes using the Chainpoint v3 protocol.

STEP 2. The element proof is retrieved from the 2 Chainpoint nodes, generate the proof verification, and stored against the hashed elements in the corresponding MongoDB database of each element.

STEP 3. The monitoring service filters the events according to the involved parties and submit them in to each party dashboard.

STEP 4. Any party is then able to retrieve the proofs of an element and the verification proof.

In the prototype, the proofs and the verified proofs are displayed alongside the corresponding elements (Figure 22 and Figure 23). The verified proof, generated with Chainpoint, proves that the hash has been included in the Merkle tree and includes the timestamp of when it was submitted to the Chainpoint V3. The information about the Merkle root is also attached to the verified proof.

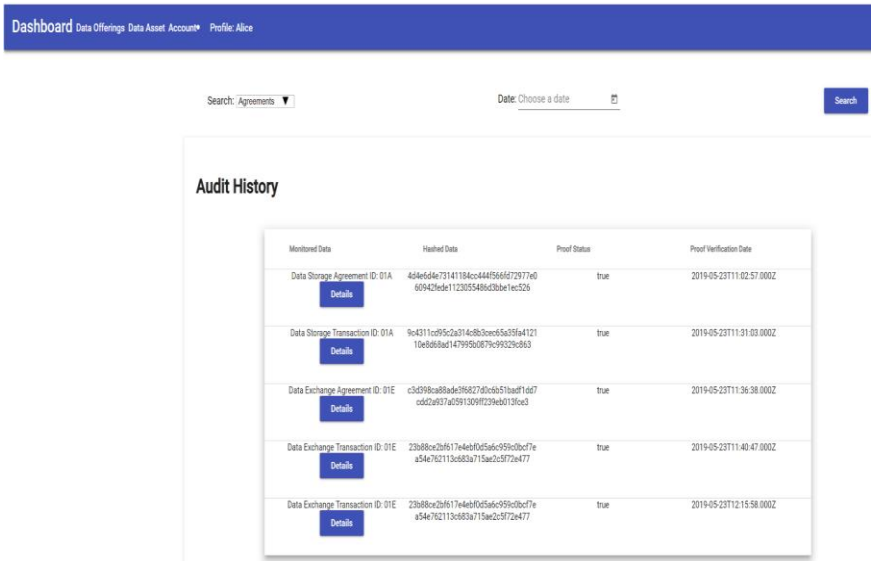

Figure 22: Audit trails example

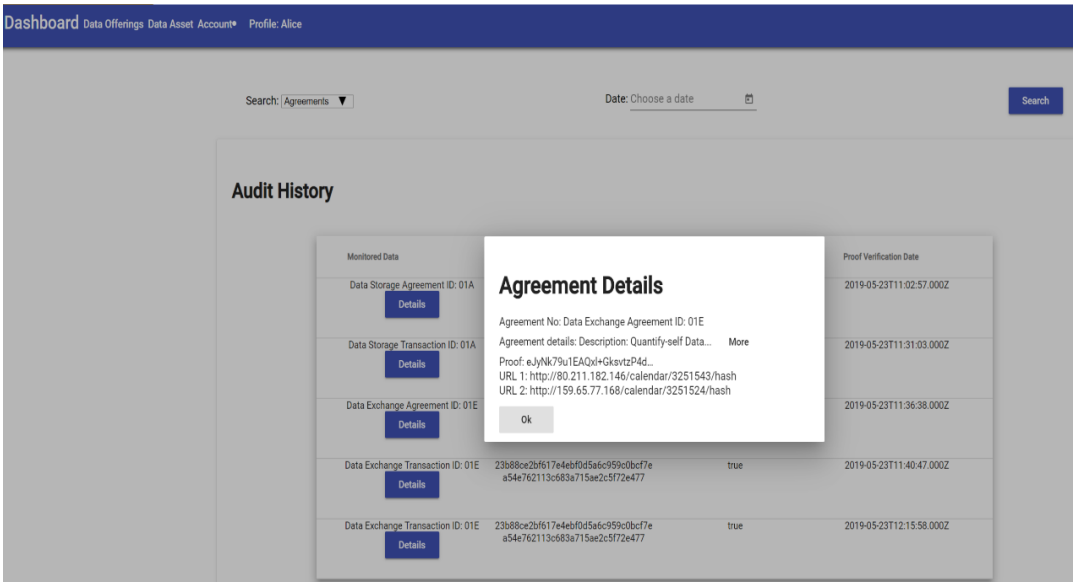

Figure 23: Details of an audit element

Data provisioning

We use the RabbitMQ message broker to support the data provisioning between the value-added service, the data custodian, and the data consumer. RabbitMQ is an open-source enterprise messaging system modeled on the Advanced Message Queuing Protocol (AMQP) standard. It is used to design a data bus provided by the marketplace. To guarantee the security and the confidentiality of data asset transfer, we configure RabbitMQ to handle TLS connections. One could enhance this feature by transferring only encrypted data.

9.3 Summary

In this chapter, we described the implementation of the prototype of our framework. First, we identified the requirements for technical implementation along with the enabling technologies. Second, we create a scenario which enable the implementation of each part of the framework. Finally, we described the implementation steps and the technologies used. In the next chapter, we evaluate the prototype, discuss the possible enhancement of the framework and the limitation of this work

Chapitre 10. Evaluation and discussions

Our general framework enables to define fair and responsible cooperation between marketplace participants. The main components were successfully implemented in our platform prototype. Our functional platform is available for demonstration purpose (Appendix B: https://github.com/sabrina-ossey/MarketFramework). The source code of the platform is also available as open source under general public licenses. The prototype demonstrated the feasibility of our design process and the designed artifact. The design of this prototype platform is based on the predefined requirements through which we can analyze, design and effectively implement the components of data market ecosystems. During the design process and the implementation phase, the ultimate goal was to provide a different approach for data exchange. Therefore, the evaluation of our design artifact is chosen accordingly.

The design science research specifies the need for validation of the research outputs, especially, the models and instantiations designed as part of the research. Different approaches have been proposed in the scientific literature for the evaluation of information technology artifacts. One evaluation approach explained by Hevner et al. (2004) can be achieved in terms of functionality, completeness, consistency, accuracy, performance, reliability, usability, fit with the environment, and other relevant quality attributes. They suggest five evaluation methods outlines in Table 18 that should be matched appropriately with the designed IT artifact:

Design Evaluation Methods	
Observational	Case Study: Study artifact in depth in business environment
	Field Study: Monitor use of artifact in multiple projects
Analytical	Static Analysis: Examine structure of artifact for static qualities (e.g.complexity)
	Architecture Analysis: Study fit of artifact into technical IS architecture
	Optimization: Demonstrate inherent optimal properties of artifact or provide optimality bounds on artifact behavior
	Dynamic Analysis: Study artifact in use for dynamic qualities (e.g. performance)
Experimental	Controlled Experiment: Study artifact in controlled environment for qualities (e.g., usability)
	Simulation - Execute artifact with artificial data
Testing	Functional (Black Box) Testing: Execute artifact interfaces to discover failures and identify defects
	Structural (White Box) Testing: Perform coverage testing of some metric (e.g., execution paths) in the artifact implementation
Descriptive	Informed Argument: Use information from the knowledge base (e.g., relevant research) to build a convincing argument for the artifacts utility
	Scenarios: Construct detailed scenarios around the artifact to demonstrate its utility

Tableau 20: Design Evaluation Method (Hevner et al. 2004)

March and Smith (1995) state that "*the evaluation of constructs tends to involve completeness, simplicity, elegance, understandability, and ease of use*." The evaluation of models should be done in terms of their fidelity with real-world phenomena, completeness, and level of detail, robustness, and internal consistency. Furthermore, to inform researchers in the field, the new model must be positioned with respect to existing models. Evaluating instantiations are complicated because it is difficult to separate them from constructs, models, and methods which they embody. And finally, March and Smith (1995) mention that in design science "*evaluation is complicated by the fact that performance is related to the intended use, and the intended use of an artifact can cover a range of tasks*". Vaisnavi and Kuechler (2007) also identified a set of patterns for the evaluation and validation of the research artifact which are: demonstration, experimentation, simulation, metrics usage, benchmarking, logical reasoning and mathematical proofs. A specific pattern may be used according to its appropriateness and the strength with which it proves the validity of a designed solution.

Table 19 illustrates how the two basic activities of design science, build and evaluate are implemented in this research. The building is the process of designing constructs, models, methods, and instantiations according to initial goals. Evaluating is the process of determining how well the constructs, models, methods, and instantiations perform compared to the initial goals and by using a set of metrics.

	Goal	**Build Outcome**	**Evaluation Metrics**	**Methodology /patterns**
Construct	Identify the relevant data categories	Data taxonomy	Completeness and understandability	Literature reviews
Model	Describe the core element of a data market ecosystem	Framework for data market ecosystem	Fidelity with real word phenomena, completeness, internal consistency	Literature reviews, instantiation
Instantiation	Apply the framework to an application.	Prototype	Applicability, more to be explored and tested in further research.	Demonstration

Tableau 21: Evaluation and validation of research outputs

Evaluating our framework for data market ecosystem can be done through four direct qualitative methods:

- Compare the framework with the literature,
- Evaluate the framework by practitioners
- Test the framework with use cases

March and Smith (1995, p.260) indicate that "*building the first of virtually any set of constructs, model, method, or instantiation is deemed to be researched, provided the artifact has utility for an important task. The research contribution lies in the novelty of the artifact and in the persuasiveness of the claims that it is effective. Actual performance evaluation is not required at this stage*" as argued

by Vaishnavi and Kuechler (2007) who state that "*The demonstration pattern is appropriated if the solution is novel and solves a problem for which no solution exists".*

10.1 Evaluation

The objective of the evaluation is to demonstrate that our framework is defendable and realizable for a set of predefined situations. At this stage, our framework is a prototype and as such represents a proof of concept. Therefore, evaluation methods intended to optimize systems or evaluate performances would not be appropriate at this stage. Moreover, the application is not an innovation in terms of technologies or communications protocols. It is an innovation in terms of design elements that introduce fairness, trust, and responsibility in the data market ecosystem. As the acceptance of the data market ecosystem depends on several prerequisites (structured data model element, acceptance of individual into the process of data exchange), it is not possible to conduct user test and apply observational evaluation methods. For these reasons, we have chosen the use of the *demonstrative pattern*. As Vaishnavi et al. (2007) argued, the construction of a prototype demonstrates that an artifact is reasonable for a set of predefined situations.

10.1.1 Evaluation by prototype demonstration

Based on a scenario defined in section 9.2.1, we designed a prototype to demonstrate that our theoretical framework is achievable and valid in the predefined boundaries. In evaluating the utility of our proposed artifact, we will use the following criteria:

- The usability and usefulness of the prototype for data exchange tasks based on the scenario
- The extent to which the data taxonomy cube enables the generation of suitable agreement based on the data category
- The extent to which the prototype helps in empowering data subjects and leverage trust, fairness and responsibility in the data market ecosystem

We argue that the prototype enables to build a data market ecosystem where the actors interact in a transparent manner with respect to these criteria because it provides an environment where any action is monitored and open for auditing tasks.

Instantiation of the scenario

By instantiating the predefined scenario of section 9.2.1, we provide data exchange use cases between a data subject, a data custodian, a quantified- self service provider, and a practitioner service. The quantified-self service describes its service's APIs based on the service definition model of section 7.2.2. We assume that the data subject has subscribed to this service. The quantified-self service API's are connected to the marketplace via the express API gateway, which enables them to push the data of the data subject in the marketplace data bus whenever new data assets are produced.

The quantify-self service monitor the daily activity of the data subjects. We simulated daily data production by using data generated from a Fitbit service

(Appendix D). Then dataset are packaged into JSON format and submit to the data custodian for storage (Figure 24).


```
[
 { "B": "ID", "C": "Name", "D": "Date", "E": "Day of Week", "F": "Is Weekday", "G": "Is Weekend",  "H": "Calories Burned", "I":
  "Calories BMR",
  "J": "Steps", "K": "Distance (Km)", "L": "Elevation (Ft)", "M": "Resting Heart Rate", "N": "Floors", "O": "Minutes Sedentary", "P":
    "Minutes Lightly Active",
  "Q": "Minutes Fairly Active", "R": "Minutes Very Active", "S": "Activity Calories", "T": "Active Score", "U": "Cardio minutes", "V"
    : "Cardio calories", "W": "Fat Burn minutes",
  "X": "Fat Burn calories", "Y": "Peak minutes", "Z": "Peak calories", "AA": "Normal Cardio calories", "AB": "Normal Cardio minutes",
    "AC": "Sleep Efficiency", "AD": "Minutes Asleep",
   "AE": "Minutes to fall asleep","AF": "Sleep Start time", "AG": "Sleep End time", "AH": "Time in bed", "AI": "Minutes Deep sleep",
    "AJ": "Deep sleep count", "AK": "Minutes Light sleep",
   "AL": "Light sleep count", "AM": "Minutes REM sleep", "AN": "REM sleep count", "AO": "Minutes Awake", "AP": "Minutes Awake count",
    "AQ": "% Deep sleep", "AR": "% Light sleep","AS": "% REM sleep"
 },
 { "B": "ID9865", "C": "Alice", "D": "2018-06-24", "E": "7", "F": "false", "G": "true", "H": "1996", "I": "1690", "J": "3367", "K": "3
  .01", "L": "0", "M": "59", "N": "0", "O": "1377","P": "47",
   "Q": "2", "R": "14", "S": "362", "T": "-1","U": "11", "V": "133.10892", "W": "14", "X": "89.2088", "Y": "0", "Z": "0", "AA": "446
    .51352", "AB": "284", "AC": "92", "AD": "379", "AE": "0", "AF": "2018-06-24T22:17:00.000", "AG": "2018-06-25T05:28:30.000",
   "AH": "431", "AI": "56", "AJ": "2", "AK": "244", "AL": "26", "AM": "79", "AN": "6", "AO": "52", "AP": "25", "AQ": "13", "AR": "57"
    , "AS": "19"}]
```


Figure 24: Example of quantify-self data packaged into JSON format

Using the data taxonomy cube, the data asset is classified as:

- Data content: ***personal data*** because the data asset identifies an individual
- Data staticity: ***dynamic*** because the data asset is produced on a daily basis
- Data sensitivity: ***sensitive*** according to GDPR because the data asset is composed of biometric data, more precisely, behavioral characteristics of the data subject that enables the unique identification of that person.

The data taxonomy cube is a sufficient decision tool that enables the generation of data storage and exchange agreement template on the basis of the data category. Two agreements have been created for this scenario. The first is a data storage agreement between the data subject and the data custodian. The second is the data exchange agreement between the data subject and the practitioner service.

The Figure 25 shows an example of a data exchange agreement template, which is generated by the data taxonomy cube based on the three dimensions of data category. With our solution, more constraints can be added to the predefined agreement to capture any use case that we miss in our work.

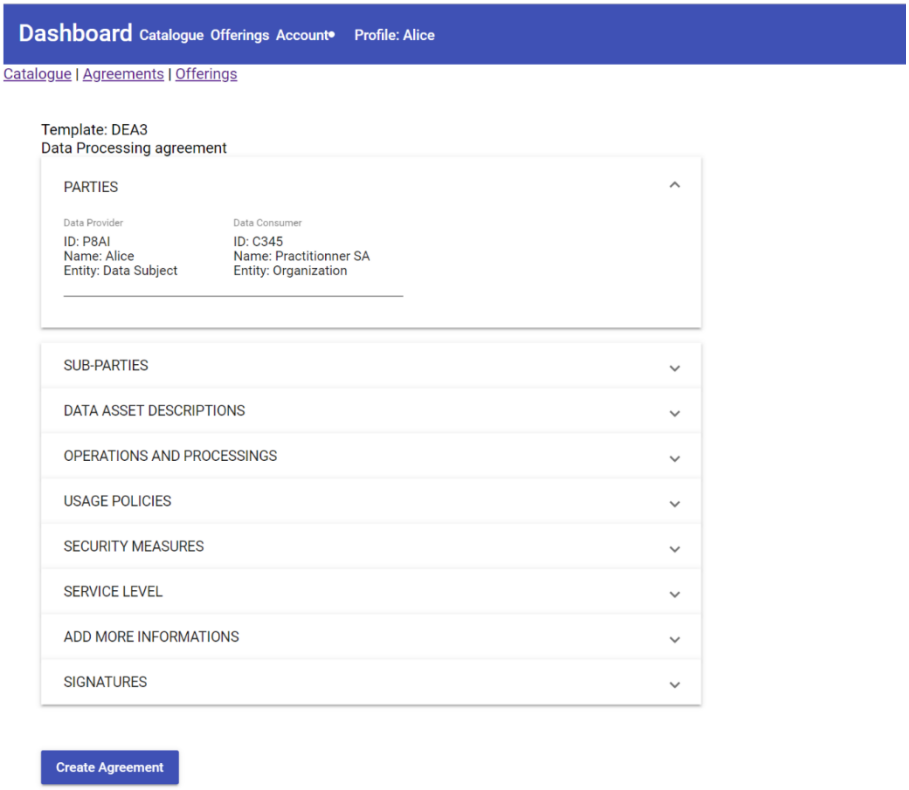

Figure 25: Data exchange agreement template

For data exchange between the data subject and the practitioner, the usage of tokens shows that the practitioner needs to rely on the data custodian for accessing accurate data asset, because of its dynamic character. However, there is no guarantee that after the agreement termination, the practitioner will delete the collected data asset. These issues are discussed in the research of Bhaskaran et al. (2018), where there is no guarantee that the data asset once collected by a data consumer will be deleted after the agreement termination. However in our data market environment, one may assume that the delegation of data storage to data custodians or data providers is a benefit for the data consumers for reducing their storage cost and rather invest into the processing activities, and rely on the marketplace for data access.

Evaluation of Data subjects control over data collection and sharing

From an applicability perspective, the prototype provides to the data subject mechanisms for the GDPR compliance. The data subject has control over her data exchange, and the design concept enforces the following GDPR rights:

- “Right of access”, and “right of data portability”: These rights have been enforced in the prototype by enabling the data subject to collect her data by default from the quantified-self services.

- “Right to be informed” and “Right of restricted processing”: the data subject manages data usage policy through the data exchange agreement. Data exchange required the marketplace to grant permission to data consumer through the delivery of access tokens always by verifying the data processing agreement between the data subject and the practitioner service. Each transaction such as token delivery is hashed and stored in the Chainpoint V3, thus observables by the data subject.
- “Right of rectification”: We did not implement the process of data alteration by the data subject. We have discussed the condition under which it might be implemented in section 8.5, in such a way that enables the achievement of the right of rectification in a responsible approach.
- “Right to be forgotten”: In the prototype, any agreement creation and transaction are hashed and then stored in the chainpoint V3 DLT, hence we mitigate the violation of the “right to be forgotten” as we do not considered hashed data as personal data.

The risk mitigation of GDPR non-compliance
This is partially achieved for the quantify-self service and the practitioner service. By default, these services process personal data only if authorized for specific usage timeline and purpose based on the corresponding agreements. By default, these services data processing activities are transparent for the data subject. Moreover, all the audit trail of agreements and transactions are available for auditors that can compare the audit trail to the data processing activities of these services.

Security considerations
Security of the identity, authentication, and authorization mechanisms, which depends on the security of the cryptographic primitives, are assumed to be secure. Operations (e.g., generate access token, verify access token) are authenticated, authorized and executed by invoking the corresponding agreements and verifying the signatures for ensuring that the processes are executed and not compromised by any parties.

The tokens generated and signed by the marketplace are sent over secured channel during transmission through TLS in such a way that prevent attackers from reading. This mitigates the security issue associated with the token. For data provisioning, we used a one-time access token which adds an additional security level. As the tokens are valid for one use, an attacker cannot reuse the token for another usage in case of tokens leak.

Privacy and data protection considerations
Personal data are disclosed only to the authorized party and others such as the marketplace is only in charge of the transportation of encrypted data.

For data exchange, the data asset are transported over a secure channel. For additional security, one might consider the definition of a proxy-re-encryption scheme (Ateniese et al. 2009) where the quantify-self service encrypted by default the data subject’s data with his public key before sending it to a data custodian. Hence a data custodian will not be able to manipulate the data it received for storage. The data custodian will, therefore, transfers the data via the marketplace

which will play the role of the third-party proxy. The marketplace will send the re-encrypted data to the practitioner who will be able to decode them with his private key.

Services and data exchange auditability
By generating transaction logs to the involved actors, the marketplace enables each actor to have the visibility on the transactions happening in the marketplace. The integration of DLT enables the addition of a trust layer in the data market ecosystem.

Information about marketplace operations and data exchange transactions, including information on who, what, when, why, are hashed and immutably recorded in the chainpoint V3. Consequently, the proposed solution forces the marketplace, the data custodian, the data subjects, and the practitioner to be responsible by default and design for data processing transparency. Therefore, any unauthorized transactions initiated by any of them can be always be retrieved. Furthermore, the investigation for compliance is empowered as all activities logged in the DLT can be traced back. The signaling of a non-compliant activity could trigger official investigation and auditing of a party by an auditor. The decisions could be made based on whether an authorized transaction is recorded in the log ledger or if there is no proof of a transaction that respects the associated agreement in the DLT. In this regard, the DLT can be considered as legal grounds for compliance.

10.1.2 Discussion
This thesis addresses the issues of fairness and responsibility in data market ecosystems. Different requirements have been proposed as the foundation for the elaboration of a framework in this context. The framework provides a set of services that support the collaboration of the ecosystem participants. Moreover, it proposes an agreement management mechanism for supporting these collaborations.

Blockchain system design should preserve the idea of decentralized digital transaction processing, adapting it into a permissioned network, while centralizing some aspects of regulatory compliance and maintenance activity as needed for an enterprise context.

Storage consideration
The marketplace needs to support high transactions volume for agreement creations, validations, verifications, and audit and transaction logs; hence design considerations must be evaluated or weighted to design the storage for the marketplace by deciding for instance: what information is best stored by the marketplace database and what information should be stored in the market participant store. A full analysis of application storage is beyond the scope of this research but some measurements would indicate that disk usage will be growing with a significant overhead because of the need for verification. Thus supporting the idea of data custodians.

Real-time data collection
The data consumers rely on data custodians or directly to the data provider (in the case of an organization) for data access and usage. However, in the case of real-time data asset that needs to be consumed in a relatively short time after their creation, new considerations must be taken for ensured that the data provider is able to deliver such data and guarantee the timely delivery of such data. In that case, a direct link should be created between the service that generates these data and the marketplace pipeline without relying on a custodian service. Therefore, the service will be in charge to guarantee the SLA part of the data processing agreement.

Cost pricing and incentives model
While useful for the framework adoption, pricing and cost models has not been carried out in this prototype. For instance, the cost of data storage and transactions handled by the marketplace must be viable for the market participants in order for them to operate on it. Pricing models can rely on an existing mechanism such as micro-payments, subscription model, etc., depending on the value of service and data asset exchange in the marketplace. We might assume that based on the demand in the ecosystem, the marketplace may help service provider and data provider adjust their model accordingly or propose new models depending on their strategy for consumer acquisition. Incentives models for personal data asset are discussed in the future research directions section of the conclusion as it is a new field to investigate.

10.2 Limitations

The main limitation of this research is based on the fact that there is no application of this framework in a real-world situation which will enable to evaluate the level of data subject empowerment, the compliance of the services operating in the ecosystem and the desirability of such approach for engaging each party in this environment.

Furthermore, we have not addressed the role of data brokers in practice as it is difficult to stimulate in our prototype settings. Such difficulties challenge the real need of this entity in our context. Probably, this role can be endorsed by a third-party service for delivering search and negotiation services. It is however an important part of the design. We did not addressed the compensation model and service in our prototype.

Finally, the performance and scalability of the proposed prototype have not been evaluated at this stage despite the fact that this solution is expected to serve a large number of clients accessing data simultaneously. The measurement of performance and scalability of the prototype should be carried out before deployment.

10.3 Summary

This chapter covered evaluation aspects with respect to the design science methodology. The evaluation is mainly based on the prototyping of the framework proposed in Chapter 7. As results, we demonstrate the feasibility of the

implementation of our general framework and showed that our prototype is a valid implementation. Following, we define the main limitations of this thesis.

The next chapter concludes this dissertation and provides some ideas to further develop data market ecosystems.

Chapitre 11. Conclusion and future works

In this thesis, we used a design science approach for addressing the issues of fair and responsible data market ecosystem. Based on a holistic approach, we considered the main parties and their role in data exchange and service provisioning. Our observations current research directions and practices show that imbalance power and business models in this area create an opaque and unfair business environment. Hence, much attention has been on the empowering data subjects, the weakest party in this environment, with their data by enhancing their rights through the GDPR and a set of initiatives for personal data management.

We argued for an ecosystem approach for responsible and fair data exchange and propose a framework to enhance the collaboration between the main parties. Building on existing knowledge of the domain, the framework describes the parties, the components and their relationship in the context of data market ecosystems.

The results of this research have shown that data market ecosystems have the potential to be further explored and enhanced. Above all, the ability to create a transparent ecosystem for data sharing and to provide value-added services in this multi-stakeholders environment seems is interesting to provide sustainability for data exchange. Furthermore, responsible data sharing can improve the innovation in the community and the participation of the individuals. However, this approach depends on the engagement of all the identified parties and their full compliance with the GDPR.

11.1 Contributions

This thesis is built around the following elements:

- The requirements for designing fair and responsible data market ecosystems. These requirements allow the definition of new concepts by analogy to the existing one. Thus, we define the concept of responsibility by design and by default by the main actors. We also introduce the concept of Informed Trusted Party which is necessary for interfacing all parties in an ecosystem.
- A Taxonomy for data in our context which enables to capture the main characteristics of data that can be exchanged in a data market ecosystem and that are sufficient to elaborate agreement clauses.
- The main contribution of this thesis is the framework which is designed around the predefined design requirements and data taxonomy. The framework suggests the design of the marketplace, which plays the role of Informed Trusted Party in an open way to enable its evaluation by other parties. Furthermore, the trust of this platform is enhanced by the integration of a DLT that capture all the transactions. Furthermore, the framework defines the core design elements for favoring trust and fairness by the ecosystem participants. These design elements were successfully implemented in our prototype which validates our framework.

Qualitative and descriptive analysis demonstrate the utility and feasibility of the framework.

11.2 Future research directions

We outline some possible future research that draws from and build on the research described in this thesis. As data market ecosystems are a very broad domain and still a young research stream this list of applications is of course not exhaustive. It contains some research directions that may be worthy pursue.

Evaluation of cost and benefit for the market participants
An area of research could be the evaluation of cost and the gain with our approach in terms of storage and compliance to GDPR for the market participants, in particular for data consumers and service providers who for the majority rely on data acquisition that create duplicated data and obsolete data store on multiples servers.

New business model including non-pecuniary incentives
It is important to explore new business models in the digital world. The probity of pecuniary payments, compensations, or incentives offered for personal data is highly debated as it constitutes an inalienable part of a person in some regulations such as the European human rights. Also, previous studies have demonstrated the insignificance of amounts paid in data access and usage which fails to actually attract data subjects as well as data consumers. Therefore, the study of the non-pecuniary incentives for personal data processing and the impact of pecuniary and non-pecuniary incentives for attracting data subjects and data consumers to enhance partnership and sustainability in data market ecosystems and how the regulatory body may participate in the regulation of these incentives.

Certification system for data market ecosystems and trust measurement
The establishment of credible systems and services certification in data market ecosystem may enable the design of systems and services operating on it to be measurable and controllable in order to assess the conformity to the general ecosystem requirements. The certification system will rely upon the three acceptable functions such as:

- Standard Setting: the definition of certification requirements will be elaborated in collaboration with the ecosystem participants and coordinated by a standardizing body such as regulators or marketplace auditors. key areas of certification are the transparency, the data protection, privacy, and security.
- Certification: each services operating in an ecosystem will be check for accuracy by a Certification Body against their fulfillment of the certification requirements.
- Accreditation: the competence of the certification body will be assessed by an accreditation body.

The certification standard will help ensure the safety of online transactions and personal information exchanged between services and systems and also facilitates secure and reliable collaboration in this multi-stakeholder environment. Moreover, it is necessary to develop criteria for the measurement and the

attestation of trust and loyalty of systems and services in a data market ecosystem.

Evaluation of performance and scalability of the proposed prototype
Future work in this domain is to carry out the measurement of performance and scalability of the prototype and deploy it based on a real use case.

Bibliography

Abood, T. 2018. “The Role Of Trusted Third Parties” available from: https://accessventures.org/blog/the-role-of-trusted-third-parties/, accessed 2019.

(AEDH). 2017. “Human rights and the propertisation of personal data”, accessed http://www.aedh.eu/en/human-rights-and-the-propertisation-of-personal-data/#_ftn3.

Alturki, A., Gable, G. G., and Bandara, W. 2013. The design science research roadmap: in progress evaluation. PACIS 2013 Proceedings. Retrieved from http://eprints.qut.edu.au/61626/

Androulaki, E. et al., 2018. “Hyperledger Fabric: a distributed operating system for permissioned blockchains”. EuroSys.

Ashok, V.G., Navuluri, K.,Alhafdhi, A., and Mukkamala, R. April 2015.“Datalessdata mining: Association rules-based distributed privacy-preserving datamining,” inInformation Technology - New Generations (ITNG), 201512th International Conference, pp. 615–620.

Ateniese G., Benson K., Hohenberger S. 2009 “Key-Private Proxy Re-encryption.” In: Fischlin M. (eds) Topics in Cryptology – CT-RSA 2009. CT-RSA 2009. Lecture Notes in Computer Science, vol 5473. Springer, Berlin, Heidelberg.

Azure, accessed from https://azure.microsoft.com/en-us/solutions/big-data/, Retrieved 14.02.2019.

Bailey, K.D. 1994. “Typologies and taxonomies: An introduction to classification techniques”, Thousand Oakes:SAGE.

Belleil, A. 2009. “ La régulation économique des données personnelles ? ”, Legicom , n°42, pp. 143-151.

Big Data definition, Gartner, Inc. [Online]. Available: http://www.gartner.com/it-glossary/big-data/ accessed on 2019.

Borgman, C. L., Wallis, J. C., and Enyedy, N. 2007. “Little science confronts the data deluge: habitat ecology,embedded sensor networks, and digital libraries”, Int J Digit Libr 7:17–30 DOI 10.1007/s00799-007-0022-9.

Bradley, J., Sakimura, N., and Jones. M. 2015. “JSON web signature (JWS)”. Technical Report 7515, RFC Editor, Fremont, CA, USA, 7.2.

Bradley, J-C., Lancashire, R. J., Lang, A. S., and Williams, A. J. 2009. “The Spectral Game: leveraging Open Data and crowdsourcing for education”, Journal of Cheminformatics20091:9https://doi.org/10.1186/1758-2946-1-9, accessed on 09.06.2019.

Brickley, D., and Miller, L. Sept 2004.“FOAF Vocabulary Specification”, Namespace Document 2, FOAF Project, http://xmlns.com/foaf/0.1/, accessed on 17.01.2019

Buck, J., 2019. “Constantinople Cancellation Exhibits Ethereum’s Centralized Side,” https://beincrypto.com/constantinople-cancellation-exhibits-ethereums-centralized-side/.

Bundesblock 2018. “Blockchain, data protection, and the GDPR.” Withe paper, VR 36105 B, 27/ 661 /52176.

Bundesdatenschutzgesetz (BDSG). 1970. retrieved 14.02.2019.

Calimaq, 2017. “Evgeny Morozoz et le domain public des données personelles”, https://scinfolex.com/2017/10/29/evgeny-morozov-et-le-domaine-public-des-donnees-personnelles/, accessed on 17.01.2019

California Online Privacy Protection Act (CalOPPA), 2004. available from : https://privacypolicies.com/blog/caloppa/, accessed on 14.02.2019.

California S.B. 1386, 2003, Available from : http://www.leginfo.ca.gov/pub/01-02/bill/sen/sb_1351-1400/sb_1386_bill_20020926_chaptered.pdf, accessed on 14.02.2019.

Campbell, D. 1988. “Somebody's listening”, The New Statesman.

Cao, Q. H., Khan, I., Farahbakhsh, R., Madhusudan, G., Lee., G. M., and Crespi, N. 2016. “A trust model for data sharing in smart cities”, in IEEE International Conference on Communications (ICC).

Carr, N. G. 2003. “IT Doesn't Matter,” Harvard Business Review (81:5), pp. 41-49

Carr, N. G. 2005. “The End of Corporate Computing,” Sloan Management Review. 46.

Chang, E., and West, M. 2006. “Digital Ecosystem - A next generation of the collaborative environment”, presented at iiWAS Yogyakarta.

Chainpoint 2017 https://chainpoint.org/, accessed 2019.

Christopher, R. 2013. “Who owns our data?” ,Computer Law & Security Review, 30.

Cloud Standards Customer Council. 2015. “Practical Guide to Cloud Service Agreements,” available from: https://www.omg.org/cloud/deliverables/CSCC-Practical-Guide-to-Cloud-Service-Agreements.pdf, accessed on 14.02.2019

Cobb, S., and CISSP. 2016. “Data privacy and data protection: US law and legislation”. ESET.

Conseil national du numérique (CNNum). 2015. “Ambition numérique- Pour une politique française et européenne de la transition numérique”, Conseil national du numérique.

Consumer Privacy Protection Act of 2015. 2015. H.R. 2977, 114th Cong.

Cortes, P. 2010. Online Dispute Resolution for Consumers in the European Union Routledge.

Cosy, available from: https://cozy.io/en/, accessed on 14.02.2019.

Council of Europe. 2018. “Convention 108+ Convention pour la protection des personnes à l’égard du traitement des données à caractère personnel” accessed from: https://rm.coe.int/convention-108-convention-for-the-protection-of-individuals-with-regar/16808b36f1, Retrieved 14.02.2019.

Crabtree, A., Lodge, T., Colley, J., Greenhalgh, C., Mortier, R., and Haddadi, H. 2016. “Enabling the new economic actor: data protection, the digital economy,

and the Databox", Personal and Ubiquitous Computing, vol. 20, no. 6, pp. 947–957.

Crain, M. 2016. "The limit of Transaparency", new media & society, Vol. 20(I), pp. 88-104, DOI:10.1177/1461444816657096 SAGE.

Cryptonomist. 2019. "Telefónica: a blockchain platform for selling personal data," accessed from: https://cryptonomist.ch/en/2019/02/22/telefonica-blockchain-personal-data/, Retrieved 14.02.2019.

Cutrell, E., Robbins, D., Dumais, S., and Sarin, R., 2006. "Fast, flexible filtering with phlat" in Proceedings of the SIGCHI conference on Human Factors in computing systems, 261-270.

Dan, A., et al. 2004 "Web services on demand: WSLA-driven automated management", IBM Syst. J. 43(1), 136–158.

Databroker DAO. 2018. https://databrokerdao.com/, accessed 2019.

DataMarket, https://www.datamarket.com/, accessed on 09.01.2019.

DataPortability Project, accessed from: http://dataportability.org/, Retrieved 14.02.2019.

Data-XC. 2019. "Data exchange consortium" available at: http://www.data-xc.jp/. Accessed on 09.01.2019

DataWallet. 2014. https://datawallet.com/home# , accessed 2019.

De Groot, J. 2019. "The History of Data Breaches" accessed from: https://digitalguardian.com/blog/history-data-breaches, Retrieved 14.02.2019.

De Hert, P., Papakonstantinou, V., Malgieri, B., Sanchez, I. 2018. "The right to data portability in the GDPR: Towards user-centric interoperability of digital services", Computer Law & Security Review, Vol. 34, Issue 2, pp. 193-203.

Deloitte, 2014. "New business models with data," accessed from: https://ec.europa.eu/futurium/sites/futurium/files/deloitte_pov_-_new_business_models_with_data.pdf , Retrieved 14.02.2019.

Doc Searl 2018" We can do better than selling our data". doi: 10.1109/APSCC.2011.59

Dong, X., Guo, B., Duan, X., Shen, Y., Zhang, H., and Shen,Y. 2016 "DSPM: A Platform for Personal Data Share and Privacy Protect Based on Metadata", 13th International Conference on Embedded Software and Systems (ICESS) DOI: 10.1109/ICESS.2016.10; ISBN: 978-1-5090-3727-8.

Draskovic, D., and Saleh, G., 2017 "Datapace Decentralized data marketplace based on blockchain" Withe paper.

Dumais, S., et al., "Stuff I've seen: a system for personal information retrieval and re-use," in SIGIR '03 Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, 72-79, 2003.

Dumitru, R., and Gatti, S. 2016. "Towards a reference architecture for trusted data marketplaces: the credit scoring perspective", In Irfan Awan and Muhammad Younas, editors, 2nd International Conference on Open and Big Data, OBD 2016, Vienna, Austria, pp. 95-101, IEEE Computer Society, 10.1109/OBD.2016.21.

Egea, M., Matteucci, I., Mori, P., and Petrocchi, M. “Definition of Data Sharing Agreements: The Case of Spanish Data Protection Law,” A4Cloud 2014, LNCS 8937, pp. 248–272, 2015.

Egorov, M., and Wilkison, M. 2016 .“ZeroDB white paper”, In: CoRR abs/1602.07168, URL: http://arxiv.org/abs/1602.07168.

Elsweiler, D., Ruthven, I., and Jones, C. 2005. “Dealing with fragmented recollection of context in information management” in Context-Based Information Retrieval (CIR-05) Workshop in Fifth International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT-05),

Elsweiler, D., Ruthven, I., and Ma, L. 2006. “Considering human memory in pim” in SIGIR 2006 Workshop on Personal Information Management, 10-11.

European Parliament (europarl). 2010. “The Charter of Fundamental Rights of the European Union”, accessed from http://www.europarl.europa.eu/charter/pdf/text_en.pdf, Retrieved 26.06.2019.

European Policy Center (EPC). 2010. “The economic impact of a European digital single market”, European Policy Centre, Copenhagen Economics.

Evidon. 2016. accessed from: https://www.evidon.com/solutions/gdpr/, Retrieved 14.02.2019.

Express gateway 2018, accessed from: https://www.express-gateway.io/docs/, 2019.

Federal Trade Commission (FTC). 2014. “Data brokers: a call for transparency and accountability”, Available at: https://www.ftc.gov/system/files/documents/reports/data-brokers-calltransparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf, accessed on 25 August 2015.

Federal Trade Commission (FTC). 2010. “Fair Information Practice Principles (FIPs), 2. Choice/Consent.” Archived March 9, 2010, at the Wayback Machine.

Freedombox, available from: https://www.freedomboxfoundation.org/, accessed on 14.02.2019.

Gandy, O. H. 1989. “The Surveillance Society: Information Technology and Bureaucratic Social Control”, Journal of Communication, Vol. 39, Issue 3, pp. 61–76, https://doi.org/10.1111/j.1460-2466.1989.tb01040.x

GDPR. 2018. “Rights of data subject”, https://www.gdpr-info.eu/chapter-3/, accessed on 09.01.2019.

GOV.UK. 2012. “Open data white paper: unleashing the potential”, the Stationary Office, ISBN: 9780101835329, retrieve from http://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf.

Goodman, L.M. 2014. “Tezos: A Self-Amending Crypto-LedgerPosition Paper”.

Graeff, H. 2002. “Collecting and using personal data: consumers’ awareness and concerns”, Journal of Consumer Marketing, Vol. 19 Issue: 4, pp.302 – 318.

Greenleaf, Graham, (2017a) 'Renewing Convention 108: The Coe's 'GDPR Lite' Initiatives,' Privacy Laws & Business International Report142: (2017), 14-7.

Haenni, R. 2017. “datum network: the decentralized data marketplace.” Withe paper V14.

Heipke, C. 2010. “Crowdsourcing geospatial data”, ISPRS Journal of Photogrammetry and Remote Sensing, Vol. 65, Issue 6, pp. 550-557, ISSN 0924-2716.

Hevner, A.R., et al., 2004, “ Design Science in Information Systems Research”, MIS Quaterly:p.75-105.

Hub-of-All-Things (HAT), Available from: https://static1.squarespace.com/static/5b5988f9b105985261c0a722/t/5c179df3f950b7d4a378c11b/1545051654456/hatpresentation1.7PDF.pdf , accessed on 14.02.2019.

Hutton, L., Price, B. A., Kelly, R., McCormick, C., Bandara, A.K., Hatzakis, T., and Nuseibeh, B. 2018. “Assessing the Privacy of mHealth Apps for Self-Tracking: Heuristic Evaluation Approach”, JMIR mHealth and uHealth.

Iapp, 2019. “A brief history of the general data protection regulation”, iapp.org, https://iapp.org/resources/article/a-brief-history-of-the-general-data-protection-regulation/. Accessed on 17.01.2019

IMB. 2009. “The strategic importance of OLAP and multidimensional analysis”, White paper, retrieve from http://www.inteligencija.com/download/cognos/wp_the_strategic_importance_of_olap_and_multidimensional_analyis.pdf.

Immonen, A., Palviainen, M., And Ovaska, E. 2014. “Requirements of an Open Data Based Business Ecosystem.”

IOTA data marketplace. 2015. https://data.iota.org/#/ , accessed 2019.

Iyilade, J., and Vassileva, J. 2013. “A framework for privacy-aware user data trading”, In Proceeding of User Modeling, Adaptation, and Personalization (UMAP), pp. 310–317. Springer.

James, B. 2013. "NSA's Prism Surveillance Program: How It Works and What It Can Do – Slide from Secret PowerPoint Presentation Describes How Program Collects Data 'Directly from the Servers' of Tech Firms – Obama Deflects Criticism over NSA Surveillance". The Guardian. Retrieved 14.04. 2019.

Jones, M.L., Kaufman, E., Edenberg, E. 2018. “AI and the ethics of automating consent”, IEEE Secur. Priv. 16(3).

Jones, W., and Teevan, J. 2007. “Personal Information Management”. University of Washington Press, Seattle.

Kaggle, https://www.kaggle.com/, accessed on 09.06.2019.

Krutz, R. L., and Vines, R.D. 2010. “Cloud security: a comprehensive guide to secure cloud computing”, Indianapolis: Wiley

Kunewa, Meglena, 2009. “Keynote Speech”, retrieve from http://europa.eu/rapid/press-release_SPEECH-09-156_en.htm.

La Fing. 2012. “MesInfos: Cahier d’Exploration”, available from: http://fing.org/?Cahier-d-exploration-MesInfos, last accessed on accessed on 14.02.2019.

Landreau, I., Peliks, G., Binctin, N., and Pez-Perard, V. 2018, “Mes data sont à moi”, Génération Libre.

Laudon, K., and Laudon, J. P. 2014. “Managing the digital firm”, Chapter 4, Ethical or Social Issues in Information Systems, pp. 418 – 420.

Liang, X., Zhao, J., Shetty, S., Liu J., and Li, D. 2017. “Integrating blockchain for data sharing and collaboration in mobile healthcare applications” IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, 2017, pp. 1-5. doi: 10.1109/PIMRC.2017.8292361

Liu, F., Tong, J., Mao, J., Bohn, R. B., Messina, J. V., Badger, M. L., and Leaf, D. M. 2011. “NIST cloud computing reference architecture: version 1”, National Institute of Standards and Technology, Gaithersburg.

Locher, T., Obermeier, S., and Pignolet, Y. A. 2018. “When Can a Distributed Ledger Replace a Trusted Third Party?”

Ludwig, H., Keller, A., Dan, A., King, R. P., and Franck, R. 2003. “Web service level agreement (wsla) language specification,” IBM Corporation, pp. 815–824 .

Lyon, D. 2014. “Surveillance, Swoden, and Big Data: Capacities, consequences, critique”, Big Data & Society, 1(2), pp. 1-13.doi: 10.1177/2053951714541861

MacAskill, E., Borger, J., Hopkins, N., Davies, N., Ball, J. 2013. “GCHQ Taps Fibre-Optic Cables for Secret Access to World's Communications – Exclusive: British Spy Agency Collects and Stores Vast Quantities of Global Email Messages, Facebook Posts, Internet Histories and Calls, and Shares Them with NSA, Latest Documents from Edward Snowden Reveal”. The Guardian. Retrieved 14.02.2019.

Macnish, K. 2014. “Ethics of Big Data and Social Media - Written Submission to Parliamentary Committee on Science and Technology”, academia.edu, SMD0001.

Mahony, R. 2016. “Telefonica seeks to broker new relationships with consumers and Internet players”, https://ovum.informa.com/resources/product-content/telefonica-seeks-to-broker-new-relationships-with-consumers-and-internet-players accessed on 18.01.2019

Malik, M., Ghazi, M., and Ali, R., 2012 “Privacy preserving data miningtechniques: Current scenario and future prospects,” inComputer andCommunication Technology (ICCCT), 2012 Third International Confer-ence , 2012, pp. 26–32

Manson, R. O. 1986. “Four Ethical Issues of the Information Age”, MIS Quarterly (10:1), pp. 4-12.

Martinelli, F., Saracino, A., and Sheikhalishahi, M. 2016. “Modeling privacyaware information sharing systems: A formal and general approach,” in15th IEEE International Conference on Trust, Security and Privacy inComputing and Communications,

McCarthy, J. 1992.“Reminiscences on the history of time sharing.”IEEE Annals of the History of Computing,Vol.14, No.1, pp.19–24.

Midata, 2014. “Personal data :Review of the midata voluntary Program”, Department for Business, Innovation and Skills, BIS/14/941.

March, S. T., and Smith, G. F. 1995. “Design and natural science research on information technology” ,Decision Support Systems, 15(4):251-266.

MedRec: Using Blockchain for Medical Data Access and Permission Management

Meeco. 2018. “Zero Knowledge Proofs of the modern digital life for access, control, delegation and consent of identity and personal data.” Meeco Planet Pty Ltd Technical Whitepaper Version 1.0, available from: https://meeco.me/whitepaper.html. Retrieved 14.02.2019.

Mitchell, A., Henderson, I., and Searls, D. 2008. “Reinventing direct marketing — with vrm inside”, Journal of Direct Data and Digital Marketing Practice, 10(1):3–15.

Mitchell, J. 2017. “Unlocking the value of our data: The individual as the point of integration”, Mydex Data Services CIC, available from: https://mydex.org/sites/mydex.org/files/assets/unlocking_the_value_of_our_data_-_mydex_cic.pdf, accessed on 14.02.2019

Mont, M.C., Pearson, S., and Bramhall, P. 2003. “Towards accountable management of identity and privacy: sticky policies and enforceable tracing services”, In Database and Expert Systems Applications, . Proceedings. 14th International Workshop , pp. 377–382.

Mortier, R., Haddadi, H., Henderson, T., McAuley, D., and Crowcroft, J. 2014. “Human-data interaction: The human face of the data-driven society”, Social Science Research Network, doi:10.2139/ssrn.2508051.

My Cloud, available from: https://www.mycloud.com/#/, accessed on 14.02.2019.

Narayanan, A., Toubiana, V., Barocas, S., Nissenbaum, H., and Boneh, D. 2012.“A Critical Look at Decentralized Personal Data Architectures” , arXiv preprint arXiv: 1202.4503.

National Research Council (NRC). 1995. “On the Full and Open Exchange of Scientific Data”, Washington, DC: The National Academies Press. https://doi.org/10.17226/18769.

Nextcloud, available from: https://nextcloud.com/, accessed on 14.02.2019.

Ni, Q., Bertino, E., Lobo, J., Brodie, C., Karat, C.-M., Karat, J., Trombetta, A. 2010. “Privacy-aware role-based access control”. ACM Trans. Inform. Syst. Secur. 13(3), 24:1–24:31

Nickerson, R. C., Muntermann, J., Varshney, U., and Isaac, H. 2009. “Taxonomy development in information systems: developing a taxonomy of mobile applications”, In Proceedings of the European Conference on Information Systems.

Noorian, Z., Iyilade, J., Mohkami, M., and Vassileva, J. 2014. “Trust mechanism for enforcing compliance to secondary data use contracts”, In: IEEE 13th International Conference Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 519–526.

Nwatchock A Koul A, S., Morin, J-H. 2016. “Towards a Taxonomy of Data and Guiding Principles for Data Markets”, Proceedings of the 11th Pre-ICIS Workshop on Information Security and Privacy, Dublin, Ireland.

Nwatchock A Koul A, S. Morin, J-H. December 09 2017. “Towards a framework for a fair and sustainable data market ecosystem”, Proceedings of the 12th Pre-ICIS Workshop on Information Security and Privacy, Seoul, South Korea.

Nwatchock A Koul A, S. Morin, J-H. December 09 2018. “Agreements framework for data market ecosystem”, Proceedings of the 13th Pre-ICIS Workshop on Information Security and Privacy, San Francisco, December 13, 2018.

OECD. 2013. “Exploring the Economics of Personal Data: A Survey of Methodologies for Measuring Monetary Value,” OECD Digital Economy Papers, No. 220, OECD Publishing .DOI: 10.1787/5k486qtxldmq-en3

Onetrust. 2019. accessed from: https://www.onetrust.com/products/gdpr-compliance/, Retrieved 14.02.2019.

OpenID Authentication 2.0, OpenID Foundation. 2007. http://openid.net/specs/openid-authentication-2_0.html.

Peres, E. (Conseil économique, social et environnementa) 2015. “Les données numériques : un enjeu d’éducation et de citoyenneté”, Economic social and environmental concil, accessed from: http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/154000045.pdf

Purtova, N. 2017. “Do property rights in personal data make sense after the Big Data turn? Individual control and transparency”, 10(2) Journal of Law and Economic Regulation.

Rehnquist, W.H. 1969. “FOIA Policy”, Justice Department memorandum to all federal agencies.

Rustom Al Nasar, M., & Mohd, M., and Nazlena, M. A. 2011. “Personal information management systems and interfaces: An overview”, International Conference on Semantic Technology and Information Retrieval, STAIR 2011. 197 - 202. 10.1109/STAIR.2011.5995788.

Schomm, S., and Vossen, G. 2013. “Marketplaces for data: an initial survey”. ACM SIGMOD Record, 42(1): pp.15–26.

Shrier, A. A., Chang, A., Diakun-thibault, N., Forni, L., Landa, F., Mayo, J., van Riezen, R., and Hardjono, T. 2016. “Blockchain and Health IT: Algorithms, Privacy, and Data”, White paper.

Simonite, T. 2014. “Sell your personal date for $8 month”, MIT Technology Review, accessed from: https://www.technologyreview.com/s/524621/sell-your-personal-data-for-8-a-month/, retrieved 18. 01 2019.

Spiekermann, S., Novotny, A. 2015. “A vision for global privacy bridges: Technical and legal measures for international data markets”, Computer Law & Security Review, Vol. 31, Issue 2, pp. 181–200.

Stahl, F., Schomm, F., and Vossen, G. 2014. “The Data Marketplace Survey Revisited”, Working Paper No. 18, ERCIS, ISSN 1614-7448.

Steem, 2018. “Steem An incentivized, blockchain-based, public content platform.” accessed from: https://www.steem.com/steem-whitepaper.pdf, Retrieved 14.02.2019.

Streamr. 2017. “Unstoppable Data for Unstoppable Apps: DATAcoin by Streamr”, https://marketplace.streamr.com/, retrieved 14.02.2019.

Szabo, N. 1997. “Formalizing and Securing Relationships on Public Networks.” Accessed from https://firstmonday.org/ojs/index.php/fm/article/view/548/469.DOI:http:/dx.doi.org/10.5210/fm.v2i9.548, retrieved 14.02.2019.

Teperek, M. 2016. “Why data should be the new oil and how to make data the new water”, talk, Swiss research data management day 2016, https://www.dlcm.ch/swiss-research-data-management-day-2016/#1480580687673-261443ef-c3ed

The economist. 2015. “The business of data”, https://eiuperspectives.economist.com/sites/default/files/images/Business%20of%20Data%20briefing%20paper%20WEB.pdf.

Things, http://culturedcode.com/things/, accessed on 09.06.2015.

Thuraisingham, B. 2015. “Database security: Past, present, and future,” inBigData (BigData Congress), 2015 IEEE International Congress , pp. 772–774.

Transparency Market Research (TMR). 2018. “Global Data Broker Market Size, Status and Forecast 2018-2025”.

Truong, H. L., and Dustdar, S. 2009. “On analyzing and specifying concerns for data as a service”, in APSCC, M. Kirchberg, P. C. K. Hung, B. Carminati, C.-H. Chi, R. Kanagasabai, E. D. Valle, K.-C. Lan, and L.-J. Chen, Eds. IEEE, pp. 87-94.

Truong, H. L., Comerio, M., De Paoli, F., Gangadharan, G. R., and Dustdar, G. R. 2012. “Data contracts for cloud-based data marketplaces.” IJCSE 7, pp. 280-295.

Truong, H., Dustdar, S., Gotze, J., Fleuren, T., Muller, P., Tbhariti, S-E., Mrissa, M., and Ghedira, C. 2011. “Exchanging Data Agreements in the DaaS Model”, IEEE Asia-Pacific Services Computing Conference, Jeju Island, , pp. 153-160.

Trustarc. 2019. accessed from: https://www.trustarc.com/products/gdpr-compliance/, Retrieved 14.02.2019.

Turow, J., Hennessy, M., and Draper, N. June 2015. “The tradeoff fallacy: how marketers are misrepresenting American consumers and opening them up to exploitation”, Report, Annenberg School for Communication, University of Pennsylvania, Pennsylvania PA,

Uber. 2019. “Open data Portal” available from: https://movement.uber.com/?lang=en-US, accessed on fevrier 2019.

UK Government. January 2016. “Distributed Ledger Technology: beyond block chain”, Office for Science. Retrieved 29 August 2016.

Vaishnavi, S. V., and Kuechler, W. 2007. “Design Science Research Methods and Patterns: Innovating Information and Communi cation Technology”, New York: Auerbach Publications, Taylor & Francis Group.

Vu, Q. H., Pham, T.-V., Truong, H.-L., Dustdar, S., and Asal, R. 2012. “DEMODS: a description model for data-as-a-service”, In AINA, pp. 605– 612.

Vu, Q.H., Pham, T.V., Truong, H.L., Dustdar, S., and Asal, R. 2012. “DEMODS: a description model for data-as-a-service”, In: IEEE 26th International Conference on Advanced Information Networking and Applications (AINA), pp. 605–612. Fukuoka, Japan. doi:10.1109/AINA.2012.91.

Xiginite, accessed from: https://www.xignite.com/, Retrieved 14.02.2019.

Xu, X., Pautasso, C., and Zhu, L. 2016. “The Blockchain as a Software Connector”, 13th Working IEEE/IFIP Conference on Software Architecture.

Zhan, Z. J. 2006. “Privacy-preserving collaborative data mining,” Ottawa, Ont.,Canada, aAINR18614.[17].

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S.P., et al. 2017. “Ten simple rules for responsible big data research”. PLoS Comput Biol 13(3): e1005399. https://doi.org/10.1371/journal.pcbi.1005399

Zyskind, G., O. Nathan, and A. Pentland, “Enigma: Decentralized Computation Platform with Guaranteed Privacy”, Tech. Rep. (2015).

Zyskind, G., Nathan, O., and Pentland, A. May 21-22, 2015. “Decentralizing Privacy: Using Blockchain to Protect Personal Data”, in 2015 IEEE Symposium on Security and Privacy Workshops, SPW 2015, San Jose, CA, USA, pp. 180–184.

Sören, Ö. "Implementing Data Protection in Law" (PDF). Accessed from http://www.scandinavianlaw.se/pdf/47-18.pdf, Retrieved 10 May 2017.

Appendix

Appendix A: Glossary

API. An application programming interface is a set of subroutine definitions, communication protocols, and tools for building software.

Advanced Message Queuing Protocol (AMQP). An open source published standard for asynchronous messaging by wire.

Blockchain. The first fully functional Distributed Ledger Technology.

Data element. A unit of data for which the definition, identification, representation, and permissible values are specified by means of a set of attributes.

DLT. Distributed Ledger Technology describes technologies which store, distribute and facilitate the exchange of value between users, either privately or publicly.

GRAKN keyspace. A keyspace is the outermost container for data in a Grakn knowledge graph, corresponding closely to a relational database.

IOT. A system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction.

Knowledge Graph: A model of a knowledge domain with the help of machine learning algorithms. It provides a structure and common interface for data and enables the creation of smart multilateral relations between data.

Message broker. An intermediary program that translates messages from the formal messaging protocol of the publisher to the formal messaging protocol of the receiver.

Merkle tree. A hash-based data structure that is a generalization of the hash list. It is a tree structure in which each leaf node is a hash of a block of data, and each non-leaf node is a hash of its children.

Microservice architecture. The term describes the design of software applications as suites of independently deployable services.

Appendix B: Prototype code

The prototype implementation code is available in the repository: https://github.com/sabrina-ossey/MarketFramework and the execution process is detailed in the README.md file of the repository.

- Technology Stack

We are using a NodeJS service with a MongoDB for our backend.

- NodeJS 8.10
- MongoDB 3.4
- Docker for Ubuntu
- Angular 6
- RabbitMQ 3.7.7
- Chainpoint V3
- Grakn 1.4.3

The implemented microservices are:

Third-party services:

- quantifySelfService,
- custodianService,
- practitionnerService.

Marketplace services:

- CatalogService: data Asset and Service catalogue,
- HashingService: service for agreement and transactions hashing,
- AgreementManagerService: Service for agreement creation and management,
- MonitoringService: Service for handling monitored elements and submit into the Chainpoint Calendar Blockchain,
- DatabusService: service for handling data transfer,
- APIGatewayService: Implement a service which is the entry point into others microservices

- Marketplace User Interface: MarketUI

Appendix C: Data processing agreement Template from: https://gdpr.eu/data-processing-agreement/

This Data Processing Agreement ("Agreement") forms part of the Contract for Services ("Principal Agreement") between

(the “Company”) and

(the “Data Processor”) (together as the “Parties”) WHEREAS

(A) The Company acts as a Data Controller.

(B) The Company wishes to subcontract certain Services, which imply the processing of personal data, to the Data Processor.

(C) The Parties seek to implement a data processing agreement that complies with the requirements of the current legal framework in relation to data processing and with the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

(D) The Parties wish to lay down their rights and obligations.

IT IS AGREED AS FOLLOWS:

Definitions and Interpretation

Unless otherwise defined herein, capitalized terms and expressions used in this Agreement shall have the following meaning:

"Agreement" means this Data Processing Agreement and all Schedules;

"Company Personal Data" means any Personal Data Processed by a Contracted Processor on behalf of Company pursuant to or in connection with the Principal Agreement;

"Contracted Processor" means a Subprocessor;

"Data Protection Laws" means EU Data Protection Laws and, to the extent applicable, the data protection or privacy laws of any other country;

"EEA" means the European Economic Area;

"EU Data Protection Laws" means EU Directive 95/46/EC, as transposed into domestic legislation of each Member State and as amended, replaced or superseded from time to time, including by the GDPR and laws implementing or supplementing the GDPR;

"GDPR" means EU General Data Protection Regulation 2016/679;

"Data Transfer" means:

a transfer of Company Personal Data from the Company to a Contracted Processor; or

an onward transfer of Company Personal Data from a Contracted Processor to a Subcontracted Processor, or between two establishments of a Contracted Processor,

in each case, where such transfer would be prohibited by Data Protection Laws (or by the terms of data transfer agreements put in place to address the data transfer restrictions of Data Protection Laws);

"Services" means the Company provides.

services the

"Subprocessor" means any person appointed by or on behalf of Processor to process Personal Data on behalf of the Company in connection with the Agreement.

The terms, "Commission", "Controller", "Data Subject", "Member State", "Personal Data", "Personal Data Breach", "Processing" and "Supervisory Authority" shall have the same meaning as in the GDPR, and their cognate terms shall be construed accordingly.

Processing of Company Personal Data

Processor shall:

comply with all applicable Data Protection Laws in the Processing of Company Personal Data; and

not Process Company Personal Data other than on the relevant Company's documented instructions.

1.1 The Company instructs Processor to process Company Personal Data.

Processor Personnel

Processor shall take reasonable steps to ensure the reliability of any employee, agent or contractor of any Contracted Processor who may have access to the Company Personal Data, ensuring in each case that access is strictly limited to those individuals who need to know / access the relevant Company Personal Data, as strictly necessary for the purposes of the Principal Agreement, and to comply with Applicable Laws in the context of that individual's duties to the Contracted Processor, ensuring that all such individuals are subject to confidentiality undertakings or professional or statutory obligations of confidentiality.

Security

Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of Processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, Processor shall in relation to the Company Personal Data implement appropriate technical and organizational measures to ensure a level of security appropriate to that risk, including, as appropriate, the measures referred to in Article 32(1) of the GDPR.

In assessing the appropriate level of security, Processor shall take account in particular of the risks that are presented by Processing, in particular from a Personal Data Breach.

Subprocessing

Processor shall not appoint (or disclose any Company Personal Data to) any Subprocessor unless required or authorized by the Company.

Data Subject Rights

Taking into account the nature of the Processing, Processor shall assist the Company by implementing appropriate technical and organisational measures, insofar as this is possible, for the fulfilment of the Company obligations, as reasonably understood by Company, to respond to requests to exercise Data Subject rights under the Data Protection Laws.

Processor shall:

promptly notify Company if it receives a request from a Data Subject under any Data Protection Law in respect of Company Personal Data; and

ensure that it does not respond to that request except on the documented instructions of Company or as required by Applicable Laws to which the Processor is subject, in which case Processor shall to the extent permitted by Applicable Laws

inform Company of that legal requirement before the Contracted Processor responds to the request.

Personal Data Breach

Processor shall notify Company without undue delay upon Processor becoming aware of a Personal Data Breach affecting Company Personal Data, providing Company with sufficient information to allow the Company to meet any obligations to report or inform Data Subjects of the Personal Data Breach under the Data Protection Laws.

Processor shall co-operate with the Company and take reasonable commercial steps as are directed by Company to assist in the investigation, mitigation and remediation of each such Personal Data Breach.

Data Protection Impact Assessment and Prior Consultation

Processor shall provide reasonable assistance to the Company with any data protection impact assessments, and prior consultations with Supervising Authorities or other competent data privacy authorities, which Company reasonably considers to be required by article 35 or 36 of the GDPR or equivalent provisions of any other Data Protection Law, in each case solely in relation to Processing of Company Personal Data by, and taking into account the nature of the Processing and information available to, the Contracted Processors.

Deletion or return of Company Personal Data

Subject to this section 9 Processor shall promptly and in any event within

10 business days of the date of cessation of any Services involving the Processing of Company Personal Data (the "Cessation Date"), delete and procure the deletion of all copies of those Company Personal Data.

Processor shall provide written certification to Company that it has fully complied with this section 9 within 10 business days of the Cessation Date.

Audit rights

Subject to this section 10, Processor shall make available to the Company on request all information necessary to demonstrate compliance with this Agreement, and shall allow for and contribute to audits, including inspections, by the Company or an auditor mandated by the Company in relation to the Processing of the Company Personal Data by the Contracted Processors.

Information and audit rights of the Company only arise under section 10.1 to the extent that the Agreement does not otherwise give them information and audit rights meeting the relevant requirements of Data Protection Law.

Data Transfer

The Processor may not transfer or authorize the transfer of Data to countries outside the EU and/or the European Economic Area (EEA) without the prior written consent of the Company. If personal data processed under this Agreement is transferred from a country within the European Economic Area to a country outside the European Economic Area, the Parties shall ensure that the personal data are adequately protected. To achieve this, the Parties shall, unless agreed

otherwise, rely on EU approved standard contractual clauses for the transfer of personal data.

General Terms

Confidentiality. Each Party must keep this Agreement and information it receives about the other Party and its business in connection with this Agreement ("Confidential Information") confidential and must not use or disclose that Confidential Information without the prior written consent of the other Party except to the extent that:

disclosure is required by law;

the relevant information is already in the public domain.

Notices. All notices and communications given under this Agreement must be in writing and will be delivered personally, sent by post or sent by email to the address or email address set out in the heading of this Agreement at such other address as notified from time to time by the Parties changing address.

Governing Law and Jurisdiction

This Agreement is governed by the laws of .

Any dispute arising in connection with this Agreement, which the Parties will not be able to resolve amicably, will be submitted to the exclusive jurisdiction of the courts of , subject to possible appeal to .

IN WITNESS WHEREOF, this Agreement is entered into with effect from the date first set out below.

Your Company

Signature Name: Title: Date Signed:

Processor Company

Signature	Name	Title	Date Signed

Appendix D: Fitbit Data

Accessed from https://github.com/yashatgit/fitbit-analyzer/tree/master/data retrieved 11.10.2018