POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Multi-sided non transactional platforms in healthcare: Value Capture from Real World Data

TESI DI LAUREA MAGISTRALE IN
MANAGEMENT ENGINEERING
INGEGNERIA GESTIONALE

Author: **Filippo Palmieri**

Student ID: 10800397
Advisor: Emanuele Lettieri
Co-advisor: Alessandro Carrera
Academic Year: 2021-22

Abstract

Background: Due to the enormous amount of health data generated, and with the advent of new technologies, Real-World Data (RWD) are increasingly being used by healthcare companies to develop new products and treatments. In this context, multi-sided non transactional platforms are emerging: on one side, they are collecting RWD from patients; on the other sides, they are manipulating and sharing the RWD collected to pharma and tech companies, research institutes, and healthcare operators. This research aims at investigating how multi-sided non transactional platforms are capturing value through Real-World Data in the healthcare industry.

Materials and Method: A multiple case study method was used to address the research question mentioned above. Four private companies were studied thanks to the triangulation of interviews, documentations and direct observation. Then, the coding methodology from Grounded Theory was used to test the hypotheses developed from the research model built on the Systematic Literature Review conducted.

Results: The results obtained from the research support the hypothesis developed. It was found that multi-sided non-transactional platforms are mainly leveraging on Secondary Sources of RWD as a key value driver, while having to develop different value propositions for the end users and business sides. Moreover, the dominant monetization model was found to be strongly skewed towards the B2B sides. Lastly, the development of a closed ecosystem was found to be the best way to ensure health data privacy and security.

Discussion: Platform providers willing to create and capture value through Real-World Data in the healthcare industry should focus on the collection of Secondary Source of RWD, being able to create and sustain different value propositions for each of the side affiliated to the platform. It is also important to build internal technological capabilities to unlock the power of RWD, while also preserving high standards for Privacy and Security.

Key-words: multi-sided non transactional platforms, Real-World Data, RWD, business model, value creation, healthcare, Life Sciences

Abstract in italiano

Background: Grazie alla considerevole quantità di dati sanitari generati e con l'avvento delle nuove tecnologie, i Real-World Data (RWD) sono sempre più utilizzati dalle aziende sanitarie per sviluppare nuovi prodotti e trattamenti. In questo contesto, stanno emergendo le cosiddette piattaforme multi-sided non transazionali: da un lato, esse raccolgono RWD dai pazienti; dall'altro, esse rielaborano e condividono i RWD raccolti con aziende farmaceutiche e tecnologiche, istituti di ricerca e operatori sanitari. Questa ricerca si propone di indagare come le piattaforme multi-sided non transazionali stiano creando valore attraverso i Real-World Data nel settore sanitario.

Materiali e metodo: Per rispondere alla domanda di ricerca definita sopra, è stato utilizzato il metodo dei casi studio multipli. Sono state studiate quattro aziende private grazie alla triangolazione di interviste, documentazioni e osservazioni dirette. Successivamente, è stata utilizzata la metodologia di coding della Grounded Theory per testare le ipotesi sviluppate grazie al modello di ricerca costruito sulla base dell'analisi della letteratura condotta.

Risultati: I risultati ottenuti dalla ricerca supportano le ipotesi sviluppate. È emerso che le piattaforme multi-sided non transazionali fanno principalmente leva sulle fonti secondarie di RWD come principale driver di valore, pur dovendo sviluppare proposte di valore diverse per gli utenti finali e per le aziende. Inoltre, il modello di monetizzazione dominante è risultato fortemente sbilanciato verso il lato B2B. Infine, lo sviluppo di un ecosistema chiuso è risultato essere il modo migliore per garantire la privacy e la sicurezza dei dati sanitari trattati.

Discussione: I fornitori di piattaforme che intendono creare e catturare valore attraverso i Real-World Data nel settore sanitario dovrebbero concentrarsi sulla raccolta di fonti secondarie di RWD, essendo in grado di creare e sostenere diverse proposte di valore per ogni lato affiliato alla piattaforma. È inoltre importante costruire capacità tecnologiche interne per estrarre massimo valore dai RWD, pur mantenendo elevati standard di privacy e sicurezza.

Parole chiave: piattaforme multi-sided non transazionali, Real-World Data, RWD, business model, creazione di valore, healthcare, Life Sciences

Contents

1 Executive Summary

1.1. Introduction

Today, life sciences companies are navigating unprecedented times: the collision of scientific progress, technology disruptions, and innovation is having tremendous potential to improve patients' lives and create corporate value. On one hand, thanks to the advent of new technologies like AI and biosensors, companies can collect and extract value from an increasing number of patients' health data. On the other hand, multi-sided platform business models are emerging in the industry as a way to aggregate and collect patients' health data, creating new value for patients and the industry as a whole.

The aim of this thesis is to provide a comprehensive view on the impact of Real-World Data on the Life Sciences sector, by investigating on the business model of multi-sided non-transactional platforms and on the way in which they create value.

More in particular, the main research question pursued is to describe how multi-sided non-transactional platforms are allowing the collection and valorization of Real-World Data, with a specific focus on the following three dimensions:

- The types of Real-World Data used, and the opportunities and challenges brough by such data.
- The technologies used to collect and valorize Real-World Data.
- The business model characteristics through which multi-sided non-transactional platforms are creating value in the industry.

Starting from these reasonings, a model has been developed to address the research question by also leveraging on multiple qualitative case studies.

1.2. Theoretical Background

Real-World Data (RWD) are data relating to patients' health status and/or the delivery of health care collected from a variety of sources and allow the generation of evidence on the effectiveness of new products, therapeutics, and medical devices.

Real-World Data can be classified into two categories, depending on their source: Primary, which are collected specifically for a clinical study (like patients' surveys); Secondary, which are collected for other purposes (like EHR's data).

Multi-sided non-transactional platforms, instead, are platforms with two or more affiliated sides, where the relationship among the sides is not transactional. As example, a platform where the platform provider supplies users' data to another affiliated sides is a non-transactional platform, as the sides do not directly and transactionally interact with each other.

The four business model constituents of multi-sided non-transactional platforms, which have been studied in this Thesis, are the following:

- Value creation: how the platform creates and captures value, the role of network effects, and value propositions.
- Sides management: number and type of affiliated sides, the incentive mechanisms and relationships among the sides.
- Revenue model: pricing models and differentiation, subsidization strategies.
- Governance and control: degree of openness and control over the platforms' assets, activities and stakeholders.

At the intersection of RWD and the platforms under study, the Systematic Literature Review conducted on 655 papers sheds light on the current status of the managerial research, along with the research gaps in academia.

Firstly, the most advanced technological and business applications of RWD on multi-sided non-transactional platforms have been found in the literature to serve multiple purposes, which can be divided in two main categories:

- Disease diagnosis, treatment, and drug development. Under this purpose, the platforms allow to leverage on Real-World Data to understand the evolution of certain diseases and develop new treatments.
- Improving health research. Under this purpose, platforms and Real-World Data are means to improve the processes and outcomes of health research.

Secondly, along with above mentioned applications, Real-World Data also pose some managerial challenges and concerns for platform providers. Among all, privacy and ethical issues are the most pressing. With the rise of cyberattacks, and with the vulnerabilities of the existing technologies, managerial operators must design their platforms with the highest data security standards, complying with the existing regulations on the subject matter.

However, despite the richness and depth of the research at the intersection between RWD and platforms, from the Literature Review it seems that the academia focused more on the study of such platforms in research and public contexts. A research gap exists therefore on the application of such platforms in business contexts, to understand how such platforms are helping Life Sciences stakeholders to collect and valorize Real-World Data.

Using as a main theoretical framework the theory of platforms and their business models, this thesis has the objective of addressing the gap found.

1.3. Research Model and Hypotheses

Integrating all the theoretical contributions on Real-World Data and multi-sided non-transactional platforms, a theoretical framework has been built as to address the research question under study.

The figure below represents the key components and subcomponents of the theoretical framework, that organizes and links together the areas of investigation for the empirical and qualitative research.

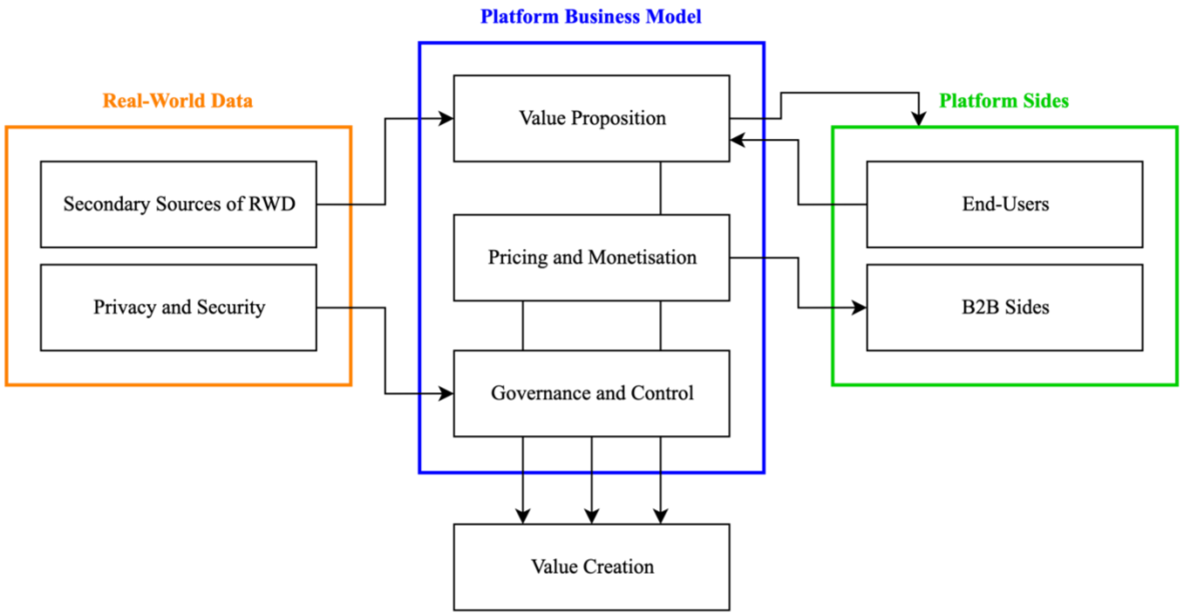


Figure 1.1: Research Model.

The model proposed has three main constituent macro areas: Real-World Data, Platform Business Model, and Platform Sides. At the center of the framework, the Platform Business Model macro area is the main component: a business model in fact allows a platform to create and capture value through the orchestration of external inputs (Real-World Data) and stakeholders (Platform Sides). The framework developed served as a reference to structure and conduct the qualitative analysis, and to finally develop the research hypothesis below.

Table 1.1: List of Hypotheses.

Hypothesis ID	Description
H1	Multi-sided platforms use mainly Secondary Sources of RWD as part of their value proposition, as Secondary Sources of RWD are major enablers for value capture.
H2	Multi-sided platforms in the Life Sciences industry need to formulate two different value propositions – one for the end-user side and one for the business sides.
H3	In multi-sided platforms in the Life Sciences sector, end-users are part of the value proposition for business customers.
H4	In multi-sided platforms in the Life Sciences sector, the monetization of the business model is B2B oriented.
H5	Due to Privacy and Security concerns, multi-sided platforms in the Life Sciences sector develop a closed ecosystem where access is restricted.

The above hypotheses have been tested for validity through the research methodology outlined in the following section.

1.4. Research Methodology

In order to address the research question of this thesis, and given the peculiarities of the topics under study, the research methodology chosen is of qualitative and descriptive nature.

More in particular, the research methodology chosen is the exploratory multiple-case study. According to Yin, a case study is "an empirical enquiry that (i) investigates a contemporary phenomenon within its real-life context, especially when (ii) the boundaries between phenomenon and context are not clearly evident."

The multiple-case study has been conducted on 4 multi-sided non-transactional platforms, chosen to be part of the sample through the principle of purposeful sampling. The companies are:

- Evidation Health
- MedM
- Withings
- Elysium

For the data collection part, the data triangulation principle was followed to enhance the accurateness of the results. Therefore, multiple sources of evidence were used to evaluate the hypotheses and conduct the exploration on the four companies involved: interviews, direct observation, and documentation.

The sources of evidence were analyzed through the Grounded Theory principle, following therefore the three coding steps required. First, open coding was performed to derive first order concepts with the lowest level of abstraction. Second, axial coding was used to group the first order concepts found through open coding under more abstract second order concepts or categories. Third, selective coding was the last step consisting of connecting all the categories created during axial coding to their relative core categories. The core categories defined in this process are the ones representing the pillars of the research and its main contributions to the literature.

1.5. Results

The results analysis has been structured as to both describe the outcomes of the interviews and test the hypotheses developed.

Through coding, three core categories have emerged for the interviews analysis: Real-World Data, Technologies used, and Business Model components.

On Real-World Data, it is worth mentioning that all the multi-sided platforms involved in the research are mainly leveraging Secondary Sources of Real-World Data, and in particular Patient Generated Health Data. This is mainly due to the fact that the Real-World Data collected by these platforms should serve multiple research and business purposes.

On the technologies used to collect and valorize Real-World Data, the coding results show that all the companies interviewed are leveraging on mobile apps as a means for collecting Real-World Data, where users of the platform can input or share their Health Data through wearable integrations, documents scanning, or other upload methods. On the other hand, AI is the most utilized technology to valorize Real-World Data collected, as it’s a means to draw insights and find patterns among the considerable amount of health data gathered.

On the business model side, results were found for all the different business model components in scope of the research model. Starting from value propositions, it is important to mention that all the platforms involved had to develop and sustain different value propositions for the B2C and B2B sides, as these side types have different needs to be satisfied. Moreover, end-users’ health data have found to be an important value driver comprised in the value proposition offered by platforms for the B2B sides. Another interesting finding was the ability of platforms to lessen the need for the development of network effects by licensing out their platform and technology to B2B sides who can use it for other purposes outside of the platform itself. Regarding pricing, it has been found that the main monetization source comes for the B2B sides, who are intrinsically more willing to pay for the value taken out of the platform affiliation. This, in some cases, manifests through a cross-subsidization strategy, where the platform's service is offered to end users for free, being subsidized by B2B sides who pay for the service obtained. Finally, regarding the governance component of the platform's business model, the degree of openness of the platform was an area that provided a major insight: all the companies studies have developed a closed ecosystem, where the data privacy and security is prioritized through consent mechanisms and access constraints. The Data Privacy and Ethical issues attached with Real-World Data have been found to be major drivers of this business model decision.

Summing up all the coding results, it is shown that all the five hypotheses developed in the research model are supported, as the codes representative of these hypotheses were present in the interviews of all the platforms involved in the study.

1.6. Theoretical and Managerial Contributions

The study conducted presents various theoretical contributions and managerial implications.

Given the lack of papers on private for-profit multi-sided platforms found in the literature, this research provides for additional theoretical concepts in the field, given the hypothesis supported. First, thanks to the testing of hypothesis H1, it was possible to prove that multi-sided platforms mainly use Secondary Sources of Real-World Data as part of their value proposition, especially Patient Generated Health Data. These sources have found to be the major enablers of value creation for the companies studied. Second, thanks to the testing of hypothesis H5, it was also proved that when faced with a trade-off of keeping the platform opened or closed, these types of platforms tend towards building closed ecosystems, in order to better protect the Privacy and Security of the Real-World Data collected.

Along with the theoretical contributions, managerial implications were also developed. Given the importance of Secondary Sources of Real-World Data for value creation, platform providers should invest their resources towards collecting these types of health data, considering integrations with wearable technologies that can automate the collection of PGHD. Moreover, platform providers should also invest in developing strong technological capabilities inside the company, given the importance of mobile apps and AI found for the collection and valorization of Real-World Data. On the business model capabilities, platform providers should be able to develop and sustain different value propositions to attract both B2C and B2B sides, considering that the value proposition for the B2B sides may contain end user's data as a main source of value. In order to sustain the business and monetize the model, the pricing should be skewed towards the B2B sides, in some cases considering a cross-subsidization strategy that can allow to offer the service to B2C users for free and hence growing the user base quickly. Lastly, platform providers should develop a secure infrastructure that allows to protect the health data collected from patients, as Privacy and Security are one of the key concerns in the field.

1.7. Limitations and Future Research

Despite the previous considerations, there are additional insights coming from the results obtained that are worth being explored by future research. Moreover, within the present study, we also identified intrinsic weaknesses that suggest eventual next developments and improvements.

Starting from the findings, a future study should try to address not only the role of private companies as platform providers, but also the role of public or non-profit platform providers used by governments and other public entities to leverage on Real-World Data. During the Literature Review, in fact, many applications of multi-sided non-transactional platforms in the public sector have been found, and the study of them could bring further insights on the way in which they have structured their business model to create value for the stakeholders involved.

Moving to the limitations of the study, a first weakness can be found in the way the sampling has been conducted. Even if purposeful sampling was adequate to conduct the multiple case study because of its focus on finding information rich case for in-depth study, its major weakness is the presence of bias in the sampling due to the lack of randomness. This bias can limit the generalization of the results found.

Another limitation of the study is related to the interview process. Even if I, the researcher, have been careful in not biasing the respondents with the questions made during the interviews, some degree of bias may still be present. The bias from the interviews may impact the quality of the responses received.

Moreover, the focus on exploratory and qualitative methods brought a lack of quantitative methods and causal relationships. This lack can harm the external validity of the findings, and their general application to the whole population of multi-sided non-transactional platforms in the healthcare sector.

Finally, the last limitation can be detected in the instrument adopted to recruit interviewees for the empirical interviews. Indeed, the channel used, LinkedIn, has some intrinsic characteristics that on one side can been seen as potentialities but, on the other, might limit the possibility to generalize the results obtained. It is possible to detect a sort of homogeneity in the backgrounds and roles of the respondents.

2 Introduction

2.1. Research Objective

As a student of Management Engineering and a passionate about the Life Sciences sector, I decided to merge the two topics into a research in which I will try to understand what are the new technological trends in the industry, and how are they impacting the business processes of corporations and start-ups active in the field.

One major innovation trend within the Life Sciences sector is the advent of new technologies that are unlocking the true potential of big data, which can be leveraged by companies to conceive and market innovative products, with the aim of improving patients' health situations.

This new trend is called Real World Data, and I would like to understand how it is shaping the future of the Life Sciences industry, and how are firms embracing it to deliver new value for patients.

More in particular, I would like to focus my attention at the intersection of Real-World Data (RWD) and platforms, trying to understand how multi-sided non-transactional platforms are allowing the collection and valorisation of Real-World Data in the Life Sciences sector, favouring the development of innovative medical products or treatments. In doing that, the objective of this research is to describe, through a qualitative study, how are such platforms creating value for all the stakeholders in the Life Sciences sector through Real-World Data, with a particular focus on:

- The types of Real-World Data used, and the opportunities and challenges brought by such data.
- The technologies used to collect and valorise Real-World Data.
- The business model with which such platforms are creating value in the industry.

To achieve this objective, I will start by defining what are Real-World Data and the Real-World Evidence that is derived from the former, using definitions to let the reader understand why they are so powerful, but also which are the attached challenges.

Consequently, I will investigate on which are the new technologies that today allow and facilitate the extraction of Real-World Data, like wearables, IoMT, and AI among all. The advent of those new technologies is the driver unlocking the true potential of Real-World Data usage.

Finally, I will leverage on the two previous points to address the main focus of the research. I will describe in which way multi-sided non-transactional platforms are allowing the collection and valorization of Real-World Data, favouring the development of innovative medical products or treatments.

2.2. Research Question

The main objective of this research is to analyse the impact of Real-World Data on the Life Sciences sector.

More in particular, the main focus of my research is to investigate on how multi-sided non-transactional platforms are valorising Real-World Data, allowing the main stakeholders in the Life Sciences sector to access and use them to develop innovative medical products or treatments.

Stemming from what was stated above, the main scientific and technological question that I would like to address within my study is: how multi-sided non-transactional platforms are allowing the collection and valorisation of Real-World Data, creating and capturing value in the healthcare sector.

The research question will be addressed by analysing three main components:

- Real-World Data: the types and sources of RWD treated by such platforms
- Technologies: the technologies used to collect and valorise the RWD treated
- Business model: the business model components that allow the creation of value in the healthcare sector.

In pursuing the main research question, therefore, many complementary research questions will be faced within the purpose of the research:

- What are "Real-World Data" (RWD) and Real-World Evidence (RWE), and how are they impacting the Life Sciences industry?
- How multi-sided non-transactional platforms are allowing the collection and valorization of Real-World Data, to favour the development of innovative medical products or treatments from Life Sciences companies?
- Which are the new emerging technologies that those platforms are using, that are allowing and facilitating the collection and elaboration of Real-World Data?
- How is the business model of such platforms structured as to create value in the Life Sciences ecosystem?

To address such questions, I will first define the context and make clarity on the definitions, and therefore in the first part the aim will be to understand what we mean by Real-World Data and Real-World Evidence, why can they be beneficial for companies and how can they spur innovation in the Life Sciences sector. In that section, the main sources of Real-World Data will be defined, along with the opportunities and challenges that Real-World Data bring to the Life Sciences sector.

Another key point would be then to define and study the leverages that platforms can use to unlock the full potential of RWD: the new technologies. The aim will be to understand the new technological advancements that are allowing the collection and elaboration of RWD, and most importantly which is the role of each technology with respect to the topic. The main opportunities and challenges of such technologies will be presented.

To finally close the loop, and truly achieve the objective of the research, I will look at how RWD and the related technologies are applied in the real context by multi-sided non-transactional platforms to allow the collection and valorisation of Real-World Data. The aim is to understand how such platforms operate, and which are the characteristics and peculiarities of their business models.

All the above concepts will serve the main research question of my thesis, which is to describe how multi-sided non-transactional platforms are allowing the collection and valorisation of Real-World Data, analysing the types of data they use, the technology they leverage, and the business model through which they create value for the entire Life Sciences ecosystem.

Table 2.1: Research question overview.

Research Question	Components	Sub questions	Variables
How multi-sided non-transactional platforms are allowing the collection and valorization of Real-World Data, creating and capturing value in the healthcare sector?	RWD	Which are the types of RWD collected?	Sources of RWD
	Technologies	How are RWD collected and valorized?	Technology categories
	Business Model Components	How are platforms creating and capturing value?	Value Propositions
			Monetization
			Sides Management
			Governance

2.3. Research Background

The Life Sciences industry comprises companies operating in the research, development and manufacturing of pharmaceuticals, biotechnology-based food and medicines, medical devices, biomedical technologies, nutraceuticals, cosmeceuticals, food processing, and other products that improve the lives of organisms.

Today, Life Sciences companies are navigating unprecedented times: the collision of scientific progress, technology disruptions, and innovation is having tremendous potential to improve patients' lives and create corporate value. Those organizations have recognized that the digital revolution is presenting vast opportunities: they are in fact onboarding top executives from leading digital companies and are collaborating with start-ups to enhance the value coming from technology.

The digitalization trend has also brought another key dimension: big data and data analytics. With an unprecedented amount of information available to Life Sciences organizations, in fact, big data is playing an increasingly important role in shaping the future of the industry. On one hand, new technologies like AI and Cloud Computing are enhancing the ability of companies to store and process huge amounts of data, allowing them to finally unleash the potential of big data. On the other hand, new devices like biosensors and wearables are giving access to an additional and extremely valuable source of health data: real-time patient-generated data.

All the above factors, coupled with an increasing acceptance of data-driven approaches from regulatory authorities, are creating the right context for Real-World Data to flourish.

In this context, multi-sided platform business models are emerging as a way to valorize Real-World Data coming from patients, in such a way that companies in the Life Sciences sector can use them to accelerate the development of innovative medical products or treatments. Platforms, in fact, are emerging as a player who can aggregate and collect different types of Real-World Data, and then manipulate them in a way that allows for Real-World Data to be truly valorized by other Life Sciences companies.

The current research, as of now, is very much fragmented on the topic: there is no comprehensive research on how multi-sided platforms are allowing the collection and valorization of Real-World Data, covering the types of RWD used, the technologies leveraged, and the business model characteristics of such platforms. This is in part due to the fact that the discipline is new, and the trend of Real-World Data in Life Sciences is seeing its main developments in the recent period.

With this Thesis, I want to deliver a comprehensive and holistic view of the impact of Real-World Data on the Life Sciences sector, by investigating the characteristics of multi-sided non-transactional platforms and the way in which they create value.

2.4. Research Significance

Regarding the theoretical significance, I think that my research will bring the following:

- A comprehensive study on Real-World Data impact on the Life Sciences sector, which comprises the newest technologies and the types of Real-World Data used in real settings. This work will be useful to anyone who would like to have an holistic view on the industry trends, in relation to the topic of Real-World Data.
- The research will offer also a rich portion of case studies and empirical data, coming from the research methodology I have chosen. With the support of Tongji University and the Life Sciences Innovation Observatory of my home university, I have developed a database of multi-sided non-transactional platforms in the Life Sciences sector, and I have also created multiple case studies. The empirical nature of my study adds significance to the research, as it allows to study what is happening in the industry, directly from the companies involved.

Moreover, my research will also bring the following application values:

- Companies within the Life Sciences sector will be able to have a comprehensive view of the Real-World Data impact on their industry, that will allow them to understand the current trends and how to embrace some of them. In other words, it can favour innovation in the Life Sciences sector.
- Multi-sided platform providers will be able to understand the business model characteristics of current players, which are the types of Real-World Data valorized, and which are the technologies used.
- Finally, I strongly think that this research will benefit also new start-ups in the landscape, who would like to enter in the market with new innovative products that leverage Real-World Data to create value.

3 Literature Review

3.1. Systematic Literature Review

3.1.1. Research Methodology

As the first step of this research, I have conducted a Systematic Literature Review to further explore the concept of Real-World Data in the context of multi-sided non-transactional platforms. This review aims at uncovering the academic achievements in this field, along with the areas not yet covered by current researchers. In this section, transparency is made on the process and steps followed to conduct the literature review.

The Systematic Literature Review was conducted based on principles outlines in Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA), which I used to transparently report the process followed and the findings.

To conduct the literature review, the Scopus database was used as a source to find abstracts and articles relevant to the topic investigated. The keywords used to query the relevant articles have been the following: Real-World Data; Big Data; Healthcare; Multi-Sided Platform.

The detailed query used on Scopus also contains some variations of the above keywords, aiming at obtaining a more comprehensive result that covers also possible variations in terms and academic definitions used by researchers. For example, the following variations have been included in the Scopus query for the Multi-Sided Platform keyword: Multisided Platform; Platform; Multi Sided Platform; Multi-sided Platform; Non Transactional Platform; Non-transactional Platform; Nontransactional Platform; Orthogonal platform.

Additionally, in order to gather a broad yet up-to-date collection of articles, the research has been limited to papers published within the time horizon starting from the beginning of 2010 to the latest available of 2020. Some further limitations have been applied as to collect papers written only in English and related to the following subject areas: Engineering; Social Sciences; Decision Sciences; Business, Management and Accounting; Economics, Econometrics and Finance.

With all the above parameters, the final Scopus query used is the following:

TITLE-ABS-KEY ((("multisided platform") OR (platform) OR ("multi sided platform") OR ("multi-sided platform") OR ("non transactional platform") OR ("non-transactional platform") OR ("nontransactional platform") OR ("orthogonal platform")) AND ((health) OR (healthcare) OR (health-care) OR ("life science*")) AND (("big data") OR ("real-world data") OR ("RWD") OR ("real world data"))) AND (LIMIT-TO (PUBSTAGE , "final")) AND (LIMIT-TO (SUBJAREA , "ENGI") OR LIMIT-TO (SUBJAREA , "SOCI") OR LIMIT-TO (SUBJAREA , "DECI") OR LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "ECON")) AND (LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "ch") OR LIMIT-TO (DOCTYPE , "re") OR LIMIT-TO (DOCTYPE , "bk")) AND (LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012) OR LIMIT-TO (PUBYEAR , 2011) OR LIMIT-TO (PUBYEAR , 2010)) AND (LIMIT-TO (LANGUAGE , "English"))

The query used above returned 655 results.

After the definition of the Scopus query, some refinements and screenings have been made to the results to derive the final list of papers selected to be part of the Literature Review.

The first refinement has been made on papers' sources, with the aim of selecting only the sources with a certain degree of theoretical robustness and reliability, and articles that are published in top journals. For this step, the robustness of the source has been identified through the Shimago Institution Rankings (SJR), a publicly available portal that includes the journals and country scientific indicators developed from the information contained in the Scopus database. After this first refinement, 487 papers have been excluded, as per Table 3.1 below.

Table 3.1: Exclusion criteria after the first literature screening.

Exclusion Criteria	# of Documents
Source's Ranking	162
Source Not Ranked	160
Source Not Present	165
Total	**487**

The exclusion criteria adopted in this refinement were the following:

- Source's Ranking: The SJR ranking for a specific source is given in Quartiles, ranking from Q1 to Q4, where the former is the highest ranking. Per this criterion, the sources ranked in Q3 and Q4 were excluded.
- Source Not Ranked: The specific source was present in the SJR database but was not ranked for any particular reason.
- Source Not Present: When searched on SJR database, the source was not found as not present in the database.

Out of the remaining papers, the second refinement was conducted by reading all the abstracts to identify the interesting articles and excluding the ones which are not relevant for the research's purpose. After this second refinement, 98 papers have been excluded, as per Table 3.2 below.

Table 3.2: Exclusion criteria after the second literature screening.

Exclusion Criteria	# of Documents
Platform Characteristics	22
Technical Focus	27
Out Of Scope	11
Data Characteristics	23
Industry Characteristics	15
Total	**98**

The exclusion criteria adopted in this refinement were the following:

- Platform Characteristics: the characteristics of the platform involved in the study or being studied were not aligned with the boundaries and definition of multi-sided non-transactional platforms.
- Technical Focus: the paper was more focused on the science and technicalities behind the platform being studied, rather than focusing on its managerial and business implications.
- Out Of Scope: this exclusion criterion encompasses multiple possible reasons for which the paper was not falling inside the scope of the work being conducted in this research. As example, the paper objective and its research question were not compatible with my research.
- Data Characteristics: the characteristics of the data being used for the study were not aligned with the boundaries and definition of Real-World Data or was not health data.

- Industry Characteristics: the industry where the platform under study was operating were not within the boundaries of the Life Sciences and Healthcare sectors.

After the two refinements outlined above, 70 papers have been selected for a deeper analysis. These papers have been fully read to derive the final selection of articles to be included. After such deeper analysis, a total of 39 articles have been selected, while 31 articles have been excluded.

Table 3.3: Exclusion criteria after the final literature screening.

Exclusion Criteria	# of Documents
Narrow Perspective	6
Technology description	13
Out Of Scope	9
Real-World Data usage	3
Total	**31**

The exclusion criteria adopted in this refinement were the following:

- Narrow Perspective: the paper was focused on a too-narrow perspective within the topic addresses. For example, the article was focused only on privacy implications for the use of Real-World Data in business contexts.
- Technology description: the articles were focused mainly on describing the big data technology beyond the platform under study.
- Out Of Scope: even if with the abstract analysis the articles seemed to be in scope, with a deeper analysis of the entire paper it turned out that it was out of scope.
- Data Characteristics: the actual data used by the platforms under study were not falling in the scope of Real-World Data, even if Real-World Data were at least mentioned within the text and abstract.

After the final refinement, the articles selected to be included are 39. Below is visually described the funnel used to select the articles to be included in this research.

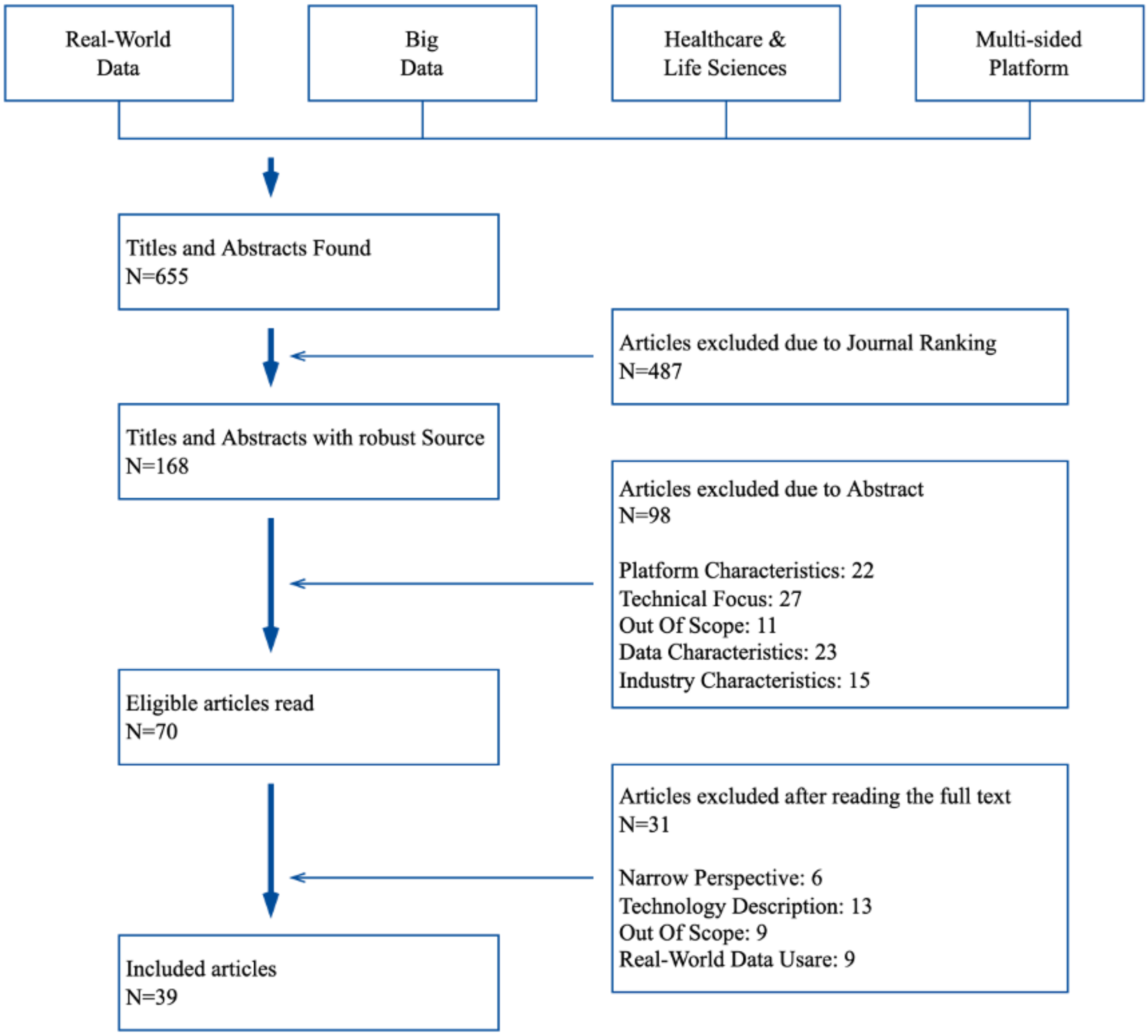
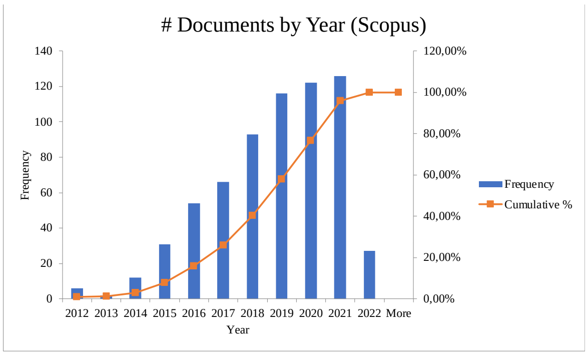


Figure 3.1: Funnel of the selection process for the Literature Review.

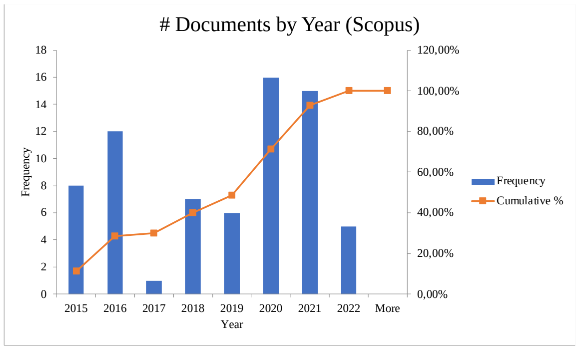
3.1.2. Descriptive Analysis

By looking at the publication years for the articles involved in the research, we can identify the trends related to the topic of interest along with the differences among the articles found with the first Scopus research and the articles selected to be fully read.

By looking at Figure 2, we can see a strong and constant increase in the interest in Real-World Data and Multi-Sided Platforms in the Life Sciences sector starting from 2013 onwards. It is important to mention that the data for 2022 are limited to the month of March, hence we can confidently assume that the positive trends may continue also for the current year. This positive trend is also partly seen in the analysis of publication years for the articles selected to be fully read after all the refinements made (Figure 3.2a), where the years 2020 and 2021 return the highest frequency of papers in the selection. However, the frequency of articles published in 2017, 2018, and 2019 is lower than expected given the macro trend found in Figure 3.2b.



3.b(a) Yearly distribution of the 585 selected articles.



3.2(b) Yearly distribution of the 70 selected articles.

The orange line in both the above images indicates the cumulative percentage, showing how each year contributes to the total number of documents by year. This metric was inserted to show the trends of documents published for each cohort of papers.

The second descriptive analysis has been done on the keywords' frequency, where Image 3.3 below presents the frequency distribution of the main keywords shared by the authors of the 70 selected papers.

As can be seen, the most frequent keywords reflect the central role of Big Data in Healthcare, and also the importance of use of the latest technologies to enable value capture from such health data. Interestingly, the keyword "COVID-19" appears to be one of the main keywords mentioned in the selected papers, given the importance of the topic in academia since the pandemic started, and given the relevancy of the topic in the context of this research which covers the healthcare sector and health data.

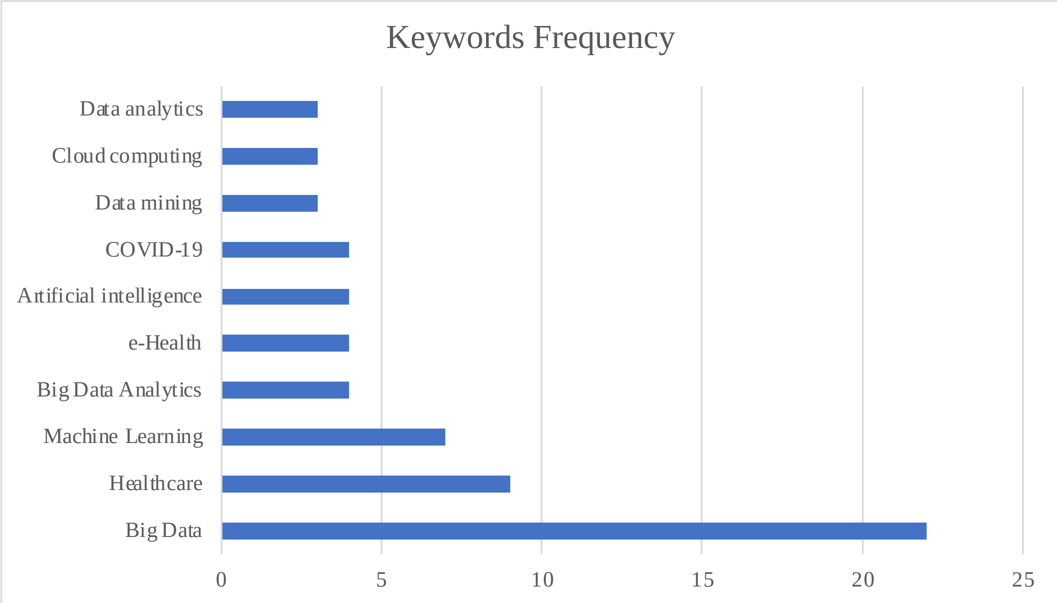


Figure 3.3: Frequency distribution of the main keywords in the 70 selected papers.

3.2. Real-World Data and Real-World Evidence: definitions and sources

In this section, I will highlight the literature review on the definitions of Real-World Data and Real-World Evidence, and on the sources of Real-World Data.

3.2.1. Definitions

At the core of the Thesis, there are two critical concepts: Real-World Data and Real-World Evidence, which as we will see are strongly interconnected.

According to the U.S. Food and Drug Administration [3], “Real-World Data (RWD) are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources.”

Going in more depth with the literature review, however, consensus on the precise definition of RWD and its sources is lacking [4], and a certain degree of disparity remains among different stakeholders when it comes to thoroughly define RWD [5].

A study conducted by Amr Makady et al. [4], based on a literature review and stakeholders interviews, differentiates the various definitions of RWD into four main categories:

1. Data collected in a non-RCT setting (i.e., all health data except those collected in the setting of a conventional RCT setting): where RWD are

defined as "data used for decision-making that are not collected in conventional RCTs" [6].

2. Data collected in a non-interventional/non-controlled setting (i.e., data collected without interference with treatment assignment, and/or patient monitoring/follow-up, and/or selection of study population): where RWD are defined as "observations of effects based on what happens after a prescriptive (treatment) decision is made where the researcher does not, or cannot, control who gets what treatment and does not, or cannot, control the medical management of the patient beyond observing outcomes" [7].
3. Data collected in a non-experimental setting (i.e., in a setting in which the investigator has no control over any of the conditions and not de novo data collection occurs on the basis of a pre-established study protocol): where RWD are defined as "data that are not collected under experimental conditions, but data generated in routine care" [8].
4. Others (i.e., none of the aforementioned).

The emergence and availability of Real-World Data, coupled with the need to find new ways of generating evidence on the effectiveness of medical treatment in addition to conducting Randomized Controlled Trials (RCTs), has favoured the development and growth of Real-World Evidence.

According to the U.S. Food and Drug Administration [9], Real-World Evidence is "the clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of Real-World Data."

Real-World Evidence, therefore, is evidence obtained form Real-World Data, and is aimed at providing a more comprehensive understanding of how a new therapeutic option will work in the "real world".

Going further with the literature review, we can better define Real-World Evidence as "a form of evidence (along with RCTs, health economics studies, etc.) generated to answer a question or test a hypothesis. It is derived from primary or secondary real-world data (RWD) sources, often data from computerized medical record (CMR) systems. It includes appropriate and rigorous design and analyses, generally set out in a protocol in advance of conducting the study. It provides evidence about patient populations' diseases, medicines, and health care that will inform clinical practice. It generates further research questions." [10]

Another definition describes Real-World Evidence as "the technology-facilitated collation of all routinely collected information on patients from clinical systems to a comprehensive, homogeneously analyzable dataset (big data) that reflects the treatment reality in the best possible and comparable manner." [11]

From the various definitions of RWE, we can appreciate its differences from RCTs. In general, RWE studies are complementary to RCTs in the generation of scientific evidence, but they also overcome some of the limitations of RCTs. One of the major weaknesses of Randomized Controlled Trials is their lower generalizability and external validity, as the inclusion and exclusion criteria of RCTs for patients participating in clinical trials usually create idealized conditions which are often not generalizable. Following their definition, RWE studies provide insights into the routine clinical setting, and hence may benefit from greater generalizability and external validity compared to RCTs [12].

RCTs are also relatively time and resource-intensive, while RWE studies have the promise of being conducted significantly faster and more resource-efficiently once the necessary structures have been established in the centers and institutions.

Despite that, RCTs remain until now the gold standard for the generation of clinical evidence, but Real-World Evidence studies can yield important additional insights into research and clinical care, and Real-World Data are increasingly being adopted by companies in the Life Sciences sector.

3.2.2. Sources of Real-World Data

Focusing on the Real-World Data, it is important to know which are the sources through which they can be collected, or in other words where can RWD be extracted and found.

First, we have to distinguish between two types of Real-World Data: primary Real-World Data and secondary Real-World Data. The former can be defined as Real-World Data that are collected specifically for the RWE study. The latter, instead, can be defined as Real-World Data that are collected for other purposes and studies, but used for the RWE study. [10]

Stemming from the two definitions made above, we can therefore distinguish between primary sources and secondary sources of Real-World Data.

Among the primary sources of Real-World Data, there are:

- Prospective patient registries. The most appropriate definition of a patient registry is “an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes.” [15]
- Prospective observational or longitudinal cohort studies. Longitudinal studies “employ continuous or repeated measures to follow particular individuals (cohorts) over prolonged periods of time—often years or decades. They are generally observational in nature, with quantitative and/or qualitative data being collected on any combination of exposures and

outcomes, without any external influence being applied." [17] Observational cohort studies are instead purely observational.

- Pragmatic clinical trials. Clinical trials can be designed to be either pragmatic or explanatory. [19] Pragmatic trials are designed to find out how effective treatment actually is in routine, everyday practice. Explanatory trials are designed to find out whether a treatment has any efficacy (usually compared with a placebo) under ideal, experimental conditions. [20]
- Patient and caregiver surveys. The surveys can be of various types, being qualitative or quantitative and related to different topics, but for the purpose of this research are all created specifically for the RWE study that has to be conducted, as they are primary sources of Real-World Data.
- Hybrid studies. A hybrid trial is a clinical trial that includes both traditional and pragmatic clinical trial elements. It begins as a traditional clinical trial where participants are randomized to different intervention groups with standardized procedures. The remaining data are Real-World Data collected through routine healthcare visits via sources such as EHRs (Electronic Health Records, which will be defined later in the section), medical claims, pharmacy databases, etc. By doing that, this design preserves the benefit of randomization and provides real-world outcome data while potentially accelerating product development and lowering the cost of data collection and patient follow-up. [21]

Among the secondary sources of Real-World Data, there are:

- Retrospective databases. A retrospective study is a study performed a posteriori, using information on events that have taken place in the past. Health-related retrospective databases are datasets containing information on past health-related events. Retrospective databases provide a relatively inexpensive and expedient approach for answering the time-sensitive questions posed by decision makers. [22]
- Personal health records. The first definition of Personal Health Record (PHR) frames PHR as "an electronic application through which individuals can access, manage and share their health information, and that of others for whom they are authorized, in a private, secure, and confidential environment." [24]
- Genetic and biomarker databanks. Genetic databases can be defined as one or more sets of genetic data (genes, gene products, variants, phenotypes) stored together with software to enable users to retrieve genetic data, add genetic data and extract information from the data. Genetic databases bring together several streams of data about individuals: molecular genetic data;

high-quality standardized clinical data; data on health, lifestyle, and environment; and in some cases, genealogical data. [26]

- Electronic Health Record (EHR). An EHR is a digital version of a patient's paper chart. An EHR can contain a patient's medical history, diagnoses, medications, treatment plans, allergies, and test results. EHRs are built to share information with all the clinicians involved in a patient's care.
- Medical claims and billing data. They both are documents that must be submitted from a medical practitioner to the health insurer of the patient for reimbursement of the expenses suffered by the patient. Those documents can be a source of Real-World Data because they include information about the type of medical treatment made by the medical practitioner on the patient, like procedures, exams, diagnoses, prescriptions, and medical supplies.
- Product and disease registries. Registries are special databases that contain information about people diagnosed with a specific type of disease. Most disease registries are either hospital-based or population-based. A hospital-based registry contains data on all the patients with a specific type of disease diagnosed and treated at that hospital. A population-based registry contains records for people diagnosed with a specific type of disease who reside within a defined geographic region.
- Patient Generated Health Data (PGHD). "PGHD are health-related data created, recorded, gathered, or inferred by patients to help address a health concern" [86]. PGHD are distinct from data generated in clinical settings and through other providers in two important ways. First, patients, not providers, are primarily responsible for capturing or recording these data. Second, patients direct the sharing or distributing of these data to healthcare providers and other stakeholders.
- Social Media. Social media sites, such as Facebook, Twitter, Redditt, and LinkedIn, have significant potential to harness the patient's opinion and, as such, become a valuable source of Real-World Data.

All the above definitions are summarized in the below table.

Table 3.4: Exclusion criteria after the final literature screening.

Source Category	Source Type
Primary Sources of RWD	Prospective patient registries
	Prospective observational or longitudinal cohort studies
	Patient and caregiver surveys

	Pragmatic Clinical Trials
	Patient and caregiver surveys
	Hybrid Studies
Secondary Sources of RWD	Retrospective databases
	Personal Health Records
	Genetic and biomarker databanks
	Electronic Health Record (EHR)
	Medical claims and billing data
	Product and disease registries
	Patient Generated Health Data
	Social Media

3.2.3. Opportunities and Challenges of RWD in the Life Sciences industry

It is crucial that developers of innovative products in the Life Sciences sector understand the potential for Real-World Data and Real-World Evidence to inform their operation. To understand this potential, the opportunities of RWE and RWD must be viewed in balance with their limitations.

In light of everything said above, the major challenges introduced by Real-World Data and Real-World Evidence are [44]:

- Bias and Confounding. Observational analyses, like the ones using RWD and RWE, are inherently vulnerable to selection biases and confounding. Questions are often raised about internal validity. The potential for bias is the biggest concern in the use of RWE [45], as they are not considered to meet the methodological rigor of RCTs, that leverage randomization to eliminate bias.
- Incomplete data. Datasets, particularly RWD, are vulnerable to systematic omissions or misclassification due to the manipulation that happens in real-world scenarios outside controlled environments. In addition, there are often gaps in the data. Data gaps are particularly prevalent when relying on patients or physicians to submit their own data, rather than when it is proactively collected by researchers.
- Data mining. The concern in the context of RWD and RWE is that organizations can continue to reanalyze datasets using different modeling approaches until preferential outcomes are identified. This highlights the vulnerability of RWE.

- Data access. Sharing of data across different Life Sciences organizations is not common, leading to gaps in data. The challenge is strengthened by regulatory measures that restrict data sharing and access to patient-identifiable information. These problems often arise because RWD is being used for purposes beyond those for which it was originally collected [46].
- Lack of universally accepted methodological standards. Many of the challenges outlined so far are strengthened by a lack of universally accepted standards or principles for the design, conduct, analysis and/or reporting of RWD and RWE.
- Lack of expertise. It is important that operators understand RWD well in order to be able to interpret it properly and adjust for systematic omissions and confounding biases appropriately. Being a relatively new discipline, there is a perceived lack of expertise in this area, which is an important challenge because it erodes trust in RWD and RWE and undermines their potential.

Despite the above-mentioned challenges, RWD and RWE bring valuable opportunities to the stakeholders in the Life Sciences sector, that are driving adoption and usage:

- Real-time evidence-based medicine. The biggest potential benefit of Real-World Data may be health systems' ability to combine this data with analysis that is translated into protocols and guidelines for health professionals that enable them to actively manage patients with specific characteristics. Handheld devices can have software that categorizes patients based on their specific characteristics and identifies the relevant set of customized interventions.
- Real-time monitoring of patients. A survey found that 70% of American adults are tracking at least one indicator of their health [47]. Real-time monitoring of patients can be viewed as an opportunity that will (i) reduce the costs of collecting evidence; (ii) expand evidence as to how patients are responding to drugs, and (iii) allow for different types of remote monitoring.
- Accelerated access to innovative therapies and products. The increasing access to a vast amount of RWD will allow companies in the Life Sciences sector to develop innovative products faster, with meaningful consequences for patients. Considering accelerated pathways alongside improvements in study design, RWD and RWE provide a new paradigm to get evidence on the comparative effectiveness and cost–effectiveness of the drugs within the health systems, in response to the greater use of accelerated access regulatory pathways.
- Cost reduction. Many Life Sciences decision-makers are developing policies that allow the integration of evidence from Real-World Data to analyze the

effectiveness of innovative medical treatments and various combinations of them because of the cost advantages they bring. Real-World Data allow multiple treatments to be evaluated simultaneously and offer flexible features such as dropping treatments for futility, declaring one or more treatments superior, or adding new treatments to be tested during the course of a study [48]. This comes at lower costs for Life Sciences companies and higher effectiveness.

- Personalized medicine. Precision medicine is "an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person." [50] RWD can be exploited to better understand these important differences between patients and develop innovative treatments that are customized to each specific peculiarity of cohorts of similar patients.
- Social listening. Social listening is the mining of social media to collect valuable insights into a patient's real-life use of therapeutics and medical products. The FDA has already initiated digital-listening programs that automatically collect and analyze social media information about drug safety events and experiences with medical products. Social listening is a powerful, low-cost, real-time and real-world data source; however, the use of the data presents also some limitations with regard to data validation analysis and manipulation [49].

3.2.4. Emerging technologies in the Real-World Data landscape

In this section, I will highlight which are the newest and most promising technologies that are both allowing and facilitating the collection and elaboration of Real-World Data: Sensors, the Internet of Medical Things, and Artificial Intelligence.

First, sensors are devices that detect physical, chemical, and biological signals and provide a way for those signals to be measured and recorded. Physical properties that can be sensed include temperature, pressure, vibration, load or weight, the flow rate of gases and liquids, amplitude of magnetic and electronic fields, and concentrations of many substances in gaseous, liquid, or solid form. [87] Among the various types of sensors, biosensors are the most promising and remarkable in driving innovation. Biosensors are analytical devices used for the detection of a chemical substance, for example the glucose level in the blood of a patient with diabetes.

Within the Life Sciences sector, sensors and more specifically biosensors can have a multitude of applications. but the two most promising ones are: wearable and implantable sensors.

1) Wearable Sensors

Wearables are non-invasive, non-intrusive sensors that can monitor an individual's health or wellness status without interrupting their daily activities.

An example of the application of wearable sensors is Remote Patient Monitoring (RPM), that is enabling the monitoring of patients' health outside of conventional clinical settings through sensors that are connected via Internet with care providers.

2) Implantable biosensors

Implantable biosensors are biosensors that are implanted within the body of a patient. It is an important class of biosensors because of their ability to provide continuous health data. This enables health trends and changes over time to be monitored without any need for intervention from either the patient or clinician. Implantable biosensors have therefore great potential in the diagnosis, monitoring, management and treatment of a variety of disease conditions.

In the near future, implanted electronics will in fact be an important tool in Life Sciences, since they can provide a clearer picture of the cascade of events occurring inside the body in a certain period of time, helping monitoring chronic diseases, or progress after treatment and/or surgery. [33] An important example of application of the implantable technology is the continuous glucose monitoring: a biosensor implanted under the skin that constantly reads blood glucose concentrations and signals a warning in case of dangerous changes.

Secondly, the Internet of Medical Things (IoMT) is a healthcare application of the IoT technology and comprises a network of connected devices that sense vital data (Real-World Data) in real-time. The Internet of Things (IoT) ecosystem is "a very complex architecture, in which multiple components interact with each other to enable various solutions for the end user. This is an interdependent system, which enables real-time data acquisition, device connectivity, data transfers, and analytics to control end user applications. IoT provides the connected environment, comprising the cyber physical systems, which integrates human intervention with computer-based systems and facilitates data-driven decision processes." [89]

A complete and comprehensive definition of the Internet of Medical Things (IoMT) defines it as "the interconnection of communication-enabled medical-grade devices and their integration to wider-scale health networks in order to improve patients' health." [35]

The IoMT combines therefore both the reliability and safety of traditional medical devices and dynamicity, genericity and scalability capabilities of traditional Internet of Things (IoT), which is characterized by the features defined above.

From the definition, we can understand that the IoMT is the interconnection between not only numerous personal medical devices but also between devices and health care providers, such as hospitals, medical researchers, or private companies.

The advent of the IoMT is mainly caused by increase in use and development of connected and distributed medical devices. Moreover, the development of smart sensors, smart devices, and advanced lightweight communication protocols created the possibility of interconnecting medical things to monitor patient's biomedical signals, contributing to the development of IoMT technologies.

Despite the challenge of reliable connectivity, the IoMT has the potential to disrupt business processes through data driven health prediction, real-time intervention, and increased efficiency.

Thirdly, Artificial Intelligence can be defined as "a system's ability to correctly interpret external data, to learn from such data, and to use those learnings to achieve specific goals and tasks through flexible adaptation." [90]

A definition of AI which is more focused on the Life Sciences sector defines Artificial Intelligence as "the use of complex algorithms and software to emulate human cognition in the analysis of complicated medical data and analyze the relationships between prevention or treatment techniques and patient outcomes." [38]

Among all the different types of Artificial Intelligence technologies, one of the most commonly used in the Real-World Data landscape is Machine Learning (ML), which can be defined as "a statistical technique for fitting models to data and to learn by training models with data". [39]

The increasing usage of Machine Learning in Life Sciences has been driven by three main trends:

- A focus on RWE and its potential role in generating high-quality evidence previously reserved for Randomized Clinical Trials (RTCs).
- Dramatically expanding amount of Real-World Data on patients' health (some of it in curated research data assets), most in need of a great deal of work before being suitable for research.
- The latest advances in Artificial Intelligence and more specifically in machine learning methods.

In fact, several aspects about the changing Life Sciences data landscape - the rapid growth in the volume of healthcare data, the fact that much of it is unstructured, the ability to link different types of data together (claims, EHR, sociodemographic characteristics, genomics), the speed with which data are being refreshed - create serious challenges for traditional statistical methods. As a result, there is growing interest in the use of Machine Learning to help address these analytic challenges. [41]

Within the Life Sciences, and especially now with the latest technologies, a lot of Real-World data are created. However, processing and storing them requires a lot of power. The advent of AI, ML, and Cloud has solved this problem, and now companies can implement them within their processes.

Many Life Sciences companies have in fact begun to invest in resources, technologies, and services, especially in generating and assembling data sets to support research in Artificial Intelligence and Machine Learning, and many of those data sets are from RWD sources.

3.3. Multi-sided non-transactional platforms

In this section, I will present and analyze the concept of multi-sided platform and in particular multi-sided non-transactional platform, seeing it as the major means through which Life Sciences companies are collecting Real-World Data to create innovative products.

3.3.1. Multi-sided platforms: definitions and typologies

Despite their extraordinary penetration and development, there is still a high degree of ambiguity when trying to give a definition to platforms.

It all started with the first definitions of two-sided markets, considered as those markets "characterized by the presence of two distinct sides whose ultimate benefit stems from interacting through a common platform". [51]

Putting further clarifications on the concept, two-sided platforms have been characterized for presenting three necessary conditions: (1) the existence of two (or more) groups of customers, (2) linked through indirect network externalities and (3) with a platform provider able to internalize (at least partially) the externalities. [52]

In this context, indirect network externalities are defined as a phenomenon happening when the utility of a product increases with the greater availability of compatible complementary products. [53] As example, as more travelers join Airbnb, the hosts get greater value from the platform.

In 2006, Rochet and Tirole started to pair the concept of two-sided markets to the broader concept of multi-sided markets, providing for the possibility of having also more than two sides, in the following definition: "two-sided (or, more generally, multi-sided) markets are roughly defined as markets in which one or several platforms enable interactions between end-users and try to get the two (or multiple) sides 'onboard' by appropriately charging each side". [54]

Moving forward, in fact, a new trend happening relates to how successfully two-sided platforms tend to evolve toward more complex configurations, multi-sided platforms, adding new sides and other players that can exploit the value of data generated through the platforms. Possible examples are the abovementioned Airbnb and Uber, which added for examples experiences (Airbnb Experiences) and restaurants (Uber Eats) generating new transaction lines. [55] In defining a multi-sided platform, however, the previously cited definitions present some weaknesses in terms of clarity and scope. Trying to better define multi-sided platforms, the vast majority of the literature has opted for using a simpler and straightforward definition (also highlighted by Rochet and Tirole) based on the presence of indirect network effects, a concept defined previously. Although this definition is very broad, most of the authors consider the presence of indirect network externalities as a fundamental element in the definition of multi-sided markets. [56]

In their paper, Hagiu and Wright [57] criticize the approach related to the presence of indirect network externalities, stating that "multi-sided platforms have two key features beyond any other requirements (such as indirect network effects or non-neutrality of fees):

- They enable direct interactions between two or more distinct sides.
- Each side is affiliated with the platform."

From a literature review perspective, the definition from Hagiu and Wright [57] seems the most accurate and comprehensive in clarifying the concept of multi-sided platforms.

Despite the various differences and approaches to define multi-sided platforms and markets, from a comprehensive literature review we can understand that there is a general agreement on what is relevant to define multi-sidedness, and on the three pillars that determine the boundaries of multi-sidedness [58]:

- Network effects. Multi-sidedness requires the presence of noninternalized externalities among end-users (i.e., indirect network effects). The concept has been previously defined and is considered to be present when the utility from a product (the platform) for one side increases as the presence of the other side in the platform increases.
- Price structure. Multi-sidedness requires the possibility of cross-subsidizing different categories of end-users (i.e., price structure matters). In this context, cross subsidization is the practice of charging higher prices to one side of the platform to artificially lower prices for another side. For instance, the TikTok platform applies cross-subsidization because the application is free for users, but advertisers pay to participate in the platform.
- Property rights. Multi-sidedness requires that prices must reflect that end-users that are parties to a transaction retain control over essential terms of the interaction (i.e., control rights).

The failure of any of these conditions can impair the presence of multi-sidedness. For example, if network effects are not found, or the platform is price-taker on one side, or the platform behaves as a reseller, traditional one-sided analysis is better suited. [59]

As it is pointed out by Filistrucchi et al. [60], at first sight, it appears to be still some debate on the exact definition of a multi-sided market, but the different definitions appear consistent enough to allow the practical identification of multi-sided markets. In fact, when we combine the different approaches found in the literature, it seems that a clear concept of multi-sidedness emerges around price structure, network effects, and control rights.

3.3.2. Multi-sided platforms: types

In the multi-sided platform literature, more than 80 different models have been developed, and they can be classified in several ways: according to the nature of fees (membership or use), the number of platforms (monopoly, duopoly, or N-platforms), the possibility of being in one or several platforms at the same time (single-homing vs multi-homing), etc. [56]

Several authors have in fact tried to classify models and platforms with regard to their characteristics. The most relevant categorization has been brought by Filistrucchi [62], who defines two main categories of two-sided markets using the observability of the transaction between the two sides as the main rationale:

- Two-sided non-transaction markets, or Media type. In these markets, the transaction is not present, or it is unobservable. These markets only set membership fees. For example, newspapers. Readers read the newspapers with their ads, but the newspaper does not know if the ads are generating transactions for the advertiser. Non-transaction Two-Sided Platforms look like traditional "one-sided" businesses, having a straightforward customer side to address and managing the value creation process toward them completely. Nevertheless, they have the ability to manage also a second value proposition and to exploit the value generated by the first side users while they enjoy the service.
- Two-sided transaction markets, or payment card type. These markets are characterized by the observability of transactions between the sides, like payments with credit/debit cards. The platform can monitor the transaction, and it can set both transaction and adoption fees.

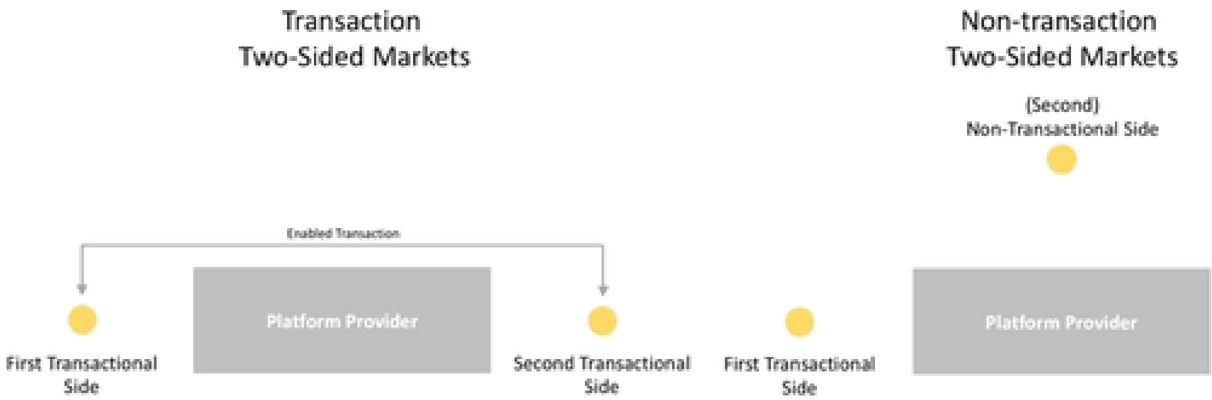


Figure 3.4: Graphical representations of platform typologies.

The latter category has been the object of many studies. In the field of the former category, instead, scholars have focused mainly on two topics. First, the pricing dynamics have been considered extremely relevant. In this case, the basic idea of the Gundlach et al. [63] model is to use the second side, namely the advertisers, as the subsidizer for the first one, namely the readers who do not pay or pay less than the cost price to receive the service. The second area of study was related to the "quality", in particular, the number of advertising messages, the perception of quality by end-users, and the overall platform quality related to the reduction of the entry barriers.

Starting from the traditional concept of a two-sided market, Trabucchi et al. [64] identified three strategies to foster innovation on a two-sided platform, that eventually lead to the emergence of three different categories of multi-sided platforms. The three strategies are:

- Supply (Side) Extensions. This strategy enables the platform to identify new meaningful transactional sides that may be linked to the first transactional side to enlarge, de facto, the supply side and define different levels of transactions.
  In summary, this strategy maintains the essential transactional relationship between the two seminal sides and adds a second level of the transaction, involving a third side and linking it with the seminal first side. This strategy exploits the value of the customers who join the platform on the first side to attract a new (third) transactional side, building a new type of multi-sided platform which is categorized by the transactional nature of the relationship among the sides.
- Transactional advertising. In this strategy, the platform provider now offers advertising mechanisms to the supply side within the platform. It utilizes the end-users' data gathered through the activity on the platform as a source to provide the supply side with advertising services. In this way, the transactional relationship expands towards non-transaction mechanisms. The platform provider has the chance to create a second relationship with a

specific side, moving from the pure transactional perspective (i.e., matching the first and the second side) to a non-transaction relationship, using the first side both as the target and as a source. The result of that is the creation of a multi-sided platform in which the nature of the transactions among the participants are both transactional and non-transactional.

- Data trading. Within the context of this strategy, platforms can identify what may potentially be a new side interested in the data and able to extract the value embedded in those data. The overall aim of the strategy is to find additional ways to capture the value embedded in the data they own, which are generated through the already existing services on the platform. Therefore, in this strategy, the platform owner aims at finding a market opportunity for selling the gathered data, usually to create a new revenue inflow. Within this context, it is important to highlight the role of privacy policies in implementing this strategy. Usually, companies reserve the rights to use the gathered data, even if the data is aggregated in an anonymized way, to avoid violation of the privacy of the players on the various sides.

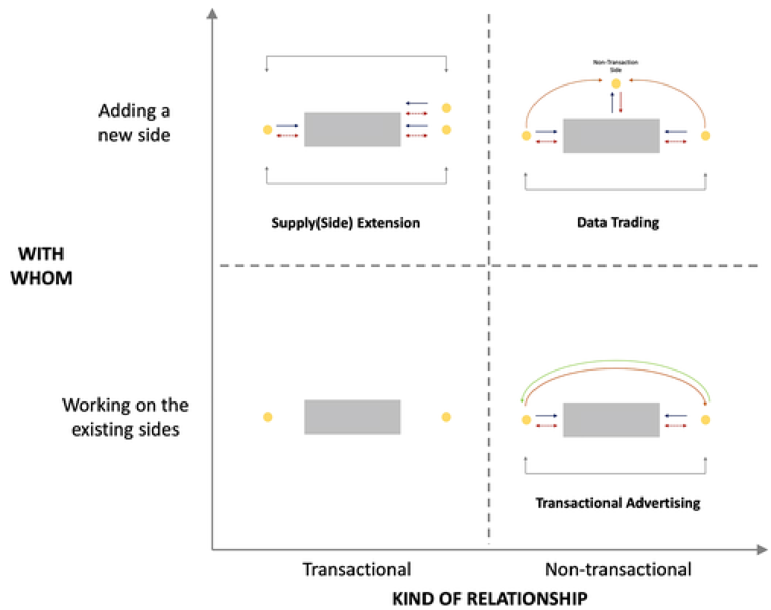


Figure 3.5: Graphical representations of platform strategies.

The main difference among Data Trading and Transactional Advertising strategies is that in the latter, a side is offered both a transactional and non-transactional relationship with another side simultaneously, while in the former the new side added benefits only from a non-transactional relationship with the other sides.

Evidence shows that platforms can pursue many of the above strategies simultaneously, by finding increasing ways of creating value through their networks and solving unmet needs of additional groups of customers which eventually affiliate to the platform, creating another side.

This evolution of two-sided platforms can result in what Trabucchi et al. [55] define as hybrid multi-sided platforms, complex platforms that evolve by adding new supply sides and even orthogonal players that can exploit the value of data generated through the platforms. The key point in the evolution of platforms towards hybrid multi-sided platforms seems to be the platform provider's ability to see that the basic two-sided structure embeds untapped value that can be exploited for further expansion of the business model. An important example of this relates to Facebook, that having only end-users on one side and advertisers on the second became a hybrid multi-sided platform introducing companies, developers, sellers (with the marketplace) and many other sides. This example shows how an existing company has assets, existing relationships, knowledge, a customer base and many other resources that can be re-arranged in a hybrid multi-sided platform to unlock new value flows and innovation opportunities.

Therefore, hybrid multi-sided platforms can exploit some of their basic underlying mechanisms by relying on the opportunities that emerging technologies and needs may provide, by exploring a new level of business model innovation. By doing that, hybrid multi-sided platforms become able to manage many value flows, mainly coming from outside their boundaries, but having the power to orchestrate the flows and to set the rules of the games. They may in fact evolve to include many different types of groups of customers (referred to as sides), which interact with the platform owner and with other sides through a combination of complex transactional and non-transactional relationships.

Within the objective of this thesis, the type of platform that will be studied is the kind that presents, among all the types of transactions characterizing the platform, at least one non-transactional relationship between the various sides. Following the categories outlined above, therefore, the study will comprise platforms that either present only non-transactional relationships, or more complex and hybrid platforms that allow for both transactional and non-transactional relationships.

3.3.3. Multi-sided non-transactional platforms: business model

In this section, the focus will be put on multi-sided non-transactional platforms, where at least one non-transactional relationship is present within the platform's business model. The main components of the business model characterizing such platforms will be presented and explained, along with the strategic decisions that platforms need to face in each particular business model area.

The business model constituents that are characteristic of multi-sided non-transactional platforms and that will be analyzed in the following sections are:

1. Value creation

2. Sides
3. Revenue model and pricing
4. Governance & control

1) Value creation

A multi-sided non-transactional platform can provide value and extract value from its users in a variety of ways. From a literature review, the value creation seems to be related both to quantity and quality aspects.

On the quantity aspect, the role of network externalities (or network effects) is key to generate and extract value from the platform. Previous research on multi-sided markets indicates that the various user groups affiliated to a platform can exhibit different kinds of network effects. Users may derive positive cross-side network effects (also called indirect network effects) from the participation of members on the other side of the market, which means the larger the installed user base on one side of the platform, the more attractive the service for the opposite side's users. So, in this case, indirect network effects exist if the number of users on one platform side influences the utility of the opposite groups' users. Network effects can also emerge within one user group, known as same-side network effects (or direct network effects). In this other case, instead, direct network effects exist if a user's utility is affected by the installed user base of his or her own user group. [67]

As the presence of network effects is considered a characterizing element of a multi-sided platform, it is also an important way through which platforms create and capture value, and it is a meaningful driver of growth. Network effects can create in fact create high barriers to entry, which explains why successful MSPs occupy privileged and often hard to reach positions in their respective industries. Sometimes, strong barriers to entry and network effects can lead to market dominance.

Despite the quantity-driven strategy for platform value creation is considered to be the most significant and is also the most studied, a quality-driven strategy has been shown to be also important for growth and value creation. Studies have in fact shown that product quality exerts a significant, positive influence on market share, return on investment, premium prices charged, advertising, perception of quality, and stock market returns. [68]

Another study made by Lau et al. [69] investigates on the quality perspective of non-transactional platforms in the context of media companies. Building on the previous theoretical models that cast doubts on the notion that advertiser-supported media results in an optimal array of media programming, they conclude that advertising-funded media may undercut quality. Even if the already mentioned theoretical models have argued that viewer-supported media may be more likely to result in more optimal outcomes from the standpoint of the consumer and quality of the platform, a tendency has been found to persist for free (advertiser-sponsored) pricing aimed at undercutting viewer-supported media in some media markets.

Taking a comprehensive view, both the quality and quantity strategies seem not only to allow platforms to create and capture value, but they are also interconnected in that they increase the switching costs perceived by users affiliated to the platform [70]. Cross-side network effects alone do not guarantee high barriers to entry: for a multi-sided platform to keep rivals and new entrants away, high switching costs or high costs to belong to more than one competing network are also necessary on one or all sides of the platform.

Strictly connected with the concept of network externalities and on how to capture value through them, an important concept is the one of the chicken and egg problem. This problem arises when, for a non-transactional multi-sided platform with indirect network effects, there is the need for a critical number of users from one side to attract users from the other sides, but the former side will only adopt the platform and invest in it if they expect a sufficient number of the latter on the other side to join. Or, in short, there is no economic value for one side to join the platform when there are no users from the other side, and vice versa.

In their paper, Stummer et al. [71] provide for six platform strategies to avoid the chicken-and-egg dilemma and achieve network effects to enable value capture:

1. Single Target Group. To this end, multi-sided platforms may start, for example, with a single city or industry. By reducing the total market size and the required critical user mass, multi-sided platforms require fewer resources and less time to reach the critical inflection point from which the multi-sided platforms can grow to other market segments. The first and single target group on which a multi-sided platform can focus should be formed by the users whose participation brings extraordinary value for other platform users, or by the most loyal users.
2. Platform staging. With the platform-staging strategy, a multi-sided platform evolves in two distinct steps from a traditional vendor-based business model in the first stage to a platform business model in the second stage after reaching the critical user mass.
3. Subsidizing. This kind of strategy requires the multi-sided platform to typically have a ‘subsidy side’ that allows the use of the platform with discounts or even for free, and a ‘money side’ that is charged for participation or transactions. The idea behind that is to subsidize one side of the market to attract the ‘money side’ of the platform until the critical inflection point is reached. This strategy is very common for non-transaction multi-sided platforms.

4. Platform Envelopment. The platform envelopment strategy aims at leveraging the shared relationships with other established platforms and their networks.
5. Exclusivity agreements. Signing exclusivity agreements on one market side can attract other users on both market sides and help overcome the chicken and egg problem.
6. Side switching. The strategy behind the side-switching strategy is to make a multi-sided platform one-sided temporarily by finding a platform design that allows users to fill both market sides of the MSP at the same time.

Regardless of the strategy used to solve for the chicken-and-egg problem and enable network effects to capture value, every multi-sided platform is faced with the need to define a clear and valuable value proposition for each side of the platform, and this need introduces the challenges of crating multiple value propositions at the same time. The key question is to determine why any party might join the platform.

For the sides containing groups of consumers, the motives can be as varied as the benefits offered by the multi-sided platform; for the sides containing groups of companies, instead, the motives are generally linked to the size of the audience, its particular characteristics and/or the usefulness of the data collected from this audience. For example, Fish [72] states that B2B companies involved in multi-sided markets will benefit from consumers' private data (the "privacy capital"), i.e. businesses advertising on Facebook do so because they can micro-target their audience based on the personal information (age, gender, interest etc.) provided through this audience. In any case, it seems that the typical multi-sided non-transactional digital business model sees sides made of consumers as loss leaders (and therefore they get the service for free) and business participants as subsidizers (they pay to reach the audience of end-users).

In their paper, Muzellec et al. [73] formulate a proposition stating that multi-sided platforms need to formulate different value propositions, one for each different side affiliated to the platform. The literature on multi-sided platforms provides little information about the role of each of those sides, but the concept of reciprocal value propositions offers some further insights on this issue. The concept of reciprocal value propositions represents a recent development introduced by Glaser [74], claiming that if participants in the value creating process recognize that their objectives are complementary rather than antagonistic, the value outcomes for all parties are likely to be enhanced. Value in this sense is not so much a strategy or a set of customer benefits but an all-inclusive process, where participants share in the creation of value. This means that in any proposed marketing exchange there will be at least two negotiators, and their assessments of value become linked in reciprocal promises. While the concept of a reciprocal value proposition may apply in various contexts, multi-sided platforms add a level of complexity as the business customers and the actual end-user of the service are different and in the non-transactional case do not even interact and negotiate among each other. Therefore, in order to qualify as a reciprocal exchange of value, the reciprocal value proposition should be transparent about to whom that value should flow, as well as being perceived as a fair exchange of value. This is the main task with which multi-sided platforms are faced when designing their value proposition for each side.

The main challenge that multi-sided non-transactional platforms face is that often the end users' data become then the item being exchanged on the business side. Hence, one could say that end users are in fact the value proposition, albeit for a totally different segment. The value of this proposition resides in the very nature of the end-user being an audience who can be monetized either because of its size, hence the focus on the number of unique visitors (a value proposition that is extremely attractive for mass advertisers), or because of the demographic, psychographic or behavioral characteristics of this audience. Therefore, the common strategy that multi-sided non-transactional platforms is to formulate a value proposition focused on the free service for consumers, while businesses end up paying for participating in the platform where the value comes from the possibility of using users' data collected.

2) Sides

One of the main business model components of multi-sided non-transactional platforms is related to the sides affiliated to the platform: how many sides participate in the MSP, which are their roles, which are the incentive mechanisms used for their engagement and how the relationship with them is managed by the platform.

In some cases, the answer is obvious and constrained by the choice of industry. Sometimes, however, MSPs face a real choice when it comes to the number and identity of the sides to attract, and the choice presents some important trade-offs.

In 2011 LinkedIn, the world's leading professional networking service and one of the major examples relating to strategic choices on the different sides, was running a three-sided platform that connected individual users (professionals), recruiters and advertisers. The company derived significant revenues from all three sides: 20% of revenues came from premium subscriptions, 30% from advertising solutions and 50% from recruiting solutions. The company was at that time attempting to attract two additional sides: corporate users (company HR departments that would set up LinkedIn profiles to interact with their employees) and application developers. The challenge was that some individual users might not have welcomed the presence of corporate users (their employers) and that applications would have to be strictly restricted to a professional context. Thus, while adding two more sides could potentially help LinkedIn grow, it also increased the risk of friction between the multiple sides and thereby LinkedIn's costs of operation.

Looking at the above example, the trade-off involved in choosing whether to attract more or fewer sides becomes apparent [65]. More sides lead to potentially larger cross-side network effects, larger scale and potentially diversified sources of revenues (as corporate users or application developers). But, on the other hand, there are two good reasons that may favor the presence of fewer sides. First, it may not be economically viable for one (or several) side to exist independently. Second, even if attracting many sides is possible, doing so carries the risk of creating too much complexity and even conflicts of interest between the multiple sides and the MSP (as with LinkedIn's efforts to attract employers as a new side).

Adding more sides can also cause a "lowest common denominator" issue, in that the need to please many different and heterogeneous platform constituents greatly constrains an MSP's ability to innovate by introducing truly ground-breaking features.

The above mentioned "lowest common denominator" problem is partly connected with the former proposition stating that too many sides with conflicting interests can increase the risk of creating too much complexity. When too much complexity is created, the multi-sided platform owner may lose its ability to freely introduce new features or innovations that would benefit specific sides but may conflict with some others' interests. It is therefore strategically important for a multi-sided platform to consider such trade-offs that can emerge when additional sides bring increasing complexity, as this can impact the capability of the platform to provide value to all the affiliated sides.

Finally, even if it makes sense to attract more sides in the long run, some MSPs find it easier to solve the initial chicken-and-egg problem by starting with fewer sides and at least partially vertically integrating into some of the missing sides. This kind of partial vertical integration presents the opportunity to reap higher returns by owning some of the most profitable complementary products or services.

Another important strategic aspect that multi-sided non-transactional platforms face and that can potentially impact their business model is the segmentation within each side [66]. The platform, in fact, may create a segmentation of different types of users within each side, consisting of differentiating among various sub-groups. For example, a platform can allow users of a particular side to become premium users, providing them with additional functions or facilitations under the payment of an additional fee. The result of this segmentation is the creation of user sub-groups within a specific side of the platform, where each sub-group has a different type of affiliation with the platform as it has access to different services or features.

According to Ardolino et al. [66], platforms can segment each side in different sub-groups depending on different criteria and applying different strategies:

- Premium segments. In this case, for instance, the platform may create a "premium segment" to allow users to benefit from enhanced services in exchange for a (higher) fee. LinkedIn is a remarkable case with regard to this strategy, as there are multiple premium sub-groups of customers that have access to different features. With the Sales Navigator subscriptions, premium users get access to an exclusive platform called Sales Navigator, in which increasing data points can be extracted on LinkedIn users.

- Increased visibility. In other cases, a segmentation in the supply side might be aimed at enhancing the visibility of a group of users towards the demand side. For example, a specific subgroup of advertisers can be granted increasing visibility on the other users.
- Quality. Another way to segment a side may be related to the achievement of a specific objective of quality concerning the services provided and can therefore access to additional benefits or features.

Summing up all the above-mentioned concepts, owners of multi-sided non-transactional platforms face many different and complex strategic decisions regarding the number of sides and the type of affiliation relationships with each of them. All the above decisions affect the business model and value creation dynamics of the platform and can affect its growth as well as its development stages.

3) Pricing

In competitive industries, prices are largely determined by the marginal cost of producing an extra unit, and margins tend to be thin. In industries with high barriers to entry, the price ceiling is set by customers' willingness to pay, and margins are more likely to be fat. For multi-sided platforms, pricing is a more complicated challenge. Platform providers have to choose a price for each side, factoring in the impact on the other side's growth and willingness to pay. Because multi-sided platforms serve multiple types of customers, they potentially have multiple revenues and profit sources. In reality, however, most MSPs have discovered that they have to offer their services for free or at subsidized prices to at least one side of the platform and derive their profits on the other side. Typically, therefore, multi-sided non-transactional platforms have a "subsidy side," that is, a group of users who, when attracted in volume, are highly valued by the "money side," the other user group. Because the number of subsidy-side users is crucial to developing strong network effects, the platform provider sets prices for that side below the level it would charge if it viewed the subsidy side as an independent market. Conversely, the money side pays more than it would if it were viewed as an independent market. The goal is to generate cross-side network effects: if the platform provider can attract enough subsidy-side users, money-side users will pay considerable amounts to reach them. Cross-side network effects also work in the reverse direction. The presence of money-side users makes the platform more attractive to subsidy-side users, so they will participate in greater numbers. The challenge for the platform provider with pricing power on both sides is to determine the degree to which one group should be encouraged to swell through subsidization and how much of a premium the other side will pay for the privilege of gaining access to it. The challenge is therefore that is not always obvious which side, if either, the platform should subsidize and which it should charge, and how much should they charge each side relative to the others. Pricing structures have been the first and dominant focus of the economics and strategy work on multi-sided non-transactional platforms to date.

In their paper, Eisenmann et al. [75] define a model containing the factors influencing pricing decisions for platform owners. The factors presented are:

- Ability to capture cross-side network effects. A potential subsidize in the form of free pricing for a side of the platform will be wasted if the platform's subsidy side can transact with a rival platform provider's money side. That's what happened to Netscape, which subsidized its browser to individuals in the hope of selling Web servers to companies operating Web sites. However, Web site operators didn't have to buy Netscape's server in order to send pages to Netscape's big base of users; they could buy a rival's Web server instead.
- User sensitivity to price. Generally, it is more effective to subsidize the network's more price sensitive side and to charge the side that increases its demand more strongly in response to the other side's growth. In this context, the strategy relies on treating each side of a multi-sided platform independently of the others. The price sensitivity on any given side of a multi-sided platform can be estimated by the availability of substitute services — or simply by the bargaining power that the multi-sided platform has over that particular group. Usually, online newspapers follow this pricing rule. Their network consists of two sets of stakeholders: advertisers, who create commercials and want to leverage on reader's data to achieve their business outcomes, and readers, who read the articles and also receive the commercials made by advertisers them. Readers are very price sensitive, so they pay nothing for reading the news. If readers were charged even a small amount, the platform's user base would be much smaller. Advertisers, who greatly value the platform's audiences, pay a fee for their advertising activities. If the platform reversed its approach, charging readers and subsidizing advertisers, its network would collapse. Advertisers are less price sensitive, so free advertising services would not dramatically boost their numbers.
- User sensitivity to quality. High sensitivity to quality also marks the side a multi-sided platform should subsidize. This pricing strategy may seem at first counterintuitive: rather than charge the side that strongly demands quality, the platform should charge the side that must supply quality. Such a strategy is evident in video games. To deliver compelling quality, game developers incur enormous, fixed costs. To amortize these costs, they must be assured that the platform has many users. Hence the need for a consumer subsidy. Platform providers make sure game developers meet high quality

standards by imposing strict licensing terms and charging a high royalty. This royalty is not passed through to consumers: developers charge the highest prices the market will bear, regardless of the royalty rate. However, the royalty helps to avoid games of marginal quality. Once the royalty is added, titles with poor sales prospects cannot generate enough contribution margin to cover their fixed costs, so they never get made in the first place and quality is assured within the platform. The same dynamic can be seen in social media platforms, where the auction pricing strategy that advertisers face ensures that high quality advertisements are cheaper than low quality advertisements.

- Value extracted by customers. If there is no priced transaction between the sides, which is the case of multi-sided non-transactional platforms, then the platform provider should charge more to the side that stands to benefit more from the presence of the other side or sides.
- Users' brand value. The participation of "marquee users", defined as users who are influential in their respective markets, can be especially important for attracting participants to the other side of the network. A platform provider can accelerate its growth if it can secure the exclusive participation of marquee users in the form of a commitment from them not to join rival platforms. It can however be expensive, especially for small platforms, to convince marquee users to forfeit opportunities in other networks. When the participation of a few large users is crucial for mobilizing a network, conflict over the division of value between platform providers and large users is common.

Due to the factors outlined before, which are the main drivers affecting the pricing structure of multi-sided non-transactional platforms, a common feature in this type of platforms is the prevalence of heavily skewed pricing strategies in which price markups are much higher on one side of the market than the other. [76] And, often, the side for which price markups are lower is the consumer side.

4) Governance & Control

Platform-based markets as multi-sided platforms are characterized by indirect network effects, where the demand for the platform on one side of the market will subsequently affect the demand for the platform on the other side of the market. Thus, a platform with greatest pool of complementary products and sides should attract most of the new end users which then stimulates further support by complementors, eventually resulting in self-reinforcing demand dynamics. The strong role of indirect network effects in the growth of platforms is such that platform-based markets are called winner-take-all markets.

In their paper, Pontus et al. [77] challenge this concept proving that the winner does not always take it all, and there are other factors affecting platform adoption and market dominance, other than indirect network effects. Since a multi-sided platform enables both direct and indirect interactions between different users, control mechanisms should be set to prevent inappropriate behaviors and actions by the users that can damage the image and reputation of the platform. In their study, in fact, one of the first incorporating demand-side dynamics on the micro-level, they prove that quality control mechanisms help enhance adoption and growth of the platform. As the adoption decisions of selectively attentive consumers are affected by the changes in complementary product quality in the market, this implies that the competitive advantage of a multi-sided platform is partially tied to its ability to renew the pool of complementary products and ensure its quality. That is, if another platform is better in renewing the pool of complementary products and ensuring their quality relative to a specific platform, the former can obtain more adopters. In other words, not merely the size, but the change in size and quality of the pool of complementary products is a key factor affecting platform competitiveness.

In their research, Ardolino et al. [66] provide for three key business model factors with which multi-sided platforms can control the quality the complementary products or services included in the platform:

- Control mechanisms. The mechanisms arranged by the platform aim at controlling the behavior and the activities of the users as well as the contents provided through the platform.
- Rating and review system. The presence of a rating and review system helps both users in choosing the best match for their need and the platform manager in verifying potential incorrect behaviors.
- Exclusive agreements and contents. The presence of exclusive agreements between the platform manager and users allows the former to provide exclusive services or products so users are forced to join that platform.

Another important aspect which affects the business model and value creation mechanisms of multi-sided platforms is the governance mode of the platform, that can be distinguished between open and closed governance platforms. Broadly speaking, openness relates to the easing of restrictions on the use, development, and commercialization. The polar extremes of openness can be understood in relation to property rights. Closed technologies are wholly owned, proprietary, vertically integrated, and controlled by a single party. By default, the owner of a closed platform fully restricts outsiders from the technology through patents and copyrights, secrecy, or other means. In contrast, a purely open platform is placed in the public domain, neither owned nor controlled by any party, thus accessible to all. Opening key enabling assets thus allows free entry into the platform. Open policies sometimes go even further by guaranteeing rights to modify, transform, and build upon previous developments in an unfettered and non-discriminatory fashion [78]. A platform considering whether to pursue an open governance strategy faces a trade-off that has come to be known as "adoption versus appropriability". Pursuing an open strategy reduces the platform's share of profits by lowering entry barriers and introducing intra-system competition. On the other hand, all else being equal, opening might encourage wider adoption of the platform. Multi-sided platforms should therefore take into account this trade-off when designing the business model of their multi-sided non-transactional platform.

3.4. Managerial research on RWD platforms in the healthcare industry

In the following sections, I will focus on the applications of RWD multi-sided non-transactional platforms in the healthcare industry, presenting also the main managerial problems emerged in academia. I'll then conclude with the research gaps found with the literature review, which will be addressed in my thesis.

3.4.1. Multi-sided non-transactional platforms: business model

Based on the literature review, the main managerial and academic applications of multi-sided non-transactional platforms leveraging Real-World Data and new technologies can be grouped in the below categories. The criteria to group the various managerial and academic applications found has been the purposes of use of RWD, which means the objective with which such health data are used by the multi-sided non-transactional platforms investigated.

1) Disease diagnosis, treatment and drug development

In their paper, Venkata et al. [1] propose a multi-sided platform for better mining, integration and visualization of translational medicine big data aimed at improving disease diagnosis and the development of new treatments. Translational medicine big data include biomedical dataset, genomics databanks, and public health data.

Taking a more focused approach to disease diagnosis, XX propose a platform and AI framework to predict the risk of renal failure directly from a Real-World Data repository of chronic disease population. The aim of the platform is to leverage Real-World Data for routine chronic disease management procedures and for more preemptive, widely-covered screening of renal risks, which would in turn reduce the damage caused by the disease through timely intervention.

In their paper, Rajal et al. [28] describe a multi-sided platform with the aim of improving stroke research and diagnosis. Its main functions are the following: (i) store, manage, process, and facilitate sharing of high-value stroke imaging data sets, (ii) implement automated computational methods to extract image characteristics and disease-specific features from contributed images, (iii) facilitate integration of imaging, genomic, and clinical data to perform large-scale analysis of complications after stroke;

and (iv) develop a collaborative platform aimed at both data scientists and clinical investigators.

From disease diagnosis to drug development, the platform proposed by Harding [29] under the project Orchid is a conceptual Clinical Intelligence Exchange and Virtual Innovation platform utilizing an Open-Source approach to support clinical innovation efforts and multi-national collaboration that can be locally sustainable for low- and middle-income countries. This platform wants to leverage Real-World Data to enable low- and middle-income countries research organizations to accelerate their clinical trial process maturity in the field of drug discovery, population health innovation initiatives and public domain knowledge networks.

2) Improving Health Research

The second category of RWD platforms applications relates to the purpose of improving health research, both in academia and industry. Loryana et at. [30] describe a significant military-civilian collaboration for a big data business intelligence platform called the Person-Event Data Environment (PDE). The PDE is a consolidated data repository that contains unclassified but sensitive training, financial, health, and medical records covering U.S. Army personnel, civilian contractors, and military dependents, and is designed to bring researchers and military scientists to a single computerized repository rather than porting vast data resources to individual laboratories. In the paper, researchers from the University of Pennsylvania leveraged the PDE to learn more about relations between psychological and health assets and health outcomes, including healthcare utilization and costs. PDE studies have the potential to provide much more detailed insight into health-related questions of broad societal concern, improving health research globally thanks to a military-civilian collaboration.

Following a different approach for the collection of Real-World Data, Chetan et al. [31] leverage on Twitter data to develop a big data analytics platform aimed at improving health research and patient understanding through their Twitter activity. The platform uses machine learning (ML) and artificial intelligence (AI)-based systems for trend and sentiment identification, and for healthcare-related research.

In their paper, Pala et al. [32] present the Participatory Urban Living for Sustainable Environments (PULSE) project, a data analytics platform designed to provide public health decision makers with advanced approaches to jointly analyze maps and geospatial information along with healthcare and Real-World Data. Through this platform, Real-World Data are paired with geospatial data through Artificial Intelligence to improve health research.

Finally, Ye et al. [34] describe a multi-modal sensor platform with wearable and non-wearable sensors to develop a predictive system to identify early symptoms of dementia in a hospital-based dementia behavioral care unit. Using a different approach, researchers created a continuous monitoring system that can detect and predict the onset of early dementia symptoms, as opposed to basing on observation. Thanks to their innovative approach, the platform created allowed improvements in health research thanks to a continuous, more comprehensive, and data-driven monitoring of people living with dementia (PLwD).

3.4.2. Managerial challenges and Ethical concerns

According to the literature review, the application of Real-World Data mining for multi-sided non-transactional platforms brings many managerial challenges, that are preventing companies to fully unlock the potential of Real-World Data.

The first managerial challenge is represented by data governance. According to Kruse et al. [36], data governance will need to move up on the priority list of organizations, and it should be treated as a primary asset instead of a by-product of the business. Data ownership and data stewardship should create new roles in business that consider big data analytics, and new partnerships will need to be brokered when sharing data.

Another important challenge is the lack of appropriate technological skills to extract value from Real-World Data. Kruse et al. [36] state that health care workers should be also kept up to date with the use of constantly changing technology, techniques, and a constantly moving standard of care. Due to the constant evolution of technology, there exist populations of individuals lacking specific skills; as such this is also a significant continuing barrier to the implementation of Real-World Data. Lack of technological skills, on the other side, can also impede the adoption of digital multi-sided non-transactional platforms using Real-World Data. Wake et al [42] describe MyDiabetesMyWay (MDMW), an award-wining national platform for diabetes patients in Scotland. One of the major barriers to adoption found by researchers is the lack of technological skills and capabilities in using a computer and digital devices, which is causing some patients to be left behind from the usage of such platforms.

Another critical managerial issue relates to intellectual property rights, and the appropriation of benefits coming from Real-World Data among data owners, patients, and secondary users of such health data. Kalkman et al. [43], emphasize the need for establishing adequate systems for recognition, ownership and attribution, that are designed in such a way that due credit and acknowledgment is given to all who contributed to the results. To these principles between data holders and secondary users, researchers call upon the application of intellectual property (IP) laws to data access arrangements. According to some, policy should make sure to cover benefit sharing and IP issues as transparently as possible, and for it to be communicated appropriately. However, other sources point out that exclusive ownership runs counter to the goals of data sharing initiatives among various stakeholders. This would hold for individuals whose data is being shared but also for other actors involved in data sharing activities. A solution recommended includes inserting a “perpetuity” clause as a condition for making data available in a data sharing platform. The clause would only allow withdrawal of the data in case the grounds for making them available have changed.

According to most of the papers analyzed, however, data privacy along with the ethical and legal issues attached to it, are the major managerial challenges faced by platforms dealing with Real-World Data. According to Dang et al. [105], who conducted a research on IoMT applications in the healthcare sector, the development of these technologies leads to the sharp rise of cyber-attacks, so that hackers can exploit a system and aim for the most precious data. The information the hackers gain from attacking IoT medical devices helps them successfully infiltrate the hospital network or making devices malfunction and affect patient care. However, a collaboration between providers, vendors and security experts can prevent cyber-attacks by reinforcing standards and normalizing secure protocols. Thus, facilities that want to utilize IoT and cloud computing in healthcare must be fully aware of existing vulnerabilities and threats and design a security model to protect networks and devices from potential cyber-attacks.

Along with security of health data, some ethical issues are arising. Stoeklé et al. [106] state that platform dealing with Real-World Data and new technologies pose several ethical issues, some of which are already of importance, including the reasons for communicating with patients and how best to do so, particularly as concerns the biobanks, databases, genomic or bioinformatic processes, clinical trials and retrospective studies developed through these flows. Another ethical issue is represented by the risk of unequal access to such technologies on the basis of socioeconomic or genetic criteria.

Moreover, McKeown et al. [107] investigate the ethical and legal issues of privacy related to sharing and re-use of health data, where platform-based approaches require a new thinking about consent requested to patients, which should contain also some conditions of data reuse for research purposes.

3.5. Results of the Literature Review

3.5.1. Summary of the Literature

The Literature Review made for the purpose of this Thesis has been focused on the two main topics: Real-World Data and Multi-Sided Non-Transactional platforms.

Real-World Data (RWD) are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. They are therefore collected outside of traditional and more controlled setting like the Randomized Controlled Clinical Trail (RCT). Thanks to the peculiarities of Real-World Data, Life Sciences companies can today integrate a new complementary means of generating evidence on the effectiveness of their new products, therapeutics, and medical devices. This new type of evidence is called Real-World Evidence, which is the clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of Real-World Data.

Real-World Data can be collected by Life Sciences companies through a variety of sources, both primary and secondary, including patient registries; observational cohort studies; patient surveys; hybrid and pragmatic studies; retrospective databases and databanks; electronic health records and personal health records; patient generated health data and social media.

The advent of Real-World Data brings meaningful opportunities for Life Sciences companies, that come however with new challenges. Among the main opportunities, RWD are favouring real-time evidence-based medicine, real-time monitoring of patients, accelerated access and product development cycle of innovative products and innovative therapies, cost reductions for companies and personalized medicine. The new challenges introduced by RWD, however, include the possibility of incomplete data, bias and confounding within datasets, challenges with data access and data mining, lack of generally accepted methodological standards and lack of expertise due to the novelty of the subject matter.

Within the Real-World Data landscape, the new technologies being leveraged on by Life Sciences companies are various, but the three most relevant trends are: sensors, Internet of Medical Things and Artificial Intelligence. Many companies in the Life Sciences sector begun to invest their resources with the objective of collecting large Real-World Data databases and then using artificial intelligence to improve their processes of innovative product development.

The second major topic covered in the Literature Review was multi-sided non-transactional platforms and their business models. Multi-sided non-transactional platforms are multi-sided platforms where two or more sides are affiliated to the network, and the nature of the transactions among the sides are non-transactional, which means that no direct transaction is present between the parties involved. This type of multi-sided platforms is characterized by the presence of two or more groups of customers interacting in a non-transactional way thanks to the platform provider, and by the presence of strong indirect network externalities. The business model constituents that are characteristic of multi-sided non-transactional platforms and that have been analyzed in this Literature Review are the following:

- Value creation and value propositions. The driver of value creation for a multi-sided non-transactional platform can be both quantitative, with respect to the creation of direct and indirect network effects, and qualitative, with respect to product quality. To solve the chicken-and-egg problem, however, platform providers need to be able to create and maintain different value propositions for each side of the platform.

- Sides. A platform provider's decision on the number of sides that will be affiliated to the platforms is critical for its business model and comes with important trade-offs.
- Pricing. Strategic decisions about pricing can largely affect the revenue model and growth of the platform. The correct pricing strategy depends on many factors, but evidence has shown that there is a tendency towards subsidizing consumers sides and extracting value from business sides.
- Governance and control. Apart from indirect network effects, a strong driver of growth and hence a key strategic business model decision for platforms relates to whether to keep the platform closed or open, and to how to enforce control mechanisms aimed at ensuring overall quality.

At the intersection among the two concepts, the most advanced technological and business applications of RWD on multi-sided non-transactional platforms have been found in the literature to serve multiple purposes, which can be divided in two main categories:

1. Disease diagnosis, treatment, and drug development. Under this purpose, the platforms analyzed in the Literature Review act as a bridge among patients and companies who are trying to leverage on Real-World Data to better understand the evolution of certain diseases and develop new treatments accordingly.
2. Improving health research. Under this purpose, platforms and Real-World Data are means to improve the processes and the outcomes of health research, both in academic and business settings.

Focusing more on the business and managerial aspects, the main concerns related to Real-World Data in the context of platform utilization are privacy and ethical issues. With the rise of cyberattacks, and with the vulnerabilities of the existing technologies, managerial operators must design their platforms with the highest data security standards, complying with the existing regulations on the subject matter. Along with privacy, other ethical issues arising in terms of patient communications, equal access to advanced technologies and their benefits, and with health data re-use.

3.5.2. Research Gap

From the papers chosen and reviewed, it can be seen that the academia focused more on the study of multi-sided non-transactional platforms using Real-World Data only in the research and public fields. There's instead a lack of studies on such platforms applied in the private and industrial sectors. This research gap is filled and addressed by my Thesis, which wants to expand knowledge on how private companies are using the multi-sided platform business model to leverage on Real-World Data and new technologies for value creation in the healthcare industry.

Regarding the application of multi-sided non-transactional platforms in the public sector, Štufi et al. [108] describe, also from a technological perspective, a Big Data Analytics (BDA) platform powered by Artificial Intelligence to meet the requirements of the Czech Republic National Health Service and enabled by the usage of Real-World Data. The reported Big Data Analytics platform is transferrable to healthcare systems in other countries interested in developing or upgrading their own national healthcare infrastructure in a cost-effective, secure, scalable and high-performance manner.

Moreover, as another application example of platforms in the public sector, Tupasela et al. [109] analyzed the centralized healthcare data platforms developed in the European Nordics countries, especially Finland and Denmark. The mentioned countries have extensive nationally maintained and centralized registers, health data records, as well as, numerous biobanks, where the entire population becomes a study cohort. The Nordic countries have envisioned their platforms as a pervasive form in the organization of social activities to the extent that it has penetrated economic thinking as well. In these visions, the vast public data collection machineries, such as population registries, health data collected from primary and specialist healthcare services, are made increasingly available and productive for secondary purposes. In this platform economy approach, data from the population serves as the basis for secondary use by extractive industries.

Lastly, on public health management, Bo et al. [109] describe a service platform for college students' physical health to solve the storage, processing and mining of health Real-World Data. The experiment result shows that the platform can effectively complete the processing and analysis of the big data of College students' physical health, which has a certain reference value for college students' physical health monitoring during the COVID-19 epidemic. Especially for novel corona virus asymptomatic infections, the initial analysis of physical health data can help to detect the possibility of virus infection to some extent, and the platform addresses the lack of an effective storage, management, query and analysis present in traditional hospitals and other medical institutions.

At the intersection of research and public health management, in their systematic literature review Kalkman et al. [43] describe the Innovative Medicines Initiative's (IMI) BigData@Heart, an example of a consortium that is currently designing an international data sharing platform to stimulate drug development and personalized medicine for cardiovascular disease. This multi-stakeholder initiative has been funded to develop a data-driven translational research platform to improve patient outcomes and reduce the societal burden of specific disease areas in the European Union (EU). Such platform is enabled by the sharing, across the multiple stakeholders affiliated, of Real-World Data on EU patients' health, like medical and biomedical records.

As can be seen from the papers cited above, most of the research in academia has been focusing on public health applications of multi-sided non-transactional platforms leveraging on Real-World Data in the healthcare sector. There's therefore a lack of study on private for-profit platforms and their business models in the healthcare sector. This gap will be addressed by my thesis research, with the purpose of extending the academic knowledge on the topic mentioned.

3.5.3. Conclusions and Consideration

From the Literature Review on Real-World Data and Multi-Sided Non-Transactional platforms, we can draw some considerations that will drive the research methodology and the type of research performed.

Real-World Data, paired with the advancements in new technologies that allow data collection and manipulation, are becoming an increasingly used source of value to create evidence on the effectiveness of new medical products within the Life Sciences sector, and are contributing to meaningful improvements of the business processes of innovative product development for such companies, in terms of both effectiveness and efficiency.

Given the opportunities coming from the use of Real-World Data, however, some new challenges are emerging with respect to data governance, privacy and the correct usage of Real-World Data datasets.

In the context of Real-World Data usage, Multi-Sided Non-Transactional platform business models are emerging in the Life Sciences industry as a means to connect all the stakeholders and facilitate the collection and valorization of such patients' data. Such platforms are today able to collect, aggregate, and manipulate Real-World Data from various sources, allowing then companies and other stakeholders to access such data and valorize them in the pursuit of their own goals about innovative product development.

However, through the analysis of the Literature Review and the articles selected, it seems that the academia focused more on the study of such platforms in research and public contexts.

Additional research must be done in private business contexts for understanding how such platforms are helping Life Sciences stakeholders to collect and valorize Real-World Data. In particular, the peculiarities of the business models with which platforms pursue such objectives and the strategic business model decisions made should be analyzed to have a more complete understanding of the subject matter.

Using as a main theoretical framework the theory of platforms and their business models, the next section will outline the research methodology used to address the above-mentioned research question and to conduct the analysis.

4 Research Methodology

4.1. Theoretical Framework

4.1.1. Introduction

The analysis of the literature has been carried out focusing mainly on the following areas:

- Real-World Data. We defined Real-World Data and Real-World Evidence, specifying also which are the types and the different sources of Real-World Data currently available in the Life Sciences sector. We then clarified which are the opportunities and challenges associated with Real-World Data.
- New technologies emerging in the Real-World Data landscape. We identified and described the new emerging technologies that are fostering the development and utilization of Real-World Data, along with the incremental challenges and benefits they bring.
- Multi-sided non-transactional platforms. Starting from an overview of multi-sided platform, we clarified the characteristics and defining elements of multi-sided non-transactional platforms.
- Business model of multi-sided non-transactional platforms. Having defined their nature, we then looked at the main business model components of multi-sided non-transactional platforms, which influence the strategic decision of companies and the growth of the platform.

Stemming from the above-mentioned topics, a theoretical model has been developed to address the main research question of describing how multi-sided non-transactional platforms are allowing the collection and valorization of Real-World Data. In the following chapters, the theoretical model will be described in detail, in terms of the literature supporting the model, the rationale behind the creation of the model, and the main components of the model.

4.1.2. Theoretical Bases of the Model

As the literature review has focused on Real-World Data, new technologies in the landscape, and business model of multi-sided non-transactional platforms, in this section the main theoretical bases used for the construction of the model will be presented. It is important to clarify one main assumption underlying the model construction: a platform will be considered multi-sided non-transactional platform if, among all the types of transactions characterizing the multi-sided platform, at least one non-transactional relationship between the various sides exists. Therefore, the study will comprise multi-sided platforms that either present only non-transactional relationships, or more complex and hybrid platforms that allow for both transactional and non-transactional relationships.

In the following table, the main theoretical bases used for the development of the model will be presented, relating to each of the elements analyzed in the literature review:

Table 4.1: Theoretical bases for model development.

Area	**Theoretical Concepts Used for the Model**	**Authors and References**
Real-World Data	Real-World Data (RWD) are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. There may be many types of RWD, and the main classification distinguishes them among primary and secondary sources, each containing specific types of RWD. The main challenges brought by RWD are incompleteness of data, bias, data collection and manipulation procedures, and privacy issues. The benefits, instead, are real-time monitoring and patients, acceleration and cost reduction of innovative product development processes, and personalized medicine.	[3]; [4]; [10]; [44]
New technologies in the RWD landscape	Three main types of technologies have been identified as emerging within the RWD landscape. They are: (1) Sensor, and more specifically Wearables technologies; (2) Internet of Medical Things; and (3) Artificial Intelligence. These technologies enable the collection and manipulation of RWD, as to allow platforms to aggregate and valorize patient's health data coming from the real-world.	[87]; [35]; [38]
Multi-sided non-transactional platforms	Multi-sided platforms non-transactional platforms have two defining components. First, they are multi-sided, which means that there are two or more distinct sides affiliated to the platform. Second, the nature of at least one transaction among the affiliated sides is non-transactional, which means that the transaction is not present, or unobservable.	[55]; [57]; [62]; [64]

Business model of multi-sided non-transactional platforms	Four main components of the business model characterize multi-sided non-transactional platforms and affect the strategic decisions of the platform provider and the growth of the platform. The four components are: (1) value creation: managing different value propositions, enabling indirect network externalities, modes of value creation; (2) sides: number of sides involved, relationship with each of them and segmentation; (3) pricing: cross-subsidization, managing different pricing schemes for each side, and the factors affecting them; (4) governance: data governance, ensuring quality of the platform, open or closed ecosystem.	[65]; [66]; [67]; [71]; [73]; [75]; [78]

Stemming from the above analysis, therefore, the managerial and organizational aspects that the model will tackle and try to describe (directly or in-directly) are:

- RWD. Types of RWD analyzed by the platforms, its challenges and opportunities, and the technologies used in the process of collecting and valorizing RWD.
- Business model. Strategic decisions made by platforms providers with respect to the four main areas identified:
  - Value creation, value propositions and network effects.
  - Sides and stakeholders.
  - Pricing structure and strategies.
  - Governance. Data governance, and openness of the ecosystem.

4.1.3. Theoretical Bases of the Model

Following what said above and integrating all the major theoretical contributions identified for the creation of the theoretical framework, it is possible to draw a model that will be used to address the key research question of describing how multi-sided non-transactional platforms are facilitating the collection and valorization of Real-World Data, allowing their stakeholders in the Life Sciences sector to develop innovative medical products.

The structure and key components of the developed model are represented in Figure X. Three macro areas have been identified: Real-World Data, Platform Business Model, and finally Platform Sides. These macro areas represent the constituents of the model. At the centre of the model, the Platform Business Model macro area is the main component: a business model in fact allows a platform to create and capture value through the orchestration of external inputs (Real-World Data) and stakeholders (Platform Sides).

The model represented has been created by combining both existing theories and academic studies, which have been adapted to the context of this research, and also novel concepts developed from the Literature Review conducted, which have been created to fit the purpose of this research.

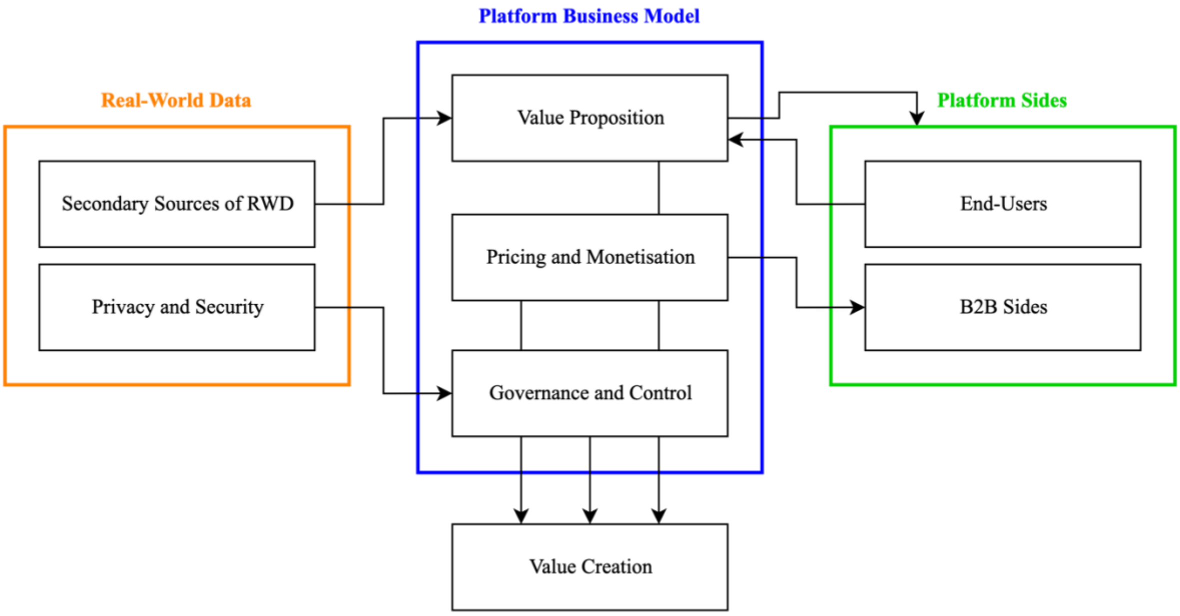


Figure 4.1: Research Model.

Following the Literature Review, the theoretical bases used, and the research model created, some hypotheses have been created as to address the research question and investigate on how multi-sided non-transactional platforms in the Life Sciences sector are creating value through Real-World Data and new technologies.

HYPOTHESIS H1: Multi-sided platforms use mainly Secondary Sources of RWD as part of their value proposition, as Secondary Sources of RWD are major enablers for value capture.

Hypothesis H1 finds its roots in the Literature Review performed, where a trend was identified such that the platforms studied have been mainly focused on the collection and valorization of different types of Secondary Sources of Real-World Data.

In their paper, as example, Chetan et al. [31] developed a big data analytics platform developed that allows to gain access to Twitter healthcare-related research data. As can be understood in the Literature Review, social media represent a novel and important category among the Secondary Sources of Real-World Data. In the article mentioned, the extraction of RWD from social media is a key value driver for the stakeholders using the platform itself.

As another example, Tupasela et al. [109] described in their paper the government owned platforms used by the Nordic countries in Europe, especially Finland and Denmark, which routinely collect patients' health data from various sources. These Secondary Sources of Real-World Data have been a major value driver for innovation and health research advancements in these countries, who were able to create longitudinal biobanks containing many useful health data on the population.

However, as per the main research gap found in the Literature Review, also in this context the academia has been focused mainly on platforms applied in research and public settings. As there's a lack of studies on private platforms active in the business environment, the hypothesis H1 was developed as to test on private for-profit platforms what was found in the literature for mostly public and academic platforms, related to using Secondary Sources of Real-World Data as part the platform's value proposition to the affiliated sides.

HYPOTHESIS H2: Multi-sided platforms in the Life Sciences industry need to formulate two different value propositions – one for the end-user side and one for the business sides.

In their paper, Muzellec et al. [73] state that the literature on multi-sided markets provides little information about the role of each of the sides affiliated to the platform. However, they propose that multi-sided platforms have to formulate different value propositions for B2B and B2B sides, as these sides have different needs, and both participate in the co-creation of value for the other side. These concepts have been used to formulate the hypothesis H2, with which the need for different value propositions has been tested in the context of the Life Sciences sector and Real-World Data.

Through the Literature Review, in fact, many examples of multi-sided non-transactional platforms providing different value propositions for B2C and B2B sides were found. Inn their paper, Vicdan H. et al. [111] describe PatientsLikeMe, a platform using social media technologies for social networking, participation, and collaboration among healthcare actors, and connecting over 750,000 patients suffering from severe or chronic illnesses and collecting self-reported health data about over 2,800 conditions. These data are then compiled by PatientsLikeMe and used for scientific and commercial research. Therefore, by analyzing the value propositions for the end users and for the business sides, it is evident that the platform formulated two different but correlated value propositions. For end-users, PatientsLikeMe enables patients to manage their care by tracking their disease's evolution through their profiles and a variety of self-reporting and datafication tools. For the business sides, instead, patient-reported data are continuously and systematically recorded, pooled, and shared with partners for medical research.

From the example above, and from others already described in the literature review, the hypothesis H2 was developed.

HYPOTHESIS H3: In multi-sided platforms in the Life Sciences sector, end-users are part of the value proposition for business customers.

In close connection with the previously developed hypothesis H2 and with what said above, hypothesis H3 finds its roots in the paper written by Muzellec et al. [73], which investigates not only on the presence of different value propositions, but also on the fact that end-users provide for a key part in the value proposition for business customers.

As a consequence of the reciprocal value propositions concept, in many instances the value that is being exchanged through the platform is access to data related to the end-users. In other words, the end users provide personal data in exchange for which they can use the service for free. From recipients of the value proposition on the B2C side, end-users become then the item being exchanged on the B2B side. Hence, end-users become a key component of the value proposition, for the B2B segment. This resides in the very nature of the end-user being an audience which can be monetized either because of its size, or because of the demographic, psychographic or behavioral characteristics of this audience.

Starting from this concept, the hypothesis H3 is developed by expanding the scope of the original proposition as to encompass also end users' health data (RWD), and testing the validity of the hypothesis also in the Life Sciences sector, which is intrinsically complex and highly regulated.

HYPOTHESIS H4: In multi-sided platforms in the Life Sciences sector, the monetization of the business model is B2B oriented.

As stated by Muzellec et al. [73], for many multi-sided platforms, the end-users do not reciprocate financially the value extracted from the platform. Often, end users are the recipients of a B2C value proposition for which they are not required to give any money in exchange. In this context, multi-sided platforms engage in a cross-subsidization strategy, where the monetization of the business model comes from extracting monetary value from the business sides, who generally have higher willingness to pay than end users. To achieve this goal, platform providers charge higher prices to one side of the platform to artificially lower prices for another side.

For instance, the TikTok platform applies cross-subsidization because the application is free for users, but advertisers pay to participate in the platform.

Despite a lack of research on the private for-profit platforms in the Life Sciences sector, a few examples emerged from the literature that empirically show a similar trend in the industry. As one of these examples, The PatientLikeMe platform mentioned above, and studied by Vicdan H. et al. [111], provides value for end users at no cost, while monetizing its business model from the business sides affiliated who are using Real-World Data to achieve their business and medical objectives.

Starting from these examples and from the proposition outlined by Muzellec et al. [73], the hypothesis H4 was developed.

HYPOTHESIS H5: Due to Privacy and Security concerns, multi-sided platforms in the Life Sciences sector develop a closed ecosystem where access is restricted.

The last hypothesis H5 finds its roots in the Literature Review and has been develop as novel. No similar theoretical model, hypothesis or proposition has been seen through the literature analysis performed. However, a major trend has been found related to Data Security challenges and Ethical issues for Real-World Data.

As Morrison et al. [112] point out, big health data and new technologies are transforming and reconfiguring the boundaries between patients, research participants and consumers, between research and clinical practice, and between public and private domains. New configurations of technologies, service providers and users are challenging existing regulatory categories, presenting novel opportunities and risks, and raising important ethical questions. Increased sharing of personal medical and biological information (Real-World Data) and increasingly international movements of these data raise issues of privacy and security. Moreover, technology is also posing some security concerns. In their paper, Dang et al. [105] raise privacy and security concerns about platforms in the healthcare sector leveraging IoT technologies to collect RWD. Even if valuable technologies, these can be target of hackers who can leverage on the still present weaknesses of IoT devices to provide serious harm, like for example infiltrating the hospital network or causing devices malfunctions that affect patient care.

The ethical and security issues outlined above have been coupled with the model developed by Boudreau K. [78], which investigates the degree and mode of openness of a platform. In this regards, hypothesis H5 has been created as to test whether the ethical and security issues which are characteristic of Real-World Data can cause platform providers to opt for a more closed platform, where access to it is controlled and restricted so as to protect data security.

Table 4.2: List of Hypotheses.

Hypothesis ID	Description
H1	Multi-sided platforms use mainly Secondary Sources of RWD as part of their value proposition, as Secondary Sources of RWD are major enablers for value capture.
H2	Multi-sided platforms in the Life Sciences industry need to formulate two different value propositions – one for the end-user side and one for the business sides.
H3	In multi-sided platforms in the Life Sciences sector, end-users are part of the value proposition for business customers.
H4	In multi-sided platforms in the Life Sciences sector, the monetization of the business model is B2B oriented.
H5	Due to Privacy and Security concerns, multi-sided platforms in the Life Sciences sector develop a closed ecosystem where access is restricted.

4.2. Research Design

As thoroughly presented in the above sections and in the theoretical framework, with this Thesis I will try to investigate on the impact of Real-World Data in the Life Sciences sector, by studying which are the emerging technologies that allow the use of Real-World Data, and how are multi-sided non-transactional platforms facilitating the extraction and valorization of Real-World Data.

As the theme under analysis is not yet fully mature, and RWD applications in business are at their early stages, the focus of my research will be on describing the phenomena happening within the environment, rather than trying to develop a new theory or establishing a cause-effect relationship among specific variables.

The research, in fact, has been designed as to allow me, the researcher, to analyze and describe comprehensively how is the Life Sciences industry evolving, and most of all what are multi-sided non-transactional platforms doing to allow Life Sciences companies to leverage on RWD in their processes to develop new medical products.

With all that in mind, my research method will mix both descriptive and observational methodologies, with qualitative data as the main fuel and enabler of the research.

For what regards the kind of knowledge my research aims to contribute, the research type is:

- Basic. As opposed to applied (developing new techniques or procedures), my research aims at expanding and developing new knowledge on the research question specified.
- Exploratory. As opposed to explanatory (defining causality inferences), my research aims at exploring and describing the main aspects of the research question under examination.

With regards, instead, to the types of data used and how will they be collected, the research type is:

- Primary. Since most of the data will be collected directly by me, the researcher, and the collection methodologies will be created specifically for this research, the study will be primary.
- Descriptive. Data will be gathered without controlling any variable, and the aim will be to find common patterns rather than causal inferences.

Within the research design, one important actor I have taken into consideration as a facilitator of this research is the Life Sciences Observatory of my home university, Polytechnic of Milan. The Observatory is a newly established branch of my home university, that focuses on managerial research within the Life Sciences sector, that business professionals and partners can use to understand the current trends in the industry and innovate their companies.

The Observatory has established sponsorship and partnership relationships with some companies, associations, and other actors within the Life Sciences industry, and some of them have been extremely valuable resources for my research. I had in fact been able to leverage on the Observatory's network of companies for my primary research, following the procedures explained in the next sections.

Putting all the above pieces together, the research methodology has been designed to fit with the research question under study. With this research, in fact, I would like to describe how platforms are allowing, through real applications and their business model, the use Real-World Data that Life Sciences companies use to develop new products. The new knowledge that I will bring will serve as an additional piece to form evidence of the power of RWD in the Life Sciences sector.

Given everything said above, the research methodology I have chosen is the exploratory multiple-case study. According to Yin [94], a case study is "an empirical enquiry that (i) investigates a contemporary phenomenon within its real-life context, especially when (ii) the boundaries between phenomenon and context are not clearly evident."

Therefore, Yin suggests to use the case study method if the researcher deliberately wants to cover contextual conditions, believing that they might be highly pertinent to the phenomenon under study. In an experiment, for instance, the researches does not consider the context and distinguishes it from the phenomenon under study, so that the attention can be focused on only few variables which are typically controlled. Because phenomenon and context are not always distinguishable in real-life situation, a new set of technical characteristics and data analysis strategies are needed, giving birth to the case study method. According to Yin, following from the above, the case study inquiry has three characteristics: (i) it copes with the technically distinctive situation in which there will be many more variables of interest than data points, and as one result (ii) relies on multiple sources of evidence, which data needing to converge in a triangulating fashion, and as another result (iii) benefits from the prior development of theoretical propositions to guide data collection and analysis.

According to Yin, the three main criteria I have used to choose the case study method have been:

- Type of research question. A basic categorization scheme for the types of research question is the following: “who”, “where”, “what”, “how”, “why”. In this context, “how” and “why” questions are more explanatory and likely to lead to the use of case studies.
- Extent of control over behavioral events. Here, the case study is recommended when there is no need to control behavioral events.
- Degree of focus on contemporary as opposed to historical events. According to this criterion, the case study is preferred in examining contemporary events, but when the relevant events and behaviors cannot be manipulated.

Below, a table from Yin’s paper is derived to show the framework used in the selection of research methodology:

Table 4.3: Research Methodologies comparison.

Strategy	**Form of Research Question**	**Requires Control of Behavioral Events?**	**Focuses on Contemporary Events**
Experiment	How, Why?	Yes	Yes
Survey	Who, What, Where, How many, How much?	No	Yes
Archival analysis	Who, What, Where, How many, How much?	No	Yes/No
History	How, Why?	No	No
Case study	How, Why?	No	Yes

As can be understood from the previous sections, the research question and characteristics of my study satisfy the criteria and requirements of the case study method, so that is why this specific methodology has been used as a research strategy comprising an all-encompassing method.

Going further, there can be multiple types of case studies. The first criterion to categorize case studies is the amount of case studies collected, where we can distinguish among (i) single-case studies and (ii) multiple-case studies. The second criterion is the type of research and application performed, among which we can distinguish the following categories:

- Explanatory. To explain causal links in real-life interventions.
- Descriptive. To describe an intervention and the real-life context of occurrence.
- Illustrative. To illustrate certain topics within an evaluation in a descriptive mode.
- Exploratory. To explore those situations in which the intervention being evaluated has no clear, single set of outcomes.

Given the peculiarities of my research questions and the intervention being studied, the type of case study chosen is the multiple-case exploratory study.

The case study was designed following the five components of research design outlined by Yin in his paper. The five components are the following:

- Study's question. The initial task is to clarify the nature of the study questions which are appropriate for the type of study being made, following the framework of "why", "what", "how", "where", "who".
- Study's prepositions. Following directly from the research question, each preposition created directs attention to somethings that should be examined within the scope of the study. The study's prepositions have the important function of helping the researcher in moving towards the right direction. As stated by Yin, exploratory case studies may not have prepositions, as long as the purpose of the exploration and the research are made clear.
- Unit of analysis. This part is aimed at defining what the case is. In this case, the study prepositions help in correctly defining and making clarity on the unit of analysis or units of analysis to be studied. The unit of analysis can be an individual, a company, an event, a specific policy, an industry, a country, or an entity that is less well defined than a single individual. Of course, each different type of unit of analysis will require a customized and appropriate data collection strategy.
- Logic linking data to the prepositions. The data collection strategy should be coherent with the prepositions made and the units of analysis selected. Data

should be collected and analyzed in a way which allows to verify and test the prepositions stated.

- The criteria for interpreting the findings. Once data have been collected and analyzed, some specific criteria have to be set in order to interpret the findings and understand whether the propositions made were tested or not. Clear criteria allow for better interpretations of the case studies.

The last key concept when designing the research for a case study is triangulation, which is a rationale for using multiple sources of evidence. One of the major strengths of the case study method is the possibility of using many different sources to generate evidence on the units of analysis. Given to its characteristics, the need to use multiple sources of evidence in the case study method is stronger than that the need in other research strategies. The data triangulation method will be further described in the data collection section, where the criteria used to conduct such triangulation will be outlined.

5 Data Collection and Analysis

5.1. Data collection strategy

The data collection strategy chosen follows the three principles of data collection outlined by Yin in his paper. The three principles are the following:

1) Triangulation: use multiple sources of evidence

Given the characteristics of the case study method, it is possible and is also recommended to use multiple sources of evidence to evaluate the propositions and conduct the exploration. The use of multiple sources of evidence in case studies allows the investigator to address a broader range of historical, attitudinal, and behavioral issues. However, the most important advantage presented by using multiple sources of evidence is the development of converging lines of inquiry. Thus, any finding or conclusion in a case study is likely to be much more convincing and accurate if it is based on several different sources of information. There are four different types of triangulations: (1) data triangulation; (2) investigator triangulation; (3) theory triangulation; and (4) methodological triangulation. The type of triangulation chosen for this research is the data triangulation method, which provides for the use of different data sources among the six sources specified below:

- Documentation. This type of if information can take many forms and shole be the object of explicit data collection plans. It may include letters, communications, announcements, written reports, articles, administrative documents, etc.
- Archival records. Unlike documentations, the usefulness of archival evidence can vary from case to case. Records may comprise service records, organizational records, maps and charts, survey data or personal records.
- Interviews. They are one of the most important sources of case study information. Case study interviews are most commonly of an open-ended nature. Another type of interview can be focused interview, where the interviewer is following a more strict set of questions and protocol.
- Direct observation. It is conducted by making a field visit to the online or physical site of the unit of analysis, to grasp some important behaviors or conditions about the unit under analysis.

- Participant-observation. It is a special mode of observation where the researcher is not merely a passive observer but assumes a variety of roles to participate in the events being studied.
- Physical artifacts. A physical or cultural artifact can be a technological device, a toll, an instrument, a work of art or some other types of physical evidence. However, they have less potential relevance in most of the case studies.

2) Create a study database

This second principle governs and relates to the way in which data collected are organized and documented for the case studies. Following these principles, the documentation collected will then consist of (1) the data or evidentiary database and (2) the reports of the investigator.

The study database is created under the form of a secure Dropbox folder, where only I, the researcher, have access.

3) Maintain a chain of evidence

An important principle to be followed to increase the reliability of the information in a case study is to maintain a chain of evidence. The principle states that the investigator has to allow external observers to follow the derivation of any evidence, ranging from initial research questions to ultimate case study conclusions, and also the steps followed in either direction.

5.2. Data sources

Among the six data sources specified by Yin, three of them were selected to implement the data triangulation strategy aimed at increasing the accuracy and power of the conclusions made.

The three data sources selected, which will be collected for each unit of analysis involved in the case studies, are: Documentations, Interviews, Direct Observations.

5.2.1. Documentations

Documentations from the companies involved in the case studies had served as a powerful integration to the interviews made, with the objective of verifying the content of the interviews through evidence coming from the documentations released by the unit of analysis in question. As provided for by Yin, the different types of documentations analyzed include press releases, research papers, communications materials published by companies, public administrative documents, letters, and documentations for public authorities in case the companies are publicly listed in the stock market.

Through such documentations, the following topics has been analyzed for the purpose of this research: (1) business model components of the company in question (value propositions, governance model, pricing, presence of network effects, stakeholders involved); (2) modes of collection and valorization of Real-World Data from patients; (3) types of technologies utilized to deliver on the value propositions proposed and to allow the collection and manipulation of Real-World Data.

5.2.2. Interviews

The interviews conducted with the companies under analysis for the case studies will be of semi-structured type. Semi-structured interviews are defined as "verbal interchanges where one person, the interviewer, attempts to explicit information from another persona by asking questions. Although the interviewer prepares a list of predetermined questions, semi-structured interviews unfold in a conversational manner offering participants the chance to explore the issues they feel are important." [80]

The questions prepared for the semi-structured interview and posed to each of the interviewees are the ones shared in the table below, presenting the framework used.

Table 5.1: Interview Questions.

Area of study	**Questions**
General Information	*"Can you please describe your company, and which is your role inside the organization?"*
Real-World Data: types of data collected, technologies, opportunities, and challenges	
Purpose and types of data collected	*"Which is the purpose / Which are the purposes of use of the patient's data that you collect?"* *"Which are the types of data you collect from patients, coming from real-world settings?"*
Technologies used to collect and manipulate patient's data	*"Which are the various technologies you leverage on to collect and manipulate patient's data? (E.g., wearables, sensors, EHRs, AI, etc.)"*
Opportunities and challenges of RWD	*"Which are the barriers and challenges you are facing in (1) collecting and (2) valorizing patient's data?"*

	"Which are the benefits and opportunities coming from patient's data for you and your stakeholders?"
Business Model components	
Governance rules	*"Are there any rules regulating access, governance and treatment of patients' data?"*
Platform's sides	*"Which are all the stakeholders interacting and engaging with your company? How are they interacting with your company?"*
Value creation and network externalities	*"Which are the value propositions you offer to each stakeholder engaging with your company? How do you manage to sustain them?"* *"How were you able to overcome the chicken-and-egg problem so that each stakeholder could have derived value by joining your ecosystem? Which stakeholder did you prioritize first to join the ecosystem and which were the incentives / strategies you used to bring them on?"*
Pricing strategy	*"How is the pricing strategy structured for each of the stakeholders mentioned before?"*

5.2.3. Direct Observation

As for the Documentation data source, the Direct Observation has been a data source used to further validate the statements of interviewees during the semi-structured interviews. In particular, the direct observation has been implemented through the analysis of companies' online websites, in order to analyze the statements and information shared publicly. As for such documentations, the following topics have been analyzed for the purpose of this research: (1) business model components of the company in question (value propositions, governance model, pricing, presence of network effects, stakeholders involved); (2) modes of collection and valorization of Real-World Data from patients; (3) types of technologies utilized to deliver on the value propositions proposed and to allow the collection and manipulation of Real-World Data.

5.3. Sampling and cases description

5.3.1. Sampling strategy

Given the characteristics of the research question I am pursuing, and the qualitative nature of my research, the sampling strategy used to generate the multiple case studies is the purposeful sampling. Purposeful sampling is one of the core distinguishing elements of qualitative inquiry, as qualitative inquiry typically focuses in depth on relatively small samples, even single cases (n = 1), selected purposefully. [81]

Purposeful sampling is part of the non-probability sampling methods, in which not all members of the population have an equal chance of participating in the study because the researcher selects samples based on subjective judgment, and are opposed to probability sampling. The latter method is defined as the selection of a sample from a population when this selection is based on the principle of randomization, that is, random selection or chance.

Purposeful sampling is a technique widely used in qualitative research for the identification and selection of information-rich cases for the most effective use of limited resources, and involves identifying and selecting individuals or groups of individuals that are especially knowledgeable about or experienced with a phenomenon of interest. [103]

In addition to knowledge and experience, Bernard [104] notes also the importance of availability and willingness to participate, and the ability to communicate experiences and opinions in an articulate, expressive, and reflective manner. In contrast, probabilistic or random sampling is used to ensure the generalizability of findings by minimizing the potential for bias in selection and to control for the potential influence of known and unknown confounders.

While the purpose of probability-based random sampling is generalization from the sample to a population, what would be 'bias' in statistical sampling, and therefore a weakness, becomes the intended focus in qualitative sampling, and therefore a strength. The logic and power of purposeful sampling lies in selecting information-rich cases for study in depth. Information-rich cases are those from which one can learn a great deal about issues of central importance to the purpose of the inquiry, thus the term purposeful sampling. Studying information-rich cases yields insights and in-depth understanding rather than empirical generalizations.

Stemming from the above concepts, purposeful sampling can be defined as a strategy that focuses on selecting information-rich cases whose study will illuminate the questions under study.

The advantages of purposeful sampling are:

- Economical. Due to its characteristics, purposeful sampling is less costly and less time consuming than other sampling methods. In fact, the flexibility of purposive sampling allows researchers to save time and money while they are collecting data. It offers a process that is adaptive as circumstance change, even if they occur in an unanticipated way. A researcher can meet multiple needs and interests while still maintaining the foundation of a singular focal point. Moreover, the non-random approach used allow to generate results that can then provide more information about future decisions that need to be made.
- Avoidance of irrelevant items. Due to the knowledge of the researcher applied in sampling, it prevents unnecessary and irrelevant items entering the sample per chance.
- Intensive study. It ensures intensive study of the selected items, which fits with the purpose of qualitative studies that want to investigate the cases in depth. This intensive study allows to better describe and explore a phenomenon.
- Lower margin of error of the information collected. When researchers approach a population group with a random survey, then the margin of error on their conclusions can be significant. Researchers achieve a lower margin of error using the purposive sampling approach because the information they collect comes straight from the source. Each person has identifiable characteristics that place them into the same demographic, avoiding to poll from a random sample.

The limitations of purposive sampling, instead, are:

- Bias. Purposive samples can be highly prone to researcher bias. The idea that a purposive sample has been created based on the judgement of the researcher may make it difficult to alleviate possible researcher biases, especially when compared with probability sampling techniques that are designed to reduce such biases. However, this judgmental, subjective component of purpose sampling is only a major disadvantage when such judgements are ill-conceived or poorly considered; that is, where judgements have not been based on clear criteria.
- Lower generalizability. The subjectivity and non-probability-based nature of unit selection (i.e., selecting people, cases/organizations, etc.) in purposive sampling means that it can be difficult to defend the representativeness of the sample. In other words, the criteria for selecting case-studies are subjective, and hence may differ from researcher to researcher. This causes the fact that the results obtained from the study may suffer from low theoretical, analytic, and logical generalizability. If, for example, a different

researcher would have used different criteria to select the cases, the results might have been different.

- Invalid for large population groups. Purposeful sampling is at its most effective when there are a limited number of individuals or units who possess the specific traits that are being studied. This is the case for the research conducted, as in the Life Sciences sector the number and diversity of multi-sided non-transactional platforms is limited.
- Reliability of the expert. It is usually very challenging to evaluate the reliability of the authority involved or the experts who are performing the purposive sampling. That means it is difficult to determine if there is a sampling error that is present in the information presented by the researcher, and there is also room to question the interpretation of the results.

Given its strength and limitations, purposeful sampling has been chosen due to the purpose and characteristics of the research question being studied.

Following the strategy of purposeful sampling, a database has been developed containing startups and corporations that match the following criteria: (1) collect, manipulate or share Real-World Data; (2) have a platform-based business model, respecting the criteria of multi-sided non-transactional platforms. For the objective of this thesis, it is worth clarifying again that the type of multi-sided non-transactional platform considered is the one that presents, among all the types of transactions characterizing the platform, at least one non-transactional relationship between the various sides. Therefore, the sample comprises platforms that either present only non-transactional relationships, or more complex and hybrid platforms that allow for both transactional and non-transactional relationships.

Under the principles of purposeful sampling, the final sample for the case studies has been selected from a database which I have created, containing a list of multi-sided non-transactional platforms active in the Life Sciences industry. The database is the result of a research where the multi-sided non transactional platforms have been identified and selected through the public information available on their websites or other public digital resources. The criteria to include in the database the companies found through my research were: (i) usage of Real-World Data, as defined in the Literature Review; (ii) platform business model, and the presence of the platform characteristics outlined in the Literature Review; (iii) presence of at least one non-transactional relationship among the sides affiliated to the platform; (iiii) operating in the Life Sciences industry.

The database resulting from this research and the criteria mentioned is the following:

Table 5.2: Database of multi-sided non-transactional platforms.

Company	Website	Employees	Country of Incorporation
Empatica	Website	50-100	USA
Evidation Health	Website	100-500	USA
Elysium	Website	1-10	Italy
PicnicHealth	Website	100-500	USA
Truveta	Website	100-500	USA
Hugo Health	Website	10-50	USA
Flatiron Health	Website	1000-5000	USA
Embleema	Website	1-10	USA
Prognos Health	Website	100-500	USA
Clinerion	Website	10-50	Switzerland
IQVIA	Website	50000-100000	USA
Clarivate	Website	5000-10000	USA
QArdio	Website	50-100	USA
Medtronic	Website	50000-100000	USA
MedM	Website	10-50	USA
Withings	Website	100-500	France
Vitls	Website	1-10	USA
100Plus	Website	50-100	USA
Huma	Website	100-500	London
Athelas	Website	100-500	USA

As can be seen, the database contains companies which differ in terms of scale, measured with the number of employees, and also country of incorporation, which also affect the market of activity of the company.

From the above database, four companies have been selected for conducting the multiple case studies of in-depth study. The companies sampled are also diverse as to represent the diversity of the database in terms of scale and country of incorporation. Following the sampling strategy principles described before, the final sample extracted from the research database is such that the companies selected are heterogeneous in terms of number of employees and country of incorporation. The final sample is composed by the following companies:

- Evidation Health:
  - Number of Employees: 100-500.
  - Country of Incorporation: USA.
- MedM.
  - Number of Employees: 10-50.
  - Country of Incorporation: USA.
- Elysium.
  - Number of Employees: 1-10.
  - Country of Incorporation: Italy.
- Withings.
  - Number of Employees: 100-500.
  - Country of Incorporation: France.

In the following sections, the four companies sampled will be presented, with a high-level analysis of their mission and vision, the characteristics of their platforms and business models, and the stakeholders involved in the ecosystem.

5.3.2. Case 1: Evidation Health

The first case study created for the purpose of this research regards Evidation Health. Evidation Health is a company incorporated in the USA, counting today more than 300 employees and active mainly in the US market. Evidation Health's mission is to enable and empower everyone to participate in better health outcomes. As can be understood through the company's website, Evidation measures health in everyday life and enables users to participate in health research and programs. Evidation's platform is used by millions of individuals, generating data coming from real world settings. Evidation partners with healthcare companies, public organizations and research entities to understand health and disease outside the clinic environment, and therefore in the real world.

Evidation Health is a non-transactional multi-sided platform as it interacts with various stakeholders (sides), that however do not engage directly into monetary transactions. Through Evidation Health's mobile application and platform, patients can monitor and store various types of health data, collected through wearables, sensors, EHRs, and other sources. Various stakeholders, among which pharmaceutical companies, technology companies, public organizations and research entities, can then access the health data shared by patients to achieve their business outcomes. The purposes of use of patient's Real-World Data can be various, as the company developed a diverse range of services for their paying customers, including research and development of new medical products, post-market surveillance, population research, development of health programs, etc.

Given to its business model, its platform's characteristics, and the use of Real-World Data, Evidation Health has been chosen as the first case study for the research.

5.3.3. Case 2: MedM

The second case study selected for the purpose of this research regards MedM. MedM is a company incorporated in the USA, counting today more than 25 employees and active mainly in the US and EU markets.

MedM's vision is to enable connected health for service providers, device vendors, developers, and for people. Their mission is to help improve the quality of care and patient satisfaction, while reducing also the overall cost of healthcare. To deliver on this mission, MedM's goal is to drive interoperability in the healthcare ecosystem, improving the speed of information exchange between all parties involved through their services and platform.

Through MedM's mobile app and software platform, more than 18 different types of Real-World Data can be collected from more than 600 Health IoT devices, peripheral sensors, and wearables. MedM's ecosystem comprises various actors, ranging from patients, healthcare service providers, device vendors and developers, who can interact in various ways: on one hand, patients can share their Real-World health Data with other users (peer-to-peer) or with their doctors; on the other hand, healthcare service providers can use the white-labeled software provided by MedM to autonomously manage the interaction with their patients and the collection of Real-World Data.

Given to its business model, its platform's characteristics and types of interactions, and the use of Real-World Data, MedM has been chosen as the second case study for the research.

5.3.4. Case 3: Withings

The third case study selected for the purpose of this research regards Withings. Withings is a company incorporated in France, counting today more than 300 employees and active mainly in the EU and US markets.

Withings' mission is to bridge the gap between patients and their care teams by continuously and effortlessly providing healthcare professionals with medical-grade data generated by patients from an ecosystem of connected devices.

Withings is adopting a different business model with respect to the other case studies under analysis, which is also a reason why it has been involved in the research. Withings is in fact able to monetize from all the sides affiliated to its platform. Withings, differently from the other case studies, also produces and sells its products to end users. On the consumer side, in fact, Withings offers a portfolio of products including connected scales, blood pressure monitors, an advanced sleep system, a smart temporal thermometer and hybrid smartwatches.

On the business side, instead, multiple types of stakeholders are affiliated, including research organizations, pharmaceutical companies, medical practices, and digital health programs organizers. Through its ecosystem of interconnected devices, Withings allows healthcare professionals to access patient's health data coming from the real world, enabling chronic disease prevention and management, remote patient monitoring, clinical research and more services.

Withings has therefore developed a complex ecosystem comprising multiple sides who are affiliated to its platform. Given to its business model, its platform's characteristics and types of interactions, and the use of Real-World Data, Withings has been chosen as the third case study for the research.

5.3.5. Case 4: Elysium

The fourth and last case study selected for the purpose of this research regards Elysium. Elysium is a company incorporated in Italy, counting today about 10 employees and active mainly in the EU markets.

Elysium's mission is to become the platform to quickly obtain accurate and suited data for medical and scientific research. The Real-World Data collected through Elysium's platform, in fact, are aimed at optimizing business strategies, speeding up clinical trials and finally developing innovative medical products.

Elysium's ecosystem comprises a variety of different sides, among which patients, research organizations, pharmaceutical companies, and insurance companies.

Real-World Data on patient's health are collected through Elysium's mobile application, called Medori, with which patients can manage their medical history, update their health status and insert their Real-World health data. The Elysium's data platform is then used by other sides of the platform to achieve their goals in terms of business strategy and innovative product development.

Given to its business model, its platform's characteristics and types of interactions, and the use of Real-World Data, Elysium has been chosen as the fourth case study for the research.

5.4. Data analysis strategy and method

Within the data analysis stage, all the three sources of evidence mentioned in the data collection section (documentation, interviews, observation) have been grouped and analyzed for each of the unit of analysis. In particular, the responses from the interviewees were recorded and fully transcribed. If any information was still unclear and/or more data was needed, the informants were later contacted by telephone to ask for clarification. For documentation and observation, the materials analyzed were already in a written format, so no transcription was made before the data analysis was initiated.

In this section, the data analysis strategy and method are described.

5.4.1. Within-case data analysis

Following the recommendations of Eisenhardt [82], a within-case data analysis was carried out to generate the necessary insight into the issues under scrutiny for each of the four companies. A subsequent cross-case analysis was implemented to make a comparison between the different documentations, observations, and responses given by the interviewees from the four different cases.

Concerning the within-case analysis, content analysis on the sources of evidence used was performed by implementing the practices stemming from the Grounded Theory methodology (Glaser & Strauss [83]; Strauss & Corbin [84]). The Grounded Theory methodology provides for 3 steps to follow for content analysis:

1) Open Coding

Open coding is a method suitable to study complex phenomena through a clearly defined procedure based on coding (labels, concepts and words) used to produce theory from interviews, rather than the mere finding of facts.

Open coding can be defined as the analytic process through which concepts are identified and their properties and dimensions are discovered in data. At a high level, during open coding, data are broken down into discrete parts, closely examined, and compared for similarities and differences. Events, happenings, objects, and actions/interactions that are found to be conceptually similar in nature or related in meaning are grouped under more abstract concepts termed "categories." Closely examining data for both differences and similarities allows for fine discrimination and differentiation among categories. During this process, conceptualization is used, where a concept is a labelled phenomenon. A concept is an abstract representation of an event, object, or action/interaction that a researcher identifies as being significant in the data. The purpose behind naming phenomena is to enable researchers to group similar events, happenings, and objects under a common heading or classification. Although events or happenings might be discrete elements, the fact that they share common characteristics or related meanings enables them to be grouped into concepts.

Following the first step of Open Coding, the texts coming from interview transcriptions have been analyzed. Subsequently, codes were created for each different concept through conceptualization. The codes of the interviews for each company, identified during this step of Open Coding, were iteratively contrasted and compared in order to group them into sets of first order concepts with the lowest level of abstraction.

2) Axial Coding

As second step of the Grounded Theory methodology, axial coding is the process of relating categories to their subcategories. It's called axial because coding occurs around the axis of a category, linking categories at the level of properties and dimensions.

In the course of doing open coding, however, an analyst might derive dozens of concepts. Therefore, the analyst needs to group certain concepts under a more abstract higher order concept, based on its ability to explain what is going on. That higher order concept is called a category. Grouping concepts into categories is important because it enables the analyst to reduce the number of units with which he or she is working. In addition, categories have analytic power because they have the potential to explain and predict.

Following this second step, therefore, the codes developed in the Open Coding through an analysis of interviews' transcriptions were organized and linked together into categories constituting the second-order concepts. The categories were used to draw connections among the codes found, grouping the ones which were similar in terms of properties, concept, and dimensions. The first order concepts were therefore further grouped around a set of second order themes or categories, allowing us to view the data at a higher level of abstraction (as specified in the paper of Clark et al. [85]).

3) Selective Coding

Selective coding is the last step in the Grounded Theory, consisting of connecting all the categories created during axial coding to their relative core categories. The core categories defined in this process are ones representing the pillars of the research and its main contribution to the literature.

In this step, the second order themes were grouped into overarching dimensions that captured the most important steps and constituent elements of the research performed, in light of the research question and theoretical framework defined.

Summing up, the Grounded Theory method was used to perform the within-case analysis and to analyze the different sources of evidence involved, first by labelling different words into concepts, and then grouping similar concepts into categories. This process has been structured following the above-mentioned instructions and definitions of concepts and categories.

5.4.2. Cross case data analysis

For what regards, instead, the cross-case data analysis, I have looked for similarities and differences between Cases 1, 2, 3, and 4 with reference to the first order concepts, second order themes and, above all, the overarching dimensions. This concluding procedure allowed to contrast and compare the way in which different multi-sided non-transactional platforms collect and valorize Real-World Data, for the benefits of all the sides affiliated to the platform.

The method used to conduct the cross-case analysis was the cross-care synthesis. This technique is relevant only for multiple case studies and makes the findings more likely to be robust. In this technique, at first each case study is treated individually (as explained and done in the within-case analysis), and then the findings are aggregated through specific criteria. In this case the theoretical framework created in the previous sections and the coding tree designed for the within-case analysis have served as criteria to aggregate the findings. For each section of the coding tree, the content has been compared against all the cases, to find similarities in terms of patterns and concepts. In this way, this analysis helped to understand whether the different cases shared some similarities or if there were any differences. The objective of this analysis is therefore to find patterns of concepts shared by the different unit of analysis involved in the multiple case study conducted.

6 Findings and Results

6.1. Within-case analysis results

6.1.1. Coding tree resulting from within-case analysis

After the analysis of all the sources of evidence employed, and the several coding techniques applied to the interviews according to the Grounded Theory (open, axial, selective), an inductive coding tree was built as a result. Fine-grained codes were transformed into aggregated concepts, and the real-world content obtained from the qualitative interviews enabled me to proceed with the abstraction aimed at addressing the research question under analysis.

The table below shows the coding tree resulting from for the findings. It depicts the three main dimensions that emerged from the analysis, as well as their constituent second-order themes, and the first-order concepts that led to the formation of these themes (middle and left side of the figure, respectively). The overarching emergent dimensions include Real-World Data, Technologies for Real-World Data, and Business Model components of multi-sided non-transactional.

Table 6.1: Coding tree structure of the results.

First Order Concepts	Second Order Concepts	Dimensions
Primary Sources of Real-World Data collected and analyzed	Types of Real-World Data collected and analyzed	Real-World Data
Secondary Sources of Real-World Data collected and analyzed		
Opportunities coming from Real-World Data		

Challenges coming from Real-World Data	Opportunities and challenges of Real-World Data	
Categories of technologies (AI, sensors, wearables, mobile apps, etc.)	Technologies used to collect Real-World Data	Technologies for Real-World Data
Categories of technologies (AI, sensors, wearables, mobile apps, etc.)	Technologies used to valorize Real-World Data	
Network Effects	Value Creation	Business Model components of multi-sided non-transactional platforms
Value Propositions		
Value creation mechanics		
Number of sides affiliated	Sides Management	
Types of sides affiliated		
Cross-subsidization strategy	Pricing	
Pricing scheme for each side		
Data governance and privacy	Governance	
Degree of openness of the platform		

Now that the inductive coding tree resulting from the within-case analysis has been presented, in the next section I have presented the coding process in greater detail. In particular, I have first presented the pieces of interview transcripts that were falling under the concepts and dimensions described in the table above. Secondly, I have then outlined a map showing all the codes identified during the coding process and their relationships with second-order concepts and the overarching categories.

6.1.2. Interviews' coding results

1) Case 1: Evidation Health

For Evidation Health, the documents analyzed to collect evidence and implement a data-triangulation strategy were:

- Documentation: various papers published by the company, and various company documents shared by the interviewer with which the interview was conducted, PR communications and social media announcements.
- Observation: website of the company.
- Interview: the interview was conducted via Zoom call, it was registered and transcribed. The interview has been conducted with Andrew Goldstein. Andrew is the Vice President of Business Development and Head of Sales of the company and manages the relationships with key stakeholders affiliated to Evidation's platform.

Putting together all the evidence sources, the results of the within-case study are the following:

Table 6.2: Case 1 data analysis.

1st Order Concept	Evidence from the sources analyzed	Codes
	Dimension: Real-World Data	
	Second Order Concept: Types of Real-World Data collected and analyzed	
Primary Sources of Real-World Data collected and analyzed	*"We also collect Patient Reported Outcomes (PRO) data, which are essentially surveys. These types of data are regularly collected in clinical trials."*	PRO
Secondary Sources of Real-World Data collected and analyzed	*"We connect to wearables devices that users may possess, we also connect to fitness apps like diet tracking or activity tracking. From them, we collect sleep data, steps data, fitness data, heart-related data, etc."* *"When we move users into a study environment, we connect patients data that we collect to other data, like EHR data, claims data, Bluetooth-enabled device data."*	PGHD, Claims Data, EHR

Second Order Concept: Opportunities and challenges of Real-World Data		
Opportunities coming from Real-World Data	*"Today, there's a lot of Real-World Data that are generated daily but are missed by the healthcare system, by only looking at the Real-World Data that is happening at the interaction point of patients with the healthcare system itself."* *"The benefit of PGHD is contextualization. You can understand the context of what is happening in real-life contexts between one interaction point and another interaction point with the healthcare system, where the system is already able to collect health data."*	Uncover new health data, Contextualization
Challenges coming from Real-World Data	*"One of the biggest challenges was to understand how to look at the huge amount of data collected through wearable devices and make sense of it. At first, companies were not able to understand what to do with those data."* *"We had to figure out how to parse health data in such a way that you can make up for missingness of data, and what missingness is acceptable or not acceptable."* *"We had to understand how to manipulate, treat, and deal with those data and how to deal with outliers, in terms of what outliers can be snipped off or not."* *"We had to understand how to look at and understand wearables data, clear out the noise, and then connect them with other sources of data in a meaningful way."*	Missing Data, Bias and Confounding, Data valorization
Dimension: Technologies		
Second Order Concept: Technologies used to collect Real-World Data		
Categories of technologies (AI, sensors,	*"We connect to wearables devices that users may possess, we also connect to mobile fitness apps like diet tracking or activity tracking."*	Wearables, APIs,

wearables, mobile apps, etc.)	*"We use APIs to connect to the devices created by other companies or by our sponsors to access patient's health data."* *"We've created a consumer engagement mobile app, where we serve content to patients to be engaged and to collect health data."*	Mobile app
Second Order Concept: Technologies used to valorize Real-World Data		
Categories of technologies (AI, sensors, wearables, mobile apps, etc.)	*"We've created a digital platform where sponsors can access patients' data, under the model of SaaS."* *"We can offer data analysis services for sponsors, where we use statistical methods and AI to analyze data."*	Digital platform, AI, Big Data Analytics
Dimension: Business Model components of multi-sided non-transactional platforms		
Second Order Concept: Value Creation		
Network Effects	*"We didn't prioritize any side to solve the chicken-and-egg problem. We prioritized the technology piece first. We developed first the technology piece of our platform, and we proved that it was clinically viable first to then scale the business. In this way there is value in the platform itself, even if the community doesn't exist. So, a third party could have used our service without our users, by recruiting its own participants outside the Evidation's environment."* *"The more participants we have in the community, the more costly is our community."*	No sides prioritization, Technology prioritization
Value Propositions	*"The ability to collect all those data in a platform allows for a far lower investment of energy for the participants, lowering the burden on them."* *"Our consumer mobile app is gamified, so that users gain points by reading certain contents, sharing data, or*	Gamification, End Users Data is part of B2B proposition,

	participating in surveys, and earn money through the points collected." *"We allow users, mainly those who are interested in their health or those who have diseases, to be part of research. And we give them easy access to it."* *"The value that our clients are getting is being able to make decisions on who medications are right for, making better recommendations to physicians and regulatory bodies as to when medications should be used, and what is experience that patients have that would lead them to the need of a medication or a therapy."* *"For technology companies, we have a ready cohort of participants to test things on in a quick way. Companies are also trying to prove that their technology is solving a health-related problem thanks to our cohorts."* *"For Life Sciences companies, we offer a way to collect evidence in a decentralized way, saving money, to understand the effect of their treatments on patients."*	Research contribution, Personalized medicine, One-stop-shop for health data, Cost reduction, New Product/Service development, Different Value Propositions for B2C and B2B users
Value creation mechanics	*"We enable the collection of a variety of different health data and the permissioning of the collection of those data from our participants (users), which are then used by, after collection, identification, and normalization of data, the different sponsors to answer specific healthcare- or business-related questions."*	Health data collection for business or medical purposes
Second Order Concept: Sides Management		
Number of sides affiliated	*7 sides, stemming from the "types of sides affiliated" section.*	7 affiliated sides
Types of sides affiliated	*"We've created an environment which is engaging with patients on one hand and engaging with many different types of sponsors on the other hand. Those sponsors are*	Pharmaceutical companies,

	pharmaceutical companies, non-for-profit organizations, foundations, research organizations, government agencies, and technological companies"	Nonprofit orgs, Foundations, Research orgs, Gov agencies, Tech companies, Users
	Second Order Concept: Pricing	
Cross-subsidization strategy	*"The more participants we have in the community, the more costly is our community. So, there's a cost to have that people on board."* *"The mobile app is free for users, and they also earn money thanks to the points accumulated, which are paid by our clients."*	Free service for users, B2B clients pay
Pricing scheme for each side	*"Technology companies and pharmaceutical companies are similar in terms of pricing structure. We have a licensing business model, like a SaaS, which is subscription based depending on the lengths and entity of the project."* *"Because we staff people that are clinical in nature and are able to conduct trials and run studies, we charge clients for the cost of their efforts in case they are needed by the sponsor."* *"Depending on the service requested by the sponsors, we can charge fees for the effort of clinical staff, fees for data analysis, and other pass-through costs. Those costs are additional to the standard pricing scheme of licensing of the platform."* *"We didn't stratify the client base in terms of pricing strategy, and we charge the same treatment to all of them. Sometimes, if we work with associations, we charge the fees*	Licensing, Subscription, Clinical services, No pricing differentiation

	additional to the platform licensing at cost, because we want to be part of that study for scientific and business reasons."	
	Second Order Concept: Governance	
Data governance and privacy	*"When a participant downloads our app, they don't agree for Evidation Health to share such data with third parties. At first, only Evidation can use such data internally. Whenever a third party wants access to health data, we always ask consent to the users another time after explaining how the data will be."* *"We tokenize data and anonymize them."* *"We've created a security mechanism for which data are safely stored."*	Consent required to share heath data with third parties, Data anonymization, Secure storage
Degree of openness of the platform	*The platform is closed as the patient's data are stored by Evidation Health, and nobody can access them without authorization of patients.*	Closed platform

2) Case 2: MedM

For MedM, the documents analyzed to collect evidence and implement a data-triangulation strategy were:

- Documentation: various papers published by the company, and various company documents shared by the interviewer with which the interview was conducted, PR communications and social media announcements.
- Observation: website of the company.
- Interview: the interview was conducted via Zoom call, it was registered and transcribed. The interview has been conducted with Victoria Krasilshikova. Victoria is the Chief Ecosystem Officer of the company and is in charge of coordinating all the stakeholders participating in the ecosystem and making sure that MedM's ecosystem develops uniformly.

Putting together all the evidence sources, the results of the within-case study are the following:

Table 6.3: Case 2 data analysis.

1st Order Concept	Evidence from the sources analyzed	Codes
Dimension: Real-World Data		
Second Order Concept: Types of Real-World Data collected and analyzed		
Primary Sources of Real-World Data collected and analyzed	*Not treated.*	
Secondary Sources of Real-World Data collected and analyzed	*"We collect 18 different types of data. It's all vital signs. It can be blood glucose, temperature, weight, etc."*	PGHD
Second Order Concept: Opportunities and challenges of Real-World Data		
Opportunities coming from Real-World Data	*"There's going to be a massive amount of patient's data collected overtime because never before have we been able to so easily collect health data, from healthy people and from different geographies, and with different lifestyles and backgrounds."* *"The healthcare system has the possibility to really shift from just hearing about the disease occurred to understand how they occur and then improve prevention."*	Disease understanding and prevention, Data collection
Challenges coming from Real-World Data	*"It's technically and legally hard to let all the stakeholders exchange data among each other. There's a lot of ambiguity and complexity."* *"There's no accepted international standard. So all the different devices are outputting data in their own format."* *"Everyone wants to create their own closed ecosystem, so*	Data sharing, Lack of international standards, Lack of regulation,

	that a user has to use branded devices, register data in the cloud owned by the company and keep the data there. Most RPM providers and device vendors are trying to keep their data for themselves." *"There needs to be more regulation on patient's health data sharing and storing."* *"It's hard to convince some device vendors that they should collaborate with us and we are not a threat to them."* *"Some healthcare institutions are scared to try and adopt new solutions. They are stressed out thinking about learning how to deal with a new system."* *"Some devices are expensive to produce, and sometimes they do not work properly. We need to make them cheap, and we need to make them work."*	Stakeholders Collaboration, Lack of expertise, Technological challenges, Costs
Dimension: Technologies		
Second Order Concept: Technologies used to collect Real-World Data		
Categories of technologies (AI, sensors, wearables, mobile apps, etc.)	*"We have integrated more than 600 different medical sensors that are used at home by patients who want to collect and record their data."* *"We made a mobile app that allowed users to automatically collect data from devices, capture them, and either store them in their phones or in the cloud."* *"We're currently Bluetooth protocols and we are working on integrating Wi-Fi devices."*	Sensors, Mobile app Bluetooth protocols, Wi-Fi devices
Second Order Concept: Technologies used to valorize Real-World Data		
Categories of technologies (AI, sensors,	*"For businesses, we offer a white labeled software for Remote Patient Monitoring providers. We are not an RPM provider, but we offer a ready-made solution that an RPM provider can purchase to use our SaaS with a license."*	White labeled software

wearables, mobile apps, etc.)	*"Through our white labeled RPM SaaS, providers can channel the data gathered to an EHR, or any other health system."*	
Dimension: Business Model components of multi-sided non-transactional platforms		
Second Order Concept: Value Creation		
Network Effects	*"For our solution to work for our clients, there needs to be a critical number of patients, which is at least 500 patients."*	Patients side prioritization
Value Propositions	*"For businesses, we offer a white labeled solution for Remote Patient Monitoring providers. We are not an RPM provider, but we offer a ready-made solution that an RPM provider can purchase to use our SaaS with a license."* *"In our B2C arm, users can share their health data to other users (peer-to-peer) or also to their doctors."* *"With MedM, users are no longer tied to a specific device vendor, because users can control their own data fully, and also choose not to have a backup in someone else's cloud. Or they can use MedM to move their data from an Apple watch to a Samsung Watch for example."* *"With MedM, doctors are able to see and understand how the disease is going, how the patient is reacting to certain medical treatment, and they can be alerted when there are some irregularities."*	Remote Patient Monitoring, End Users Data is part of B2B proposition, Health data sharing, Data control, Disease understanding and prevention, Different Value Propositions for B2C and B2B users
Value creation mechanics	*"MedM is a software company which is positioned in the middle of the healthcare ecosystem as far as patient's data is concerned."*	Being at the center of the healthcare ecosystem
Second Order Concept: Sides Management		

Number of sides affiliated	*4: users, doctors, device vendors, RPM providers.*	4 affiliated sides
Types of sides affiliated	*"We have many sides. We are in between users (patients), the device vendors who actually make the medical devices, and then the care providers. They all have to exchange information among each other."* *"Users can share their health data to other users (peer-to-peer) or also to their doctors."*	Users, Device vendors, RPM providers, Doctors
	Second Order Concept: Pricing	
Cross-subsidization strategy	*"Our B2C solution is entirely free, we make money from the B2B version."* *"The free B2C solutions allows us to keep an eye on the market and understand the needs of customers."*	Free service for users, B2B clients pay
Pricing scheme for each side	*"For businesses, the white labeled RPM SaaS is paid through a license for use."* *"The pricing is on a per patient per month structure."*	Licensing, Subscription
	Second Order Concept: Governance	
Data governance and privacy	*"In our B2C arm, the data belong to end users. They can delete it at any time, they are anonymized and twice encrypted."* *"We do not as of now analyze any patient's data, as we are not sure how ethical this is."* *"Patients can share their own data with other users, doctors, or RPM providers only with their permission."* *"User can invite other to access their health data, grant access if requested, and revoke that access if no longer needed."*	Data encryption, Consent required to share heath data with third parties, GDPR, HIPAA

	"We abide to the GDPR in Europe, which says that users need to be able to understand the uses of their data, where they are stored, and must be also able to delete them." *"In the US, we abide to the HIPAA rules, where data has to be shared in secure servers."*	
Degree of openness of the platform	*The platform is closed, and nobody can access patients' data without authorization of patients.*	Closed platform

3) Case 3: Withings

For Withings, the documents analyzed to collect evidence and implement a data-triangulation strategy were:

- Documentation: various papers published by the company, and various company documents shared by the interviewer with which the interview was conducted, PR communications and social media announcements.
- Observation: website of the company.
- Interview: the interview was conducted via Zoom call, it was registered and transcribed. The interview has been conducted with Vincent Vercamer. Vincent is the Head of Market Access and Public Affairs of the company. He is in charge of managing and developing partnerships with key stakeholders affiliated to the platform and managing the relationships with healthcare professionals.

Putting together all the evidence sources, the results of the within-case study are the following:

Table 6.4: Case 3 data analysis.

1st Order Concept	Evidence from the sources analyzed	
Dimension: Real-World Data		
Second Order Concept: Types of Real-World Data collected and analyzed		
Primary Sources of Real-World	*Not Treated.*	

Data collected and analyzed		
Secondary Sources of Real-World Data collected and analyzed	*"Withings devices allow tracking more than 20 data points through a complete ecosystem of connected devices."* *"Withings devices can collect data about weight, sleep, temperature, hearth rate, oxygen saturation level, ECG, blood pressure, BMI and body fat mass."*	PGHD
Second Order Concept: Opportunities and challenges of Real-World Data		
Opportunities coming from Real-World Data	*"User data drives Withings innovation and product improvements from the beginning."* *"For example, when we launched the world first connected scale in 2009, we discovered that out users usually weight themselves around 7:30 but usually check their weight history in the app around 11:30. This tells a lot about cognitive mechanisms of someone to their own health."* *"We have also published research with researchers and doctors during the COVID-19 pandemic to better understand human health."*	Health research improvement, Business and product innovation
Challenges coming from Real-World Data	*"Common understanding of GDPR and other regulations by all stakeholders is a real challenge, we are doing our best when negotiating research partnerships but we often loose months talking with legal departments of universities because of misunderstanding of the definition and applications of the regulations. Everyone has its own interpretation of the law in addition to custom policy."*	GDPR, Lack of common interpretation of regulations
Dimension: Technologies		
Second Order Concept: Technologies used to collect Real-World Data		

Categories of technologies (AI, sensors, wearables, mobile apps, etc.)	*"The Withings portfolio includes connected scales, blood pressure monitors, an advanced sleep system, a smart temporal thermometer and hybrid smartwatches. It also includes data connectivity options (such as cellular devices) as well as a remote patient monitoring platform."* *"We use wearable sensors (watches, Blood Pressure Monitors, thermometers) and non-wearable sensors (under-the-mattress Sleep Analyzer, or Scales)."*	Sensors, Digital platform
Second Order Concept: Technologies used to valorize Real-World Data		
Categories of technologies (AI, sensors, wearables, mobile apps, etc.)	*"We are experts in ML and edge-computing, so all our AI algorithm are running "on edge" with neural networks and other ML technologies directly embedded into products and executed on small MCU (low power, low battery consumption)."* *"We do big data analysis to constantly improve the products, medical knowledge and our user's health."*	AI, Machine Learning, Big Data analytics, Edge-computing
Dimension: Business Model components of multi-sided non-transactional platforms		
Second Order Concept: Value Creation		
Network Effects	*Not present, as different and independent value propositions are offered to the stakeholders interacting with the company.*	No network effects
Value Propositions	*"We serve healthcare professionals across chronic disease prevention and management, remote patient monitoring, clinical research and more."* *"Withings Health Solutions extends this expertise to the healthcare industry to remove friction in the patient's journey and allow for telehealth to expand."*	Disease understanding and prevention, End Users Data is part of B2B proposition,

	"For patients, the main purpose is to make the data available to for them to take control of their health by having a broader and longer view of their health data."	Disease management, Remote Patient Monitoring, Telehealth, Health control, Different Value Propositions for B2C and B2B users
Value creation mechanics	*"Our missions is to bridge the gap between patients and their care teams by continuously and effortlessly providing healthcare professionals with medical-grade data generated by patients from an ecosystem of connected devices."*	Health data collection for business or medical purposes
Second Order Concept: Sides Management		
Number of sides affiliated	*6: DHP providers, RPM providers, CROs, Pharmaceutical companies, researchers, users.*	6 affiliated sides
Types of sides affiliated	*"We work with Digital Health Programs providers, Remote Patient Monitoring providers, Pharmaceutical companies, Clinical Research Organizations (CROs), Researchers."*	Digital Health Program providers, RPM providers, Pharma companies, CROs, Research orgs, Users
Second Order Concept: Pricing		

Cross-subsidization strategy	*Not present, because users pay for the smart devices bought on Withings e-commerce.*	No cross-subsidization strategy, No free service for users
Pricing scheme for each side	*"For the B2C, users pay for devices which are in the high quality & high prices categories. Mobile app & cloud services are free forever. Additional health programs or services could be proposed."* *"For the B2B, we offer volume discount for big orders and additional services regarding logistics and data connectors like API, SDK, Data Hub, and cellular connectivity."* *For the healthcare professionals, we offer the MED PRO CARE for doctors, who pay a monthly fee per active patient and additional costs. per shipped devices to the patients."*	Pay per device, Licensing, Subscription
Second Order Concept: Governance		
Data governance and privacy	*"We have the ISO13485 and ISO27001 certifications."* *"We have internal policies regarding employee training, access to patient data and pseudonymization or anonymization processes."* *"Beyond the main purpose of patient's data, such data are also used for (i) average or categorization of our users data to help them comparing their results to their pairs; (ii)communicating with customer support; (iii) improving our products and services; (iv) marketing, research and recommendations; (v) clinical research project with universities.*	ISO certifications, Data anonymization, Consent required to share heath data with third parties

Degree of openness of the platform	*The platform is closed, and nobody can access patients' data without authorization of patients.*	Closed platform

4) Case 4: Elysium

For Elysium, the documents analyzed to collect evidence and implement a data-triangulation strategy were:

- Documentation: various papers published by the company, and various company documents shared by the interviewer with which the interview was conducted, PR communications and social media announcements.
- Observation: website of the company.
- Interview: the interview was conducted via a Microsoft Teams call, it was registered and transcribed. The interview has been conducted with Ahmed Abdel Rahman. Ahmed is the CEO of the company. He is in charge of making major corporate decisions, managing the overall operations and resources of a company, acting as the main point of communication between the board of directors and corporate operations, and being the public face of the company.

Putting together all the evidence sources, the results of the within-case study are the following:

Table 6.5: Case 4 data analysis.

1st Order Concept	**Evidence from the sources analyzed**	**Codes**
Dimension: Real-World Data		
Second Order Concept: Types of Real-World Data collected and analyzed		
Primary Sources of Real-World Data collected and analyzed	*Not treated.*	

Secondary Sources of Real-World Data collected and analyzed	*"We extract information from patient's health data, that means they take a photo of their health papers, or they upload already digital medical data on the mobile app. We then extract health data from those documents as we label the exact information contained in the document depending on the patient's health situation (rare disease, chronic disease, or acute damage) and the researcher's needs."*	EHR, PGHD, Health papers
Second Order Concept: Opportunities and challenges of Real-World Data		
Opportunities coming from Real-World Data	*"We are collecting submerged information from patients. Patients have a lot of health data which are no more possessed by hospitals, because after 10 years they must delete and cancel such data. That means that if I don't have any copy of my documentation, such health data will be missed. That's why we collect data directly from patients. Those data allow to have a clear overview of the patients."* *"Another benefit from health data is new, automated, and better services for healthcare."*	Submerged health data collection, Business and product innovation
Challenges coming from Real-World Data	*"The most important barrier is the normative barrier. We have GDPR in Europe and HIPAA in the US. We are now in Italy and we are collecting data in Italy, where we have another framework other than GDPR. So, the most important barrier is related to the different regulation frameworks in each country or region. But it is ok for us because we want to preserve ethics and privacy of our customers."* *"Another barrier is to protect information of users, because there are hackers, so we just protect their valuable data."*	Regulations, GDPR, HIPAA, Data security
Dimension: Technologies		
Second Order Concept: Technologies used to collect Real-World Data		

Categories of technologies (AI, sensors, wearables, mobile apps, etc.)	*"We use a mobile app, Medori, through which we collect health information via screenshot made by users or via digital images uploaded by them."*	Mobile app
Second Order Concept: Technologies used to valorize Real-World Data		
Categories of technologies (AI, sensors, wearables, mobile apps, etc.)	*"We then extract health information from images thanks to OCR in the first instance, then we label automatically the data thanks to HL7 and other vocabularies. Then, we use supervised learning to automate and complete the process."*	OCR, Supervised learning, AI
Dimension: Business Model components of multi-sided non-transactional platforms		
Second Order Concept: Value Creation		
Network Effects	*"The chicken-and-egg problem is a huge problem we faced. B2B customers are there, and they need the data we provide as submerged data are not in the market."*	Patients side prioritization
Value Propositions	*"The purpose to collect health information from patients is to provide the right information to improve clinical and scientific research."* *"We provide to the patient the Medori app, in which they can visualize their health data. Not only because you can collect the data, but also because they can visualize the data so they can understand the behavior of their disease."* *"For B2B, the value proposition is to give opportunity to access submerged data, we are squeezing the time needed to collect them by 53% and we are hence reducing the cost."*	Health research improvement, End Users Data is part of B2B proposition, Disease understanding and prevention, Submerged health data access, Cost reduction, Different Value

		Propositions for B2C and B2B users
Value creation mechanics	*"The purpose to collect health information from patients is to provide the right information to improve clinical and scientific research."*	Health data collection for business or medical purposes
Second Order Concept: Sides Management		
Number of sides affiliated	*The number of sides affiliated are: users, pharma companies, CROs, and insurance companies.*	4 affiliated sides
Types of sides affiliated	*"We clustered patients into three groups: patients with chronic diseases, rare diseases, and with acute damages."* *"We have two different customers. The first is the patient, which is a data provider. On the other side, we have pharma companies, insurance companies, and CROs."*	Users, Pharma companies, CROs, Insurance companies
Second Order Concept: Pricing		
Cross-subsidization strategy	*"We provide to the patient the Medori app, which is completely free. Pharma companies, insurance companies, and CROs pay for the service"*	Free service for users, B2B clients pay
Pricing scheme for each side	*"Pharma companies need to do their research and they enter out platform and they search for keywords."* *"We are still working on finding the right pricing model for B2B sides."*	Pricing model under discovery
Second Order Concept: Governance		

Data governance and privacy	*"We have GDPR in Europe and HIPAA in the US. We are now in Italy and we are collecting data in Italy, where we have another framework other than GDPR."* *"We are not profiling our customers at all. We know the health data but we don't know who is the user, because we cannot connect the specific customer with the health data."*	GDPR, HIPAA, Data anonymization
Degree of openness of the platform	*The platform is closed, and nobody can access patients' data without authorization of patients and platform provider.*	Closed platform

6.1.3. Data structure of the results

As a result of the coding process conducted on all the interview transcripts and described in the sections above, the figure below shows the data structure summarizing all the findings. As can be seen, the result of the open coding step is shown on the left column, where all the codes related to the first-order concepts are listed. In the central column, the result of the axial coding is presented, where the second order concepts represent the categories that link together the similar codes found in the previous step. Lastly, the right column is dedicated to the overarching categories resulting from the selective coding, where the categories related to the same dimensions have been aggregated.

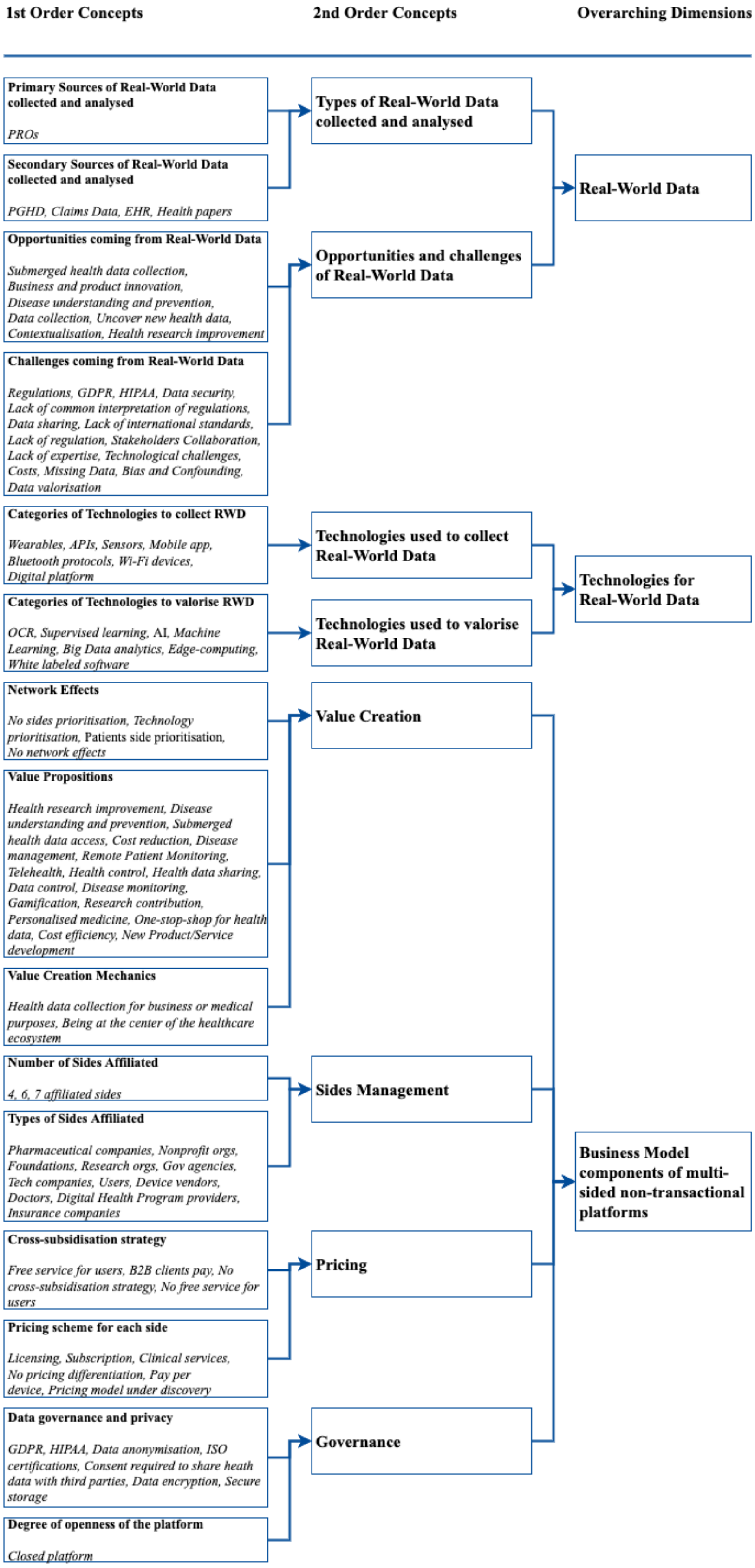


Figure 6.1: Data Structure of the Results.

6.2. Cross-case analysis results

6.2.1. Coding results from cross-case analysis

A cross-case comparison was also performed to complement the within-case analysis and underscore the main similarities and differences between the four cases and search for any patterns followed by the multi-sided non-transactional platforms in collecting and valorising Real-World Data for the benefits of the Life Sciences stakeholders.

The cases were compared with reference to the first order concepts, second order themes and, most importantly, overarching dimensions, following the framework outlined in the previous sections. The objective has been to identify any possible pattern match or mismatch. This goal was operationally achieved by merging the coding trees of the four cases and comparing them to find for similarities and differences.

More specifically, each code and concept found for a single case thanks to the within-case analysis has been related to the codes and concepts found for the other cases, in pursuit of some commonalities or differences which can shed light on the research question under study. The objective of this analysis is to offer a qualitative description of how such companies are operating, in terms of Real-World Data usage, the technologies involved, and the business model characteristics of such companies.

The results and findings of the cross-case analysis performed have been reported in the table below, where each first order code found was compared against the four case studies involved in the research.

Table 6.6: Cross-case analysis results.

Codes	Evidation Health	MedM	Withings	Elysium
Dimension: Real-World Data				
Second Order Concept: Types of Real-World Data collected and analyzed				
1st Order Concept: Primary Sources of Real-World Data collected and analyzed				
PROs	X			
1st Order Concept: Secondary Sources of Real-World Data collected and analyzed				

PGHD	X	X	X	X
Claims Data	X			
EHR	X			X
Health papers				X
Second Order Concept: Opportunities and challenges of Real-World Data				
1st Order Concept: Opportunities coming from Real-World Data				
Submerged health data collection				X
Business and product innovation			X	X
Disease understanding and prevention		X		
Data collection		X		
Uncover new health data	X			
Contextualization	X			
Health research improvement			X	
1st Order Concept: Challenges coming from Real-World Data				
Regulations				X
GDPR			X	X
HIPAA				X
Data security				X
Lack of common interpretation of regulations			X	
Data sharing		X		

Lack of international standards		X		
Lack of regulation		X		
Stakeholders Collaboration		X		
Lack of expertise		X		
Technological challenges		X		
Costs		X		
Missing Data	X			
Bias and Confounding	X			
Data valorization	X			
Dimension: Technologies for Real-World Data				
Second Order Concept: Technologies used to collect Real-World Data				
1st Order Concept: Categories of technologies (AI, sensors, wearables, mobile apps, etc.)				
Wearables	X			
APIs	X			
Sensors		X	X	
Mobile app	X	X		X
Bluetooth protocols		X		
Wi-Fi devices		X		
Digital platform				X
Second Order Concept: Technologies used to valorize Real-World Data				
1st Order Concept: Categories of technologies (AI, sensors, wearables, mobile apps, etc.)				

OCR				X
Supervised learning				X
Machine Learning			X	
AI	X		X	X
Big Data analytics	X		X	
Edge-computing			X	
White labeled software		X		
Dimension: Business Model components of multi-sided non-transactional platforms				
Second Order Concept: Value Creation				
1st Order Concept: Network Effects				
No sides prioritization	X			
Technology prioritization	X			
Patients side prioritization		X		X
No network effects			X	
1st Order Concept: Value Propositions				
Health research improvement				X
Different Value Propositions B2C and B2B	X	X	X	X
Disease understanding and prevention		X	X	X
Submerged health data access				X
Cost reduction	X			X

Disease management			X	
End Users Data is part of B2B proposition	X	X	X	X
Remote Patient Monitoring		X	X	
Health control			X	
Health data sharing		X		
Data control		X		
Telehealth			X	
Gamification	X			
Research contribution	X			
Personalized medicine	X			
One-stop-shop for health data	X			
New Product/Service development	X			
1st Order Concept: Value Creation Mechanisms				
Health data for business or medical purposes	X		X	X
Center of the healthcare ecosystem		X		
Second Order Concept: Sides Management				
1st Order Concept: Number of Sides Affiliated				
4		X		X
6			X	
7	X			

1st Order Concept: Types of Sides Affiliated				
Pharmaceutical companies	X		X	X
Nonprofit orgs	X			
Foundations	X			
Research orgs	X		X	
Gov agencies	X			
Tech companies	X			
Users	X	X	X	X
Device vendors		X		
Doctors		X		
Digital Health Program providers			X	
RPM Providers		X	X	
Insurance companies				X
CROs			X	X
Second Order Concept: Pricing				
1st Order Concept: Cross-subsidization strategy				
Free service for users	X	X		X
B2B clients pay	X	X	X	X
No cross-subsidization strategy			X	
No free service for users			X	
1st Order Concept: Pricing scheme for each side				

Licensing	X	X		
Subscription	X	X		
Clinical services	X			
No pricing differentiation	X			
Pay per device			X	
Pricing model under discovery				X
Second Order Concept: Governance				
1st Order Concept: Data Governance and Privacy				
GDPR		X		X
HIPAA		X		X
Data anonymization	X		X	X
ISO certifications			X	
Consent to share heath data with third parties	X	X	X	
Data encryption		X		
Secure storage	X			
1st Order Concept: Degree of openness of the platform				
Closed Platform	X	X	X	X

6.2.2. Hypotheses evaluation

After a cross-case analysis, it was possible to test the hypotheses developed from the research model build in the previous sections.

The table below summarizes the results, where all the hypotheses created are verified from the cross-case analysis. The method used to test each hypothesis relies on the identification of a particular code that is able to explain the hypothesis. For each code, its frequency was taken into account and taken from the cross-case analysis, to check for how many companies out of the total sample that specific code was mentioned in the within-case analysis of the interviews. Finally, based on the code frequency, the hypothesis was either deemed as supported or not supported.

As we can see below, all the hypothesis developed are supported.

Table 6.7: Hypotheses Testing from cross-case analysis.

Hypothesis ID	Code	Frequency	Hypothesis
H1	PGHD	100%	Supported
H2	Different Value Propositions B2C and B2B	100%	Supported
H3	End Users Data is part of B2B proposition	100%	Supported
H4	B2B clients pay	100%	Supported
H5	Closed Platform	100%	Supported

7 Discussion

7.1. Discussion of the results

As stated in the proper section, the objective of the research is to describe how multi-sided non-transactional platforms are allowing the collection and valorization of Real-World Data, creating and capturing value in the Life Sciences sector. In particular, the research focused on the types of Real-World Data treated by such platforms, the technologies they use to collect and manipulate data, and the business model characteristics allowing them to create value for the entire Life Sciences ecosystem.

Stemming from the above-mentioned objective, this section will present a discussion of the results from the multiple case studies conducted. For clarity, the findings will be divided into three main parts related to Real-World Data, technologies, and business model.

7.1.1. Real-World Data: usage, opportunities, and challenges

The first area of analysis regards the purposes and types of Real-World Data collected by multi-sided non-transactional platforms. Taking as references the categorization of Real-World Data made by De Lusignan et al. [10], the companies involved focus mainly on the collection of secondary sources of Real-World Data, which are not used specifically and exclusively for a unique Real-World Evidence study. More specifically, the main types of Real-World Data collected are Patient Generated Health Data (PGHD). PGHD are health-related data created, recorded, gathered, or inferred by patients. Thanks to a widespread use of wearable technologies, such platforms are able to collect a wide variety of PGHD, relating to various categories of vital signals and biomarkers like heart rate, weight, ECG, temperature, blood glucose, etc.

Other than PGHD, such platforms also allow the integration and collection of other secondary types of Real-World Data: Electronic Health Records (EHRs), medicals claims and billing data.

The data collected from such platforms are secondary sources of Real-World Data as they are collected for general purposes, in such a way that they can be used by multiple types of stakeholders for multiple purposes. Each company affiliated to the platform then uses such data for their own business and medical purposes, after some processes of manipulation and adjustments as to fit their needs.

Regarding the opportunities brought by Real-World Data, the main concepts mentioned by the involved companies are:

- Abundance of data. Today, thanks to the latest technological advancements, companies are able to collect and have available a vast amount of data relating to patient's health coming from real world settings, which were not previously available. Such data can be leveraged to reduce the cost and accelerate the processes of research and development of new medical products or treatments.
- Contextualization. Thanks to RWD, all the stakeholders in the Life Sciences sector can switch from just acknowledging that a disease occurred, to understanding how it occurred and what happened during the process. Companies can now understand what is happening in real-life contexts between one and another interaction point with the healthcare system, augmenting and integrating the already existing knowledge about patient's conditions.

The main challenges mentioned, instead, are:

- Confounding. As RWD are gathered for general purposes, companies are finding difficult to understand how to make sense of the data they collect from the various sources, and how to also deal with outliers as to make the results of their findings more externally valid.
- Incomplete data. Often, there may be gaps in the data collected, or there may be some missing data for a specific individual, which can pose incremental challenges for companies.
- Data access. Due to a lack of regulation on data sharing, it is today challenging to create a unified ecosystem where all the stakeholders involved, who own different types of data, collaborate and share patient's data as to benefit the ecosystem itself. Some players are in fact willing to create closed ecosystems, not accessible by other stakeholders.
- Lack of regulation and universally accepted standards. There are no common standards used for the sharing, design, collection and analysis of RWD. Platforms and companies are finding it difficult to standardize the data collected and some data manipulation is required as to fit such data to the purpose of each study.

7.1.2. Technologies: types of technologies used to collect and valorize RWD

The second section of analysis regards the new technologies used by multi-sided non-transactional platforms to collect and valorize Real-World Data. As for the collection of Real-World Data, the technologies comprised are those that allow the collection, monitoring, and tracking of user's health data coming from the Real-World. As for the valorization of Real-World Data, the technologies comprised are those that allow the manipulation and analysis of such data, as to make them available to Life Sciences stakeholders in such a way that the latter can leverage them to achieve their medical and business goals.

With focus on the category of technologies used to collect Real-World Data, sensors are the most used technology. Both wearable and non-wearable sensors have been mentioned by the companies involved, that use devices that can be either worn directly by users or also be present in the real-life contexts where users spend their time and use such devices on occasion. In particular, among all the sensors mentioned, wearable sensors are the most powerful and used means to collect Real-World Data. Blood pressure monitors, thermometers, smart watches are only some of the wearable sensors utilized by the platform providers analyzed, as they allow the continuous tracking and monitor of some important patient's vital signs.

The Real-World Data collected, mainly through wearable devices, are stored in the cloud or into mobile phone's local memories. Another commonality among all the platforms involved is the use of mobile applications as a way to let users track, monitor and see the Real-World Data collected by the platform. Mobile apps are the interface offered to users and constitute a strong driver of value creation for patients, who are able to constantly track their vital signals, in most cases for free. So, to sum app, wearable devices are the most used technologies to collect Real-World Data, with mobile apps as the main technology offered to users to track their health data and have an interface where to observe the output of the wearable devices collection.

With respect to the technologies used to valorize Real-World Data, one major commonality shared by the platforms involved is the use of Artificial Intelligence and Machine Learning to perform complex data analysis and manipulation on the Real-World Data collected, as to allow the other sides affiliated to the platforms to access such data and use them to achieve their goals.

Artificial Intelligence and Machine Learning algorithms are usually embedded in the products and platforms provided by the companies analyzed to allow the valorization of user's Real-World Data, and in most cases are critical to make sense of the data collected, deal with outliers, and perform accurate data analysis on the vast datasets available thanks to the use of wearable sensors.

The interface used by Life Sciences stakeholders in the valorization phase is different than the one used by users in the collection phase. The companies analyzed have in fact developed digital platforms, offered through licensing under the form of a Software as a Service, where the output of the collection and data analysis activities can be seen by the different actors affiliated to the platform, and can be then used to achieve their goals.

7.1.3. Business Model characteristics of multi-sided non-transactional platforms

The third section of analysis regards the business model strategies applied by multi-sided non-transactional platforms to create and capture value in the Life Sciences sector. As seen from the literature review, the main components analyzed for this specific category of platforms are: value creation, sides management, pricing strategy and governance. In this section, each component mentioned will be faced, with respect to the findings of this research.

1) Value Creation

As all the platforms analyzed have multiple and different sides affiliated to them, the need to create, develop and maintain different value proposition at the same time is faced constantly. A shared pattern is for such platforms to offer a value proposition to users consisting of being able to monitor and track their health data, and to participate in research. On each B2B side, instead, the shared pattern is to offer as main value proposition the access to patient's Real-World Data that can be used for specific business and medical objectives. This is resulting in the support of the two hypothesis made regarding the need for multi-sided non-transactional platforms to create different value propositions for B2C e B2B sides, and that end users are a key part of the B2B value proposition. Interestingly, an emerging new pattern was found in this research. Paired with the access to patient's Real-World Data, most of the platforms involved in the study also offer to B2B sides the possibility to use their digital platforms as a licensed Software as a Service, which can be utilized as a standalone product and does not require the presence of patient's Real-World Data coming from the platform. In this way there is value in the platform and its software itself, even if the community of patients providing their health data doesn't exist. Thanks to this additional value proposition, a B2B party affiliated to the platform could use the platform's service without the platform's users, by recruiting its own participants outside the platform's environment. As an example, a Life Science company affiliated to the platform can use, through a license, the software offered by the platform to conduct their own study by recruiting their own users, collecting their health data, and use them to achieve their goals. All this without the need to use health data coming from users in the platform's environment.

This important finding shows that such platforms have evolved their value proposition and value creation mechanisms, in such a way that they are no more completely reliant on network effects, but instead they can provide value as a standalone service that can be used outside the platform's environment, because there is value in using the platform's software itself, even is the user base is not well developed or even absent. However, it is wort to be noted that still, for all the platforms involved, the main value driver in the value proposition comprised the access to end user data through the platform itself.

Even if a certain degree of network effects has been found, confirming the quantitative view of platform's value creation, the qualitative view is being reinforced as a way for platforms to reduce their dependency on network effects to being able to create value, because network effects are difficult to create and present a chicken-and-egg problem which can impede platform's growth. Offering this new value proposition allows platforms to partly overcome the need to achieve a critical mass of users to start creating value, in such a way that value can be created from the start and that paying customers can be attracted even without the need for a large user base.

With reference to the literature, moreover, the common strategies that emerged among the platform providers involved to face the chicken-and-egg problem are:

- Platform staging: within this strategy, a multi-sided platform evolves in two distinct steps from a traditional vendor-based business model in the first stage to a platform business model in the second stage, after the critical mass of user has been reached. This has allowed the platforms studied to create value and generate revenues from the start, without the need to attract a strong user base.
- Subsidizing: within this strategy, the multi-sided platform typically has a 'subsidy side' that allows the use of the platform for free, and one or more 'money side(s)' that are charged for participation or transactions. In the cases analyzed, the majority of the platform providers have created a subsidy side, offering their mobile apps for tracking and monitoring their health data for free. In this way, they could have attracted more and more users, so as to start building the critical mass needed to improve their value creation mechanisms.

2) Sides management

From the analysis made, all the platforms have made the strategic decision to attract more than 2 sides, and all the platforms have at least 4 different sides which are affiliated to them. It is therefore evident that the trade-off [65] involved in attracting more or fewer sides has been solved for the former strategy of attracting more sides. As discovered in the literature, this strategy has the potential benefits of leading to larger cross-side network externalities, larger scale and diversified sources of revenue. The disadvantage of having more sides is the risk of crating additional complexity, which can lead to the "lowest common denominator" problem. In the cases analyzed, the platforms providers were able to attract sides with similar interests and needs, so that their platforms and the associated services were valuable to all of them, without creating too much complexity. As those sides have similar needs, also the functionalities to serve them tend to be similar, and therefore the risk of "lowest common denominator" problem is mitigated.

The platform providers analyzed share also the types of sides attracted, which range from pharmaceutical companies, non-for-profit organizations, foundations, research organizations, government agencies, and technological companies. All those sides share the similar interest of accessing Real-World Data to solve for their business and medical challenges, and also the need for having a platform which offers them services and features to valorize such data, or use the platform as a standalone product to carry our studies without the platform's environment.

3) Pricing

The pricing strategy analysis can be divided into two parts: for the B2C side and for the B2B sides.

With respect to the B2C side, most of the platform providers are using a cross-subsidization strategy, as forecasted by the literature, to offer their mobile app and health monitoring services for free. This strategy is part of the Subsidization strategy adopted to deal with the chicken-and-egg problem, so as to attract to the platform the side which is more sensitive to price.

Looking at the literature, the heavily skewed pricing strategy that leads to subsidization was predicted, because of the characteristics of multi-sided non-transactional platforms [76]. Due to this, price markups are much higher on one side (B2B) of the market than the other (B2C). The factors influencing this subsidization strategy for B2C users are the following [75]:

- Ability to capture cross-side network effects. As the platforms studies present a certain degree of cross-side network effects, the subsidization strategy allows to gather more users, more Real-World Data, and enhance network effects.
- User sensitivity to price. Generally, it is more effective to subsidize the network's more price sensitive side. In this case, users are more price sensitive, as the B2B sides have greater willingness to pay when joining the platform, as this entails access to patient's Real-World Data and a software to valorize them.
- Value extracted by customers. For multi-sided non-transactional platforms, due to their characteristics, the platform provider should charge more the side that stands to benefit more from the presence of the other side or sides, which in this case is the B2B side comprising all the Life Sciences stakeholders who want to access patient's health data.

All those factors have influenced the subsidization strategy adopted by most of the platforms involved in the study. However, a particular case is emerged: Withings does not offer their products and services to users for free, as the company is able to monetize also the B2C sides through the sale of hardware for collecting and monitoring health data.

The strategy adopted by Withings is such that more value is offered to users, under the form of the hardware and sensors they produce. In this case, therefore, a stronger value proposition is offered to the B2C side, that allows the platform provider to monetize both the B2C arm and the B2B arm. This is achieved by creating and selling high quality sensors that allow users for a better collection and monitoring of health data. Some of the sensors include smart scales, smart blood pressure monitors, smart watches, smart sleep tracking devices and thermometers. Different from the other platform providers, Withings has moved also in the production of sensors, which is a strategy allowing them to extract monetary value from consumers. This strategy has not been pursued by the other platforms, who chose to just create integrations with already existing sensor devices, such that users could use their free app to connect their wearable devices and stand tracking their health data.

With respect to the B2B side, instead, all the platform providers analyzed adopted a strategy to extract monetary value from the Life Sciences stakeholders affiliated to the ecosystem, confirming the hypothesis stating that the monetization strategy of these platforms is mainly B2B oriented. More in particular, the pricing strategy adopted by them is mostly a licensing strategy, through which platform providers let B2B clients to access patient's Real-World Data and to use their platform by paying a license fee. This fee can be charged differently depending on the number of users, the time range, etc., but the main licensing logic is applied by all the platforms involved as a pricing strategy to extract value from the B2B arm. The strategy of targeting B2B clients as the main source of revenue is driven by the same factors that are driving a subsidization strategy on the B2C side, which can be mainly summed up in the fact that the B2B clients are the ones extracting greater value and hence more willing to pay for being affiliated to the platform.

As an additional way to extract monetary value from B2B sides, Evidation Health has developed an additional layer of services which go beyond the licensing of the software's platform: B2B clients can in fact access additional services like sophisticated data analysis, involvement of clinical staff to conduct clinical studies, etc. Those additional services are charged over the licensing of the platform, concurring to the creation of an additional stream of revenues derived from the additional services built on top of the platform.

4) Governance

With regards to the degree of openness of the platforms studied, all of them have in common the characteristic of being closed: nobody can access them, and the data collected by patients without patient authorization and without the authorization of the platform provider. This is a way to guarantee a high level of security and privacy, a theme commonly viewed as of critical importance by the platforms involved. All the companies studied, in fact, rigorously follow the EU or US guidelines on patient's data collection, storage maintenance and treatment. Practices such as anonymization, tokenization, allowing users to delete the data stored at any time, and storing patient's data in places which comply to the most stringent security requirements are applied by each platform provider.

7.2. Theoretical Contributions

The added value from the research conducted with respect to the theoretical contributions is two-fold: on one hand, it provides a comprehensive analysis of value capture from RWD in private business settings, while on the other hand it expands the existing knowledge on the topic addressed with some new hypotheses tested.

Firstly, as emerged from the Literature Review conducted, a lack of studies on the applications of RWD on multi-sided non transactional platforms in the private business sector was found. The research performed has addressed the research gap found by describing how private platform providers are capturing value from RWD, by comprehensively analysing factors such as the types of RWD used, the technologies leveraged to valorize RWD, and also the business model characteristics allowing such companies to create value in the healthcare sector.

Secondly, the hypotheses tested and supported provide new knowledge related to the research question studied.

As explained in the relevant section, H2, H3, and H4 have been developed from the research model created by Muzellec et al. [73] in their paper. These hypothesis, mainly related to the value propositions and the business model monetization strategy, were successfully tested for their validity within the healthcare sector. The fact that these hypotheses were supported further strengthens the validity of the model used as reference and proves its validity also when applied to multi-sided non transactional platforms in the healthcare sector, which is intrinsically more complex and regulated than the other industries studied.

Moreover, as pointed out in the relevant section, H1 and H5 have been developed as novel hypotheses based on the papers and articles reviewed during the Literature Review. The successful testing and support of these hypotheses contributes to novel knowledge on the subject matter.

H1, in fact, provides for the fact that platform providers leverage mainly on Secondary Source of Real-World Data as key value driver for their B2B value propositions. Therefore, it has been shown that not only platform providers should sustain different value propositions for B2B and B2C sides, but they should also include end users' health data (secondary sources of RWD) as part of their B2B value proposition.

H5, instead, provides for the fact that given the peculiarities of the Real-World Data collected, platform providers in the healthcare sector opt for closed ecosystems due to the Security and Ethical concerns faced.

7.3. Managerial Implications

For multi-sided non-transactional platforms active in the Life Sciences sector, Real-World Data present both opportunities and challenges, that must be handled at a strategic level to enable value creation.

Real-World Data, especially coming from secondary sources and directly produced by patients, enable a faster, cheaper and more effective development of new medical products or treatments. However, these advantages must be balanced with some challenges that multi-sided non-transactional platforms face in terms of privacy, bias, incompleteness of data, and a lack of standardized regulations and procedures. Being a relatively new and still developing field, platforms can benefit from growth and value creation opportunities, if the above-mentioned challenges can be solved.

The increasing adoption of new technologies and the latest advancements in the field can help platform providers to overcome such challenges and unlock value creation, especially thanks to the use of sensors, wearables, the IoMT and artificial intelligences. Incumbents and new entrants should therefore follow those technological trends and build internal capabilities to be able to leverage on them and exploit Real-World Data at best.

From a strategic and business model perspective, managers of multi-sided non-transactional platforms in the Life Sciences sector must face four critical aspects: value creation mechanisms, sides management, pricing strategy, and governance.

Due to their multi-sided nature, platform providers should be able to formulate and sustain multiple different value propositions both on the B2C and B2B sides, taking into account the presence of the chicken-and-egg problem and of network effects. To overcome such issues, a new strategy emerging is to develop a standalone platform that brings value in itself and can be used also outside the platform's ecosystem. In this way, platform providers can deliver and extract value from the start, without the need for a large user base.

Moreover, Real-World Data can bring value to multiple different stakeholders in the Life Sciences sector, which can use them to achieve their own different business and medical objectives. Despite the differences, some commonalities are present that allow the platform providers to serve all the stakeholders with the same platform and software, mitigating the risk of the "lowest common denominator" problem. Due to this, such platforms are able to attract multiple different sides to the platform without increasing complexity too much.

With respect to pricing, the characteristics of the B2C sides create the need for a cross-subsidization strategy, in which B2B sides are the easier to monetize stakeholders, due to their higher willingness to pay. Users can still be monetized, but by drafting more complete and comprehensive value propositions which can include the creation of higher quality sensors to monitor and track Real-World Data.

Lastly, the privacy and security requirements mandate platforms to keep their ecosystems closed, as to preserve and protect the Real-World Data collected. Privacy and security should be therefore top priorities for multi-sided non-transactional platforms, as to develop trust from the stakeholders involved, and to favour a sustainable value creation process.

All the above strategic and technological factors are allowing multi-sided non-transactional platforms to collect and valorize Real-World Data, enabling faster and cheaper development of new products and treatments by Life Sciences stakeholders.

8 Conclusions

8.1. Conclusions

The primary goal of this research was to investigate the impact of RWD and new technology applications on the multi-sided non-transactional platform business model in the healthcare industry.

Thanks to their business model, platform dynamics and their ability to act as a source of Real-World Data, multi-sided non-transactional platforms are emerging as a key stakeholder in the healthcare industry, connecting patients and companies in a non-transactional way as to allow companies to achieve their business and medical goals of improving patient's lives.

Given the recency of the topic of multi-sided non-transactional platforms and Real-World Data, this research aims at describing the role of multi-sided non-transactional platforms under three main lenses:

- Real-World Data: types of Real-World Data collected, opportunities and challenges faced in their collection and valorization.
- Technologies: new technologies used to collect and valorize Real-World Data.
- Business model: value creation mechanisms, pricing structure, sides management and governance strategies.

Thanks to this exploratory research conducted through the multiple case study methodology, the main research question above mentioned has been investigated with the aim of integrating the existing knowledge with new findings on how multi-sided non-transactional platforms are using Real-World Data and new technologies to create and capture value in the healthcare sector.

The literature review conducted on the topic has highlighted the various types and sources of Real-World Data, along with the opportunities and challenges they can bring in the healthcare sector. Security, missing data, bias, and a lack of regulatory framework has been found to be the main challenges to be faced. On the other hand, thanks to the new technological advancements, new possibilities to address such challenges are emerging. Sensors, wearables, Internet of Medical Things, Artificial Intelligence and Machine Learning are emerging technologies in the healthcare industry as the main enablers of collection and manipulation of Real-World Data. Such technologies are leveraged by multi-sided non-transactional platforms to create and capture value through their business models, whose critical factors are value creation mechanisms, pricing, sides management and platform governance.

Starting from the literature review as a reference, a theoretical framework and model was built to conduct the exploratory multiple case study method. The model aimed at investigating how multi-sided non-transactional platforms are creating and capturing value, in terms of (i) types of Real-World Data collected; (ii) value propositions offered to the affiliated sides; (iii) pricing structure and monetization strategy; (iiii) degree of openness of their ecosystem and platform.

The result of this exploratory research is that the multi-sided non-transactional platform providers studied are mainly leveraging the secondary sources of Real-World Data, and in particular Patient Generated Health Data, which are directly generated by patients who then decide to share them with the platform and other stakeholders. In the process of collecting such data, as predicted by the literature, regulations and data issues are the main barriers faced. The most valuable technologies used by the involved platform providers are wearable devices, mobile applications, AI and ML, which allow the real time collection, monitoring and manipulation of the vast amounts of health data generated by patients.

From a strategic and managerial perspective, moreover, the results of the research show major trends related to the business model configuration of multi-sided non-transactional platforms. In terms of value creation, the presence of cross-side network externalities and the associated chicken-and-egg problem can slow the value creation process, and require platform providers to create and sustain different value propositions for both the B2C and B2B sides affiliated. In this setting of value proposition, it was found that end users Real World Data are a key part of the B2B value proposition for the business sides.

A new trend reported from the research is for such platform providers to try to solve for the chicken-and-egg problem brought by network externalities by offering a standalone software platform that has value in itself, and that customers can use even outside the platform's environment. That software can be used through licensing by B2B customers to collect and valorize Real-World Data of patients which are recruited from the customer itself, and not recruited through the platform's environment. In this way, platforms are able to create and capture value from the start, without the need to attract a critical mass of users before being able to sustain the value proposition for B2B customers. From a sides management perspective, instead, a general trend of dealing with the number of sides trade-off is for multi-sided non-transactional platforms to attract more than only two sides, in general more than four. This is because the B2B customers have different needs but similar requirements in terms of product and service features, so that the platform is able to provide high value without the need to add complexity. Thanks to this dynamic, peculiar of the healthcare sector, the "lowest common denominator" problem is mitigated, allowing the platform provider to open its platform to many sides. In terms of pricing, the major trend is for platform providers to pursue a cross-subsidization strategy where users receive value for free, as they contribute with their health data. B2B customers, instead, who have higher willingness to pay and extract the greatest amount of value form the platform, are generally paying under the form of licensing. However, ways to monetize also from customers have been found, in particular when the platform provider pursues a more complex value propositions for which it sells to users also higher-quality hardware, like wearables and sensors to track and monitor health data. Lastly, regarding governance, all the platforms have a closed model, mainly driven by privacy and security issues. Privacy regulations have been found to be a major challenge in the process of collecting Real-World Data, mainly because each country has different rules to which platforms should adapt and which required additional effort from the platform provider. Privacy and security, however, are a top priority of platforms despite the regulations, because they are a major way to acquire and retain user's trust, which is then beneficial to feed the cross-side network externalities from the user's side. Due to the above, therefore, strict security measures are put in place by platforms providers to respect the regulations and protect user's data, for the benefit of all the stakeholders involved.

In conclusion, the research presents various strengths relying on the multiple case study method chosen, and on the comprehensive model built to address the research question.

This research offers the opportunity to identify the key factors of how multi-sided non-transactional platforms are allowing the collection and valorization of Real-World Data, from a data, technology, and business model perspective.

8.2. Limitations and Future Research

Despite the previous considerations, there are additional insights coming from the results obtained that are worth being explored by future research. Moreover, within the present study, we also identified intrinsic weaknesses that suggest eventual next developments and improvements.

Starting from the findings, a future study should try to address not only the role of private companies as platform providers, but also the role of public or non-profit platform providers used by governments and other public entities to leverage on Real-World Data. During the Literature Review, in fact, many applications of multi-sided non-transactional platforms in the public sector have been found, and the study of them could bring further insights on the way in which they have structured their business model to create value for the stakeholders involved.

Moving to the limitations of the study, a first weakness can be found in the way the sampling has been conducted. Even if purposeful sampling was adequate to conduct the multiple case study because of its focus on finding information rich case for in-depth study, its major weakness is the presence of bias in the sampling due to the lack of randomness. This bias can limit the generalization of the results found.

Another limitation of the study is related to the interview process. Even if I, the researcher, have been careful in not biasing the respondents with the questions made during the interviews, some degree of bias may still be present. The bias from the interviews may impact the quality of the responses received.

Moreover, the focus on exploratory and qualitative methods brought a lack of quantitative methods and causal relationships. This lack can harm the external validity of the findings, and their general application to the whole population of multi-sided non-transactional platforms in the healthcare sector.

Finally, the last limitation can be detected in the instrument adopted to recruit interviewees for the empirical interviews. Indeed, the channel used, LinkedIn, has some intrinsic characteristics that on one side can been seen as potentialities but, on the other, might limit the possibility to generalize the results obtained. It is possible to detect a sort of homogeneity in the backgrounds and roles of the respondents.

Bibliography

[1] Venkata Satagopam, Wei Gu, Serge Eifes, Piotr Gawron, Marek Ostaszewski, Stephan Gebel, Adriano Barbosa-Silva, Rudi Balling, and Reinhard Schneider. (2016). Integration and Visualization of Translational Medicine Data for Better Understanding of Human Diseases. Big Data, 4:2, pp. 97-108.

[2] Yujie Yang, Ye Li, Runge Chen, Jing Zheng, Yunpeng Cai, Giancarlo Fortino. (2021). Risk Prediction of Renal Failure for Chronic Disease Population Based on Electronic Health Record Big Data. Big Data Research, Volume 25, 100234.

[3] U.S. Food and Drug Administration. (2016) Real-World Evidence. 21st Century Cures Act of 2016, Section 3022, pp. 64-65.

[4] Amr Makady, Anthonius de Boer, Hans Hillege, Olaf Klungel, Wim Goettsch. (2017). What Is Real-World Data? A Review of Definitions Based on Literature and Stakeholder Interviews, Value in Health, Volume 20, Issue 7, Pages 858-865.

[5] A Makady, W Goettsch. (2015). PHP307 - Review of Policies And Perspectives on Real-World Data for Drug Development and Assessment (Imi-Getreal Deliverable), Value in Health, Volume 18, Issue 7, pp. A567.

[6] Louis P. Garrison, Peter J. Neumann, Pennifer Erickson, Deborah Marshall, C. Daniel Mullins. (2007). Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report, Value in Health, Volume 10, Issue 5, pp. 326-335.

[7] A.M. Pleil. Using real world data in pharmacoeconomic evaluations: challenges, opportunities and approaches. http://www.pharmacy.ucsd.edu/faculty/AppliedPEForum/docs/Andreas_M_Pleil.pdf.

[8] Romio, S.A., Sturkenboom, M., & Corrao, G. (2013). Real-world data from the health decision maker perspective. what are we talking about? Epidemiology, biostatistics, and public health, 10.

[9] U.S. Food and Drug Administration. (2018) Framework for FDA's Real-World Evidence Program, pp. 3-12.

[10] S. De Lusignan, L. Crawford. (2015). Creating and using real-world evidence to answer questions about clinical effectiveness. J Innov Health Inform., Vol.22(3), pp.368-373.

[11] Maissenhaelter, B.E., Woolmore, A.L. & Schlag, P.M. (2018). Real-world evidence research based on big data. Onkologe 24, pp. 91–98. https://doi.org/10.1007/s00761-018-0358-3v.

[12] Peter M Rothwell. (2005): External validity of randomized controlled trials: "To whom do the results of this trial apply?". The Lancet, Volume 365, Issue 9453, pp. 82-93.

[13] Brooke EM. (1974). The current and future use of registers in health information systems. Geneva: World Health Organization. Publication No. 8.

[14] National Committee on Vital and Health Statistics. (2012). Frequently Asked Questions About Medical and Public Health Registries.

[15] Gliklich RE, Dreyer NA, Leavy MB, editors. (2014). Registries for Evaluating Patient Outcomes: A User's Guide [Internet]. 3rd edition. Rockville (MD): Agency for Healthcare Research and Quality (US); Apr. 1, Patient Registries.

[16] Dafna D. Gladman, Vinod Chandran. (2011). Observational cohort studies: lessons learnt from the University of Toronto Psoriatic Arthritis Program, Rheumatology, Volume 50, Issue 1, pp. 25–31.

[17] Caruana, E. J., Roman, M., Hernández-Sánchez, J., & Solli, P. (2015). Longitudinal studies. Journal of thoracic disease, 7(11), E537–E540. https://doi.org/10.3978/j.issn.2072-1439.2015.10.63.

[18] Van Belle G, Fisher L, Heagerty PJ, et al. (2004). Biostatistics: A Methodology for the Health Sciences. Longitudinal Data Analysis. New York, NY: John Wiley and Sons.

[19] Daniel Schwartz, Joseph Lellouch. (1967). Explanatory and pragmatic attitudes in therapeutical trials. Journal of Chronic Diseases, Volume 20, Issue 8, pp. 637-648.

[20] Hugh MacPherson. (2004). Pragmatic clinical trials. Complementary Therapies in Medicine, Volume 12, Issues 2–3, pp. 136-140.

[21] Ming Zhu, Saranya Sridhar, Rosalind Hollingsworth, Ayman Chit, Tammy Kimball, Kerry Murmello, Michael Greenberg, Sanjay Gurunathan, Josh Chen. (2020). Hybrid clinical trials to generate real-world evidence: design considerations from a sponsor's perspective. Contemporary Clinical Trials, Volume 94.

[22] Brenda Motheral, John Brooks, Mary Ann Clark, William H. Crown, Peter Davey, Dave Hutchins, Bradley C. Martin, Paul Stang. (2003). A Checklist for Retrospective Database Studies—Report of the ISPOR Task Force on Retrospective Databases, Value in Health, Volume 6, Issue 2, pp. 90-97.

[23] Anthonisen N. R. (2009). Retrospective studies. Canadian respiratory journal, 16(4), 117–118. https://doi.org/10.1155/2009/704292.

[24] Personal Health Working Group. (2003). The personal health working group final report. Washington, DC: Connecting for Health: A Public-Private Collaborative.

[25] Paul C. Tang, MD, MS, Joan S. Ash, PhD, David W. Bates, MD, J. Marc Overhage, MD, PhD, Daniel Z. Sands, MD, MPH. (2006). Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption. Journal of the American Medical Informatics Association, Volume 13, Issue 2, pp. 121–126.

[26] Lowrance W. W. (2001). The promise of human genetic databases. BMJ (Clinical research ed.), 322(7293), 1009–1010. https://doi.org/10.1136/bmj.322.7293.1009.

[27] J. Andreu-Perez, D. R. Leff, H. M. D. Ip and G. -Z. Yang. (2015). From Wearable Sensors to Smart Implants-–Toward Pervasive and Personalized Healthcare. IEEE Transactions on Biomedical Engineering, vol. 62, no. 12, pp. 2750-2762.

[28] Mohammadian Foroushani Hossein, Dhar Rajat, Chen Yasheng, Gurney Jenny, Hamzehloo Ali, Lee Jin-Moo, Marcus Daniel S. (2021). The Stroke Neuro-Imaging Phenotype Repository: An Open Data Science Platform for Stroke Research. Frontiers in Neuroinformatics, Volume 15.

[29] Harding, K. (2016). Global Health Innovation Technology Models. Nanobiomedicine. https://doi.org/10.5772/62921

[30] Loryana L. Vie, Lawrence M. Scheier, Paul B. Lester, Tiffany E. Ho, Darwin R. Labarthe, and Martin E.P. Seligman. (2015). The U.S. Army Person-Event Data Environment: A Military–Civilian Big Data Enterprise. Big Data, 3:2, pp. 67-79.

[31] C. H. Mendhe, N. Henderson, G. Srivastava and V. Mago. (2021). A Scalable Platform to Collect, Store, Visualize, and Analyze Big Data in Real Time. IEEE Transactions on Computational Social Systems, vol. 8-1, pp. 260-269.

[32] Pala D, Caldarone AA, Franzini M, Malovini A, Larizza C, Casella V, Bellazzi R. (2105). Deep Learning to Unveil Correlations between Urban Landscape and Population Health. Sensors, 2020, 20(7).

[33] Rodrigues D, Barbosa AI, Rebelo R, Kwon IK, Reis RL, Correlo VM. (2020). Skin-Integrated Wearable Systems and Implantable Biosensors: A Comprehensive Review. Biosensors. 10(7):79.

[34] Ye, B., Khan, S.S., Chikhaoui, B. et al. (2019). Challenges in Collecting Big Data in A Clinical Environment with Vulnerable Population: Lessons Learned from A Study Using A Multi-modal Sensors Platform. Sci Eng Ethics, Volume 25, pp. 1447–1466.

[35] A. Gatouillat, Y. Badr, B. Massot and E. Sejdić. (2018). Internet of Medical Things: A Review of Recent Contributions Dealing With Cyber-Physical Systems in Medicine. IEEE Internet of Things Journal, vol. 5, no. 5, pp. 3810-3822.

[36] Kruse C, Goswamy R, Raval Y, Marawi S. (2016). Challenges and Opportunities of Big Data in Health Care: A Systematic Review. JMIR Med Inform, 4(4), p. 38.

[37] Minerva, R., Biru, A., & Rotondi, D. (2015). Towards a definition of the Internet of Things (IoT). IEEE Internet Initiative, 1(1), pp. 1-86.

[38] Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, et al. (2017). Artificial intelligence in healthcare: past, present and future. Stroke and vascular neurology, 2(4).

[39] Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future healthcare journal, 6(2), p. 94.

[40] H. Wang, C. Ma and L. Zhou. (2009). A Brief Review of Machine Learning and Its Application. International Conference on Information Engineering and Computer Science, 2009, pp. 1-4.

[41] William H. Crown. (2019). Real-World Evidence, Causal Inference, and Machine Learning. Value in Health, Volume 22, Issue 5, Pages 587-592.

[42] Wake, D. J., He, J., Czesak, A. M., Mughal, F., & Cunningham, S. G. (2016). MyDiabetesMyWay: An Evolving National Data Driven Diabetes Self-Management Platform. Journal of Diabetes Science and Technology, 10(5), pp. 1050–1058.

[43] Kalkman, S., Mostert, M., Gerlinger, C. et al. (2019). Responsible data sharing in international health research: a systematic review of principles and norms. BMC Med Ethics, Volume 20, p. 21.

[44] Grace Hampson, Adrian Towse, William B Dreitlein, Chris Henshall, and Steven D Pearson. (2018). Real-world evidence for coverage decisions: opportunities and challenges. Journal of Comparative Effectiveness Research, 7:12, Pages 1133-1143.

[45] Garrison, L.P., Jr., Neumann, P.J., Erickson, P., Marshall, D. and Mullins, C.D. (2007). Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report. Value in Health, 10, Pages 326-335.

[46] Cole, A., Garrison, L., Mestre-Ferrandiz, J., & Towse, A. (2015). Data governance arrangements for real-world evidence. London, UK: Office of Health Economics.

[47] Center, B. P. (2016). Using real-world evidence to accelerate safe and effective cures. Advancing medical innovation for a healthier America.

[48] Saville BR, Berry SM. (2016). Efficiencies of platform clinical trials: A vision of the future. Clinical Trials, 13(3), Pages 358-366.

[49] J. Wise, A. Möller, D. Christie, D. Kalra, E. Brodsky, E. Georgieva, G. Jones, I. Smith, L. Greiffenberg, M. McCarthy, M. Arend, O. Luttringer, S. Kloss, S. Arlington. (2018). The positive impacts of Real-World Data on the challenges facing the evolution of biopharma. Drug Discovery Today, Volume 23, Issue 4, Pages 788-801.

[50] Medline Plus. What is precision medicine? National Library of Medicine. https://medlineplus.gov/genetics/understanding/precisionmedicine/definition/.

[51] Rochet, J., Tirole, J. (2003). Platform competition in two-sided markets. Journal of the EuropeanEconomic Association, Vol. 1 No. 4, pp. 990-1029.

[52] Evans, D.S. (2003). Some empirical aspects of multi-sided platform industries, Review of Network Economics, Vol. 2 No. 3.

[53] Amiya Basu, Tridib Mazumdar, and S. P. Raj. (2003). Indirect Network Externality Effects on Product Attributes. Marketing Science, 22:2, pp. 209-221.

[54] Rochet, J.C. and Tirole, J. (2006). Two-sided markets: a progress report, The RAND Journal of Economics, Vol. 37 No. 3, pp. 645-667.

[55] Trabucchi, Daniel & Buganza, Tommaso. (2021). Landlords with no lands: a systematic literature review on hybrid multi-sided platforms and platform thinking. European Journal of Innovation Management, 10.1108/EJIM-11-2020-0467.

[56] Sánchez-Cartas, Juan Manuel & León, Gonzalo. (2018). Multisided Platforms and Markets: A Literature Review.

[57] Andrei Hagiu, Julian Wright. (2015). Multi-sided platforms. International Journal of Industrial Organization, Volume 43, pp. 162-174.

[58] Sanchez-Cartas, J.M. and León, G. (2021). Multisided platforms and markets: a survey of the theoretical literature. Journal of Economic Surveys, 35, pp. 452-487.

[59] Weyl, E. G. (2010). A price theory of multi-sided platforms. American Economic Review, 100(4), pp. 72-1642.

[60] Filistrucchi, L., & van Damme, E. (2013). Identifying two-sided markets. World Competition, 36(1).

[61] Evans, D. (2003). Some Empirical Aspects of Multi-sided Platform Industries. Review of Network Economics, 2(3).

[62] Filistrucchi, Lapo. (2008). A SSNIP Test for Two-Sided Markets: The Case of Media. NET Institute Working Paper No. 08-34.

[63] Gundlach, H., & Hofmann, J. (2017). Preferences and willingness to pay for tablet news apps. Journal of Media Business Studies, 14(4), pp. 257-281.

[64] Trabucchi, D, Buganza, T. (2020). Fostering digital platform innovation: From two to multi-sided platforms. Creat Innov Manag, 29, pp. 345– 358.

[65] Hagiu, A. (2015). Strategic decisions for multisided platforms. Top, 10, pp. 4-13.

[66] Ardolino M, Saccani N, Adrodegari F, Perona M. (2020). A Business Model Framework to Characterize Digital Multisided Platforms. Journal of Open Innovation: Technology, Market, and Complexity. 6(1):10.

[67] Voigt, S., Hinz, O. (2015). Network effects in two-sided markets: why a 50/50 user split is not necessarily revenue optimal. Bus Res 8, pp. 139–170.

[68] Tellis GJ, Yin E, Niraj R. (2011). Does Quality Win? Network Effects versus Quality in High-Tech Markets. Journal of Marketing Research. 46(2), pp. 135-149.

[69] Lau, ML., Wydick, B. (2014). Does New Information Technology Lower Media Quality? The Paradox of Commercial Public Goods. J Ind Compet Trade 14, pp. 145–157.

[70] Molina-Castillo, F.-J., Munuera-Alemán, J.-L. and Calantone, R.J. (2011). Product Quality and New Product Performance: The Role of Network Externalities and Switching Costs. Journal of Product Innovation Management. 28, pp. 915-929.

[71] Stummer, C., Kundisch, D. & Decker, R. (2018). Platform Launch Strategies. Bus Inf Syst Eng 60, pp. 167–173.

[72] Fish, T. (2009). My digital footprint: a two-sided digital business model where your privacy will be someone else's business. Futuretext.

[73] Laurent Muzellec, Sébastien Ronteau, Mary Lambkin. (2015). Two-sided Internet platforms: A business model lifecycle perspective. Industrial Marketing Management, Volume 45, pp. 139-150.

[74] Glaser, S. (2006). The value of the manager in the value chain. Management decision.

[75] Eisenmann, T., Parker, G., & Van Alstyne, M. W. (2006). Strategies for two-sided markets. Harvard business review, 84(10), p. 92.

[76] Wilko Bolt, Alexander F. Tieman. (2008). Heavily skewed pricing in two-sided markets, International Journal of Industrial Organization, Volume 26, Issue 5, pp. 1250-1255.

[77] Pontus Huotari, Kati Järvi, Samuli Kortelainen, Jukka Huhtamäki. (2017). Winner does not take all: Selective attention and local bias in platform-based markets. Technological Forecasting and Social Change, Volume 114, pp. 313-326.

[78] Kevin Boudreau. (2010). Open Platform Strategies and Innovation: Granting Access vs. Devolving Access. Management Science 56:10, pp. 1849-1872.

[79] Yin, R. (1996). K. (1994) Case study research–design and methods. Applied social research method series (5): Sage: London.

[80] Longhurst, R. (2003). Semi-structured interviews and focus groups. Key methods in geography, 3(2), 143-156.

[81] Patton MQ. (2002). Two Decades of Developments in Qualitative Inquiry: A Personal, Experiential Perspective. Qualitative Social Work. 1(3), pp. 261-283.

[82] Eisenhardt, K. M. (1989). Building theories from case study research. Academy of management review. 14(4), pp. 532-550.

[83] Strauss, A., & Glaser, B. (1967). Grounded Theory: The discovery of Grounded Theory. The Journal of The British Sociological Association. 12.

[84] Strauss, A., & Corbin, J. (1998). Basics of qualitative research techniques.

[85] Clark SM, Gioia DA, Ketchen DJ, Thomas JB. (2010). Transitional Identity as a Facilitator of Organizational Identity Change during a Merger. Administrative Science Quarterly. 55(3), pp. 397-438.

[86] Shapiro, M., Johnston, D., Wald, J.S., & Mon, D.T. (2012). Patient-generated health data. White paper.

[87] Wilson, C. B. (1999). Sensors in medicine. Bmj, 319(7220), 1288.

[88] Kawamura, A., & Miyata, T. (2016). Biosensors. In Biomaterials Nanoarchitectonics, pp. 157-176. William Andrew Publishing.

[89] Shelke, Y., & Sharma, A. (2016). Internet of medical things. Report published by Aranca.

[90] Visvikis, D., Cheze Le Rest, C., Jaouen, V., et al. (2019). Artificial intelligence, machine (deep) learning and radio (geno) mics: definitions and nuclear medicine imaging applications. European journal of nuclear medicine and molecular imaging, 46(13), pp. 2630-2637.

[91] Olivier Wierzba, Bernd Ziegler, Jean-François Bobier, Thomas Delano, Jean-Baptiste Vermersch. (2020). COVID-19 Opens a New Era for Real-World Evidence in Pharma. Boston Consulting Group report.

[92] Olivia Cavlan, Sastry Chilukuri, Matthias Evers, and Ann Westra. (2018). Real-world evidence: From activity to impact in healthcare decision making. McKinsey & Company report.

[93] Arnaub Chatterjee, Sastry Chilukuri, Edd Fleming, Adam Knepp, Saif Rathore, Joe Zabinski. (2018). Real-world evidence: Driving a new drug-development paradigm in oncology. McKinsey & Company report.

[94] Alex Davidson, Chris Llewellyn, Dan Tinkoff, Manola Vyaravanh. (2020). Ten battlegrounds for digital and analytics in life sciences. McKinsey & Company report.

[95] David Champagne, Alex Davidson, Lucy Pérez, and David Saunders. (2020). Creating value from next-generation real-world evidence. McKinsey & Company report.

[96] Gaurav Agrawal, John Xue, Rachel Moss, Ralf Raschke, Stephan Wurzer. (2021). No place like home? Stepping up the decentralization of clinical trials. McKinsey & Company report.

[97] Jens Lund, Karsten Reschke, Katharina Ashauer, Corinna Breyel, Laura Fritsche, Henry Habermann. (2018). Digitalization in life sciences. KPMG report, pp 10-41.

[98] Buerk, D. G. (2014). Biosensors: theory and applications. Crc Press.

[99] Majumder, S., Mondal, T., & Deen, M. J. (2017). Wearable sensors for remote health monitoring. Sensors, 17(1), pp. 130.

[100] Gray, M., Meehan, J., Ward, C., Langdon, S. P., Kunkler, I. H., Murray, A., & Argyle, D. (2018). Implantable biosensors and their contribution to the future of precision medicine. The Veterinary Journal, 239, pp. 21-29.

[101] Wise, J., Möller, A., Christie, D., Kalra, D., Brodsky, E., Georgieva, E., et al. (2018). The positive impacts of real-world data on the challenges facing the evolution of biopharma. Drug discovery today, 23(4), pp. 788-801.

[102] Singh, G., Schulthess, D., Hughes, N., Vannieuwenhuyse, B., & Kalra, D. (2018). Real world big data for clinical research and drug development. Drug Discovery Today, 23(3), pp. 652-660.

[103] Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. Administration and policy in mental health and mental health services research, 42(5), pp. 533-544.

[104] Bernard, H. R. (2017). Research methods in anthropology: Qualitative and quantitative approaches. Rowman & Littlefield.

[105] Dang LM, Piran MJ, Han D, Min K, Moon H. (2019). A Survey on Internet of Things and Cloud Computing for Healthcare. Electronics, 8(7), p. 768.

[106] Stoeklé, HC., Mamzer-Bruneel, MF., Frouart, CH. et al. (2018). Molecular Tumor Boards: Ethical Issues in the New Era of Data Medicine. Sci Eng Ethics, Volume 24, pp. 307–322.

[107] McKeown, A., Mourby, M., Harrison, P. et al. (2021). Ethical Issues in Consent for the Reuse of Data in Health Data Platforms. Sci Eng Ethics, Volume, 27, p. 9.

[108] Štufi M, Bačić B, Stoimenov L. (2020). Big Data Analytics and Processing Platform in Czech Republic Healthcare. Applied Sciences, Volume 10(5), p. 1705.

[109] Tupasela, A., Snell, K., & Tarkkala, H. (2020). The Nordic data imaginary. Big Data & Society.

[110] Bo, Yang and Chunli, Wang. (2020). Health Data Analysis Based on Multi-calculation of Big Data During COVID-19 Pandemic. Pp. 8775 – 8782.

[111] Vicdan H., Pérezts M., & Fırat A. F. (2021). License to Heal: Understanding a Healthcare Platform Organization as a Multi-Level Surveillant Assemblage. M@n@gement, 24(4), 18–35. https://doi.org/10.37725/mgmt.v24.4586

[112] Morrison, M., Dickenson, D. & Lee, S.SJ. (2016). Introduction to the article collection 'Translation in healthcare: ethical, legal, and social implications'. BMC Med Ethics 17, 74. https://doi.org/10.1186/s12910-016-0157-6

List of Figures

List of Tables

Acknowledgments

I thank Politecnico di Milano for creating an exciting and motivating place where I've learned a lot for my future career.

I thank my professors and program coordinators for always supporting where needed, and for enriching my academic knowledge.

I thank my supervisors, Emanuele Lettieri and Alessandro Carrera, for helping me in finding the right direction for this Thesis, and for supporting me in finalizing it with guidance and encouragement.