UNIVERSITÉ PARIS II
PANTHÉON-ASSAS

Data ecosystems: an exploration of actors and the business models they practice

Towards a reference topology of revenue and pricing models in data ecosystems

Mackenzie Himmelbauer

29 August, 2021

Université Paris II- Master 2
Management Stratégique et Entrepreneuriat 2020-2021
Master's thesis
Visions SAS

Table of Contents

Acknowledgements

I would like to sincerely thank all of the participants that dedicated time out of their busy schedules to share their viewpoints and professional knowledge on data ecosystems with me. It has been very exciting to learn from you about different industries and ways in which data ecosystems evolve.

A special thanks to Visions for dedicating time and resources to this project and making sure that I had all of the information and encouragement needed to make this project a reality. I'm looking forward to transforming this research into something of value for data ecosystems on a more practical level. I'm looking forward to seeing how these discoveries will evolve.

Abstract

Data Ecosystems are a new phenomenon in which data is exchanged between a variety of actors, both public and private and for a variety of purposes. With each type of data exchange there are also value exchanges that occur between actors in exchange for this data. As there are a wide variety of actors in data ecosystems, the ways in which these exchanges occur vary greatly. There is a great deal of research on business models in general and a small number of publications on business models in data marketplaces. This master's thesis aims to examine business models (revenue models and pricing models) that occur in data ecosystems in a more holistic way.

A multi-method strategy included a qualitative research approach that resulted in seven semi-directive interviews, analysis of more than ten academic publications and case studies and examination of thirty-plus online publications. Inductive analysis and industry experience contributed by participants allowed the discovery of additional revenue and pricing models not yet exhibited in academic publications.

This research characterizes business models in data ecosystems as continuously-evolving elements. Data ecosystems must first be presented and categorized, containing the primary actors and typologies. The business models of these data ecosystems experience continuous development, which is why they have been examined in supporting sections. These sections include revenue models, pricing models and non-monetary transactions that may occur in ecosystems.

The research strategy, proposed archetypes and models can guide subsequent research on the topic.

Keywords: Data Ecosystem, data exchanges, business models in data exchanges, revenue models, pricing models, data ecosystem economy.

What are data ecosystems and the business models practiced in them?

I. Introduction :

Data is clearly the most valuable resource in the world. With the evanescence of smartphones, platforms and other technologies that track our every movement, data sources have multiplied by the thousands and thus so has the amount of data that has been created. In the past year we have generated more personal data than has ever been generated in history. This translates into a plethora of opportunities, new technologies and services, revolutionary technologies and incredible advances for humanity.

Each year we use and produce more applications, devices and products that consume and process our personal data. With the 2020 global COVID-19 Pandemic, those numbers skyrocketed as the average citizen was confined to their home and spending a great deal more time on applications for both work and entertainment. This means that more and more organizations are sharing data, between sectors and between industries. This is giving birth to a new model of data sharing known as "data ecosystems". Within these data ecosystems, information is sold, exchanged, transformed and analyzed. A variety of actors contribute their products, services and data in these ecosystems. With the evolution of these ecosystems, business and pricing models for such data exchanges are also evolving.

In order to better understand these ecosystems and the value propositions that can be built from them, the goal of this paper is to respond to the following question: What kinds of business models can be used in data ecosystems ?

In order to attempt to respond to this question, 3 sections have been curated thanks to 7 semi-directive interviews, and a multitude of academic and non-academic sources.

In the first section, we will explore the data ecosystems themselves in order to better understand the value propositions made within them. We will explore the definition of an ecosystem, the ways to categorize ecosystems, the four primary archetypes of actors within them, and the forms that they can take. There will also be a specific section dedicated to Visions, a French company that works to build and create value out of these ecosystems revolving around skills data, that I currently work for.

In a second section, we will address the different types of value exchanges possible within these ecosystems. In this study, we have defined the key components of a business model: its revenue model and its pricing model. As such, this section is divided into one section for each subject. As business models are often custom-built according to the business and the clients they serve, this paper has dissected business models and has presented revenue models and pricing models in separate sections. By doing this, I aim to provide a wide variety of examples and content so that those working in the industry can mix-and-match revenue and pricing models in order to build custom business models based on their specific businesses. In the first subsection we will examine the revenue models presented, categorized within four primary archetypes exercising these business models. In this section, important questions around these business models will be brought up such as: who uses these services/products? What value does this bring for companies? Who orchestrates these kinds of exchanges? Shortly thereafter there will be a section dedicated to a variety of pricing models that are commonly practiced in data ecosystems. In addition to these pricing models we will also explore a section revolving around non-monetary exchanges that are possible between actors.

For the sake of brevity, and transmitting the best understanding of such models, all models revolving around revenue for advertising services are touched on briefly but are not exposed in great detail. These models are very particular models that are incredibly rich and often feature very specific and complex real-time exchanges, infrastructures and components. Revenue models revolving around advertising are most commonly practiced by the GAFAM (Google, Apple, Facebook, Amazon and Microsoft) and are models that have taken years to develop. I have decided to exclude this subject in this master's thesis, as it could constitute an entirely separate thesis topic.

In a final section, we will conclude this master's thesis by presenting the scope of this project and the evolutions anticipated in these ecosystems as they evolve.

Before delving into the complexities of business models practiced in data ecosystems, we must first understand them and their primary components. This is why in the next section we will explore ecosystems and their compositions.

II. Ecosystems

A.Introduction

Ecosystems are complex, multifaceted organisms that englobe an entire collection of actors and types of exchanges. In order to best understand what a data ecosystem is and how it works, I present the primary characteristics of an ecosystem. In a subsequent section, the various actors in an ecosystem are detailed. Shortly thereafter, I provide detailed examples of types of ecosystems that exist in the market today.

In order to better comprehend what an ecosystem is, we will first examine the way in which an ecosystem is defined.

B.Definition of an Ecosystem

While there exists extensive literature on data ecosystems, we often see a lack of common vocabulary.

“Nardi and O'Day, provide a broad definition of what an ecosystem is, ‘a system of people, practices, values, and technologies in a particular local environment’. More accurately, ‘ecosystems are comprised of interacting, relatively tightly connected components with substantial interdependencies. Specific components will vary from ecosystem to ecosystem (Martin, Sébastien & Turki, Slim & Renault, Samuel. (2017).”

Described by other, more business-oriented structures, data ecosystems are “a combination of enterprise infrastructure and applications (...) utilized to aggregate and analyze information” (“What Is a Data Ecosystem?”, 2021). A data ecosystem can also be characterized as “a platform that combines data from numerous providers and builds value through the usage of processed data” (Abdulla, Ahmed, et al., 2021).

From these resources, we understand that data ecosystems are environments wherein two or more actors, whether public, private, or a combination of both, come together to exchange data or services in a variety of ways.

Before exploring the business models practiced within an ecosystem, it is important to be able to classify the types of ecosystems we are observing. In order to better distinguish data ecosystems, in an upcoming section I propose an overview of two ways to classify the ecosystems into two distinct categories: public or private ecosystems and centralized or decentralized ecosystems.

C.Categorizing ecosystems

Ecosystems and the business models that are practiced within them are impacted by the type of ecosystem that they are. Though data ecosystems contain a variety of participants (public and private) two distinct criteria help us determine the type of ecosystem that we are observing. The first distinction we can make in ecosystems is between publicly-orchestrated and privately-orchestrated ecosystems. The second way to categorize ecosystems is whether they are centralized (meaning all data is uploaded to a specific platform) or decentralized (data stays in partnering platforms and API routes enable other parties to access their data).

In this first subsection we will explore the difference between publicly-orchestrated and privately-orchestrated ecosystems.

1.Public orchestrators vs private orchestrators

Business models vary greatly between public and private orchestrators, this is why it is important to be able to classify ecosystems by public or privately-orchestrated categories.

Publicly-orchestrated ecosystems

In publicly-orchestrated ecosystems, a government entity, whether small scale (regional government) or large scale (e.g. European Commission or national government), manages the creation, organization and governance of the ecosystem.

One example of a publicly-orchestrated ecosystem is an exchange in Korea between KT Corporation (Korea Telecom) and KCDC (Korea Centers for Disease Control and Prevention) to mitigate the MERS (Middle East respiratory syndrome) disease outbreak (Boral, Austin, et al., 2019) . In order to curb a MERS outbreak in 2020, the Korean government orchestrated a quick-response data ecosystem between the KCDC and the telecommunications giant KT. A publication "Data Collaboration for the Common Good" by the WEF (World Economic Forum) details the parameters of this ecosystem: "the KCDC helped government authorities coordinate their activities as they developed critical legislation to address private-sector concerns around data privacy and consumer protection. More specifically, the legislation called for the destruction of any mobility data used during the collaboration after a designated period. This guideline helped earn the trust of KT's leaders and subscribers, while improving transparency in the data collaboration. KT's effort to improve public health via its Epidemic Readiness platform has encouraged the participation of other major mobile operators to participate in the initiative" (Boral, Austin, et al., 2019).

Government-orchestrated ecosystems such as these often have very specific purposes for their creation. Such ecosystems also have specific contracts, metrics and agreements on exactly how the data will be collected, used and often destroyed after the collaboration. Business models in government-orchestrated ecosystems are defined by a contract and contain specific business and pricing propositions between actors that we will examine in subsequent sections.

Privately-orchestrated ecosystems

Privately-orchestrated ecosystems are the most common types of ecosystems and exhibit commonly-practiced revenue and pricing models that will be presented later on in this study. Privately-orchestrated ecosystems are those organized by a private company.

One example of a privately-orchestrated ecosystem is mentioned in Marcus Roth and colleagues' article Four ways to accelerate the creation of data ecosystems wherein "a digital services provider and a supermarket chain partnered to solve a perennial challenge for consumer-packaged-goods (CPG) companies: understanding how advertising affects purchases. Each partner gathers large amounts of data—the digital services provider collects advertising and audience data from millions of accounts, and the supermarket chain collects purchasing data from millions of customers. Combined, these data sources can be used to uncover the link between advertising and purchases—a significant challenge for both CPG marketers and digital advertising platforms. The ecosystem creates benefits for both partners"(Aaser, Mohammed, et al. , 2020).

In these privately-orchestrated ecosystems we see a great variety of revenue and pricing models practiced. In order to further understand the scope of these ecosystems, we further define the difference in centralized and decentralized ecosystems in the following sub-section.

2.Centralized and Decentralized Ecosystems

Centralized and decentralized ecosystems enable different kinds of business and pricing models to be applied. This is why in this section we will describe the difference between a centralized and decentralized ecosystem.

Centralized:
In a centralized data marketplace, data is offered via a centralized location (Fruhwirth, Michael & Rachinger, Michael & Prlja, Emina., 2020). This means that the data can be uploaded in a variety of formats to a centralized point such as a marketplace, a data lake or any other type of infrastructure available.

An example of a centralized data ecosystem would be the Dawex data exchange platform. In this model, Dawex acts as the one centralized marketplace for private actors to provide, purchase and exchange data from a variety of sources. The centralized model is the model most commonly practiced in data ecosystems today.

Decentralized:
The decentralized model is a newer model of the data ecosystem created in an effort to innovate, especially when using personal data (data traceable to an individual). In a decentralized data marketplace "data assets remain at the data provider, using e.g. a blockchain" (Fruhwirth, Michael & Rachinger, Michael & Prlja, Emina., 2020). This means that technical infrastructure such as API's or blockchain-technologies are set up to build pathways for databases to communicate and transmit data. This type of decentralized ecosystem requires a great deal of innovation regarding business models.

An example of this kind of decentralized ecosystem is Visions in the Grand Calais Project wherein a variety of actors, both public and private, connect to an API enabling them to send and request data from all types of sources through the common infrastructure. In addition to coherent business models, governance rules for these kinds of ecosystems are incredibly important to ensure the stability of the ecosystem.

While the types of orchestrators, and the nature of the ecosystem can have an impact on the types of revenue and pricing models selected in these ecosystems, it is also important to understand the actors that participate in these ecosystems, as they too have an impact on the ways exchanges are monetized.

D.Actors in ecosystems

Each type of ecosystem actor has tendencies to exercise certain revenue models and apply corresponding pricing models. For this reason, this section is dedicated to describing the types of actors in the ecosystem before delving into their most commonly practiced business model archetypes.

For the sake of brevity and coherence, I have broken ecosystem participants into 4 of the most common categories of actors seen in ecosystems. The most commonly exhibited revenue models corresponding to each of these actors will be presented in a subsequent section titled “Business Models: Revenue and Pricing Models” in the sub-section “Revenue Models”.

In order to provide a topology of the most common actors in ecosystems, I have divided these categories into 4 types of actors.

- Raw or unprocessed data contributors
- Data Processors
- Connectors/Orchestrators
- Service providers

This presentation of actors can be paralleled to the sort of chronology in which an average company interacts with an ecosystem: the company buys raw or unprocessed data, they hire a data processor to turn their data into a more valuable resource, they can then use a connector to interconnect and provide or exchange additional data in an ecosystem, and if needed they can hire a service provider to transform some of their information into actionable insights or other relevant services.

Moreover, we begin our exploration of data ecosystem actors with the raw or unprocessed data contributors.

1.Contributors of Raw or unprocessed data

Raw data is “the data that is collected from a source, in its initial state. It has not yet been processed — or cleaned, organized, and visually presented” (Hiter, Shelby, 2021). These ecosystem actors provide raw or unprocessed data. Such contributors can be public or private entities depending on the form that the ecosystem takes (these ecosystem forms will be discussed in the next section). Raw data contributors can provide personal or non-personal data. These actors can also be companies of all sizes that wish to contribute their data to the ecosystem, whether by allowing transfers of their data from their cloud or database (decentralized) or providing entire raw data sets for other actors to query or access directly on a platform (centralized). Raw data sets, while much less valuable than processed data sets, are often important for companies to be able to develop new services and have insights on consumer behavior. These contributors can be data providers, meaning they sell their own data, or they can be data brokers, which sell the data of other companies. While the revenue models for this kind of data provider are fairly straightforward, unprocessed data contributors can sometimes have difficulties pricing their data as it has not yet been cleaned or transformed. We will discuss pricing models commonly used in the next sections.

Companies that purchase unprocessed data often require services that process data for it to be used in any number of ways. This is why we will define in the next section the role of data processors in an ecosystem.

2.Processors

Data processors are the actors that provide services of cleaning, transforming, aggregating, exploring and mining data in order for it to become more valuable. Companies like Palantir and the Blackrock service, Aladdin, are classified in this paper as data processors. More information on the specificities of these companies and the services they provide will be mentioned in the following section "Business Models: Revenue and Pricing" in the "Revenue Models" sub-section.

Data processing is an integral part of data ecosystems, and once a company has a large amount of processed data, with the help of orchestrators/connectors they can share and monetize it with other organizations. In the next section we will see how these connectors/orchestrators are defined and what role they play in data ecosystems.

3.Connectors/orchestrators

Connectors have many names and many roles in data ecosystems. These actors provide either a place for exchanges of data to take place and/or the technical infrastructure for other companies wishing to do so. For example, the platform Harbr produces the infrastructure allowing companies to create their own privately-orchestrated, centralized data exchanges. Other connectors provide common APIs or requirements for decentralized data exchanges. A subcategory of these actors are enablers, providing technology and additional services that enable organizations to share or exchange information.

In one interview with Markus Spiekermann, he mentions that "the connector component can be seen as a gateway to each participant into the ecosystem, and what it does is that it has APIs and interfaces to exchange the data " (Spiekermann, M., 2021). The idea that connectors can open up entire ecosystems presents many opportunities and challenges when identifying specific business models.

Moreover, revenue models and pricing models for connectors can vary greatly depending on the type of ecosystem they are in. These can range from ecosystem fee models to open-source models. Connectors enable the formation of data ecosystems that can provide and share massive quantities of information. These mass quantities of information give way to an entire sub-economy in data ecosystems, the economy of service providers.

4.Service providers

A plethora of new types of services and sub-services are evolving in data ecosystems every day. A single service provider can propose a wide variety of services that can be performed once data is processed and/or connected. Some service providers practice one or several different business models based on the services they provide and the products that they generate from these services.

While service providers are numerous, it is important to also understand how these ecosystems can possibly be formed. We have now explored the meaning of an ecosystem, the distinctions between types of ecosystems and the actors that play the most important roles in the ecosystems. We are now ready to learn about the forms that such data ecosystems can take.

E.What forms can ecosystems take ?

We have now categorized ecosystems and can start to see how all of these elements come together and the forms that data ecosystems can take.

Data ecosystems can take a variety of forms and each form of ecosystem entails revenue and pricing models that are specific to their users' needs. In order to provide an overview of the most common types of data ecosystems, this paper divides them into 3 main categories:
- marketplaces
- private data circuits
- open data & data lakes

An additional section detailing Visions ecosystems has been added in order to demonstrate a new model of ecosystem currently evolving in the European Union.

In order to best understand the most common forms of these ecosystems, we first explore the most prevalent type of ecosystem: the data marketplace.

1.Marketplace

Data marketplaces are the most common types of data exchanges and as such, contain the widest collection of business and pricing models. Data marketplaces are most often privately-orchestrated and by definition are centralized platforms. These marketplaces can cater to public or private buyers, although the latter is the most common. Marketplaces are mostly populated by raw data providers, and processors that sell the processed data. To share data, ecosystem participants usually use the marketplace's connection technology. In some cases however, a marketplace can purchase their infrastructure from these orchestrators.

To begin, we explore the definition of a data marketplace as "electronic platforms that facilitate the exchange of data [...] A data marketplace ecosystem consists of data providers, data buyers, third-party service providers and a marketplace owner" (Fruhwirth, Michael & Rachinger, Michael & Prlja, Emina., 2020). Marketplaces often feature rating systems wherein the data quality and data providers are evaluated. These types of ecosystems can sometimes feature quality evaluation mechanisms or screening processes for new sources wishing to sell or exchange their data.

A great amount of research has been conducted on cataloging the most prominent data marketplaces. Markus Spiekerman, in his Intereconomics article, presents a list of 14 well-known generalist data marketplaces, some include: Dawex and Data Broker, and Qlik. Other marketplaces can be much more industry-specific like Farmobile providing agricultural data from farmers, or specific to a certain type of data they provide such as IOT data marketplaces like Terbine.

Data marketplaces, while becoming more and more well-known and accessible, face a few challenges. Data pricing can be difficult to evaluate based on the amount of time and resources that it can take to transform the data into something usable. Pricing models in this kind of ecosystem can vary greatly and will be explored later in this paper. That being said, two common pricing models that we see are data sets that have a set price or data sets that are set up with an auction-and-bid style. Data in data marketplaces can also be purchased in full, rented for a limited period of time, or charged for querying. Marketplaces can also offer additional services for cleaning and transformation of the data sets as well as other utilities. Services will be explored at a later phase in this paper.

Although marketplaces are among some of the largest and most common types of data ecosystems, the other type of major data ecosystems are private data ecosystems. We will explore what private data ecosystems look like in the next section.

2.Private data ecosystems:

Whereas data marketplaces provide access to data from a multitude of companies for monetary exchanges, some data ecosystems are formed between two or more companies for specific purposes. Private data ecosystems are most often privately-orchestrated and can be centralized or decentralized platforms. These private data ecosystems most often cater to privately-held companies, but can include public actors or data from public actor sources. Private data ecosystems are mostly populated by companies with IoT and non-personal data that they wish to exchange with other companies in order to improve processes, products or better understand certain markets. These ecosystems are orchestrated by companies, either with their own infrastructures or by purchasing products from connectors that enable them to share data.

These exchanges can happen **internally,** within one large organization wherein different branches are able to exchange data (free of charge), or **externally** and can be shared between several companies.

Internal ecosystems:
In the case of internal private data ecosystems, large corporations create data libraries wherein all branches of the company can find catalogued information. This data is often rated and categorized to optimize the usage of such data. A specific example of this was at Total Energies, giving access to data across all 4 branches of Total through their DARE project.

External ecosystems:
In the case of external private data circuits, agreements between companies are negotiated that either allow punctual exchanges of specific data sets, or free-flow of data, based on certain parameters and timelines. These processes are usually negotiated, a contract is drafted with a specific use case and POC (Proof of Concept) and a predetermined business model for a specific period of time. This practice is normally conducted with non-personal data, such as factory data or IoT data and these contracts are renegotiated frequently.

These types of ecosystems often feature: contractual obligations, specific metrics/KPI (Key Performance Indicators) defined in the contract, a specific timeline and parameters, as well as a trial period followed by continual contractual renewals. Prices in these ecosystems are typically negotiated during the contract drafting process and are progressively re-evaluated as the value and utility of the data becomes more clear. It is difficult to accurately evaluate the cost-benefit analysis of this kind of data exchange, as data must often be cleaned, transformed and analyzed in order for this to provide value for one of the partnering companies.

Overall, private data ecosystems contain several complexities that have an impact on the types of business models practiced within them. Such restrictions are much less frequent in open data exchanges and data lakes that we will see in the next section.

3.Open data exchanges & data lakes

Open data exchanges are most often facilitated by governments and launched in order to stimulate innovation and growth in an economy or respond to a specific problem. Open data exchanges are most often publicly-orchestrated and are centralized platforms. These open data ecosystems and data lakes most often provide publicly-held information for other services to use. Open data exchanges and data lakes are populated primarily with non-personal data or anonymized personal data. While open data exchanges and data lakes are united by the fact that they are often animated by public entities it is important to understand the difference between these two categories.

Open data

Open data initiatives are almost always free data sets, made accessible to all types of services and service providers, public or private. This is done in order to promote innovation, creation of new services and in some cases, stimulate government initiatives as well. In an interview with Martin de Saulles, Lecturer in Innovation and Marketing at the University of Brighton, he mentions that sometimes governments will monetize their data, such as weather forecasting data for example, with the profits of this going back into the public sector (De Saulles, M., 2021).

There are many other examples of open data ecosystems. One such example is the UK's National Health service "is probably the single biggest data set of health data in the world". While this national database provides a great amount of educational and research information generated from the healthcare system and is anonymized, Martin De Saulles explains that even if "it's anonymized, (...) it's often not that difficult to de-anonymize data and actually figure out what the where this data comes from and who it relates to" (De Saulles, M., 2021).

Another example is the Conseil Régional Île-de-France Smart Region project, which enables a large access to publicly-held information of all sorts. The Paris Region made available public data on a wide variety of topics such as administrative data, environmental data, construction data and much more. This is a trend we are also seeing across many other governments both local and nationally in many European countries.

Though open data is an ever-evolving type of ecosystem, we also note another more research-oriented type of ecosystem being presented, data lakes.

Data lakes :

Data lakes are an interesting type of ecosystem wherein public and private sector data is brought together for scientific or AI/algorithm training purposes. "Big data is often stored in a data lake (….) data lakes can support various data types and typically are based on Hadoop clusters, cloud object storage services, NoSQL databases or other big data platforms" (Botelho, Bridget, and Stephen Bigelow, 2021). In data lakes, large amounts of data are anonymized, contributed by a large number of public or private organizations and shared. This is normally done for research purposes, allowing for scientific testing and training AI and machine-learning tools. These are mostly run by the government or government-funded initiatives to improve an industry or promote scientific research/innovation.

While data lakes are most often publicly-run, a few examples of private data lakes can be found like the Amazon Web Services data lake. Another example of this was evoked during my interview with Martin De Saulles when he mentioned a different company, AOL and their research-promoting activities "AOL made millions of search queries available years ago for academic research" however great risks were exposed because "If you knew how to sort of analyze (the information in the data lake) to actually trace individual searches to to individual searches" and represented a major security concern. Unfortunately the danger of cross-referencing data, enabling companies to inference sensitive or personal information from supposedly anonymized information is known as the "mosaic effect" and is a risk in data ecosystems.

Overall, having an understanding of data marketplaces, private data exchanges and public data exchanges provides a holistic view of types of data ecosystems. In order to demonstrate a unique type of a decentralized, privately-orchestrated, ecosystem that enables public data to circulate, the next section is dedicated to presenting Visions, the company that I currently work for, and it's creation of a new model of public/private collaborations.

While we have discussed all kinds of types of data exchanges, in the next section, a concrete example of a new type of decentralized ecosystem is introduced.

F.Visions Ecosystems

In order to demonstrate a new type of data ecosystem being built, this next section will be dedicated to Visions[1], and its work in bringing a decentralized skills education ecosystem.

Visions is a company currently working on building infrastructure, tools and governance mechanisms to address these current European data challenges. Data ecosystems facilitated by visions are considered privately-orchestrated and are decentralized platforms. Visions ecosystems provide publicly-held and privately-held information that can be shared thanks to a common API infrastructure. These ecosystems are populated by public and private structures that circulate personal data.

Visions is a connector/orchestrator working to deliver not only technical solutions, such as a decentralized API infrastructure but also works collaboratively with a variety of skills providers on writing the standards in data sharing and governance. Visions' primary technology is Visions Trust, the API and connected dashboard that enables the individual to give their authorization regarding each type of data exchange for each platform that wishes to share or receive their data.

In order to build relevant tools for such data circulation that can then be adapted on a European level, Visions has been funded to build several example POC data ecosystems wherein data can freely circulate between a variety of platforms.

[1] The author of this study, Mackenzie Himmelbauer, is currently working for this company through the work-study program offered by Assas.

1.Visions Ecosystems

Visions is building several kinds of ecosystems wherein platforms of various types are interconnected in order to facilitate the mobilization of personal skills data. In these "ecosystems" where data can freely circulate, a select group of actors develop and interconnect their platforms to provide real-time updated profiles (updated with evolutions on job or training sites) and additional information about the person.

While these ecosystems can have a variety of types of actors interconnected within them, there are most often 5 types of actors woven into these ecosystems:

- Employers
- Public organizations (regional governments)
- Edtech and Jobtech platforms
- Training organizations
- Universities

These groups, thanks to a common API infrastructure, specific governance rules and personal consents, allow organizations to share data between each other. This free data circulation benefits each actor differently. The common infrastructure allows them to share and receive personal information in order to better understand the individual, provide targeted offerings, and create better products and services. In turn, students, job-seekers and employees also are able to keep their profiles up to date in real-time, and have a better understanding of their skills and capabilities.

Visions contains 3 types of actors in its ecosystems: privately-held companies, government agencies and universities/training institutions. We first explore privately-held companies in the Visions ecosystems:

2.Privately-held companies

Private companies in the Visions ecosystems usually take a few forms:

- edtechs- online services that contribute to analyzing or refining skills that can be related to education or have educational value.
- jobtechs- online services that help employees or job-seekers better evaluate and formulate their skill sets and be able to provide these to employers.

Other educational or skills-based services can also be found in the ecosystems that can take the form of : skills evaluation platforms, orientation platforms, skills matching platforms etc.These platforms both contribute and receive data in the ecosystem. They can use this data to:

- create better services for individuals
- receive information from other companies that allows them to have a more unique user experience or provide better services
- solve problems that they may be having due to a lack of information or connected services (such as sourcing new users, providing relevant content etc.)

3.Government Agencies

Visions works with the French government at two levels: national and regional. National projects engage data operators like Visions when building national infrastructure for data collection and circulation. Thanks to its permission-management technologies, Visions enables the government to have a "separation of powers" as it collects and utilizes citizen data for a variety of purposes. Visions is also integrated in Regional projects which involve the Regional governments in France to be a data provider as well as utilizing data in their government-created platforms for orientation, education and other aspects. Visions is able to integrate these actors into its ecosystems in order to contribute data as well as receive real-time data and information on its constituents.

While Visions collaborates directly with the French national and regional governments, it is also engaged with other French and European institutions. For instance the French employment agency "Pôle Emploi" is a major actor in receiving and being able to share skills data. Enabling an individual to seamlessly upload the entirety of their high school, university and skills data found on various platforms can help a job counselor easily provide relevant opportunities and dramatically reduce the time an individual is unemployed. Other state or government-run organizations can also be included in the same way for different purposes.

While these actors previously listed are the most prominent types of actors in the ecosystems, these services are interwoven thanks to a variety of back-end actors often referred to as enablers. These are partners that provide services like connectivity, analysis of complex data sets and data mapping that allow for these types of circulations to work. Actors of this type, "enablers' can be those such as mindmatcher, HeadAI etc.

4.Universities and Training Organizations

Educational institutions have a unique role to play in these ecosystems. Universities, highschools and training institutions all collect very valuable information on students as their knowledge and skills evolve. These institutions are all in search of services to enable their students to have more access to information on their skills and to be able to provide this to employers to have better opportunities. Specific university needs have led to building smaller, closed-circuit ecosystems that include a smaller number of actors that address a pain point that the university has identified. One example of this is a re-orientation package wherein 3 edtechs have come together to build a semi-automated process for when students who no longer wish to pursue an academic track, wish to transition to another field of study. The combination of these tools, in communication with the constantly changing course and professional course offerings allow for student counselors to more accurately and efficiently counsel students on their major/minor changes.

In conclusion it is important to be able to categorize these types of ecosystems, whether publicly or privately orchestrated, centralized or decentralized, or categorized as a marketplace, private ecosystem, open data ecosystem or Visions-type of holistic decentralized ecosystem. The ways in which these organizations are categorized has a large impact on the business models that they are likely to practice. Now that we have clearly defined classification methods and typologies, we can examine the types of value exchanges that occur in these ecosystems.

III. Business Models: Revenue and Pricing Models

A.Introduction

Now that we have thoroughly defined and categorized what makes up an ecosystem, we can begin to see how all classifications, actors and types of ecosystems are intertwined. All interactions in data ecosystems between actors and the clientele they serve, involve some form of revenue and pricing models. In this section we weave the business model tapestry that holds data ecosystems together by discerning revenue models and pricing models for each major archetype exhibited in ecosystems.

In this following section we will delve into the heart of this academic exercise, which is to explore the business models in data ecosystems. This chapter aims to answer the problem statement at the beginning of this thesis by providing an organized topology of business models that one may find in data ecosystems.

By definition, a business model contains the types of products and services that a company can sell to generate revenue and the pricing models used to generate that revenue. This is why we have broken the business model chapter into two primary sections: a section detailing and categorizing primary revenue models analyzed per archetype of ecosystem actor (raw data producer, processor, connector/orchestrator and service providers) and a section focused on pricing and monetization of the models practiced.

A deliberate separation has been made between the revenue models and pricing models associated with these types of actors. The reason for this separation is that for each type of revenue model there exists a multitude of pricing models used for each actor. For the sake of clarity and brevity, pricing models are mentioned apart and referred to once the primary revenue models have been identified.

To continue with this study of value exchanges in data ecosystems we dive into the revenue model sub-section.

B.Revenue Models

Business models are complex entities, often with several moving parts and simultaneous revenue models being applied one on top of the other. I have chosen to categorize these models according to the type of actor that practices them, as previously presented as "actors" in data ecosystems. As such, the composition of this subsection is as follows:

- Revenue models practiced by organizations that sell sets of raw or unprocessed data
- Models practiced by entities that process, clean or transform data
- Models that are typically employed by connectors, or actors in ecosystems that connect services or orchestrate exchanges
- Revenue models revolving around offering innovative products and services

To give perspective to the multi-layer complexities surrounding revenue models in data ecosystems, it is important to note that a majority of these revenue models can be used simultaneously by the same data ecosystem actor.

Compounding services can be offered by each aforementioned actor. To illustrate this through an example: the platform Harbr, whose primary role is to provide infrastructure as a service (connector services), also offers additional paid services for refining or processing the data (data processor services). In the attempt to simplify such complexities, this paper will address the primary business model associated with each actor.

While there are an unlimited number of other revenue models that are being created or practiced every day, the models introduced in the following subsections are those that are most commonly found in practice and found in literature.

Moreover, we will begin our exposé of revenue models with the most rudimentary actors of the ecosystem. We will begin by learning about revenue models practiced by raw data providers.

1.Raw Data Providers

Raw data providers are numerous, and can be public or private actors in any type of ecosystem (marketplace, data lake, private data exchange). Raw data sales or exchanges are most commonly seen in privately-orchestrated data marketplaces.

Before getting into the specifics of this kind of provider, we will first define what raw data is. There exists thousands if not millions of types of raw data generated every millisecond. Raw data can be manually written down or typed, recorded, or automatically input by a machine. You can find raw data in a variety of places, including databases, files, spreadsheets, and even on source devices, such as a camera. Data such as this can come from a variety of sources and industries. Some of the most commonly exchanged types of raw data include location data and IoT sensor data.

Now that we understand raw data, we can better understand why companies purchase it. A majority of the time, purchases of raw data are private-sector companies. Governments may purchase raw data for a specific purpose, but the lionshare of raw data purchases remain private companies.

Raw data is, on average, much cheaper data to acquire as it has not been processed or filtered. At a glance, raw location data looks compelling: it's less expensive to purchase than processed data, and there is much more of it to work with"("The True Cost of Raw Location Data", 2020*)*. "Businesses use this data to build new products and services, better understand and identify customer segments, and have more insights into client behavior" ("The Different Ways Businesses Can Use Raw Data", 2021).

One specific example of raw data that can be provided is location data. "Location data is normally generated from mobile data, collected from mobile signals, GPS, beacons and WiFi. This information can be incredibly useful in understanding consumer behavior and or location-based marketing and gives access to a large number of datapoints" ("Raw Location Data: Best Datasets & Providers", 2021).

Furthermore, we explore raw data as a service and a few examples of such providers.

Supplying Raw Data: (Raw Data as a Service)

We begin by defining what Raw Data As a Service is. In one of his articles, Justin Lokitz describes Data as a Service (DaaS) "DaaS hinges on a value proposition for supplying large amounts of (...) data with the idea that the customer's job-to-be-done is to find answers or develop solutions for their customers. The customers in this case may be solution providers looking to use (...) raw data to enhance their own offerings (i.e. value proposition) or even developers wanting to develop niche applications to address consumer pains" (Lokitz, 2021).

Companies whose primary source of income is generated from raw data can be known as data producers (providing their own data from their own sources) or data brokers, assembling raw data sets for resale. Companies that contribute raw data in ecosystems rarely use raw data as their primary source of revenue. Martin de Saulles presents this during our conversation, the fact that "quite a few companies will be thinking (they) could get a bit of incremental revenue through selling this data that they've got sitting around" (De Saulles, M., 2021). This type of information is typically used to provide supplemental revenue streams. While this kind of contribution of a company's raw data contains a few competitive risks and potential disadvantages, it is still a very common practice in data marketplaces and private data ecosystems and can also be known as contributing "basic data sets" (Wolfert, S., 2021).

Raw data producers or brokers such as Lifesite, Locationscloud, or start.io can provide voluminous raw data location sets in exchange for monetary exchanges, making the raw data producers revenue models the most straightforward.

More often than not, once companies have purchased raw data, they must process the data in-house thanks to skilled data scientists and specialists they have hired. However, for those that don't have access to these resources, companies can hire data processors for these services. This is why in the next section we will examine what revenue models look like for data processors.

2.Data Processors

Data processors are the vital organs of a data ecosystem. They process, refine, aggregate and clean the data so that it can provide sustainable results for companies. Some companies will have internal processing teams, and other companies make their revenue by selling data-processing services, software or infrastructure. Processors can also rework this data for direct clients, selling it directly or selling it to a data broker who will then resell it again.

Data processors are numerous, and are 99% of the time private-sector companies. Governments can pay for processing fees in order to encourage initiatives, or hire not-for-profit associations that act as processors in order to facilitate certain initiatives. Processors are a part of the larger data ecosystem and are necessary, complementary components to marketplaces and private data exchanges. Data processors are needed in both public and privately-orchestrated data ecosystems.

In the scope of this thesis I have identified two main types of data processors, data explorers/miners and data cleaners.

Data exploration and data mining as a service:

Data explorers and miners have the technical capacity to aggregate a plethora of data sources and pull poignant insights from these massive data sets. Data mining and data exploration is not only just providing insights for companies. It is the art form of combining multiple sources of complex data and crunching this into actionable processes or information that can be used strategically. In this subsection we examine data exploration and data mining a bit more closely.

With a large variety of data sets, data exploration is the necessary first step to pulling insights from data. "Data exploration is the initial step in data analysis, where users explore a large data set in an unstructured way to uncover initial patterns, characteristics, and points of interest" ("What Is Data Exploration?", 2021). During the data exploration phase, the data can be visualized and sold. The goal of data exploration is to identify larger trends that will enable a company to identify which trends to study in greater detail and eliminate non-relevant data.

In this next section, we aim to explore what data mining entails. Similar to data exploration, data mining is the use of specially-developed software to sift through massive batches of raw data spread across several different sources, find patterns and develop strategic insights. One example, mentioned by Sarb Mann provides a data-mining example as “an individual developer builds a data product by reaping freely available geology data that can be highly valuable to oil and gas companies” (Mann, 2018).

After having a better understanding of data mining and exploration, we explore why this is useful for companies. Data mining and data exploration are vital to companies in their growth and development. Combining multiple data sources enables companies to better understand markets, products, consumer behavior and much much more. In order to be able to do this though, highly technical infrastructure, trained AIs, cloud computing and machine learning mechanisms need to be calibrated for specific purposes. These kinds of resources and *savoir-faire* are often not available internally to companies. Sarb Mann draws our attention to the difficulties that companies face conducting data mining internally “It would not be easy to find valuable data from (a) massive data reservoir acquired from diverse sources. The exploration requires tremendous effort and there will be an opportunity for players and service providers who can choose an area or segment(s) and build competency” (Mann, 2018). Even some of the largest companies or organizations in the world seek other organizations to help them tackle their biggest data challenges.

For this reason, data exploration and mining businesses are incredibly lucrative and are often practiced by mastodons in the data industry. Palantir is a well-known example in the data exploration and mining industry. Boasting clients like the U.S. Army, U.S Navy, Airbus, IBM, and Amazon, Palantir guarantees military-grade security and technologies that can combine expansive data sets and provide immediately actionable insights. For example, in one of Palantir’s technologies, Gotham use cases, a simulation of a marine national security threat mobilizes 5 different complex data sources. In this use case, satellite imagery is used to detect a potential attack, machine learning models pull data from potential nautical routes and calculate arrival time, the U.S. automated aircraft database is activated to find and identify the ship, all while the coded human-deployment options are being analyzed from a database. This operation is all made possible thanks to data mining and exploration services created by palantir. Palantir’s revenue model is interesting in that it hinges on building a pilot project at a loss, in order to have a validated proof of concept for customers at a reasonable price. Once this phase is validated Palantir then expands vertically within an organization and finally scales projects up to a larger size ensuring very large profits for the company.

Another excellent example of data mining and exploration is a comprehensive tool in the financial industry. This data mining tool, Aladdin, was built to help facilitate investment processes and decision-making by providing tools that virtually interconnect several thousands of financial resources. This service is provided mainly for institutional investors and wealth managers, and combines information on public and private markets to bring clarity to and highlight a number of investment opportunities. Aladdin’s revenue model is building and refining the Aladdin tool and delivering regular investment insights, and suggestions for customers.

Pricing models in data mining and exploration vary greatly and are often tailored to specific projects or objectives that a company may have. We may observe pricing models

Where data mining and data exploration are flashy, well-known industries with big players, data cleaning is something much more rudimentary yet incredibly important for companies, especially when handling raw data.

Data cleaning as a service:

Data cleaning entails taking the raw data, eliminating errors and changing the formats so that it can be easily sold, manipulated by AI or machine learning or more readable for other sources exploiting the information. In his linkedin publication Sarb Mann notes “Data cleaning is an important aspect of data management & transformation, and (the) cleaning phase is tightly integrated with other phases (Mann, 2018).

Companies are in a constant battle with data: how much time or investment will it take me to collect/purchase and use the data versus the ROI (return on investment) of this data. Nicolaus Henke Senior Partner of McKinsey & Company in London, mentions that for companies “the fundamental problem...is from an economics standpoint, it's like a tragedy of the commons problem in most ecosystems. The efforts, and the benefits from sharing the data are normally asymmetric. For example, a retailer and a telco. The telco has a lot of location transactional data, there's a lot of very useful data which retailers have with which the retailer could use for promotional reasons, but in order to make them useful for the retailer the telco needs a lot of work, and you already have a long backlog. So do you spend time cleaning data or spending time essentially, you know, focusing on what's in it?” (Henke, N., 2021). This is clearly a problem that many companies are facing and data cleaning is a very valuable service for those companies that don’t have the time and resources to do so. That being said, for smaller scale companies, data cleaning is becoming more and more commoditized. “With today’s technologies the business model of cleaning as a service is no longer reserved to large organizations, Sarb Mann notes that “, in the new paradigm entrepreneurs can voluntarily clean and curate data sets and sell these to data merchants and brokers with unique requirements.” So now that we understand what data cleaning is and looks like, let’s explore a few businesses that provide data cleaning.

Data cleaning services are often software as a service made available by a variety of actors. and revenue models differ with each service. Companies like Trifacta, TibcoClarity, Openrefine and DemandTools are types of software services with common revenue models of selling data cleaning services. We examine the pricing models of these companies in a subsequent section.

In a very interesting real-life application, we have also seen data cleaning as a possible value exchange or commodities swap wherein certain databases will accept to share their data with another actor in exchange for it to be cleaned. This was demonstrated in an interview with Sjaak Wolfert wherein Farmobile, offers a sort of **commodities swap:** wherein farmers IOT data is collected by boxes that farmers attach to their tractors, IoT data is aggregated and in exchange for this it is cleaned by Farmobile. This cleaned data can then be resold and farmers receive a small commission fee on the sale, in addition to curated reports for farmers on their data. This case demonstrates a diversity in types of pricing models but is a fascinating real-world application to this kind of business model (Wolfert, S., 2021).

Although data processors are increasingly more important with more and more raw data being collected each year, so are the connectors. Connectors are the glue that holds ecosystems together by providing products and services enabling actors of all types to interconnect. In the next section we explore some of the business models that connectors exhibit.

3.Connectors

Connectors are the body of the data ecosystem. They build technical infrastructure and tools to interconnect ecosystem members so that data can be exchanged. Connectors can take many different forms: data intermediaries, software providers, tools developers. At this exact point in time, connectors/orchestrators are almost exclusively private entities versus Government-held entities. Connectors are also a part of the larger data ecosystem and are necessary, complementary components to constructing marketplaces or virtually any other type of ecosystem. As connector business models remain the most complex business models to be determined, we will categorize them as either orchestration as a service, Infrastructure as a service or enablers. Before exploring models we will first define what a connector is and who uses these services.

We begin by defining what a connector is. “There are many ways that the connector function can be defined: leader, architect, instigator, strategist, facilitator, or simulator. The main specificity of the stimulator [or connector] function is that it involves thinking about and influencing the ecosystem” (Martin, Sébastien & Turki, Slim & Renault, Samuel., 2017). Connectors build technical architectures, infrastructure and enabling services on top of that. Connectors/orchestrators also generally have governance or rule-enforcing responsibilities within ecosystems that they create. They are typically providing additional services on top of their architecture.

Simply put, without connectors/orchestrators, ecosystems and the entire economy around it would not exist. Connectors provide ways in which companies can share, buy, and work with data. Enabling these connections enables the entire data ecosystem economy, and thus enables a great deal of businesses both large and small.

In this next session we will explore one of the most common business models for these connectors, orchestration as a service.

Orchestration of ecosystems as a service:
Orchestrators provide technical infrastructure to facilitate interconnection of services, manage the ecosystem and its components to ensure value for each of its participants. Orchestrators are responsible for establishing the ecosystem, the governance principles and enforcing the rules commonly established by the group. The BCG Article, "Where is Data Sharing Headed" states that orchestrators "are the organizers and managers of the ecosystem. They set the governance and value capture rules, provide the platform for data sharing and innovation, coordinate the activities of participants, and provide a channel for the products and services of contributors" (Russo, Massimo, and Tian Feng, 2021).

Other examples of orchestrators can be found throughout a variety of use case projects. For example in the project Boost 4.0, "will lead the construction of the European Industrial Data Space to improve the competitiveness of Industry 4.0 and will guide the European manufacturing industry in the introduction of Big Data in the factory, providing the industrial sector with the necessary tools to obtain the maximum benefit of Big Data" (Boost 4.0, 2021). In this data space, the orchestrator or coordinator in this situation is the association Innovalia. While Innovalia is the organizer of this use case, it does not provide the technical infrastructure that this project uses to circulate data. This is an example of orchestrator as a service, wherein the revenue model is a funding model, with the funds being issued from the European Commission. This type of orchestrator activity is frequently funded through government initiatives or other funding mechanisms.

Decentralized ecosystems:
Orchestrator organizations are much more common among decentralized ecosystems wherein the data stays in the producers databases. In decentralized ecosystems, based on the authorization infrastructures, parties wishing to exchange information request and query data directly from others' databases. Decentralized orchestrators are a new kind of model currently evolving in Europe. Companies that provide these kinds of services are Vastuu, Visions, OneCub. The majority of decentralized ecosystem orchestrators operate on the separation of powers principle, meaning that they do not have access to the data that circulates within the ecosystem, but rather provide the technical infrastructure and bring together relevant actors to circulate data.

Centralized ecosystems:
For centralized ecosystems "There are numerous well-known orchestrators managing platform-based ecosystems engaged in social media, e-commerce, transportation, banking, and even mining. These include tech leaders Google, Amazon, Facebook, and Apple, as well as longer-established companies, such as Maersk and Cisco. And there are many emerging leaders among new startups, including the Switzerland-based credit platform MoneyPark, and GlobalSpec, a US-based information platform serving the engineering, industrial, and

technical communities" (Dexheimer, Maximilian, et al., 2020). The centralized orchestrator as a service revenue model often resembles the marketplace typology but is also present in private exchanges, public exchanges and public/private exchanges.

Overall orchestration as a service, though slightly different models of funding, goes hand-in-hand with the Infrastructure as a service model that we exhibit in the next section.

Infrastructure as a service:

Infrastructure as a service is a new type of business model wherein these orchestrator organizations provide a common infrastructure or technology that connects a multitude of businesses, services, or data providers to enable exchanges of data. This infrastructure can take different forms, such as providing APIs to promote data circulation or providing the technical infrastructure so that other companies can develop their own data sharing ecosystems or marketplaces. "This business model is chosen by enterprises acting as intermediaries that facilitate the access to (...) resources by profit-oriented developers or scientists" (Ferro, Enrico & Osella, Michele, 2013). Infrastructure as a service clients can be varied; Orchestrators can be clients of Infrastructure as a service and can also provide infrastructure as a service, which we will see in the next section. BCG studies state that "Orchestrators typically leverage platform and infrastructure-as-a-service solutions from hyperscalers" meaning that the orchestrating platforms are typically infrastructure-as-a-service solutions providers for organizations as large as Amazon or Microsoft.

Companies of all sizes buy infrastructure as a service. "Larger companies "oftentimes, just use the software as a service level. They want to have a cloud provider that has the software but also manages all the infrastructure for them". On the contrary, "the small and medium enterprises, they have their own small networks, so you have one or two technical guys in their company, and they just want to want to use their own infrastructure because infrastructure services cost too much, but then they need the software" provided by IaaS companies (Spiekermann, M. 2021).

One may ask, what value does Infrastructure as a Service truly offer? To provide a certain degree of simplicity we see this divided into two components: infrastructure built for a company/group and tools created for companies to build their own custom-tailored infrastructure.

For those orchestrators/connectors providing general infrastructure for a group, a McKinsey study reveals that "This archetype builds a core infrastructure and tech platform on which other companies establish their ecosystem business. Examples of such businesses are data-management platforms and payment-infrastructure providers" (Abdulla, Ahmed, et al., 2021). In an explanation of a Google "Open Data on the Web" Workshop, the group concludes that one of the primary value propositions that such Infrastructure as a service business model proposes “hinges on an attractive, inexpensive or free initial offer (...) that encourages continuing future purchases of follow-up items or services (...) that are usually consumables characterized by inelastic demand curve and high margins” (Ferro, Enrico & Osella, Michele., 2013). One example of this kind of service is the International Data Space Connector. This solution “is the central technical component for secure and trusted data exchange. The connector sends your data directly to the recipient from your device or database in a trusted, certified data space, so the original data provider always maintains control over the data and sets the conditions for its use. The connector uses technology that puts your data inside a sort of virtual “container,” which ensures that it’s used only as agreed upon per the terms set by the parties involved ”(*IDS Components, 2021).*

Build-your-own infrastructure:
Infrastructure for connecting companies can be incredibly complex to build in-house. Even within large corporations, creating infrastructure for data sharing between branches can be overwhelming. This is why there exists a service provider that creates such a ready-to-use infrastructure. One specific example of an Infrastructure as a service provider is the company Harbr. This is a privately-owned company that provides the software (Software as a Service) for joining or creating custom enterprise or commercial data exchanges. In an interview with Markus Spiekermann, he explains that Infrastructure as a service relies on having “products, and a good service to build such ecosystems.” Infrastructure-as-a-service must support all “participants, and then try to build service offerings, descriptions, and have a subscription fee that the operator would collect in order to meet ecosystem needs” (Spiekermann, M., 2021).

While Infrastructure as a service both ready-made and build-your-own versions are the central components to this connector business model archetype, it is important to remember that once you build the ecosystem, there are a number of features and additional services that can be offered. This is why ecosystems also contain Enablers and enabler-type business models.

Enablers and services they provide:

In addition to these Infrastructure as a service organizations, another interesting revenue model in ecosystem services are the enablers. A publication exploring business models in data ecosystems explains this model in that “the enabler acting as intermediary provides developers with easier access to (...) resources...(...) and datasets are subsequently catalogued using metadata, harmonized in terms of formats and exposed through APIs, making it easier to dynamically retrieve data in meaningful way” (Ferro, Enrico & Osella, Michele., 2013).
Enablers can come in many different forms: data intermediaries, governance and standards specialists, data-mapping technologies etc. These companies “provide the ecosystem with infrastructure and tools, such as connectivity, security, or computing resources. Multiple enablers will be needed to support the different models” (Russo, Massimo, and Tian Feng, 2021).

Intermediaries as enablers

For example data intermediaries are enablers added into ecosystems. They provide permissions management for personal data and offer a permissions management dashboard so that individuals can manage how their data is shared. Visions is also an example of a data intermediary, which provides a mechanism for individuals to give their permission for their data to be shared between service providers. This permission can later be revoked on a personalized dashboard. Data intermediaries often practice pricing models such as yearly operating fees based on the number of people that they manage their data for.

Gaia-X is also an enabler working towards common standards for data sharing at a European level. "GAIA-X is a project initiated by Europe for Europe and beyond. Its aim is to develop common requirements for a European data infrastructure. Therefore openness, transparency and the ability to connect to other European countries are central to GAIA-X" (*GAIA-X - Home, 2021).* This is done by establishing standards for defining metadata and descriptions for data contributors so that data can be easily tracked and catalogued when used. The overall aim of this is to enable cross-national ecosystems to operate in a transparent and more efficient way. Gaia-X is a non-for profit organization that many actors of all backgrounds support to enable data exchanges and governance principles.

Data mapping is another major problem in data ecosystems that companies solve. Data providers come together with data in a variety of formats that all have different corresponding data types. For example, on LinkedIn, a user's diplomas are listed under "education" whereas on another platform like Jobteaser, it's listed as "experiences and education". If one were to automate the process of keeping each of the platforms up to date each time a student graduates from their bachelors or master's program, it would be important to link each of the corresponding data points in each site. Companies specialised in automated data mapping exist and develop tools for detecting and automating this data mapping. One such enabler is Mindmatcher, which maps educational and skills data. This business model is a sort of matching as a service and to set pricing models, evaluate the scope of the work and set a fixed price for the work mapping work to be done. Similarly, another type of Enabler are enablers that harmonize and catalogue data so that other servers, services and developers can have easier access to the data.

Overall, in an ecosystem, it is crucial to have these orchestrators/connectors and the vast amounts of services they provide and the sub-services that they enable. Similarly to the sub-services that connectors enable, a wealth of new products and services providers are appearing everyday with the growth of this data sharing ecosystem model. We will explore a few of the most common archetypes of these product and service providers in the next section.

4.New Product and Service Providers

In this last section we explore a variety of different products and services that are made possible thanks to data ecosystems. This can be considered as a kind of sub-set of services that have evolved out of these data sharing ecosystems.

In the next section, we present two types of primary services we see within data ecosystems: optimization as a service and insights as a service.

Optimization as a service:

Optimization in businesses is a key success factor as data optimization can lead a company to achieve significant savings. This is why several service providers have focused on optimizing their data and data sets in order to help a company run more efficiently.

Internal optimization: This kind of optimization as a service can come from a company's internal consulting service or can be an outside consultant that comes in to provide technical operational efficiency. "This archetype vertically integrates data within the business and the wider value chain to achieve operational efficiencies. An example is an ecosystem that integrates data from entities across a supply chain to offer greater transparency and management capabilities" (Abdulla, Ahmed, et al.*, 2021).*

Supply Chain optimization: Another example of this kind of optimization is using a combination of data sets external to an organization and internal to the organization in order to better understand and provide value to the entire supply chain. This archetype vertically integrates data within the business and the wider value chain to achieve operational efficiencies. An example is an ecosystem that integrates data from entities across a supply chain to offer greater transparency and management capabilities (Abdulla, Ahmed, et al.*, 2021)*. Sjaak Wolfert demonstrates this through his example of the milk producer supply chain. "Another interesting one is for value net creation (...) in which you share data between various actors within the supply chain. So I always have the example of a milk production, where you have farmers producing milk for the milk processor but you need food for it. (...) So, basically, seeds, food, it also needs equipment, like robots, etc. Now all these companies are all these actors I mentioned they are basically not competitors from each other. But if they could share their data, they could have added value from that because, for example, the seed supplier delivers feed to a farmer, but then, he doesn't have any data on the production of milk and the quality of that. He has no insights into the genetic resources that are used and that would be very interesting for a seed supplier. You know that's because then you can maybe optimize your feed inputs to the farmer and sell and better feed that is more suited to his or her needs" (Wolfert, S., 2021). In this example Professor Wolfert is demonstrating a specific model called Value net creation, in which the insights and efficiencies gained from such exchanges is the payment method for this type of exchange.

Insights as a Service:
"Insights" is a major buzz word in the data world. The types of insights that are drawn out are multitudinous and have been categorized into a few sub categories.

Information as a service:
Information as a service is a notion introduced in Justin Lokitz's article *Exploring Big data business models & the winning value propositions behind them*. He describes this information as a service model wherein " customers don't want to or do not have the resources to process and analyze data. Rather they are willing to exchange value for analysis from trusted parties. The IaaS business model is all about turning data into information for customers who need something – and are willing to pay for something – more tailored. The customer's job-to-be-done is more about coming up with their own conclusions or even "selling" an idea based on certain information" (Lokitz, 2021) . In this model, those that possess the data don't necessarily know what kinds of trends or information they are looking for and a firm will provide insights on certain markets or products that the company can then act upon to launch new products or services on different segments.

Answers as a service:
Whereas in information as a service, customers are paying for general insights on an amount of data, answers as a service is a situation wherein a firm can provide "high-level answers to specific questions based on data sets" (Lokitz, 2021). An example of this would be a company that wishes to plan for their next construction project and wishes to use localization data purchased on a data exchange network and combine that with geospatial data that contains information about parcels available for purchase. Answers as a service providers are going to take massive quantities of these data sets and provide answers for this specific client and propose 3 ideal locations. This is just one example of what a firm like this is capable of doing. Navteq which was acquired by Nokia in 2007 was a provider of geospatial data that could combine any number of its data sets to sell packaged information based on location data.

Overall, this brief passage about interesting products and services that have evolved out of data ecosystems vary greatly in terms of pricing models but each represent lucrative industries when properly capitalized upon.

5.Conclusion

As we have seen throughout this section, data ecosystems weave a complex web of revenue streams. As this is a nascent industry, we often see different types of actors (Orchestrators/connectors or processors) providing additional services and practicing composed business models. Keeping track of these can be complicated enough, without thinking about specific pricing models applied in addition to this. In order to delve deeper into the business models, there was a deliberate separation of pricing models. We will now transition into examining the top archetypes of pricing models and their components.

C.Pricing models

Millions of transactions take place every second in data ecosystems and for each type of transaction there are pricing models. Each kind of actor and service can practice one or several of these types of transactions and thus exponentially multiplying the types of pricing models based on the services or products that they sell in their business model. As the goal of this paper is to present a variety of business models that are being used in data ecosystems and for this to be the baseline for professionals wishing to best navigate in data ecosystems and profit from this, I felt it would be essential to mention the different pricing models that have appeared regularly in research and that seem to take place in data ecosystems.

As there are many types of pricing models practiced simultaneously by data services providers, these transactions have been divided into 4 sections:

1. Single-request transactions and the pricing methods typically associated
2. Volume packages
3. Time-based transactions
4. Other pricing methods commonly practiced in data ecosystems.

1.Single request transactions

In single-request transactions, exchanges are made punctually. An example of a single-request transaction is the purchase of a dataset through a marketplace exchange such as on the IoT data marketplace Terbine wherein companies can purchase and exchange IoT data. The data providers upload large data sets and interested parties purchase them for a fixed sum of money (similar to the amazon pricing model).

Single-request transactions can be incredibly useful for companies, wishing to pay only for exactly what they need. Overall, it enables those who use this service to pay as they go, making it often practical and affordable (Firth, Robert, 2021). This can allow for smaller companies with smaller budgets to have access to specific datasets that they need to improve their products and services.

Associated pricing models can methods include:

- One time fee and unrestricted access: Also known as one-off fees or flat fees
- Pay-per-use
- Pay per unit
- Query-based
- Customer-proposed price[2]
- Listing fee

[2] customers propose the price that they are willing to pay for the product or service and suppliers can either accept or decline the proposition

- Transaction fee
- Commission-based sales

Raw data producers' pricing models are primarily single-request transactions. Here are a few examples: Lifesight's Data assets are priced based on monthly subscription fees ranging from $2,000 / month to $60,000 / year. Locationscloud offers custom packages for specific location data needs.

While these pricing models are easily applied to fixed data sets that can be purchased in a few clicks, we explore in the next section the kinds of pricing models that are engaged with high volumes of data.

2.Volume package transactions

Volume-package transactions are those in which larger volumes of data are exchanged or purchased, or companies pay to have access to large volumes of data for specified periods of time. "Companies use big data in their systems to improve operations, provide better customer service, create personalized marketing campaigns and take other actions that, ultimately, can increase revenue and profits. Businesses that use it effectively hold a potential competitive advantage over those that don't because they're able to make faster and more informed business decisions" (Botelho, Bridget, and Stephen Bigelow, 2021).
Associated pricing methods include :

- Step pricing[3]
- Service fees
- Storage fees

Volume transactions can vary greatly depending on the type of ecosystem and the way in which the data is exchanged. Some of the same single-request pricing models can also be exhibited in volume pricing models such as pay per volume and pay per query. Another way of categorizing pricing models is based on time-related transactions.

3.Time-based transactions

"Time-based pricing is a method of pricing based on the time which the item or service is purchased" ("Time Based Pricing", 2021). Time-based pricing methods enable businesses to charge clients based on a specific amount of time. Businesses practice this kind of pricing model because it helps businesses "capitalize on the compounding value of customer relationships. That means that as long as customers continually see the value the company provides for them, they'll continue to pay for it" (Campbell, Patrick, 2019).

Pricing methods associated with this are:

[3] stair-step pricing, which means that once a certain usage is reached, the plan must be upgraded: https://www.billflow.io/pricing-models/step-pricing

- One time fee and unrestricted access
- Freemium-limited
- Freemium-unlimited
- Subscription fee
- Package fee
- Progressive pricing
- Hourly fees

Data processors models: Blackrock's Aladdin platform, mentioned as a data processor (data miner/explorer archetype) uses time-based pricing methods, with a freemium-type basic subscription fee offering basic insights, with access to more information and tools with each increased subscription fee.

These kinds of pricing models are becoming increasingly popular in the data exchange communities. This model hinges on the data providing consistent and constant value for consumers.

Although the most common archetypes of pricing models have been presented in the aforementioned sections, other interesting models have appeared throughout my thesis research and are published below in a more generalist section.

4.Other pricing models seen across publications

Other types of pricing models have been exposed throughout the resources mentioned in my bibliography. These pricing models can be used in conjunction with the other pricing models mentioned previously. These include:

- Prototype fees for POC construction then scaling[4]
- Project-based pricing
- Flat fee (incredibly common)
- Cross-subsidization[5]
- Fixed fee
- Proprietary licencing
- Open source licencing
- Dual licencing[6]
- Royalties

[4] practiced by Palantir
[5] The practice of using profits generated from one product or service to support another provided by the same operating entity. -https://stats.oecd.org/glossary/detail.asp?ID=4968
[6] *Dual licensing usually refers to offering software under a proprietary license and an open source license.-https://www.synopsys.com/blogs/software-security/software-licensing-decisions-consider-dual-licensing/*

- Profit sharing
- Commodities exchange[7]
- Micro services provided for ecosystem fees

Raw data producers demonstrate a few of these transactions: Start.io start.io's data pricing models are One-off purchases, Monthly Licenses, Yearly Licenses, and Usage-based pricing models.

Data cleaning services are often software as a service made available by a variety of actors. and revenue models differ with each service. Companies like Trifacta, and TibcoClarity offer codeless technologies and subscription services. Others like Openrefine offer their technology on an open source (free model) and others like DemandTools offer annual licence fees per user. Once again, these types of pricing models will be explored in another chapter.

Pricing models for connectors/orchestrators are unique to each orchestrator can be presented in a variety of ways such as charge per query (requesting service charged by providing service), monthly fees (requesting service pays monthly fees) or all-inclusive membership fees (all parties in the ecosystem pay the ecosystem fee on a subscription basis and have access to data from tools). The IDSA is an association providing free open-source tools to members within their association, specifically for orchestrators. The membership for IDSA fees range from 35 000€ for companies wishing to join, to 1 000€ for startups and non-commercial institutions. IDSA is not only a technical connector but brings lucrative data exchange projects together thanks to their rich network, and provides access and assistance to applying to funding packages (*IDSA-MembershipFeeRegulations, 2021)*.

In conclusion, there exist a variety of pricing models that can be attached to each of the aforementioned business models. Each company or service provider will apply one or several of these pricing models to their products or services.

Although such pricing models are the most commonly practiced in the data ecosystems communities, we have identified another important value proposition that is appearing in our ecosystems. Very little academic research has been conducted on this next section, and is almost entirely based on on-the-field work and best practices in functioning ecosystems. These exchanges are non-monetary exchanges and will be demonstrated in the next section.

D.Non-Monetary exchanges

In order to widen the scope of analysis of ecosystem business models and pricing models, I chose to also include a section detailing a few non-monetary exchanges that can take place within ecosystem services. Some of these service providers also practice non-monetary exchanges often to make their service offerings more attractive, or to enable non-for-profit

[7] commodities between businesses are exchanged, for example data is exchanged for reports and insights that can be produced out of this data

organizations to have access to data and resources. This practice of offering non-monetary exchanges is most often in addition to monetary exchanges but in some cases can act as substitutes for financial exchange.

In some publications, these non-monetary exchanges are referred to as "soft" value exchanges,some refer to types of exchanges as formal and others as soft (ODI-Data-Ecosystem-Mapping, 2019).

In this section we examine some of these non-monetary exchanges by category. The Open Data Institute or ODI categorizes 6 specific types of soft exchanges that can be offered in data ecosystems: insights, knowledge, feedback, advice, network and policy. In addition to this, I have contributed: data-for-data exchanges and commodities exchanges. Shortly thereafter, I explain briefly other soft-exchanges that we have identified according to different types of actors divided into 2 sub-sections: private companies in skills and public actors such as regional or national governments.

We begin this section by exploring the 8 most common soft exchanges identified thus far.

Insights:
In one of the previous examples mentioned throughout this thesis, Farmobile, an agricultural data aggregation platform, exchanges detailed reports and insights based on the data that participating farmers have contributed. Farmers can then use the insights generated to conduct more efficient farming practices, better prepare for market needs (planting relevant crops). Other insights might include those seen in the Join Data agricultural project wherein farmers contribute their raw data and the association cleans the data in order to extract industry insights and contribute insights to their farmers. For example, one farmer states "The value of data does not have to be cash, but can add value to your company. By sharing data I get better, faster and more precise information. Not for example a message 'maybe one of those 10 cows is sick'. No, I want to know 'that one cow is sick'. This ultimately results in less work, less stress, less antibiotic use and less worry. I also want to know exactly which part of the day a cow is in heat. Twenty years ago it was great progress that we knew that a cow was in heat in the past 48 hours and that you had to pay attention again in 3 weeks, now that is no longer enough" (JoinData, 2019).

Feedback:
Feedback is something that is more and more valuable for solutions of all sizes. As platforms evolve and new technologies are released, something that an ecosystem is well-positioned to provide is feedback. Feedback can come in many forms: data quality, platform efficiency. At Visions, one of our primary attractiveness factors is the fact that we can provide regular student-tester feedback on our partnering platforms. A large majority of companies don't

have the time/resources to launch a tester program to improve their solutions and thus having regular feedback is very valuable.

Advice & Knowledge:
Ecosystem connectors often have a global viewpoint that other stakeholders in the ecosystem may not. These ecosystems can provide targeted advice on improving operations, evaluating upcoming trends or improving solutions based on feedback that other platforms receive from their users and/or that other platforms have to give to each other. Knowledge and understanding on what trends may be going on in an ecosystem are something that a data orchestrator or a collective of actors in an ecosystem may provide to one another.

Network:
Ecosystems and interconnections offer great synergies and joining a data ecosystem is a fantastic way to work with complementary services. As an ecosystem is a hub with several connecting nodes and new partnerships and synergies are great value propositions of data ecosystems, providing each node to work together in a controlled environment. In addition, being familiar with and oftentimes being connected to different sources facilitates further collaborations between parties and can result in new use cases and R&D projects.

Policy :
Governance and group policies are incredibly important in data ecosystems. These ensure protected collaborations between parties and can also have important implications for new European or national policies established around data sharing regulations.

Data-for-data:
A phenomenon that exists almost exclusively in practical applications rather than academic publications is data-for-data exchange. This kind of exchange was mentioned by Christoph Mertens, a part of the International Data Spaces Association (IDSA) that works on use cases with large companies in a variety of industries. He mentions that they "also observe things like (companies contributing) this specific data set and (receiving) another data set from your side, and really have an exchange of data" (Mertens, C., 2021). While there are many challenges to evaluating the value of data in these kinds of trades, it is something that in Visions ecosystems we see happening frequently.

Commodities exchange:
During the interview with Sjaak Wolfert, the notion of "commodities exchanges" is evoked, in which data can be exchanged for other products. A few propositions of such products can be as follows:

- Raw data
- Processed data
- Reports and documents
- Physical goods
- Services

- Certificates

In conclusion, a plethora of value exchanges can be proposed in data ecosystems, but will always be tailored to specific situations. In the next subsection, we observe a few soft exchanges tailored to specific actors.

Actor-specific exchanges:

In this section we demonstrate specific non-monetary exchanges specific to certain archetypes of ecosystem participants: private companies, regional and national governments and other public actors. The following section has been created based on Visions' exchanges with each of these actors in the context of the creation of decentralized skills data ecosystems in France.

Private companies:
The major issue for private actors, small companies and larger companies is most usually relative to revenue generation. This implies that the soft exchanges possible are closely related to generating additional revenue streams or capturing new parts of the market.

For instance, what we typically see as a relevant value proposition for Edtech (educational technologies companies) that join our ecosystems is:

1. generating additional university or high school prospects
2. sourcing new students or professionals to join their platform and therefore access to new groups of people to market to
3. participation in local/regional government funding bids
4. participation in international european funding bids
5. testing/feedback on their platforms and new products and improving their services
6. user information being kept up to date and having relevant information from other platforms kept up to date in real-time

Another example of a private actor are job platforms seeking:

1. more traffic on their platforms
2. real-time updates on their platforms keeping them updated on completed training or university programs, certifications
3. real-time information about when a person is employed
4. new employment opportunities opening up within other companies

Regional & National governments:
These local or regional governments are often seeking soft exchanges revolving around statistical information, and promotion of government initiatives. Some soft exchanges that have been proposed during our work with regional governments are:

1. interconnection of various government platforms and edtech/local actors to facilitate certain initiatives[8]
2. job-placement statistics enabling the government to evaluate the efficacy of its programs
3. information allowing them to easily identify and troubleshoot problems and upcoming trends

In conclusion, it is possible to provide soft value exchanges to each kind of actor in order to promote collaboration and the construction of new projects.

IV. Conclusion

Overall, through this study we have been able to identify the key components of data ecosystems, revenue and pricing models for the primary actors in ecosystems, and how they come together to exchange information and value. We conclude this paper with final reflections on data ecosystems and on the types of business model evolutions we see. Shortly thereafter we draw final conclusions on this master's thesis by extending the scope of this project, and describing the outcomes we hope this project will have.

A.Conclusion on Ecosystems:

In order to understand the future of data ecosystems, It has been important to understand the ways in which we define data ecosystems, their components and typologies.

The future we see in public/privately-orchestrated ecosystems is the following*:* at a certain point some of these large data marketplaces may become publicly-acquired goods as they will start to contain substantial amounts of citizens' personal information. The information exchanged in such ecosystems is already incredibly valuable. Take the personal information that circulates within Google's ecosystem. What would happen if any one government acquired Google and all of its information. Would entire economies shift ? Would some industries be eliminated? Would governments use this unquantifiable amount of data to manipulate their populations or other world economies?

The future of centralized/decentralized ecosystems is quite a fascinating debate with the European community focusing heavily on the future of data ecosystems. The current trend in the skills data market is "Centralization". Many private companies, in trying to take the majority market share are attempting to become the centralized one-stop shopping for all data.The problem with this model is that, when one mastodon does this (like a GAFAM), all services come together around the same actor to produce one large catalogue of services in an organized marketplace-like manner. We see this with the Apple Store model- where Apple created the centralized place for all services to appear and be purchased. However, when

8

there are many actors competing for the same centralized model, the market gets fragmented into several actors providing the same services and competing between each other. This is what is happening with the skills data market; many actors are vying for this centralized service provider title- and are splintering the market into 10 or 15 mega-sites that don't always offer quality services.

This problem is not only unique to the private sector. The French government in particular, is ready to innovate and provide new service hubs based on their own internal innovations and new types of government applications for learning, education or skills discovery. The problem with that is that by creating a centralized platform with their own versions of edtech technologies, they are putting smaller actors out of business that actually provide much much better services than the government is capable of providing with their limited resources.

Knowing this, I predict a new wave coming, which is the proliferation of the decentralized platform model. Decentralization models enable countless actors to be interconnected- thus allowing all actors to participate and thrive and letting the consumer choose where they want to share or get their data from rather than a winner-takes-all mega-platform. In this decentralized model, we can connect government platform initiatives to the private actors that they are affecting. We do this by allowing them to have access to the data generated by the smaller actors that then help the government improve their overall offering or tools without putting anyone out of business. New legislation and European Commission incentives to build sectoral databases support this kind of decentralization model by providing funding for the creation of governance principles and technical API infrastructure.

Actors in ecosystems will also drastically evolve with ecosystems. With the evolution of technology and exchange possibilities, I predict that there will be a proliferation of types of actors, and services that each actor can propose which I will detail for each actor in the business model conclusion.

On the contrary, I believe that the types of data exchanges will stay fairly simple. I estimate marketplaces will begin to merge into large conglomerates, and will most likely be owned and operated by large companies like Amazon or Microsoft.

B.Conclusion Business models section:

Business models in general are complex entities that evolve and grow more complex over time. As ecosystems and their exchanges grow more complex, so do the types and sheer number of revenue models practiced within them. I estimate that the revenue model landscape will look completely different in 5 years, 10 years and even more so thereafter.

Furthermore, as we create more tools, applications and platforms that track our data, we multiply the amount of raw data that exists. The more raw data that exists, the more companies can learn from their consumers and from people's habits in general. Raw data providers will remain extremely relevant for a large number of companies. The way in which I see this industry evolving is that more and more companies will develop tools or mechanisms for collecting their own raw data and thus may be less numerous to pay for these raw data collection services.

On the contrary, data processors (miners, explorers, enablers) are gaining increasing importance. Now that we have massive amounts of data coming from a huge number of sources, how can we turn it all into actionable insights? Large companies appear to be producing their own internal types of data processors more and more, but this processing industry still has a huge amount of expansion ahead of it and are often dominated by large companies like Palantir that have technological and commercial advances on their competitors.

Similarly to processors, the connector industry is nascent but developing every day. With this decentralization model estimated to grow as mentioned in the previous section, connectors also become more and more important. Connectors will continue to interconnect services, create places for exchanges and additional connectivity services. This market will grow exponentially over the next few years as companies produce more and more data and wish to monetize this with other companies.

In addition, the most rapidly-evolving type of industry are th*e services and service providers.* We are starting to see service providers that innovate and build off of additional services. Take for instance, this is a service that was built on top of an original service, which was providing insights. We will see this kind of piggyback evolution take place continuously as ecosystems and their services evolve.

Overall data ecosystems and all of its components are evolving at the speed of light. This is a great time to be working in a data ecosystem and being a part of the data ecosystem sub-economy. The opportunity to innovate and be an entrepreneur cannot be ignored. Building services or components that connect or participate in these kinds of ecosystems is ideal. You have clients that have significant amounts of money, governments that will finance your prototypes, and you only need an initial technical/ecosystem expertise, which is relatively simple- as the ecosystem model is just beginning.

Pricing models will also continue to evolve as more complex services are formulated and performed for paying customers. I also firmly believe that the "soft exchanges" as exhibited in this paper will evolve and be used to add value to paid packages with increasing regularity or to enable governments to collaborate more with the private sector.

C.Where are data ecosystems headed?

So after learning about the ecosystems and the value exchange models that can occur within data ecosystems, we can now ask ourselves a very important question: where are data ecosystems headed?

Data ecosystems are nascent. Associations included in this thesis such as IDSA (International Data Spaces Association) have provided mapping tools that enable us to see the sectors and types of ecosystems that they have helped grow via their Data Space Radar. There are a variety of different data spaces that are being built linking different sectors, and countries in Europe. When I asked Christophe Mertens Head of Adoption at IDSA where he saw the future of data ecosystems, he explained that his “vision is really that the understanding of data spaces will grow, amongst people. And from what I observe currently in Europe, the protectionist view that a lot of people in different endeavors take, I really hope we will be able to overcome that with the help of technical solutions” and “if you take this protectionist view and say we want a European solution, and be in competition with the rest of the world, rather than cooperating or collaborating. I think this will fail.” His “hope is that we have international solutions that bring together international companies, supply chains (so that) international data can flow freely, regardless of boundaries, if it follows, of course, the laws and rules” (Mertens, C., 2021).

I greatly appreciated having the opportunity to explore this topic on a number of levels. I was able to do so thanks to access to a wide range of academic publications as well as real-world experience from professionals working in the industry. During my interviews I was able to ask several questions about data ecosystems, in addition to just business models. Questions were asked about a variety of topics such as:

- who pays for transformation costs in data ecosystems
- risks/benefits for data orchestrators in ecosystems
- risks/benefits for data donors in ecosystems
- mechanisms for enforcing fairplay in data ecosystems
- specific mechanisms for ensuring value for all ecosystem participants

As these topics may be considered a bit off-topic for this thesis on business models in data ecosystems, I have excluded them from the official master's thesis. That being said, access to the research done on these topics will be granted to all interview participants and any other parties interested in this research. In addition, all interview participants have been given access to a list of annotated articles and academic publications that have been considered in the scope of this research project.

This research project will not conclude with the submission of this master's thesis. A modified version of this thesis will be published by my employer Visions in preparation for our application for funds from the European Commission to further research business models. Other large organizations such as MyData Operators, ANewGovernance and the

SkillsAlliance will be called upon to contribute to the publication of this “whitepaper” form of my thesis. The research done for this project will also be used to develop a working group with the French Ministry of Education as they will be building a variant of a skills data space for French youth in the next few years. I am also planning on publishing a series of articles based on the different industry topics presented in this paper, with contributions from industry specialists.

In conclusion, I share Christoph Mertens’ vision and hope that through research on business and governance models, building technical infrastructures and collaborative projects that we will be able to bridge the gaps between regional, national and international data sharing. This will enable us to have better technologies, transparency and will support the overall evolution of the date ecosystem economy as a whole.

References

Aaser, Mohammed, et al. *"Four Ways to Accelerate Creating Data Ecosystems | McKinsey." McKinsey & Company, 23 Nov. 2020,* *https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/four-ways-to-accelerate-the-creation-of-data-ecosystems*.

Abdulla, Ahmed, et al. *"Data Ecosystems Made Simple | McKinsey & Company." McKinsey, 8 March 2021,* *https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/tch-forward/data-ecosystems-made-simple*

Attard, Judie & Orlandi, Fabrizio & Auer, Sören. (2016). Data Value Networks: Enabling a New Data Ecosystem. 453-456. 10.1109/WI.2016.0073.

Boost 4.0 | Big Data for Factories. https://boost40.eu/. Accessed 29 Aug. 2021.

Boral, Austin, et al. *Data Collaboration for the Common Good Trust and Innovation Through Public-Private. Insight Report, World Economic Forum, 2019.*

Botelho, Bridget, and Stephen Bigelow. "What Is Big Data and Why Is It Important?" Target Tech, May 2021, https://searchdatamanagement.techtarget.com/definition/big-data.

Campbell, Patrick. Subscription Business Models: A Guide to Subscription Models for 2019. https://www.priceintelligently.com/blog/subscription-business-model. Accessed 29 Aug. 2021.

Casadesus-Masanell, R., & Zhu, F. (2013). Business Model Innovation and Competitive Imitation: The Case of Sponsor-Based Business Models.Strategic Management Journal, 34(4), 464-482. Retrieved August 2, 2021, from http://www.jstor.org/stable/23471019

De Saulles, M. (2021, July 27th). Personal communication [Interview].

Dexheimer, Maximilian, et al. "Four Strategies to Orchestrate a Digital Ecosystem." BCG, 4 Sept.2020, https://www.bcg.com/fr-fr/publications/2020/four-strategies-to-orchestrate-digital-ecosystem.

Ferro, Enrico & Osella, Michele. "Open Data on the Web" Workshop; April 2013, Google

Campus, Shoreditch, London, https://www.w3.org/2013/04/odw/odw13_submission_27.pdf

Firth, Robert. “The Pay-Per-Use Business Model Explained | LinkedIn.” Linkedin, https://www.linkedin.com/pulse/pay-per-use-business-model-explained-robert-frith/. Accessed 29 Aug. 2021.

Fricker, Samuel & Maksimov, Yuliyan. (2017). Pricing of Data Products in Data Marketplaces. 10.1007/978-3-319-69191-6_4.

Fruhwirth, Michael & Rachinger, Michael & Prlja, Emina. (2020). Discovering Business Models of Data Marketplaces. 10.24251/HICSS.2020.704.

GAIA-X - Home. *https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html*. *Accessed 29 Aug. 2021.*

Ghezzi, Antonio & Georgiades, Michael & Reichl, Peter & Le Sauze, Nicolas & Di Cairano-Gilfedder, Carla & Mangiaracina, Riccardo. (2013). Generating Innovative Interconnection Business Models for the Future Internet. Info. 15. 40-65. 10.1108/info-12-2012-0054.

HECK, ERIC VANBIG DATA AND DISRUPTIONS IN BUSINESS MODELS. Revista de Administraçãode Empresas [online]. 2019, v. 59, n. 6 [Accessed 6 August 2021] , pp. 430-432. Available from: <https://doi.org/10.1590/S0034-759020190608>. Epub 10 Jan 2020. ISSN 2178-938X. https://doi.org/10.1590/S0034-759020190608.

Henke, N. (2021, August 4th). Personal communication [Interview].

Hiter, Shelby. “What Is Raw Data? | Datamation | Unprocessed Data Uses.” Datamation, 4 May 2021, https://www.datamation.com/big-data/raw-data/.

IDS Components | International Data Spaces. https://internationaldataspaces.org/use/ids-components/. Accessed 29 Aug. 2021.

IDSA-MembershipFeeRegulations-2021-01.Pdf.

JoinData. https://join-data.nl/?lang=en. Accessed 29 Aug. 2021.

Lokitz, J. (2021, July 22nd). Personal communication [Interview].

Lokitz, Justin. Exploring Big Data Business Models - Business Models Inc. https://www.businessmodelsinc.com/big-data-business-models/. Accessed 29 Aug. 2021.

Mann, Sarb. “Data Exchange and Marketplace, a New Business Model in Making.” Linkedin, 26 Mar. 2018, https://www.linkedin.com/pulse/data-bazaar-new-business-model-making-sarab-s-mann-mba/.

Martin, Sébastien & Turki, Slim & Renault, Samuel. (2017). Open Data Ecosystems: Introducing the Stimulator Function. 6th International Conference on Electronic Government and the Information Systems Perspective EGOVIS 2017. https://www.researchgate.net/publication/317278860_Open_Data_Ecosystems_Introducing_the_Stimulator_Function/citation/download

Mertens, C. (2021, August 24th). Personal communication [Interview].

ODI-Data-Ecosystem-Mapping-—-Print-at-Home-Guide-A4-—-2019-06-26.Pdf.

Raw Location Data: Best Datasets & Providers 2021 | Datarade. https://datarade.ai/data-categories/raw-location-data. Accessed 29 Aug. 2021.

Redman, T. (2021, July 23rd). Personal communication [Interview].

Russo, Massimo, and Tian Feng. “Where Is Data Sharing Headed?” France FR, 7 Feb. 2021, https://www.bcg.com/fr-fr/publications/2021/broad-data-sharing-models.

Spiekermann, M. (2021, July 28th). Personal communication [Interview].

Spiekermann, Markus. “Data Marketplaces: Trends and Monetisation of Data Goods.” Intereconomics, vol. 2019, no. 4, 2019, pp. 208–16, https://www.intereconomics.eu/.

“The Different Ways Businesses Can Use Raw Data.” CoFounder, 3 Feb. 2021, https://www.cofmag.com/2021/02/different-way-use-raw-data/.

“The True Cost of Raw Location Data.” Gravy Analytics, 1 Sept. 2020, https://gravyanalytics.com/blog/the-true-cost-of-raw-location-data/.

“Time Based Pricing.” Topio Network, https://www.topionetworks.com/topics/time-based-pricing-50aa0ea13ae4b249e60179dd. Accessed 29 Aug. 2021.

Wessels, B., Finn, R., Wadhwa, K., Sveinsdottir, T., Bigagli, L., Nativi, S., & Noorman, M. (2017). Big
data, open data and the commercial sector. In Open Data and the Knowledge Society (pp. 159-174). Amsterdam: Amsterdam University Press. Retrieved August 2, 2021, from http://www.jstor.org/stable/j.ctt1pk3jhq.13

"What Is a Data Ecosystem?" | Cognizant. https://www.cognizant.com/us/en/glossary/data-ecosystem. Accessed 29 Aug. 2021.

"What Is Data Exploration?" Sisense, https://www.sisense.com/glossary/data-exploration/. Accessed 29 Aug. 2021.

Wolfert, S. (2021, August 12th). Personal communication [Interview].