



Utrecht University
Department of Computing Sciences

Data as a Platform Ecosystems

Master Thesis defended on , by:

Yannick Mijsters
(6050662)

Supervised by:
Slinger Janssen

Cosupervised by:
Marco Spruit

CONTENTS

| | | |
|-----|---|----|
| 1 | INTRODUCTION | 1 |
| 2 | BACKGROUND | 3 |
| 2.1 | Related Work | 3 |
| 2.2 | Data Supply | 4 |
| 2.3 | Data Market | 4 |
| 2.4 | Drivers for Data Sharing | 5 |
| 2.5 | Data Sharing Issues | 5 |
| 3 | RESEARCH METHOD | 8 |
| 3.1 | Problem Description | 8 |
| 3.2 | Research Goal and Questions | 9 |
| 3.3 | Research Method | 10 |
| 4 | LITERATURE REVIEW | 14 |
| 4.1 | Systematic Literature Review Method | 14 |
| 5 | INTERVIEWS | 21 |
| 5.1 | Interview Set-Up | 21 |
| 5.2 | Incremental Interview Method | 24 |
| 5.3 | Participants | 24 |
| 5.4 | Execution | 27 |
| 6 | CASE STUDY | 31 |
| 6.1 | Case Study Set-Up | 31 |
| 6.2 | Participants | 33 |
| 6.3 | Execution | 35 |
| 7 | DEFINITION | 37 |
| 7.1 | Literature Review | 37 |
| 7.2 | Initial Definition | 40 |
| 7.3 | Interviews | 40 |
| 7.4 | Final Definition | 42 |
| 8 | CHARACTERISTICS | 45 |
| 8.1 | Literature Review | 45 |
| 8.2 | Initial Model | 49 |
| 8.3 | Interviews | 49 |
| 8.4 | Final Model | 50 |
| 8.5 | Case Study | 50 |

| | | |
|------|--|------------|
| 9 | DAAP CAPABILITIES | 54 |
| 9.1 | Systematic Literature Review | 54 |
| 9.2 | Focus Area Maturity Models | 55 |
| 9.3 | Initial Model | 56 |
| 9.4 | Interview | 58 |
| 9.5 | Final Model | 59 |
| 10 | THE DATA AS A PLATFORM FOCUS AREA MATURITY MODEL | 63 |
| 10.1 | The DaaP Focus Area Maturity Model | 63 |
| 10.2 | DaaPFAMM Method | 67 |
| 10.3 | Case Study | 69 |
| 11 | DISCUSSION | 76 |
| 11.1 | Results | 76 |
| 11.2 | Validity | 80 |
| 11.3 | Limitation | 84 |
| 11.4 | Future Research | 84 |
| 12 | CONCLUSION | 86 |
| 12.1 | Conclusion | 86 |
| A | PAPERS LITERATURE REVIEW | 89 |
| B | INTERVIEWS | 91 |
| B.1 | Interview Protocol | 91 |
| B.2 | Interview Letter | 97 |
| C | CASE STUDY | 99 |
| C.1 | Case Study Protocol | 99 |
| C.2 | Interview Letter | 106 |
| D | DAAP CHARACTERISTICS | 108 |
| D.1 | First Characteristics Model | 108 |
| D.2 | Input second SLR iteration | 110 |
| D.3 | Final Characteristics Model | 113 |
| D.4 | Definitions of Characteristics | 113 |
| E | DAAP FOCUS AREA | 116 |
| E.1 | First version DaaPFAMM | 124 |
| F | INTERVIEW UPDATES MATURITY MODEL | 132 |
| F.1 | Interview Agreement on Maturity Model | 132 |
| F.2 | Change Log Maturity Model | 132 |
| G | CASE STUDY RESULTS | 150 |
| G.1 | Case Study Results | 150 |



INTRODUCTION

Since the 1960s companies have started storing more data [van der Aalst, 2014]. With the recent trend towards Big Data, the amount of data each company possesses has skyrocketed. Davenport, Barth, and Bean [Davenport et al., 2012] describe how data is increasingly incorporated for on-demand decision making in business and organizational processes. Moreover, in recent years data has become central in the value proposition of new and improved products. This is exemplified by the large group of startups that sprouted after the Payment Service Directive 2 (PSD2) legislation went into effect, such as Circle or Trussle.

A recent study by the European Commission [Scaria et al., 2018] on Business-to-Business (B2B) Data Sharing shows that a significant amount of companies are involved in data sharing efforts. Specifically, 40% of the companies surveyed reported to share data with or use data from other companies. These companies mostly aim at using the data to improve their products or business model. Moreover, the study concludes that there are strong indications that the sharing of data will significantly grow in the near future. More concretely, within this group of data sharing companies, 46% of the companies expect that sharing data will be their main source of income within 5 years. Moreover, of the surveyed companies that currently do not share any data, one third sees the potential for data sharing activities in the near future.

However, the research by the European Commission [Scaria et al., 2018] also shows how companies are still experiencing issues with sharing data. Among these concerns, privacy and fear of sharing proprietary information are the major ones. Furthermore, related research on data collaboratives [Fisher and Davies, 2016] also names security and technical barriers as blockers.

In academic literature, there is only a limited number of papers researching inter-organizational data sharing. Related fields do sometimes touch on the topic. For example, researchers in the field of Open Government Data have looked at potential business models that could be used by companies interested in opening up their data [Zimmermann and Pucihar, 2015]. However, a formal research topic discussing commercial data-sharing platforms is still lacking.

Therefore, this research introduces Data as a Platform (DaaP) as a new research topic looking into commercial data platforms. These data platforms allow companies to share their data with third parties, that use the data to create new or improve value propositions. Additionally, the commercial aspect requires the platform to have a business model that makes sense for the data-sharing company.

To formalize the topic of DaaP, this work analyzes this type of platform from several perspectives.

First of all, the research creates a clear definition of the concept and discusses its characteristics. Subsequently, capabilities are identified for a mature DaaP platform. Finally, the capabilities are used to create a maturity model describing practices that can elevate a DaaP platform to a new level.



BACKGROUND

2.1 RELATED WORK

This research looks at DaaP ecosystems, which are ecosystems created around a single data platform, which is initiated with a commercial goal and which has a business model. As a concept DaaP ecosystems have not been discussed in the literature, however, some related concepts partially overlap with DaaPs. Therefore, this section will discuss the literature gap that is filled with the introduction of DaaPs and why there is a need for this new concept alongside the present fields.

First of all, DaaPs can be seen as an evolution of software ecosystems. Returning to the definition by Jansen [Jansen et al., 2009], which stated that software ecosystems are "a set of businesses functioning as a unit and interacting with a shared market for software and services, together with the relationships among them. These relationships are frequently underpinned by a common technological platform or market and operate through the exchange of information, resources and artefacts". This shows that DaaPs and software ecosystems do not differ in a major way. The obvious specification required for a DaaP is that the shared technological platform is specifically a data platform and that the core interaction between ecosystem partners is data sharing. Viewing a DaaP as a progression from software ecosystems has the benefit that much of the literature on software ecosystems could potentially be transferred to this topic, as ecosystem health assessments [Jansen, 2014] and openness [Knauss et al., 2014] are likely to be key topics for DaaP ecosystems, which are already widely discussed for software ecosystems.

Secondly, a DaaP can be seen as a specific type of data collaborative. Data collaboratives are partnerships between three or more companies that are based on the activity of sharing data [van den Broek and van Veenstra, 2018]. This corresponds to a DaaP, as the data platform that is part of the ecosystem is used by multiple parties. However, this definition lacks the specificity to accurately describe DaaP ecosystems. Moreover, these collaborations often occur in the context of a philanthropic goal. This is quite the opposite of the goal of a DaaP, which is often exclusively for economic gains. Furthermore, research in data collaboratives has often shied away of discussing potential economic benefits [Bertot and Choi, 2013] [Susha et al., 2017a] or has discussed business models on a level that is too broad for a clear analysis of commercial value in the context of data platforms as intended in DaaPs [Klievink et al., 2018]. Therefore, a specific definition and research into DaaPs, as seen as a specific type of data collaborative can revamp the commercial aspect that can be included in data collaboratives.

Finally, the open data field bears the closest resemblance to the DaaP concept. This field, which as discussed evolved from the concept of open government data, does generally focus on opening up their data for free in an online platform. While this does come close to the concept of DaaPs, it has some of the same limitations as the data collaborative field. First of all, coming from the open governmental data field, the literature interchangeably uses open data in the context of business data and governmental data, this introduces a lot of ambiguity in papers on this topic, as it is often unclear which type of open data is mentioned. Moreover, coming from a governmental perspective, also this topic tends to prioritize data sharing aimed at a philanthropic goal [Kassen, 2018] or value creation through the use of open governmental data platforms [Zuiderwijk et al., 2016]. Finally, this field has shied away from any analysis of specific ecosystem dynamics and is more focused on big-picture ideas with little empirical evidence [Sennaik et al., 2017] creating the need for a field that more specifically focuses on this perspective.

To summarize, the introduction of DaaPs as a new concept related to the previously mentioned three research fields can create new perspectives on how companies can use their data to create value. Looking at data platforms from a commercial perspective can help academia and businesses. It can help define value propositions that can help customers and their needs while combining those propositions with new business models that can bring additional value to the companies. Moreover, the definition, characteristics and maturity model that will be introduced in this research can be a starting for researchers to continue research into the DaaP concept. Through the analysis of other ecosystem dynamics in DaaPs, academics can further define this concept, assisting future applications of this topic, both in an academic and a business setting. Furthermore, businesses will be able to use the artefacts created in this research to start their own DaaP endeavours to eventually capture value from their data.

2.2 DATA SUPPLY

As discussed, since the 1960s data storage by companies has grown exponentially [van der Aalst, 2014]. With the arrival of Relational Databases in the 1970s and their widespread adoption by companies in the 1980s data collection and processing became significantly more simple [Harrington, 1998]. Technological improvements have led to cheaper and easier data storage. Increased data storing has sparked the necessity for improvements that were made through data warehousing in the 2000s and through cloud computing, which many companies use today. Data trends, such as Big Data, have increased the desire to gather data by companies, while data collection innovations such as IoT have given companies the ability to collect these immense amounts of data.

Data can be very valuable for the organization that collected it. However, after it is no longer relevant for its initially targeted use case, its value for that organization diminishes. Nevertheless, data is often not terminated or deleted and it is kept for any potential future use. This is a key aspect of many data lifecycle management models, as reviewed by [Ball, 2012]. There is often no real process for deleting data and they advise to try and continuously find re-use methods. The General Data Protection Regulation (GDPR) does require deletion after the use case has ended, at least for the storage of personal data. However, for other data, there is no such requirement. Another aspect that many of these models have in common is the re-use or re-purposing of data that is no longer relevant for its original use case. This is also where data sharing comes into play. Data sharing with other companies that may still have an active use case for that data may be an effective way to re-purpose data and to capture more value from the data.

2.3 DATA MARKET

Data sharing is not only an effective way of re-using data but there is also a lot of interest in data in the market. Davenport, Barth and Bean [Davenport et al., 2012] describe how companies that manage their voluminous and often unstructured data best are bound to get a competitive advantage. They describe how data should no longer only be used in decisions making by managers but should be integrated into

processes. Moreover, data is increasingly used as a key aspect of new products and services. All of this has led organizations towards a "data maximization" mindset, where they try to collect as much data as possible; even when no direct use case is clear [van den Broek and van Veenstra, 2018].

However, for many companies, it is not possible to gather all potentially valuable data by themselves. Take a simple thing as weather information, for many organizations enriching their data with weather data can be highly relevant. Whether it is about predicting sales in a shop or the maintenance of outdoor machinery, weather data can add value to many datasets. However, it does not make sense for all these organizations to install their own weather stations. Therefore, these companies have a specific need for this data. Organizations that do collect this data, therefore, would underutilize the value of the data when they do not share it with other organizations.

Data sharing between companies is a quickly growing market. According to research from the European Commission [European Commission, 2017], the European data market was already worth 300 billion euros in 2016 and this is expected to grow to 739 billion by the year 2020, as illustrated in Figure ???. They also estimate that there are currently around 255 000 'data companies'. These 'data companies' are defined as "data suppliers, organizations whose main activity is the production and delivery of digital data-related products, services and technologies." All these numbers indicate that Business-to-Business data sharing is already happening on a large scale.

2.4 DRIVERS FOR DATA SHARING

There are numerous reasons for companies to share data with other companies. As discussed by Klievink et al. [Klievink et al., 2018], the core reason for sharing data is to enable the creation of value beyond the capabilities and interests of each of the individual actors. For the data-sharing company, this value could just be monetary. Research by the European Commission on B2B Data sharing [Arnaut et al., 2018] discusses that for 40% of the data-sharing companies the main incentive for data sharing is the "economic value of the data sold (data revenues)". More often the incentive is "the possibility to develop new business models and/or products or services" at 74% or "the possibility to establish partnerships with other companies interested in my data" at 48%.

Additionally, benefits from data sharing can be found in the literature. The first benefit, discussed by Brodsky and Oakes [Brodsky and Oakes, 2017] in the context of Open Banking, is that data sharing fosters innovation. For the banking industry, they describe how opening up data can help banks become more competitive in nonbank markets. Next to direct value for data sharing companies, they describe societal benefits for third parties that come from sharing data. For example, by improving financial inclusion by sharing financial data on the less fortunate for a more correct credit score. Buda, Ubacht, Janssen and Sips [Buda et al., 2016] focus on the improved transparency that can come from opening up data, which can aid partnerships between companies. This benefit is further reiterated by van den Broek and van Veenstra [Broek et al., 2018], who discuss that data sharing can lead to better collaboration. They find that effective collaboration allows organizations to learn from each other, pool resources and eventually improve their business models.

2.5 DATA SHARING ISSUES

Next to the benefits, there are also some issues with data sharing that need to be resolved. While the market size seems to be reasonably high, the literature indicates that many companies are still hesitant to share their data [Günther et al., 2017] [Klievink et al., 2018]. Again, the research by the European Commission [Arnaut et al., 2018] discusses several scepticisms organizations have against sharing data, stemming from a survey of companies not interested in sharing data. The first one, often reiterated in other literature, is "privacy concerns" at 48%. Especially since the arrival of the GDPR, companies are reluctant to open up their data, afraid of openly displaying any privacy issues that may exist with their

| Reasons | Percentage |
|--|------------|
| Privacy concerns | 49% |
| Trade secrets / fear of misappropriation by others / considerations of commercial strategy | 33% |
| Lack of demand for my company's data | 32% |
| Uncertainty about safety, security and liability conditions related to the technical process of sharing data | 28% |
| Lack of incentives to share data | 18% |
| Lack of data skills inside the company | 15% |
| Economic costs of sharing data (e.g. costs of making the data available in the desired format, infrastructure costs related to data collection, data curation costs, etc.) | 15% |
| High efforts and burden on the company to engage in this activity (e.g. collection, analysis, etc.) | 15% |
| Uncertainty about usage rights on the data and potential reputational costs for the company in case of misuse | 15% |
| Difficulties with measuring the value of data | 11% |
| Lack of appropriate licensing conditions | 0% |

TABLE 2.1: *Issues with sharing data found by the European Commission*

internal data. Additionally, they are worried that their data can be combined with data from other companies to reidentify the individuals in the dataset.

Secondly, many companies are concerned with protecting their trade secrets. At 33%, many companies indicate their reason for not sharing data is "Trade secrets/fear of misappropriation by others/considerations of commercial strategy." This point was extended by Richter and Slowinski [Richter and Slowinski, 2019], who notes that companies may not only be scared to give up secrets but are also hesitant to give up the sole control of their data. They discuss how this is especially the case for market leaders, whose dominant position may have granted them access to most data on the market. They may be reluctant to share their data, thinking they have little to gain from other organizations doing the same thing. These other organization, on their part, are then disincentived to share, because that will give the dominant company only access to more data.

Finally, there are security and technical skills issues that are described as reasons not to share data. Again, the research from the European Commission [Arnaut et al., 2018] describes this in two different points. First of all, "Uncertainty about safety, security and liability conditions" is seen as a problem by 28% of the companies that prevents them from sharing data. Secondly, the "economic cost of sharing data", which also targets the cost of the technical implementation, is mentioned by 15% of the companies. This problem is also discussed in literature by Correa, Zandar and da Silva [Correa et al.,] in the context of Open Data portals. They discuss how various options for Open Data Portals have different rates of both adoption by share and re-use companies, possibly caused by a different level of technical skill required for using these portals. An overview of all the reasons not to share data found by the survey from the European Commission can be found in Table 2.1.

Finally, the research by the European Commission [Arnaut et al., 2018] discusses some potential solutions or improvements that could incentivize unwilling companies to start sharing their data. The biggest point found in the survey was the creation of "legal clarity about the 'ownership rights' of the data." This corresponds to two of the problems discussed earlier: the privacy concerns and control over the data. Next to that, organizations indicate that the ability to "track usage of the data after sharing"

| Factors that can increase willingness to share data | Percentage |
|--|------------|
| Legal clarity about the “ownership rights” of the data | 62% |
| Ability to track the usage of the data once it has been shared | 46% |
| Certainty about how to share data from a contractual point of view | 42% |
| Availability of the necessary technical skills inside my company to ensure the quality and security of the data shared | 29% |
| An improved framework to protect the investments made for data collection, curation, anonymisation, etc. | 24% |
| Availability of standards and/or infrastructure to facilitate the adequate storage, transfer and processing of data | 16% |
| A defined framework for liability in case of damage caused by the data that are shared | 15% |

TABLE 2.2: *Factors to improve willingness to share data found by the European Commission*

would be important for them together with a “certainty about sharing from a contractual point of view.” Only after these problems come to any technical concerns, showing that compliance and legal protection are more pressing concerns than any practical limitation. A full overview of potential changes that would incentivize data sharing can be found in Table 2.2.



RESEARCH METHOD

This chapter describes the design of the research. In general, this research looks to introduce Data as a Platform in academic literature. In this chapter, firstly, the problem that has been found to be unaddressed in literature is described and DaaPs are introduced as a potential solution to this problem. Secondly, the research questions are introduced, consisting of one main question and four sub-questions. Thirdly, the research method is laid out. This consist of a high-level research design following the Design Cycle by Wieringa [Wieringa, 2014]. This design is made concrete through a method consisting of a systematic literature review (SLR), semi-structured interviews and a case study.

3.1 PROBLEM DESCRIPTION

Organizations are sitting on large piles of data, gathered from a wide range of sources. Often it is gathered without a specific use case in mind or stored for future reference even after its direct purpose has ended. While GDPR is putting a stop to the purposeless collecting and storage of personal data, companies still collect large amounts of purposeful personal data and non-personal data. After storing, this data is often left untouched for years[Buda et al., 2016]. Therefore, a lot of potential value in this data is lost, which could be unlocked by sharing the data. While they do not have use for it, organizations are still reluctant to share the data and give up their control over it [Najjar, 2013] either in fear of giving valuable insights to competitors or compliance issues, for example, with the General Data Protection Regulation (GDPR) law.

Data sharing can not only unlock value for companies on the supply side. On the demand side, many organizations are looking to make use of data. In a world where data has regularly been named the new oil[Wang, 2018], organizations have an enormous incentive to maximize their data for use in their commercial activities. This has persuaded organizations to search for alternative data sources and gathering methods[European Commission, 2017]. However, not all relevant data can be gathered by a company internally, as shown by the weather data example in subsection 2.3.1.

The problem this research attempts to solve can be summarized in the following way: *On both sides of the market there is interest in data sharing between companies. However, little is known about how this data can be shared through a platform.* Through the introduction of DaaP as a solution for data sharing, this research aims to take away the concerns and enable the widespread sharing of data.

3.2 RESEARCH GOAL AND QUESTIONS

DaaPs are researched as commercial data platforms giving third party organizations access to data, collected by other organizations, to create new or improve existing products and services. The data is shared with a commercial goal and isolated business model. This research develops several artefacts that jointly introduce DaaPs in literature. To lay a proper foundation for the DaaP concept, adoption by both practitioners and researchers is required. Therefore, all artefacts are created with both groups in mind. The following research questions are answered in this research:

Main Question: *How can DaaP platforms be utilized by organizations to unlock the value of their data?*

This main question aims the research towards introducing DaaP as a solution for organizations to commercialize their data. The commercial focus is deliberate, as it differentiates this topic from Open Government Data and related topics. Furthermore, the main question of this research is aligned with the main goal for practitioners that consider using a DaaP. This decision was made to focus the research efforts on introducing the topic in a way that is attractive to practitioners and aims to promote the use of DaaP in practice.

As the main research question is relatively broad, the research question is answered through the creation of several artefacts. Four sub-questions are posed to answer this main question and create the artefacts.

Sub-question 1: *What defines a Data as a Platform ecosystem?*

The first sub-question is designed to introduce DaaPs to academic literature through a newly formed definition. This definition is created to explain the topic in a precise and compact manner. Moreover, it is used to express the scope and boundaries of this topic, making the field more concrete for future research. Finally, this definition answers a key aspect of the main question, explaining what a DaaP entails to both researchers and practitioners.

Sub-question 2: *What are the characteristics that differentiate various types of DaaP ecosystems?*

The second sub-question is concerned with the characteristics of DaaP ecosystems. These characteristics are used to elaborate on the concept and to differentiate between types. To illustrate the results of this question, a model is created with the characteristics and the alternatives organizations can choose from. For example, if they want to open their data for everybody or restrict access to a small group of entities.

Sub-question 3: *What capabilities are relevant for operating or utilizing DaaP ecosystems?*

The third sub-question is a survey of capabilities that are important for DaaPs. These describe various ways companies can improve their DaaP platform. Capabilities are found at two levels. First of all, there are Focus Areas, which are high-level capabilities that a company can implement. There are various ways to go about improving these Focus Areas described in capabilities and within the capabilities, different levels of maturity can be achieved. Finally, the relevance of the Focus Areas for a specific DaaP can depend on the context of that DaaP. Additionally, there are Practices. Practices, in this context, are a specific implementation of a Focus Area. Practices enable specific functionality in a DaaP or define a specific decision made for the DaaP. In the next research question, these capabilities are used to create the maturity model, where the key capabilities are set out as focus areas in the maturity model and the different maturity levels of the capabilities are determined.

Sub-question 4: *How can an organization develop the Capabilities to get them to a level that fits their objectives?*

The fourth sub-question is the design and validation of the maturity model. For this research, a focus area maturity model, as discussed by van Steenberg et al. [van Steenberg et al., 2010], is selected. An example of a focus area maturity model can be found in figure 3.1. In the results from sub-question three, the elements that are relevant for a DaaP are determined. In this model, these elements are laid out as capabilities and practices, where the capabilities are on the left-side and practices fill the cells in the models. In this question, the various maturity levels connected to the practices are determined and modelled. Moreover, dependencies between the various maturity levels of the focus areas are determined and modelled. This model answers the main research question from various sides. It considers the organizational, ecosystemic, technical and data aspects key to improving the platform. It can be used by practitioners both those starting and those already having an active DaaP to develop the right capabilities to make their platform more mature and stable for themselves and the ecosystem partners. It also shows how to get to the desired maturity level from the current point and what other capabilities are required to get there. Finally, it can be used by researchers to extend on this research field. Specifically, Santhanam and Hartono [Santhanam and Hartono, 2003] discuss how maturity models do not only assist practitioners by providing them with a step-by-step plan, they also assist researchers as they create a better understanding of a functional domain and make capabilities measurable.

| Focus Area | Maturity Level | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|--------------------|----------------|---|---|---|---|---|---|---|---|---|
| Data Actuality | | | | | A | | B | | | C |
| Data Documentation | | | A | | B | C | | | | |
| Data Availability | | A | | | B | | | | C | |
| Data Flexibility | | | | A | | | B | | C | |

FIGURE 3.1: *The Focus Area Maturity Model*

3.3 RESEARCH METHOD

In this research, the Design Cycle, as introduced by Wieringa [Wieringa, 2014], is used as a research method. The design cycle and the more extended engineering cycle are displayed in figure 3.2. The Design Cycle follows three steps. The first step is problem investigation. During this step input is gathered for a solution by analyzing the problem. This step usually consists of a review of related literature on the problem and this can be extended with a problem investigation in practice through, for example, expert interviews or an observational case study. The second step is the treatment design. In this step, the results from the problem investigation are used to create a potential solution for the problem, usually in the form of an artefact. Examples of artefacts often created in information systems research are methods, models, techniques and notations. The third step is treatment validation. In this step, the artefact, designed in the previous step, is tested. Real-world implementations of designed solutions are often difficult, expensive and time-consuming. Therefore, the validation is performed in a test controlled environment. The validation is done using research methods that try to validate it in a controlled environment. Frequently

used research methods for this step are case studies, expert opinions and experiments. The goal of these methods is to determine how an artefact would behave in practice, without implementing it. Finally, in the design cycle, the results from the treatment validation can be used as the input for a new cycle restarting the process with the problem investigation. In this research, the cycle is completed once. During this iteration, two artefacts are validated. Similar methods are used for both the characteristics model discussed in research question 2 and the capabilities and Maturity Model discussed in research question 3 and 4. In addition, the design cycle can be extended in the engineering cycle. In this cycle, the process can continue, after the validation is done successfully. This is done with a real-world implementation of the artefact or solution, in the treatment implementation step, and the subsequent implementation evaluation, monitoring the success of the solution in the problem context. These steps are not taken in this research.

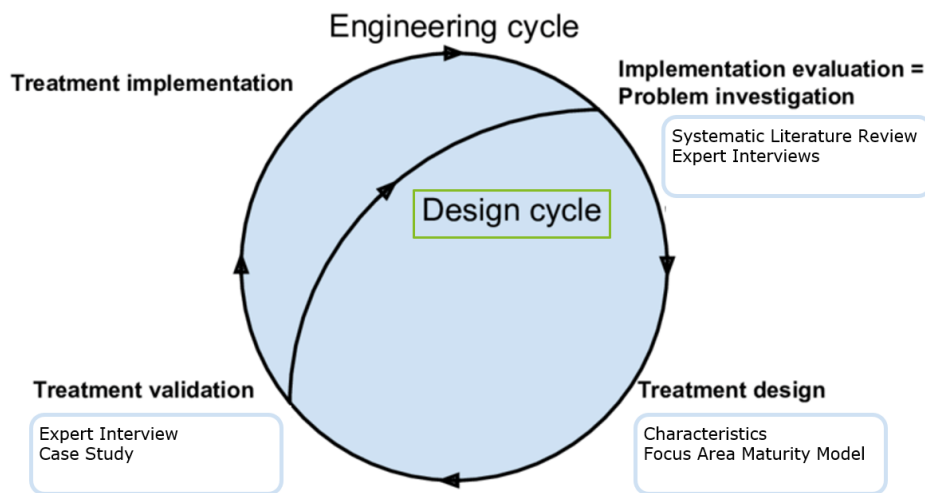


FIGURE 3.2: *The Design and Engineering Cycle by Wieringa [Wieringa, 2014]*

This research method is selected for several reasons. First of all, it is created specifically for the larger field of information system research in which this research is conducted. Wieringa [Wieringa, 2014] describes how artefacts designed in information system research are applied in the context of the design, development, maintenance and use of information systems. In this research, this is particularly the case for the maturity model, as the model can be applied during all these phases. Secondly, the various steps that should be taken in the design cycle fit the resources available for this research. The limited presence of DaaP implementations in practice makes quantitative research challenging. Wieringa describes how qualitative research methods can be used in the steps. Specifically, he states that expert interviews can be used for the problem investigation phase and that a case study can be an excellent way to validate the artefact in the treatment validation phase. These are research techniques that fit the exploratory phase, in which the concept of DaaP currently resides, as they do not require a large number of participants.

In this research, the problem investigated using the design cycle is expressed in the main question. It suggests that companies are struggling to capture value from their data and that DaaPs offer a potential solution to this. The first step towards introducing a potential solution to this problem, according to the design cycle, is to investigate the problem. This is done in research questions one to three. The first question is designed to set out a precise description of the DaaP concept and to draw the boundaries of the field. The second question is included to investigate the specifics of a DaaP platform. Determining the characteristics of a DaaP allows for an in-depth analysis of how they could potentially be used to capture value from data. Furthermore, this question is designed to gather input for the last research question that is part of the treatment design and validation phase. After the first two research questions, looking

at DaaPs as a whole, the third research question scopes the research towards one aspect of a DaaP, its capabilities and the maturity thereof. This introduces the main assumption in this research, namely, that the maturity of the capabilities of a DaaP impacts the value the operating organization can capture from it. This assumption is, subsequently, tested in the fourth research question, in which the input from the third research question is used to design a maturity model. Therefore, this question moves the research from the problem investigation phase to the treatment design and validation phase, which are both present in this research question. First, in the treatment design part, the maturity model is created. Thereafter, in the treatment validation phase, it is validated in a case study.

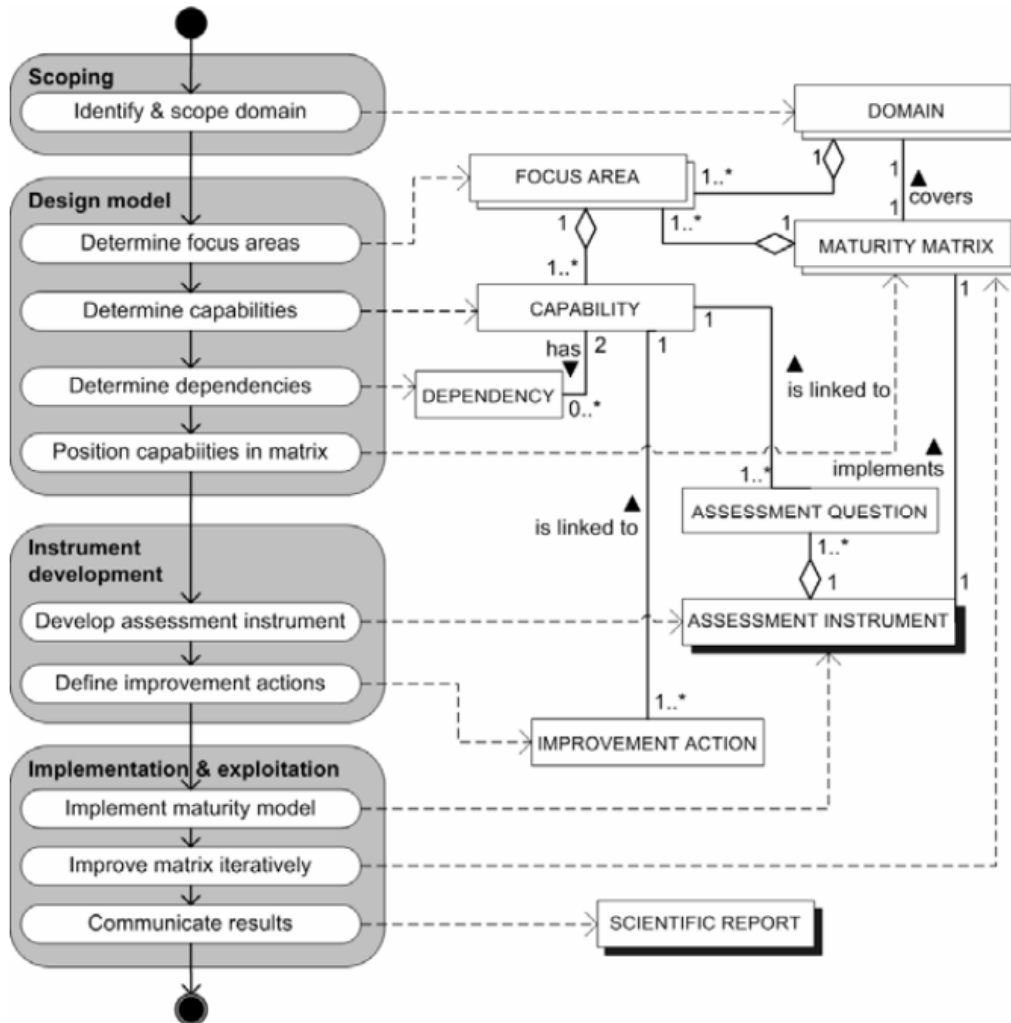


FIGURE 3.3: *Development Method for Focus Area Maturity Models based on van Steenbergen et al. [van Steenbergen et al., 2010]*

Throughout the design cycle phases, several research techniques are used to answer the research questions. During the problem investigation phase, performed in research questions 1 to 3, a systematic literature review is done and various expert interviews are conducted that only partly fall into the problem investigation phase. The literature review is done to retrieve the knowledge on this topic and on related topics that have been described in the literature. For this systematic literature review the systematic

guide to literature review development method by Okoli [Okoli, 2015] is used. An exact description of this method and the rationale behind choosing it is discussed in the next chapter. The second technique used in this phase is expert interviews, which are used for both problem investigation and treatment validation. These are used to extend the input for this research beyond the analysis of the problem from an academic perspective in literature, towards the consideration of the current business context of the problem, by interviewing experts on that side. Wieringa [Wieringa, 2014] suggests expert interviews as one of the research techniques that can be used during the problem investigation phase, especially in a semi-structured format. In this format, an interview is started off with a fixed set of questions, but it also allows for expansion on topics during the interview [Louise Barriball RGN Research Assistant and While RGN RHV Cert Ed, 1994]. To differentiate the problem investigation and treatment validation input from the expert interviews an incremental method is used for processing the results. This method is discussed in more detail in chapter 5.

Then, as the research moves to the treatment design and validation phases, different techniques are applied. First of all, in the treatment design phase, present in research questions 4, an artefact is developed. For the fourth question, this is a maturity model. The focus area maturity model will be designed based on the Development Method for Focus Area Maturity Models as introduced by van Steenbergen et al. [van Steenbergen et al., 2010]. This method is illustrated in figure 3.3. More details about Focus Area Maturity Models are discussed in chapter 9. Finally, in the treatment validation phase, both artefacts are validated through the expert interviews and a case study. For the expert interview, the feedback of the participants on the artefacts is used for their validation. Experts are asked for their opinion on the models. How exactly the expert interview results are processed for validation is discussed in chapter 5. The case study design used is an embedded single case study, as described by Yin [Yin, 2017]. This type of Case Study allows for the validation of several 'units of analysis' in a single case. In this Case Study, the units of analysis are the artefacts designed in the treatment design phase. The case itself is from an organization maintaining and maturing a DaaP. An overview of the entire method, including the research questions, the Wieringa steps and the research techniques is illustrated in figure 3.4.

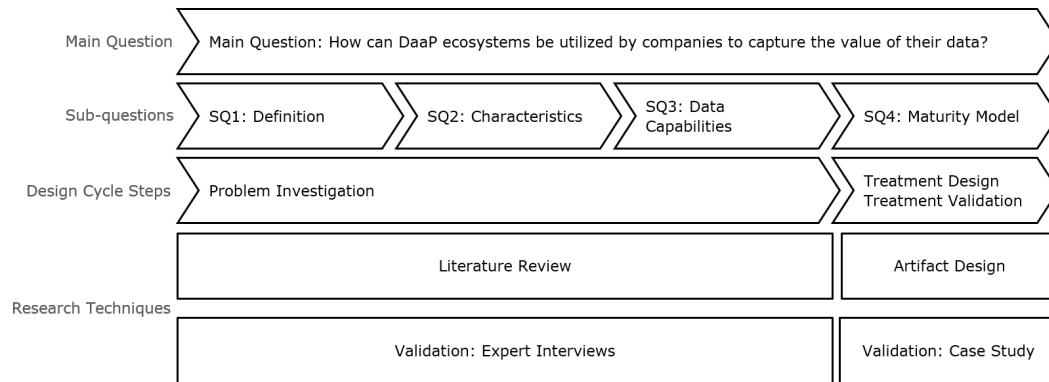


FIGURE 3.4: Complete Research Method



LITERATURE REVIEW

4.1 SYSTEMATIC LITERATURE REVIEW METHOD

In accordance with the research design, the first step in the research is the Systematic Literature Review (SLR). The goal of the literature review is to gather input for answering the research questions from the academic literature. For this review, the Systematic Guide for Literature Review Development by Okoli [Okoli, 2015] is used. Several SLR methods were considered, such as Kitchenham and Charters [Kitchenham and Charters, 2007], Eisenhardt [Eisenhardt, 1989] and Wolfswinkel et al. [Wolfswinkel et al., 2013]. However, after careful consideration, the method by Okoli was selected. First of all, the work by Okoli is the most recent publication of all methods considered. Therefore, it is considered the most relevant and up to date review method. Moreover, it discusses and integrates parts of the other methods considered, which strengthens its fit for purpose. Additionally, Okoli's method is specifically focused on information systems research, the larger field in which this work falls. Finally, Okoli specifically states that this method is useful for Thesis research, which this work is, as it gives students a foundation for solid literature analysis, something that is regularly lacking in Theses.

The Systematic Guide to Literature Review Development, as defined by Okoli [Okoli, 2015] is displayed in Figure 4.1. It consists of eight steps that are all essential for a successful SLR. The literature review in this paper, nevertheless, skips the 'Train the Team' sub-step of step 2. This is done because there is only one researcher performing the literature review and, therefore, there is no team to train.

4.1.1 PURPOSE OF THE LITERATURE REVIEW

As illustrated in Figure 4.1, the first step in conducting an SLR, according to Okoli [Okoli, 2015], is identifying and describing the purpose of the SLR. The goal of this step is to, first of all, make sure that an SLR is necessary and useful for the research. Secondly, it is used to determine the ultimate goals and audience for which the SLR is conducted.

For this research, the SLR has several purposes. The main purpose of the SLR is to answer the research questions that require a review of academic literature. For several sub-questions, it has been determined that they could potentially be answered through literature analysis. Therefore, this SLR focuses mainly on answering those questions. The research questions answered in the SLR and how they translate to research questions for this review are discussed in the next section. Additionally, the SLR has the goal to

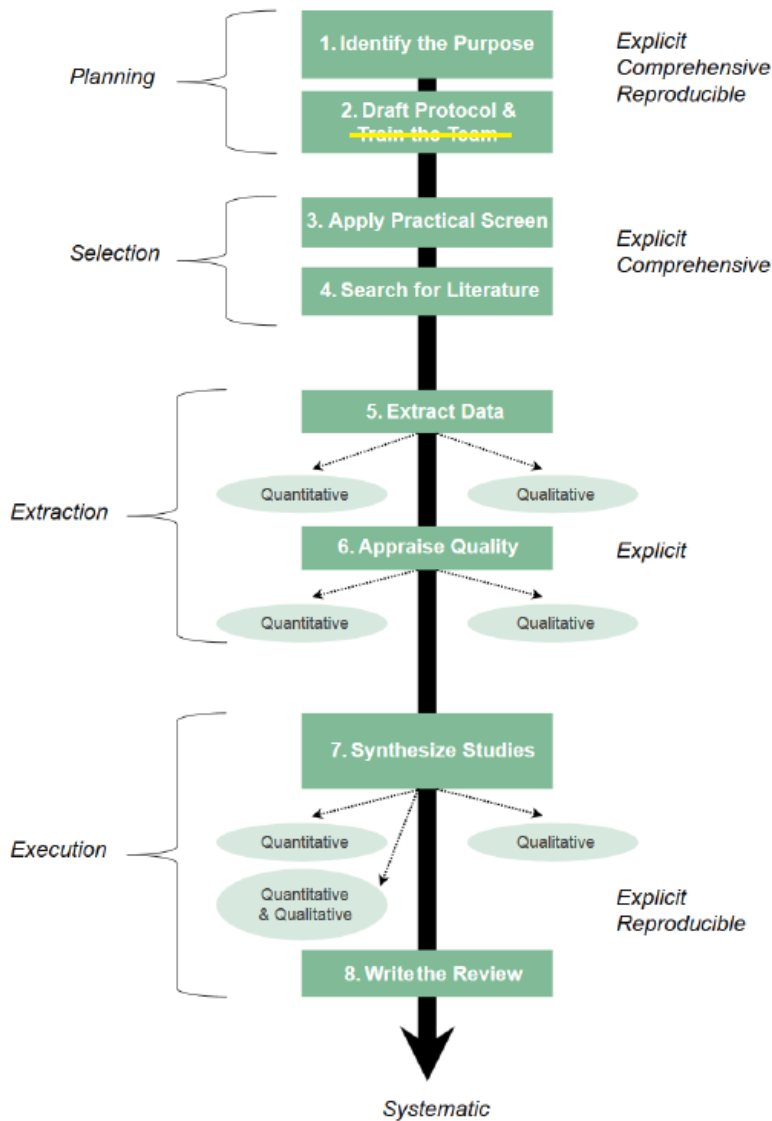


FIGURE 4.1: *The Systematic Guide to Literature Review Development (a yellow line indicates a skipped step)*

determine the place of this work in literature. As this research introduces a new topic in literature, it is key that it has not previously been introduced in the literature under a different name or in a related field. Moreover, it needs to be positioned between related topics. This helps the research by identifying topics that should be discussed in this work. Furthermore, it helps future researchers understand the context of the topic and get ideas for research questions that should be answered in future work. Both can only be achieved through a careful study of existing literature. Finally, for the understandability of the topics introduced in this work, terminology must be used that corresponds with that of related topics. The SLR can help identify the definitions and vocabulary necessary for framing the research within the field and making it understandable for the readers.

As Okoli [Okoli, 2015] discusses, it is not only important to specify why the SLR is performed when discussing its purpose, it is also key to determine its audience or the dissemination targets, as he calls them. This literature review is executed for both researchers and practitioners. It is meant for researchers, as it attempts to clearly describe the position this work takes between related research fields. This enables researchers to research this topic more easily in the future, as the SLR describes what fields could be considered for more insights into the topic and what terminology is used in those fields. On top of that, practitioners are also seen as a key target group for this SLR. For two reasons the SLR is likely to be used by them. First of all, the SLR is conducted as part of graduation research done at a major consultancy firm, whose consultants are likely to apply the results of the research in their future work. Secondly, the focus of the research is on introducing DaaP platforms as a potentially profitable business endeavour. To enable practitioners to use the results optimally, it is required that they also understand the context of the research, discussed in the SLR. This second audience group likely has limited knowledge on the context of the research, the terminology used and the related research fields. Therefore, the SLR is written to introduce topics from scratch allowing practitioners to catch up on the prerequisite knowledge.

4.1.2 PROTOCOL

The next step in the method, as identified by Okoli [Okoli, 2015], is to create a protocol and train the team. As described in the introduction, this SLR is done by a single person, therefore no team has to be trained. Nevertheless, for scientific rigour, a protocol is created for the SLR.

RESEARCH QUESTIONS

The next step in the Systematic Guide to Literature Review Development is defining the research questions for the SLR. As discussed in the previous section, the SLR has three purposes. The first is answering research questions from the overarching research, the second is positioning this work in the related literature and the third is determining the terminology that is commonly used in this topic. Therefore, the SLR research questions echo these purposes. The following SLR questions have been created:

- (A) What defines a Data as a Platform (ecosystem) or how are related topics defined?
- (B) What are the characteristics of a DaaP or related concepts discussed?
- (C) What types of DaaP or related ecosystems can be found and what differentiates them?
- (D) What capabilities are relevant for the quality of a DaaP platform?
- (E) What are related fields and what is the literature gap filled by this research?

As it is unlikely that all research questions can be answered in the context of DaaP research, each research question explicitly states that related topics should be analyzed to retrieve the knowledge from related fields. Additionally, a review of related topics allows for the introducing of the terminology used for these related fields, achieving the final one of three stated purposes. For the same reason, identification of the correct terminology is not framed as an SLR question, but is considered throughout the SLR questions. The mapping of the SLR questions is displayed in Table

SLR questions A to D have been created to answer the various research questions from the overarching research. Specifically, SLR question A corresponds directly to research question 1, SLR research questions B and C correspond to sub-question 2 and SLR question D corresponds to research question 3. Sub-question 2 has been divided into 2 SLR research questions, as it helps with separating concepts between characteristics that DaaP platforms generally have in common and the characteristics that differentiate various types. Finally, the fifth research question is designed to achieve the final purpose defined for this SLR, which is identifying the research gap filled in this research and positioning this work.

| SLR Research Question | Research Question | SLR Goal |
|-----------------------|-------------------|--------------------------|
| A | 1 | Answer Research Question |
| B | 2 | Answer Research Question |
| C | 2 | Answer Research Question |
| D | 3 | Answer Research Question |
| E | - | Describe Literature Gap |

TABLE 4.1: *The mapping of the SLR questions on the Research Questions and SLR Goals*

LOCATIONS AND SEARCH TERMS

First of all, this protocol describes the various locations used to find literature. Three libraries have been selected, as it was found that adding more libraries did not significantly diversify the papers found and the author had access to these options. The following locations are used for this literature review:

- Scopus
- Google Scholar
- Web of Science

Through a scoping search, as introduced by [Booth et al., 2016], in which various search terms and their initial results are considered, several search terms have been selected to answer the SLR research questions. The keywords are commonly used in the research field to describe these and similar concepts in literature. The following keywords are used in this SLR:

1. Inter-Organizational Data Sharing
2. Data Collaboration
3. B2B Data Sharing
4. Data as a Platform
5. Data Platform Ecosystem
6. Commercial Open Data
7. Inter-organizational Data Governance
8. API Governance

None of the keywords are specifically connected to any single SLR question and the resulting papers are checked for contents on any of the questions. With regards to SLR question E, the question discussing the related fields and the literature gap, for this question only research fields are considered that are found to be relevant for any of the other research question. Therefore, the literature gap discussed in any paper not relevant for SLR questions A to D is not considered for SLR question E.

4.1.3 APPLY EXCLUSION AND INCLUSION CRITERIA

To limit the number of papers included in this SLR, a practical screen for the selection of the final papers is applied. This 'practical screen', as it is called by Okoli [Okoli, 2015], consists of inclusion and exclusion criteria that determine what papers will end up in the final selection. Exclusion criteria are certain filters added to the online library search to reduce the number of papers that will be returned. Inclusion criteria are, subsequently, criteria that will be applied during the reading of the paper and are related to the relevance of the paper for this research.

For this SLR the following exclusion criteria are used.

- The paper must be published after January 1st 2015 and before June 1st 2019
- The paper must have been accepted for either a business, information science, computer science or engineering related journal or conference
- The paper must be written in English
- For Google Scholar: Only the 50 most relevant results are used

Moreover, the following inclusion criteria are used. A paper does not have to confer to all inclusion criteria, merely discussing one of the topics is enough for consideration of inclusion.

- Data sharing between companies is discussed
- The paper gives insights on at least one of the research questions
- A data flow in an ecosystem is considered

The inclusion criteria are applied at various points during the SLR. Specifically, the inclusion criteria are applied after reading each of the following parts of the remaining papers.

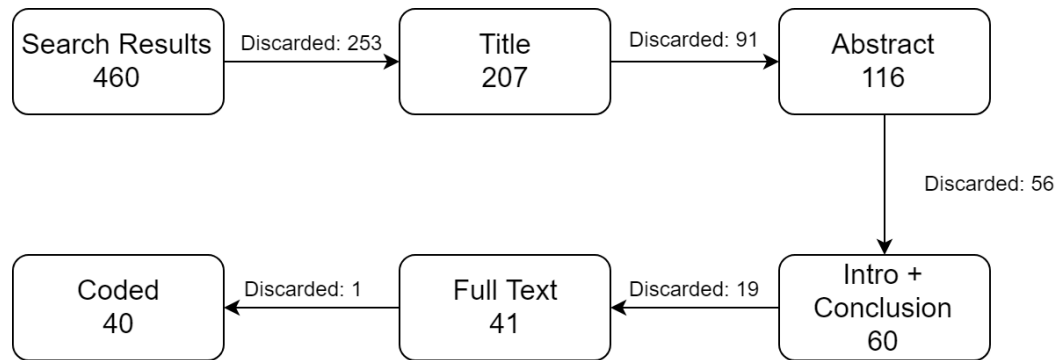
- Title
- Abstract
- Introduction and Conclusion
- Full Text

Naturally, the title and abstract may not give great insight into whether or not the paper concurs with the selection criteria, due to their short nature. Therefore, during these steps, the inclusion criteria will be applied with great leniency.

4.1.4 SEARCH FOR LITERATURE AND EXTRACT DATA

The next step, according to Okoli [Okoli, 2015], is to perform the actual search. For this step, the protocol and practical screen, as introduced in the previous sections, are applied step for step. Furthermore, it can be a good idea to perform a reference search on the landmark or particularly relevant papers. During this stage, both the references and future citations of these landmark papers are added to the paper selection and are then evaluated using the same practical screen. The papers that have been found and the progression of the filter process laid out in section 4.1.4 is displayed in figure 4.2.

After that, Okoli [Okoli, 2015] describes the next step as a crucial step in the review procedure. It requires the reviewers to retrieve the information from each paper in a systematic way, to prepare for the synthesis step. He suggests a form of data extraction where structured details can be stored and that allows additional space for comments. While Okoli does not prescribe any specific data extraction

FIGURE 4.2: *The filter process of the literature review*TABLE 4.2: *Coding Node Definitions*

| Coding Node | Definition |
|-----------------|---|
| Characteristics | An aspect of a Data as a Platform platform, or related data sharing technology, that describes a key part of that technology. A characteristic belongs to a category that can have several variations that differentiate various implementations of these technologies. |
| DaaP Capability | A high-level capability that can, depending on the context, be seen as important for the performance of a DaaP or related data sharing technology. A focus area allows for various implementations of the focus area, in the form of practices, that differ in terms of maturity. |
| DaaP Practice | A practice is a specific implementation of a focus area. A practice can be placed at a specific maturity level. |
| DaaP Definition | A definition of a Data as a Platform platform, or related concept, or a definition from a related data sharing technology concept. |
| DaaP Type | A taxonomy or description of a specific type of Data as a Platform platforms or related data sharing technology. |

method, he does point at several options. For this research, the decision has been made to code the results using Nvivo. To be specific, various broad codes were defined, centered around the various SLR research questions.

To make the coding of the selected papers repeatable, definitions for the various coding nodes are created. These definitions are displayed in table 4.2. Next to these codes, several codes were created for identifying parts of papers that could be useful. First of all, a code was introduced to identify interesting papers for further research in a reference search. Moreover, a code was made to note the benefits and issues of using DaaPs, the result of this coding can, for example, be used in the background section or for future work. Finally, a node was created to note the description of literature gaps that help place this work in literature.

4.1.5 APPRAISE QUALITY

To ensure the quality of the papers considered for the SLR, the use of a quality assessment form is suggested by Okoli [Okoli, 2015]. He suggests the use of the quality guidelines as introduced by Fink [Fink, 2005]. However, due to the limited availability of this form and its foundation in the Health Sciences,

whose rigorous standards may have a negative impact on the number of results, this work uses the guidelines by Kitchenham and Charters [Kitchenham and Charters, 2007]. They present a collection of questions, based on various methodologies, and suggests that a selection should be made based on the context of the SLR. Moreover, they differentiate between qualitative and quantitative study and have a different set of questions for either. Based on these quality appraisal questions, two papers were disregarded during the Full Paper reading phase, both papers had a complete lack of discussion on any sort of research method.

4.1.6 SYNTHESIZE STUDIES AND WRITING THE REVIEW

Okoli [Okoli, 2015] states that after selecting and reviewing all the papers, the next step is to combine the codes from the various papers to make sense of a large number of results. In this stage, the results are aggregated, discussed, organized and compared. After the codings are combined, the codes are sorted based on their topic. As can be seen from the various keywords, several topics related to DaaP have been identified and the codes with definitions, characteristics and types are sorted to fit each of the topics, allowing for the discussion of research question answer per topic.

The final step of the literature review is writing the review. For this step, Okoli [Okoli, 2015] refers, among other things, to the writing process as described by Kitchenham and Charters [Kitchenham and Charters, 2007]. Their structure for reporting on an SLR, as displayed in Table 8 in their work, will also be the structure for this review.



INTERVIEWS

This chapter discusses the interviews that were conducted after the SLR. For the interviews, the materials resulting from the SLR were used as input. Specifically, the maturity model and characteristics model were used. In total eight interviews were conducted with experts from a wide range of companies and backgrounds. An exact description of every interview participant is given later in this chapter.

The goal of the interviews was to extend and validate these models. All literature used for these models came from reference disciplines. Therefore, additional validation was required to determine if concepts found in the literature on the reference disciplines was applicable to this field. For this reason, experts were asked to give their opinion on every aspect of the initial models. Moreover, it was expected that merely translating literature from related fields would give an incomplete view of the DaaP field. Therefore, experts were asked to propose new concepts to the model where they saw fit.

This chapter discusses the interviews performed in this research in more detail. First, the set-up of the interviews is explained, including the interview protocol. After which the method used for processing the interview results is discussed. Then the interview participants are introduced. Finally, the execution of the interviews is discussed.

5.1 INTERVIEW SET-UP

Based on the goals in the introduction, a research set-up is created that forms the foundation for the interviews. This set-up consists of an interview plan, an interview protocol and an invitation letter. With these materials, the interviews are conducted. First of all, the interview plan is used to describe how potential participants are selected, what the interviews is like and how the results of the interviews are processed. Secondly, the interview protocol lays out how the interviews are conducted, describing how the topic is introduced and what questions are asked. Finally, the invitation letter is to invite participants. It is created to give a clear understanding of why and how the interviews are conducted and to give a brief introduction to the topic to allow receivers of the letter to determine their expertise on the topic.

5.1.1 INTERVIEW PLAN

The interview plan starts off describing in detail the goals for the interviews briefly discussed in the introduction of this chapter. The following goals were identified:

- Identify and evaluate a definition and characteristics of Data as a Platform
- Identify and evaluate Data Capabilities required in a Data as a Platform platform
- Identify and evaluate Business Models that can be used for DaaP platforms

Next to these formal goals, it is decided that interviewees are asked for examples from their experience. These examples can be used later in the research and in this thesis to describe findings in more detail and give additional context to them through real-world examples.

Subsequently, the interview plan discusses the procedure around conducting the interviews. The interviews are conducted in a semi-structured manner. Which means that, while there is a specific structure for the interviews with predefined questions, there is still room for off-script deep dives and continuation questions on topics the interviewer deems relevant for the research [Louise Barriball RGN Research Assistant and While RGN RHV Cert Ed, 1994]. This is preferred over a completely structured interview, because it is expected that new concepts or practices surface, which require follow-up questions. Furthermore, the interviews are to take around an hour each. This is considered a little short for the content of the interviews. However, it is projected that requesting more time from interview candidates significantly reduces the number of available interviewees, as they are expected to all have full-time jobs with busy schedules. Next to that, it is considered that it is preferable to request a follow-up interview or a continuation at a later time after the interviewees have already invested time in it.

In addition, it is decided that interviews are conducted either in person or through a video-calling service such as Skype. However, as several materials are used to assist with the interviews, notably the models, it is decided that there is a preference for in-person interviews. This allows participants to make remarks on the paper print-outs of the model, which simplifies the understanding after the interviews. Finally, the interviews are recorded, transcribed, anonymized and sent back to the interviewee for approval before they are used in the research.

This procedure is selected for several reasons. First of all, as many interviewees come from a business background and discuss active business activities, it is crucial for them and their trust that they are allowed to check if the transcript is well anonymized and no proprietary information is still present in the transcripts. Additionally, it allows participants to reconsider their explanation and phrasing of the things they discuss. In contrast to writing, spoken word, especially in an interview, does not give a person much time to consider their words carefully to express themselves precisely. Allowing interviewees to revise what they say after the fact, allows them to rephrase their words in such a way that may give additional clarity or insight to the interview.

Secondly, the interview plan describes a profile for the interviewees. First of all, the participant is supposed to work for or be very familiar with a company that is involved with a data platform. This company can be the operator of a data platform, however, can also be a regular user of one or more data platforms. In the second case, the regularity with which the data platform is used and the perceived expertise of the company is considered. Furthermore, the participant has to have a role within the company that is actively involved with the data platform. This requirement is set to make sure that the individual participating in the interview has enough relevant knowledge to contribute. Additionally, active attempts are made to interview participants with a diverse background. First of all, the group of interviewees should consist of both participants with a technical role and with a business role. Especially for the maturity model, input from both perspectives is required to ensure that the model has a broad perspective on the topic. Secondly, the group should consist of experts on various types of platforms. For this, a separation is made between commercial data platforms and open government data platforms. This separation is primarily made because it is expected that open government data interviewees are easier to find, as most well-known data platforms are from that type. On top of that, they may have an easier time discussing details, as they have fewer concerns about sharing proprietary information. On the other hand, participants from commercial data platform share insights in the commercial aspects relevant for the DaaP topic.

5.1.2 PROTOCOL AND INVITATION LETTER

Based on this interview plan, a protocol and an invitation letter are drafted. The protocol contains the specific questions that are asked to the interviewees. It can be found in Appendix B.1. The interview protocol roughly consists of 5 parts. These are the introduction, the DaaP definition, the DaaP characteristics and the corresponding model, the DaaP maturity model and the DaaP business models. Each part starts with a small explanation after which the relevant questions are asked.

The interview starts with an introduction, which includes a short explanation of the way of working. The process of transcribing, anonymizing and accepting the transcript is explained and consent is asked for the interview and recording. After that, the topic itself is introduced. The term DaaP is explained and some context is given on the goals of the research and the goals of the interviews. While the main goal of the interview is validating the models that were created, it is deemed crucial that first a shared common understanding is created between interviewer and interviewee on the topic of DaaP. As the field is new to academic literature, it is unlikely that any participant has a clear understanding of the scope of this research field. Moreover, the various artefacts that later will be discussed during the interviews are mentioned. The introduction then finishes up with some questions. Firstly, some introductory questions are asked about the interviewee and the company they work for. Secondly, they are asked to give insights into the data platforms they are familiar with and, finally, the role they have played within those platforms.

After the introduction, the interview starts off discussing the definition of a DaaP. For this question, interviewees were first asked if they fully understand the topic of DaaP as explained to them in the introduction. When this is the case, the definition made during the SLR is presented to them and they are asked to give their input on it. To make sure that they consider how the definition could be lacking from various perspectives, they are asked about it in various ways. First of all, they are asked if the definition fits the topic as described to them. Secondly, they are asked if it has all key characteristics that differentiate a DaaP from other data platforms. This is done to make sure the definition is not too broad. Finally, they are asked if they think the definition does not potentially exclude any data platforms that they think could be considered a DaaP. After all the feedback is given, they are asked to explain how the definition fits their data platform and its corresponding ecosystem, to put their perspective on the definition into more context. This is done both to explain the definition more practically, as well as to put their feedback into perspective when deciding on a final definition.

After that, the next topic discussed is the DaaP characteristics. After a brief explanation, this question immediately dives into the corresponding model. Participants are asked to give feedback on each cell of the model. As this model consists of categories of characteristics with the 'choices' a company can make within them, participants are asked if they think all characteristics are actual potential choices. Furthermore, they are asked if all characteristics apply to DaaPs. Then, when all characteristics are discussed, interviewees are asked if they have any additional characteristics for the model that they find are missing. Finally, when all their feedback has been discussed, they are asked to explain how the model fits their platform. Similarly to the question asked for the definition, this will help put their feedback into perspective and will help give examples for the various characteristics in the final version of the model.

Finally, the maturity model is discussed. This happens in a way similar to the DaaP characteristics. Initially, some more explanation is given about this model. Specifically, the differences between focus areas, capabilities and practices are explained and definitions are given for each specific level to make it easier to place the practices at a specific level. Interviewees are shown the model and are asked to give feedback on all its contents. In contrast to the characteristics model, however, here the cells do not represent choices a company can make, but ways a company can mature its capabilities. Therefore, an additional consideration is if the practices of a certain focus area are in the correct sequence and belong to the right level of maturity. Next to direct feedback on the model, interviewees are also asked how specific focus areas or practices are implemented at their organization. This is again done to give some context to the elements of the model and to put the feedback into perspective. After discussing the entire model and asking about any additional topics that it could cover, interviewees are again asked about its application to their platform.

Next to this protocol, an invitation letter is created for potential participants. This letter briefly describes relevant information for potential participants to decide if they want to participate. It explains the research by giving a brief introduction to the concept of DaaP. Then it discusses the interview, what the goals for the interview are, what the required profile is for participants and how the results of the interview will be used. Finally, some details are discussed such as the length of the interview and the period in which it will be done and how the recording and transcript are handled. This letter is sent to every potential interviewee, even if an interview was agreed upon beforehand, to make sure that every interviewee starts with at least a solid level of understanding of the research. The letter is created both in Dutch as in English and the English version is attached in Appendix B.2.

5.2 INCREMENTAL INTERVIEW METHOD

For processing the interview results, an incremental interview method is used. For several reasons, it is considered crucial that the interview results are processed throughout the interview process, instead of after they are completed. First of all, because of the limited availability of DaaPs or related data platforms in practice, a little number of potential interviewees are available. To make optimal use of the participants that can be found, it is important that participants immediately could react to the input given by their predecessors. Furthermore, as the models used as input for the interviews are created from related literature, it is expected that participants have a significant number of changes to the model, which need validation on their own. This point is further reiterated by DiCicco-Bloom and Crabtree [DiCicco-Bloom and Crabtree, 2006], who discuss that data analysis from interviews can best be performed concurrently with the data collection from the interview. Their main argument for this is that it allows investigators to use their emerging understanding to inform their questioning and the sampling of participants. They mention how this ideally leads to a point where interviews have no new information to add to the data, resulting in saturation. This theoretical saturation shows that the data collection has been completed.

For this research, this has resulted in an incremental method for processing the interview results that were concerned with the models. The first rule in this method is that each aspect of the model is to be confirmed by two or more interviewees before it becomes permanent in the final model. This was also the case for changes suggested by interview participants. This means that every change by an interviewee subsequently has to be confirmed by one other interviewee before it can be placed in the final model. Moreover, when at some point, the agreement percentage of a certain aspect of the model drops below 50%, after at least 2 interviewees have considered that aspect, it is removed from the model. This relatively strict benchmark is set to ensure the model would not get too extensive for the interviews and the final result. Furthermore, the model should remain practical for use in practice. Finally, any suggested changes are inserted in the model next to the original situation to signal changes more clearly for future participants. An overview of all methods used for processing feedback on the models is modelled in an activity model in Figure 5.1.

5.3 PARTICIPANTS

Eventually, 8 participants are found willing to participate in the interviews. These participants range over 5 different companies and are interviewed about their experience with 6 platforms. An anonymized overview of the participants and their project is displayed in Table 5.1.

PERSON A

Person A works as a consultant for a major consultancy firm, *Company A*. He currently has the job title Senior Consultant. His expertise lays with digital challengers and, specifically, the technical part of introducing new digital products or companies in the market. Additionally, he is focused on scaled agile and DevOps implementation. Person A indicated to have worked on several projects that involved a platform

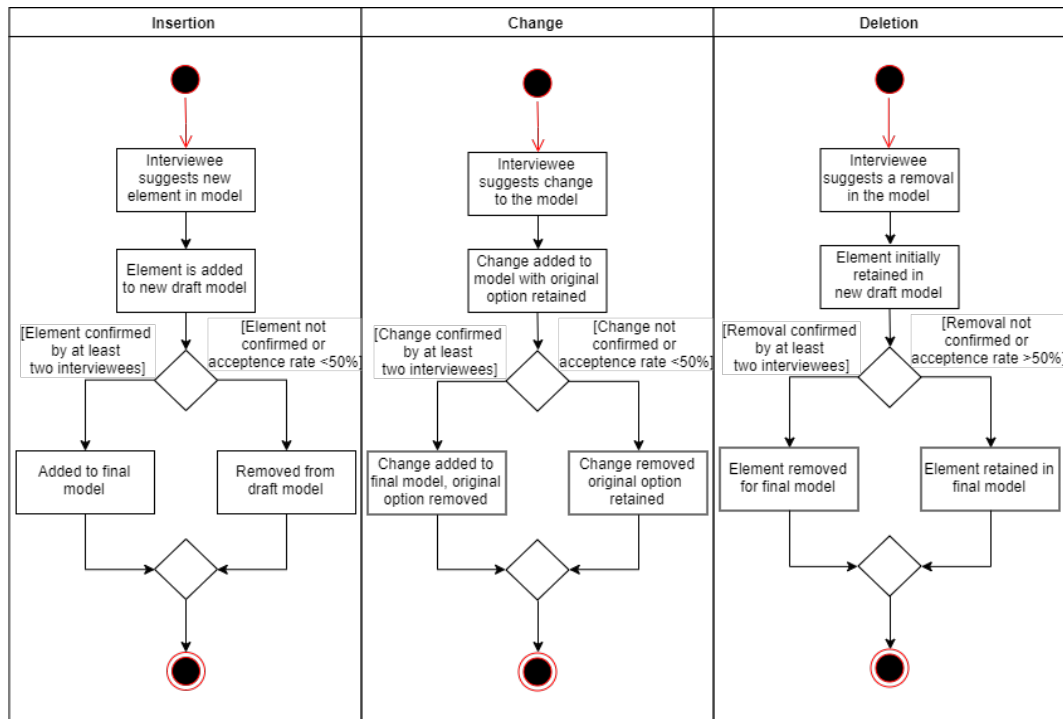


FIGURE 5.1: *The incremental model update method used in this research.*

and a data element, however, that he wouldn't call any of them a Data Platform, as the platform had more functions than merely sharing data. Nevertheless, he considers himself knowledgeable on the topic of data sharing in platforms. Therefore, he can make a valid contribution to this research. For this reason, this person is not interviewed about a specific data platform, but about his general expertise on the topic.

PERSON B

Person B works as a Manager at Company A. In contrast to Person A, however, his expertise on data platforms comes primarily from a single project. In a 5 year-long project, he helped a client company with setting up Platform A. Platform A was an, at the time new, open government data platform through which the client company would disseminate its data. This platform works consists primarily out of two APIs, one for finding the right data and another one for additional information based on the findings of the first API. While Platform A can be considered an open government data platform from a semi-governmental organisation, the data on the platform is not freely accessible. Parties interested in the data have to sign up for a subscription service that gives them access to the APIs. As the expertise from this participant primarily comes from one platform, he will be asked specifically about this platform.

PERSON C

Person C works as VP of Products at Company B. Company B's main product is Platform B and Person C is responsible for a lot of the data collection for the data that goes into the platform. Similar to what was mentioned for Person A, Platform B can't be considered a data platform, as it has more functionalities that fall outside this description. Mainly, it is more centred at reporting analysis on the data, instead of sharing it directly. Nevertheless, it is a platform in which data plays a crucial role and, therefore, relevant for this research. Furthermore, Platform B cannot be considered to be an ecosystem, as they currently do not

| Interview | Person | Company | Platform | Platform Type |
|------------------|----------------|----------------|-------------------|---------------------------------|
| Interview 1 | Person A | Company A | General Expertise | Commercial Platform |
| Interview 2 | Person B | Company A | Platform A | Governmental Open Data Platform |
| Interview 3 | Person C | Company B | Platform B | Commercial Data Platform |
| Interview 4 | Person D and H | Company A | Platform A | Governmental Open Data Platform |
| Interview 5 | Person E | Company C | Platform C | Commercial Data Platform |
| Interview 6 | Person F | Company D | General Expertise | Commercial Platform |
| Interview 7 | Person G | Company E | Platform E | Commercial Data Platform |
| Interview 8 | Person H | Company A | Platform A | Governmental Open Data Platform |

TABLE 5.1: *Overview of the Interview Participants*

have partners that use their platform to create third-party applications. However, they are considering to promote this strategy more actively in the future. In his specific role, this interviewee is also responsible for many of the data sources that enter the platform and, therefore, is familiar with data that is being shared from other organizations. For this reason, this interviewee was considered a valid participant, as he has expertise on both the sharing to and use of data from other organizations.

PERSON D

Person D works as a Senior Manager for Company A. This person was interviewed together with Person H, in the context of their current project for Platform A. Person D's expertise comes from this project with Platform A, the open government data platform described for Person B. During this interview Person D and H applied the interview materials directly for their project for their client company. During this interview, the application of the model and its specific contents on Platform A was discussed in detail.

PERSON E

Person E works as a Product Manager at a major media company. In this role, he is and has been the manager of a commercial data platform for about 4 years, after having been sales manager for a similar platform for 8 years. Platform C can be considered a commercial data platform that fits in the definition of DaaP. Using a subscription service, companies can get access to data on the platform and can use that to create new products or improve current products or services. Their platform is widely used by many companies. As one of the few interviewees that has experience with a data platform that fully falls into the DaaP definition, this participant was asked extensively about how feedback on the model related to his company's activities.

PERSON F

Person F works as an Enterprise Architect for Company D. He has been working at this company for approximately 12 years. Company D is a major retailer that is known to have access to immense amounts of data. Company D, however, cannot be considered a data platform, as a recent new article already gave away that they are not sharing any data with other parties. This was reiterated during the interview by Person F, who discussed how data sharing could hurt their competitive advantage and potentially could run into problems with privacy legislation. Nevertheless, Person F considered that he did have enough expertise on the topic to give feedback on the models presented to him during the interviews. Mostly this stemmed from the considerations that Company D did put into potentially creating a data platform.

PERSON G

Person G is an Enterprise Architect for Company E. He has been working at this company for 15 years. Currently, Person G is involved in a project for creating an architecture for a data platform that aims at unifying the data types available in their industry and making data sharing through that platform more convenient. While this platform's primary purpose is not to directly share data with other parties, this use case is considered as a potential future extension of this platform. The main reason for unifying data types is that it allows software vendors to create universal data analysis software that does not have to be customized for a single company in the industry. For this reason, Person G can give input to the model from a perspective that involves an ecosystem perspective. This data platform enables a new ecosystem, allowing software vendors to enter the market without having to create expansive customizations in collaboration with individual companies. This ecosystem perspective is extensively discussed during this interview.

PERSON H

Finally, Person H was interviewed again after the joint interview with Person D. Person H works as a Senior Consultant for Company A. Moreover, he works on a similar project for Company A as Person D. This repeat interview has taken place to get all the views of Person H on the interview question. During Interview 4, the context was specifically the current project on Platform A. However, during this second interview, the model in its completeness is applied to Platform A, including the aspects that fall outside the scope of the aforementioned project. Most of the feedback from Person H is nevertheless still based on his experience with Platform A.

5.4 EXECUTION

This section describes the execution of each of the interviews. First, the section will describe details about the execution of the interviews and will then discuss trends and what this means for the results of the interview. A full overview of the process for executing the interviews can be found in figure 5.2.

5.4.1 CONDUCTING INTERVIEWS

INTERVIEW 1

Interview 1 was initially set-up as a pilot interview to test the interview materials and to check for aspects such as the length of the interviews. However, due to the valuable input provided by Person A and the limited number of available interviewees, it was later decided, in correspondence with Person A, that the interview transcript would be used for the research. During the interview, some problems with regards to the length of the interview did rise. As was already discussed in the section on the interview protocol, the interviews were planned for an hour, but the material did not fit in an hour. During this interview, an extensive discussion on the definition for DaaP took place, leaving limited time for the rest of the materials. For this reason, merely the characteristics model and the business models were discussed.

After this interview, it was decided that the maturity model was to be given more priority, as the key artefact of this research. Therefore, the question of the characteristics model was moved to the final question of the research, which would only be discussed after all other questions were answered.

INTERVIEW 2, 5 AND 8

Interviews 2, 5 and 8 all followed a similar trend. These interviewees were all actively working on data platforms. During these interviews, the main element discussed was the maturity model. All these interviewees took longer than an hour the time to completely go through the model and gave valuable feedback

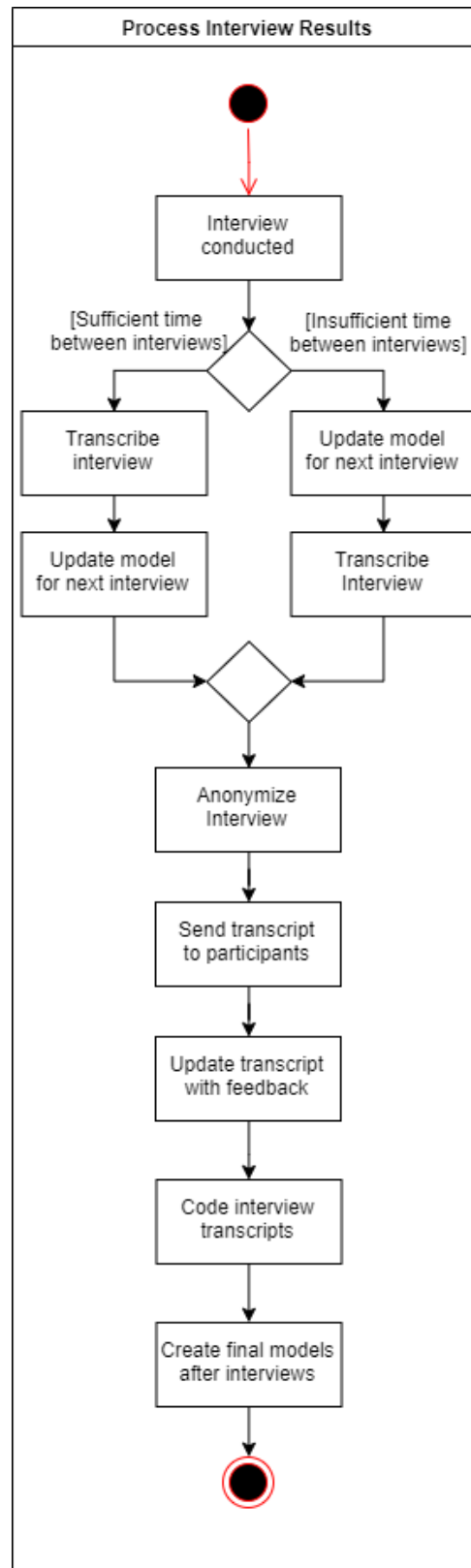


FIGURE 5.2: *The process used for processing the interviews.*

on every aspect of the model. However, this discussion of the model took too much time for a complete discussion of the characteristics model.

INTERVIEW 3 AND 6

Interviews 3 and 6 also followed a similar pattern. These interviews were both with companies that cannot merely be seen as data platforms, as their platforms do more than merely sharing data. During these interviews a longer extensive introduction and discussion on definition took place, leaving little time for the remaining interview. During interview 3 this meant that only the first part of the maturity model was discussed. During interview 6, the entire model was reviewed by the interviewee, however, this interviewee was less confident about giving feedback on the model. This resulted in only a limited amount of feedback from this interviewee.

INTERVIEW 4

Interview 4 was very different from the other interviewees, as it was a less planned and prepared interview. During this interview, the model was tested on an actual data platform part of an existing project for Persons D and H. During this interview only the maturity model was discussed and applied on Platform A, which the project is concerned with. This interview did not only result in feedback on the model itself but also helped with creating a method that practitioners can use when making use of the model.

INTERVIEW 7

Finally, interview 7 was the only interview performed through a call instead of an in-person conversation. During this interview, the definition and business models were extensively discussed. However, when the models were about to be discussed, the interviewee indicated that he would prefer to receive them by mail. Therefore, this interview merely resulted in written feedback on the models, without much opportunity for follow-up questions or extended explanation.

5.4.2 UPDATING INTERVIEW MATERIALS

After each interview, the interview materials and the models were updated based on the feedback from the interviewees. However, due to a limited timeframe for conducting the interviews, it was not always possible to go through the entire transcription process before updating the models. As can be seen in figure 5.2, depending on the time between interviews, the interviews were either first transcribed or the models were immediately updated. In either case, updates were made based on several factors. First of all, each interviewee noted comments on versions of the model, either on paper or digitally, these were used as the first source for quick updates to the model. Secondly, notes were made during the interviews that also served as input for updates of the model. Naturally, when there was enough time for transcribing between interviews, the transcripts were also used as a source for updating the models. The process used for updating the models is already discussed in section 5.3 and figure 5.1.

5.4.3 TRANSCRIPTION AND ANONYMIZATION

Transcribing the interviews happened concurrently with conducting the interviews. The OTranscribe software was used and, as automatic software was found unreliable for the Dutch language in which the interviews were conducted, the interviews were transcribed manually. Throughout the interviews, several interviewees said that the aspects they mentioned were to be kept off the record. These sentences were, therefore, skipped throughout the transcription steps.

Subsequently, the transcriptions of the interviews were anonymized. This process was also done manually. For the anonymization, all words that indicate a person, company or platform were removed.

When useful information was given that gave away the identity of any of these aspects, the text would be replaced by a summary that did not compromise the identity.

Finally, these anonymized transcripts were sent out to the participants for approval. This allowed participants to check their words and see if they formulated everything correctly. Furthermore, it allowed them to look at the anonymization to see if they agreed and that no more proprietary or identifying information was left in the transcript. Only after the transcript was accepted, it was coded for further use in the research.

5.4.4 CODING

After the transcripts are accepted by the interviews, the interviews are coded. The coding of the interview is performed using NVivo with a coding scheme that reflects the goals of the interviews. As the questions asked during the interviews are very concrete, the coding is straightforward. Codes are created to highlight the parts of the introduction. Next to that, different types of suggestions are indicated for the definition, characteristics and maturity model. Additionally, some codes are created for additional remarks that may be interesting while writing, for example, future work. A full overview of all coding nodes used with the times they appeared in the transcripts can be found in figure 5.3

| Name | Files | References |
|---|-------|------------|
| Introduction Company | | 4 |
| Introduction Interviewee | | 3 |
| Method Model Usage | | 1 |
| Maturity Model | | 0 |
| Feedback on Specific Focus Area | | 6 |
| Unclarity in the Maturity Model | | 5 |
| Addition to Maturity Model | | 4 |
| Feedback on Model in general | | 4 |
| Anecdote Focus Area | | 4 |
| Feedback Categorisation | | 2 |
| Feedback on Specific Practice | | 1 |
| Overall Research | | 0 |
| New Ideas for research | | 5 |
| Definition | | 0 |
| Feedback on Definition | | 5 |
| Anecdote for Definition | | 5 |
| Business Models | | 0 |
| Input Potential Business Models | | 5 |
| Characteristics Model | | 0 |
| Unclarity Characteristic | | 1 |
| Feedback Characteristic | | 1 |
| Feedback Model Characteristic | | 1 |
| Feedback Categorisation Characteristics | | 1 |
| Addition Characteristics | | 1 |

FIGURE 5.3: An overview of the codes used for the interview.



CASE STUDY

This chapter discusses the design of the case study performed in this research. Over the course of this research, several artefacts have been created, improved and validated using interviews. The case study is conducted to test the artifacts in practice. Specifically, the goal is to test how the artifacts can be used successfully by companies to assess and improve their data platform. The targeted result is a method for applying the results of this research in practice. The artefacts discussed in the case study are the DaaP characteristics discussed in chapter 8 and the DaaP Maturity Model discussed in chapter 9.

As stated in the paragraph above, the interviews only tested the correctness of the model. Interviewees were asked to give feedback on the model based on their personal expertise. Throughout this process, the model itself is validated by the interviewees. However, this provides an answer to the sub-questions and not to the main question of this research. When reflecting on the main question: "How can DaaP platforms be utilized by organizations to unlock the value of their data?" there is a missing step. Answering this question requires that DaaP platforms can unlock value. To achieve this, it is insufficient to have a model that shows how DaaP platforms could potentially be improved. It requires that these models can be applied in such a way that unlocks value for organizations. A method to achieve this is validated in this part of the research.

6.1 CASE STUDY SET-UP

A case study set-up is created to achieve this goal. Similar to the interview set-up, this consists of a case study protocol and an invitation letter. The protocol contains a plan describing how and why the case study is conducted. Next to that, the questions asked to the participants are specified. Finally, the models are included that are discussed during the case study. Furthermore, the invitation letter consists of a brief explanation of the research and the case study that could be sent out to potential interviewees to ask them to participate.

6.1.1 CASE STUDY GOALS

Three goals have been defined for the case study:

1. Understand how companies want to use the materials

2. Understand the motivations for using such materials
3. Validate that the materials in their current form are useful and understandable

For each of these goals, the case study protocol describes why the goal is relevant for the research and how it can be achieved during the case study. The reason behind the first goal is that understanding how companies are planning on using the materials can give great insights in how the materials can be best framed in the scientific work and outside of that. This helps to determine and validate a method for using the materials that ensures that organizations can use the models effectively and to actually unlock value through the use of DaaP platforms. This will be achieved by discussing the method for using the model with the case study participant both beforehand and after the fact. Moreover, during the assessment with the materials an initial method is used and feedback on this method is retrieved from the participants.

The second goal is to understand the motivations for using the materials, looking at why and when the materials can be used instead of how. This topic is discussed to gather insights on when the model should be used by organizations. For the final method, it enables a description of some context on when the method is best used and what environmental factors should be in place for it to be effective. These motivations are both questioned before and after the materials are used, to get insights in if motivations differ depending on the level of familiarity the participant has with the model.

The third and final goal is to validate that the materials are useful and understandable. This question is asked to support the answering of the main research question through the sub-questions belonging to the materials. Finally, during a final session, this question is discussed with the participants, asking them if they think that the overall assessment has helped their organizations unlock value from their data.

6.1.2 CASE STUDY DESIGN

Next in the protocol, the design of the case study is described. During the case study, one organization is to participate in an assessment of their DaaP. An assessment involves an interview with three people familiar with the platform, preferably in the following division:

- A person familiar with the business and organizational aspects of the DaaP
- A person familiar with the technical implementation of the DaaP
- A person that uses the DaaP, preferably from outside the organization

The goal of this division is to get input on the DaaP from various perspectives. Each participant is asked the exact same questions, however, they are able to give more specific insights on the materials in their area of expertise. The addition of the user perspective can help make the assessment more balanced, as the case study organization's participants may be biased to promote their platform. The user can also provide valuable insights for the Case Study organization of what their users think of the platform. On the other hand, however, a client may not use every functionality on the platform and may, therefore, be able to give a complete view of the platform. Therefore, the insights from the case study organization are key to get a full view of the platform.

During the interviews, all the participants are asked to assess the platform using the characteristics and the Maturity Model. After that, the input from all participants is synthesized in a single overview. This will result in three individual scores, consisting of the assessment score from the interviewees. Moreover, a combined score can be created consisting of the average of the scores from the interviewees. After all scores are determined, a final meeting is conducted with the participants from the case study organization. During this meeting the results are discussed and points of disagreement between the two participants are resolved to get a final overview of the platform. Furthermore, during the session the value of the assessment is discussed with the case study participants.

6.1.3 PROTOCOL AND INVITATION LETTER

For each participant, to start of the case study with the same background knowledge, a protocol and invitation letter are used. The invitation letter consist of a brief description of the research and the case study and ends with a request for an organization to participate. This letter can be found in Appendix C.2. The protocol describes the entire interview that is conducted with the participants. The protocol can be found in Appendix C.1. It consists of three sections.

The first one is the introduction. The introduction covers several aspects.

- A small introduction of the author and the research is given.
- The concept of DaaP is introduced and the definition that was created for it is discussed.
- The purpose and the set-up is discussed, as explained in the previous sections.
- Introductory questions are asked to the interviewee. They are asked to describe themselves, their function, the platform discussed in the Case Study and their involvement in that platform.

In the second part of the protocol is the discussion of the DaaP characteristics. In this part, the characteristics found in this research are discussed using the model that was finally created with characteristics after the interview phase. Questions corresponding to the three goals are, subsequently, discussed with the interviewees. Questions are asked about the method and motivation for using the overview and their view on the performance of the overview. The questions about the method are asked both before and after the overview is used for the assessment. Finally, during the assessment a specific predetermined method is used that is described in the chapter on the characteristics overview.

The third part of the protocol is about the DaaP Maturity Model. This part is similar to the previous part with the characteristics. Again, participants are asked about the method for using the model, the motivation for using the model and, finally, for their opinions on the model. Moreover, the method used for the assessment and for determining the result is discussed in the corresponding chapter on the Maturity Model. Finally, the Case Study session is concluded with an explanation of the next steps in the Case Study.

6.2 PARTICIPANTS

The case study was performed around a single platform. This chapter describes the companies, the participants that were interviewed and the platform discussed during the Case Study. In this chapter all participant and company names are anonymized. An overview of the participants and their companies can be found in table 6.1.

| Platform | Company | Person |
|------------|-----------|----------|
| Platform F | Company F | Person I |
| | | Person J |
| | Company G | Person K |

TABLE 6.1: *An Overview of the Case Study Participants*

PLATFORM

Platform F Platform F can be considered an Open Governmental Data platform. It is owned by a group of four governmental agencies and Company F, as one of the four agencies, is tasked with operating

the platform. For Company F it is one of several data platforms, but there is a specific team focused on this platform. The platform, at the time of writing, consist of 142 datasets. It is mostly concerned with geographical information. The platform can be found at a specific webpage, that has many functionalities, such as a data visualization tool and a showcase section. Furthermore, Platform F is not exclusively made up out of datasets from Company F, but it also distributes datasets that belong to related governmental organizations and at some point in time it even supported the uploading of datasets by users. At the time of writing, the platform the usage of the platform is significant at almost 30 million daily data calls. Most of the data is open and can be used for any use case and it is currently also being used for a wide range of applications.

ORGANIZATIONS

Company F

Company F is the operator of Platform F. It is a semi-governmental organization that executes geographical activities for the government. During the global trend towards Open Governmental Data, Company F started sharing some of the datasets they own. As time went by and the opening up of more datasets was ordered through legislation, Company F matured this capability. For this reason, Company F participants in a governance group with several governmental parties to combine their data opening efforts, resulting in Platform F. Of this group Company F is the main operator of the platform. Company F generates revenue from the other parties, who pay to have their data shared.

Company G Company G is a Dutch subsidiary from a global company specialized in geographical information systems. They sell a product that can be used for many geographical use cases and give training to support the companies that have acquired it. This company, as the responsible party for the Dutch market for the global organization, customizes the product to fit the Dutch market. For this customization, they use the data from Company F to improve their product and fit customer needs. Through this data use, they are one of the main users of the data from Company F.

PARTICIPANTS

Person I

Person I is the first interviewee from company F in the case study. This person is one of the early employees involved in Platform F. Next to that, this person is an expert on the topic of Linked Data, in which this platform has increased its efforts. After the initial development phases of the platform, this person took a step back as a part of the main development team of the platform and is currently more focused on introducing new functionalities and services in the platform. Person I was considered a fitting interviewee, as he was an employee at Company F that was very familiar with the business aspect of Platform, while he is also knowledgeable on the technical part.

Person J

Person J is the second interviewee in the case study from Company F. This person joined the development of the platform later than Person I. This person currently works in the role as product manager in the development team. This person can be considered a technical person and has more detailed knowledge about the platform than Person I. Person J was considered a valid interviewee, as he, as Products Owner of Platform F, is also very knowledgeable on both the business as the technical aspects of Platform.

Person K

Person K is the only interviewee in the case study outside Company F. This person is Product Owner and consultant and Company G. In this role, he is involved in developing new features for the product of Company G in the Dutch market. For these features, Company G also uses the data from Platform F. Person K was, because of his knowledge as an outsider, interviewed as a user of the data from Platform F and

was asked to give his perspective on the maturity of this platform. Finally, Person K was also considered a valuable interviewee, as he was very familiar with the platform, while having an outsider perspective.

6.3 EXECUTION

This section describes the execution of each of the interview. A general description is given for each of the interviews. Next to that, the process used for the entire Case Study with a description of how the results are generated and some specific feedback on the Case Study design by the interviewees is discussed.

Case Study Interview 1

The first case study interview is conducted with Person I. This interview is the longest of the three at more than 90 minutes. At the start of this interview, Person I gave a lot of context about Company F and Platform F. The main reason for this was that, as first interviewee in the case study and employee from Company F, this person had to give a lot of information on all the platform to give a more concrete view of the exact platform discussed in the case study. Especially as Company F has more platforms than Platform F that are all somewhat connected, a clear boundary had to be set for the case study. The decision to focus on Platform F specifically had two main arguments. The first was that this was the platform where most information from other parties was shared, which would give interesting insights in the ecosystem perspective of the case study. Secondly, Person J, the second interviewee in the case study, works specifically on this platform and, while he would still be able to give input on the other platforms, he would be most knowledgeable on this platform.

After a long introduction about Person I, Company F and Platform F, the assessment of the platform started. As described, this assessment involved both the characteristics and the maturity model created in this research. The results of these assessments are discussed in their respective chapters. In this section, some details about the interview are noted. One of the key discussions during this assessment was the platform that was being discussed. Details were given on the various platforms that Company F offers. It was only mid-way through the assessment that specifically Platform F was picked for the case study. While not relevant for the assessment, the discussion of the various platforms did give some interesting insights in Company F and DaaS in general. First of all, the business model of Platform F differed from that of several other platforms. Company F also offers a platform, that is more situated around a website and less around an actual platform, where data can be bought instead of shared for free. This brought some specific challenges, as the functionalities of this service were limited and some customers had very specific requests. Next to this, they describe how there were some conflicting incentives in the platform. The platform is paid for by the government with a fixed fee. While the employees of Company F are motivated to make the platform work as good as possible and to increase its usage, they are disincentivized by the additional cost that a more heavily visited platform brings. The cost of maintaining and running the platform increases as more users make use of it, but this does not influence their income. These conflicting incentives sometimes hurt the development of the platform.

Case Study Interview 2

The second case study interview is with Person J. As discussed, Person J is more actively involved with the current development of Platform F, as a Product Owner. For this reason, this interviewee provided more context on Platform F itself. First of all, this interviewee gave more details about the collaborations with the various data providers that are a part Platform F. He describes what data they provide to the platform, which is from a wide range of use cases, but usually gathered by people 'working in the field.' Next to that, this interviewee provided insights in the decision making processes and how this is mostly a collaboration by the four main data providers of the platform. In this context all four organizations make many of the major decisions, while the execution is done by Company F, who are also responsible for all operational activities. Next to that, this interviewee provides more insights in the latest developments of the platform, giving a more up to date assessment of the platform than Person I. For several practices, interviewee I indicated already that Person J would be able to tell more about the topic and this is exactly

what happened during the interview. In this way, Person I and J complemented each other quite well and there results had little differences.

Case Study Interview 3

Interviewee is the only external interviewee that was interviewed for Platform F. This interviewee looks at the platform from an outsider's perspective, but states that he is very familiar with the platform and also knows Person J personally. For this reason, this was a fitting interviewee, as he could provide the nuance of an outsider perspective, while still being knowledgeable and correct about the platform.

For a couple of years already, this interviewee is a part of a team coupling different data services to their product. As a part of this team, this interviewee has worked to establish Company G's product with data being shared through Platform F. This is mostly done either to meet a direct customer need that is requested or to show a new use case potentially interesting for current or future customers. During the assessment, this interviewee indicated many of the similar things as the other interviewees, while being a bit more strict. When some aspect were partially implemented, this interviewee considered them not implemented, as he found these things were regularly still insufficient for their use cases. Finally, this interviewee discussed how this resulted in an ecosystem around their products, where they allow partners to distribute their data through their product. These products get a compensation for this in several ways.

Joint Case Study Result Discussion Finally, the case study results are discussed with the employees from Company F. After case study interview 3, the results are calculated and a final report is created to present to Company F. This final report consists of an overview of the selected characteristics, the selected maturity levels and some questions about differences between the participants. This discussion does not have a strict protocol, but is an open forum for the participants their final pieces of feedback on the results and models.

For the results of the assessments, they discussed some misunderstandings that may have caused some of the differences in the results. First of all, Person I noted that his feedback was not always solely focused on Platform F, but that he also gave some feedback based on other platforms from Company F. This is discussed during the original interview and is considered in the results, but, through some misunderstandings, some results wrongfully made their way into the assessment of Platform F. Furthermore, they note how the platform mostly disseminates the data and, doing this barely performs any data quality checks, as this is the responsibility of the data suppliers. Therefore, even though the model may indicate that the platform may be somewhat immature in this aspect, most of these practices are performed, but earlier in the data supply chain.

For the maturity model, the participants have a few points of feedback. The main one, is that the terminology is sometimes quite challenging. This goes for both the layout of the model, as well as its contents. For the layout, the interviewees note that maturity may not be the correct word for all elements in the model. They argue that some aspects of the maturity model should be seen as decisions and choices you can make that do not necessarily have a maturity scale connected to them. Next to that, they say that additional context is needed in the model itself. The descriptions of the practices is sometimes too brief to apply it logically to a platform.

Finally, they state that they expect that they can use the results for further development of their platform. Especially some of the differences with Person K can be seen as elements where the platform can be improved on. For the materials themselves, they note how the materials can be useful when starting out, to become familiar with the options. Furthermore, they discuss how organizations should use it later on to reassess their decisions and to create an improvement plan for further developments of the platform.



DEFINITION

This chapter describes the results found for the first sub-question in this research. This sub-question is: "What defines a Data as a Platform ecosystem?" This question is answered in two steps. Firstly, input for an initial definition is gathered through the SLR in the form of definitions for related fields. These definitions are then used to form an initial definition for a DaaP ecosystem. Subsequently, this definition is discussed during the interviews to get feedback from practitioners. Based on this feedback, the definition is updated and a final version is created to answer the research question.

7.1 LITERATURE REVIEW

The first step looks for a definition of what constitutes a DaaP ecosystem in the literature. As no direct definition is currently available in literature, a definition can be created based on definitions of related topics.

On the topic of ecosystems, one of the seminal works that initially popularized the concept of biological ecosystems as a metaphor for business partnerships is that of Iansiti and Levien [Iansiti and Levien, 2004]. They discuss the analogy between biological ecosystems and business ecosystems in detail and stretch their similarities. For example, they argue how the health of an overall ecosystem has a large impact on their participants, in both types of ecosystems, a healthy ecosystem will allow its participants to thrive, while an unhealthy ecosystem makes a large part, if not all, participants suffer. Even though they do not give a specific definition, their work is considered in this section as it forms the basis for many of ecosystem related definitions.

Further definitions of two types of ecosystems are given by Jansen [Jansen et al., 2009] and Lee, Zhu and Jeffery [Lee et al., 2019] on the topic of Software and Platform Ecosystems respectively. Jansen defines software ecosystems as "set of businesses functioning as a unit and interacting with a shared market for software and services, together with the relationships among them. These relationships are frequently underpinned by a common technological platform or market and operate through the exchange of information, resources and artefacts". Even though this definition does not include a data element, it can be said that DaaP ecosystems fall in this definition. The 'common technological platform' in the definition could be a DaaP platform. However, the only exception to this categorization of DaaP as a software ecosystem is that the shared market for DaaP ecosystems is larger than merely a market for software and

services, as more traditional companies could still use the data to improve non-software products. A definition for platform ecosystems was introduced by Evans et al. [Evans et al., 2012] and stated "Platform Ecosystems are platforms which constitute two or more sided networks transacting with each other. They allow interactions between multiple groups by providing a meeting place." This definition can already include DaaP ecosystems, as it does not give any specifics on what the 'platform' or 'meeting place' are. However, the definition is quite broad, making it not directly applicable for a DaaP ecosystem definition.

Two topics discussed in the data management field approach the concept of DaaP ecosystems from a data perspective. The first is the topic of Data Collaboratives. Data Collaboratives have risen as a part of the Big Data movement. Bertot and Choi [Bertot and Choi, 2013] describe that Big Data requires a large investment in infrastructure and for gathering the data and they describe Data Collaboratives as a solution to this problem. They define Data Collaboratives as "arrangements between three or more organizations that jointly establish data protocols, data exchanges and reporting mechanisms." This definition, however, is not undisputed, as many other works on this topic have a more philanthropic perspective. For example, Susha, Jansen and Verhulst [Susha et al., 2017b] define Data Collaboratives as "a form of partnership in which a variety of parties, such as government, operators, companies and others, collaborate to exchange and integrate data to help solve public problems or create public value." Both these definitions, while approaching the concept of DaaP ecosystems fall short of defining the concept. First of all, DaaP ecosystems focus exclusively on data shared through a digital platform. While both definitions do not exclude this, they leave room for many other data collaborations between companies. For example, data sharing from company to company within supply chains, is not excluded by the definition, but this type cannot be seen as a DaaP ecosystem, as this does not involve a platform. Moreover, the philanthropic specification in the second definition is not relevant for DaaP ecosystems, as they are framed as a commercially viable endeavour with a working business model.

Finally, the research field of Open Data bears the most resemblance to the topic of DaaP ecosystems. This field, which originally was focused on governments opening up their data sources for the public to use [Zuiderwijk et al., 2015] has recently shifted focus to companies opening up their data sources in a similar fashion [Sennaïke et al., 2017]. Ojo, Porwol, Wagar et al. [Ojo et al., 2016] define Open Data Platforms as "a software platform which comprises a software ecosystem that supports different end-user interactions with open data including search and discovery of datasets, publishing of datasets, analysis and visualization of datasets as well as sharing and development of stories from datasets." This definition is extended by Sennaïke, Waqar, Osagie et al. [Sennaïke et al., 2017] by stretching their approximation to Open Governmental Data stating that "A key purpose of open data platforms is to promote access to government data and encourage development of creative tools and applications to engage and serve the wider community." This definition closely approaches the concept of DaaP ecosystems, as it stretches the opening up of datasets, intending to create an ecosystem that allows third parties to interact with the data. However, its direct comparison to software platforms and ecosystems and its broad description of 'end-user interactions' makes it stand out from DaaP ecosystems. In contrast, a DaaP ecosystem consists of a data platform without any additional software based products and services, differentiating itself from the software platform part of the definition. Moreover, it has a commercial focus that is lacking in the Open Data Platform definitions, which are, as stated, often used in the context of Open Governmental Data.

These findings show that a definition for DaaP ecosystems is not yet available in literature. While some definitions approach the core of the concept, most definitions are either too broad, missing core aspects or are focused on a different perspective. For example, several definitions are focused on public sector data, lacking a business model, or on philanthropic use cases. An overview of all definitions can be found in Table 7.1. This research will, therefore, use the definitions discussed in this literature review to form a new and specific definition on the concept of DaaP ecosystems.

| Topic | Definition | Source |
|------------------------------------|--|-----------------------------------|
| Software Ecosystems | Set of businesses functioning as a unit and interacting with a shared market for software and services, together with the relationships among them. These relationships are frequently underpinned by a common technological platform or market and operate through the exchange of information, resources and artefacts | [Jansen et al., 2009] |
| Platform Ecosystem | A platform which constitutes two or more sided networks transacting with each other. It allows interactions between multiple groups by providing a meeting place. | [Evans et al., 2012] |
| Data Collaboratives | Cross-sector (and public-private) collaboration initiatives aimed at data collection, sharing, or processing for the purpose of addressing a societal challenge. | [Susha and Janssen, 2017] |
| Data Collaboratives | A new form of collaboration, beyond the classic public-private partnership model, in which participants from different sectors – in particular companies – exchange their data to create public value. | [Verhulst and Sangkoya, 2015] |
| Data Collaborations | Arrangements between three or more organizations that jointly establish data protocols, data exchange and reporting mechanisms. | [Bertot and Choi, 2013] |
| Inter-Organizational Collaboration | Constellations of three or more autonomous organizations that collaborate to pursue collective rather than individual goals. | [Provan and Kenis, 2008] |
| Open Banking | Open banking can be defined as a collaborative model in which banking data is shared through APIs between two or more unaffiliated parties to deliver enhanced capabilities to the marketplace. | [Brodsky and Oakes, 2017] |
| Business Open Data | Data produced, collected and stored by the private organizations and made entirely or partly available to its ecosystem in a machine readable format. | [Scott, 2017] |
| Open Data | Data, which is legally accessible through the Internet in a machine-readable format. It does not have to be completely free of charge or restrictive licenses, but experimenting with the data, and running a small scale-business should be legal. | [Lindman et al., 2014] |
| Open Data | Data which is: freely accessible online, available without technical restrictions to re-use, and provided under open access license that allows the data to be re-used without limitations, including across different fields of endeavor (e.g. commercial and noncommercial alike). | [Jetzek et al., 2014] |
| Open Data Platforms | Technological infrastructure comprising of a software ecosystem that supports different end-user interactions with open data including search, discovery of related datasets, publishing, metadata management, sharing, analysis and visualization. | [Ojo and Millard, 2017] |
| Digital Platform | Digital platforms can be defined as purely technical artefacts where the platform is an extensible codebase, and the ecosystem comprises third-party modules complementing this codebase. | [Tiwana et al., 2010] |
| Digital Platform | A sociotechnical assemblage encompassing the technical elements (of software and hardware) and associated organisational processes and standards. | [Tilson et al., 2012] |
| Digital Platform | Software-based external platforms consisting of the extensible codebase of a software-based system that provides core functionality shared by the modules that interoperate with it and the interfaces through which they interoperate. | [Ghazawneh and Henfridsson, 2015] |

TABLE 7.1: *Data as a Platform Related Definitions*

7.2 INITIAL DEFINITION

Using these definitions an initial definition for a DaaP ecosystem is created. This definition is:

”A Data as a Platform ecosystem comprises a set of organizations utilizing a central data platform, a technological platform, which primary purpose is to share data for, potentially commercial, use by third parties. This central data platform is supported by an autonomous business model.”

The rest of this section describes the reasoning behind the different parts of the definition and the related definitions that were used as guidance for these parts.

The definition is based on several definitions from related fields. First of all, it uses a sentence build up that resembles that of the software ecosystem definition by Jansen et al. [Jansen et al., 2009]. At this point in the research, the ecosystem perspective is deemed crucial for the DaaP definition. Therefore, the ”set of businesses” part from the software ecosystem definition is used as a basis for the DaaP definition. Moreover, for similar reasons a continuation describing the commonality within this ”set of businesses” is chosen. Next to that, a second specification of the definition was given in a second sentence, also mirroring the definition by Jansen et al. [Jansen et al., 2009]. The main reason for this choice is the readability of the definition, as a prolonged sentence could confuse readers. The second definition used is that of Jetzek et al. [Jezek et al., 2014]. This definition for open data specifically describes that the data can be re-used for any purpose. The definition created from the SLR contains a similar descriptions with the part ”to share data for, potentially commercial, use by third parties.” This specification is added to explain what a data platform does and to specify the commercial element of DaaPs. Additionally, the definition has some overlap with the definition Ojo and Millard [Ojo and Millard, 2017]. Their definition focuses on the technological infrastructure as a key aspect of an open data platform. Our definition has a similar part specifying that a central data platform is a ”technological platform” with certain functionalities. This specification was made to describe that a data platform needs to have some central technological foundation that needs to be operated. Finally, the final part of the definition ”This central data platform is supported by an autonomous business model” was added to clarify what the commercial aspect of a DaaP entails. Especially the word autonomous is included to emphasize that the platform should have measurable benefits for the operating company.

7.3 INTERVIEWS

The first research result discussed during the interviews is the definition used for the research. An initial definition was created based on definitions of data platforms in related fields. However, to get additional input on this definition, it was discussed with the experts. Specifically, interviewees were first asked for general feedback on the definition, if they understood everything. Subsequently, they were asked if it captured all the key elements of what they understood to be a DaaP and if all elements discussed were actually crucial. Finally, interviewees were asked to, where possible, apply the definition to their data platform and see if it fit the description.

Of the eight interviewees, only five gave substantial feedback on the definition. During interview 4, the definition was not discussed, as that interview was merely focused at applying the maturity model on the current project. Moreover, at interview 3 and 6, the definition was used for an increased understanding of the topic and no real feedback was provided by the interviewees. A summary of all the feedback given by the interviewees can be found in table 7.2.

During interview 1, the definition was discussed extensively. First of all, this person was slightly confused about the commercial aspect of DaaP being mentioned twice in the definition. Firstly in the ”potentially commercial use” of the data and secondly in the ”autonomous business model” part. According to this interviewee the fact that the data was used in a ”potentially commercial” way was unnecessary, as

it was straightforward for a data platform with a business model that the data would be used for commercial purposes. Next to this, Person A commented on the ecosystem perspective taken in the definition, instead of focusing merely on the platform itself. First of all, he describes how the switch from ecosystem to platform perspective between the two sentences is rather confusing. Moreover, he questions the need to define the ecosystem instead of the platform itself, as the ecosystem is rather self-explanatory when the platform is clearly defined. Additionally, this interviewee noted that it could be made explicit if data referred to raw data or that data could be manipulated. However, as discussed later in the interview, data in a DaaP does not necessarily have to be either of these, the only requirement is that it is still data in a dataset and not some reporting on this data. Finally, the interviewee noted that it was not specified in the definition if a DaaP should have a specific use case for the data or not. However, for this remark too, it was later on discussed in the interview that both options were possible and that it was, therefore, not relevant to mention in the definition. It should be noted that after this interview the suggested removal of the phrase "potentially commercial", pointing at the use of the data in the platform, was applied, because the author agreed fully with the reasoning of the interviewee. The definition used for the remainder of the interviews was changed to:

"A Data as a Platform ecosystem comprises a set of organizations utilizing a central data platform, a technological platform, which primary purpose is to share data for use by third parties. This central data platform is supported by an autonomous business model."

During interview 2, Person B had some specific feedback on the definition. First of all, this interviewee noted that such a data platform is made up of many aspects, such as documentation, security and pricing etc. However, this interviewee agreed that it would be hard to create a complete list of aspects that would fit a definition. Moreover, this interviewee noted that it would be better to say that the "primary purpose" of a DaaP would be "monetizing data" instead of "sharing data". Noting that: "Sharing is caring I always say, however, for that we have Open Data Platforms. DaaPs are there to earn money, so the purpose is not to share, but to make money." Moreover, he questions if a DaaP does not always have a specific target audience, as in his experience they do. If this is indeed the case, then this characteristic should be included in the definition. Finally, he notes that explaining that the "central data platform" includes a "technological platform" is rather self-explanatory.

During interview 5, the interviewee was somewhat confused if the customers of the platform were part of the "DaaP ecosystems comprising of a set of organizations". However, after some discussion on this question, the interviewee agreed with this phrasing, advising to support the definition with an illustration to make it more clear.

During interview 7 some new feedback was given. First of all, in contrast to the feedback during interview 2, this interviewee argued that there was a major difference between a data platform and a technological platform. He argued that the technological platform referred to the storing of the data, while the data platform related more to the infrastructure that allowed access to the data from distant locations. Furthermore, this interviewee argued that a business model could not be autonomous. In his opinion, the fact that the business model had to be separated from any other business model in the organization was already apparent without the use of the word "autonomous".

Finally, the interviewee from interview 8 argued that the word "ecosystem" was incorrect. He argued that data marketplaces, platforms that directly sell their data to users, are not ecosystems, while they should be considered DaaPs. He suggests removing the word ecosystem and leaving the rest of the definition as is.

The feedback from the interviewees is summarized in table 7.2. This overview shows not only the feedback, but also the number of interviewees that had similar feedback on certain parts of the definition. This was the final step in which feedback was retrieved on the definition and this feedback is used to create a final version of the definition.

| Feedback | Reasoning | Interview |
|---|--|----------------------|
| Remove "potentially commercial" | Confusing combined with "autonomous business model" | Interview 1 |
| Remove ecosystem perspective | Confusing switch between ecosystem and platform perspective and not all DaaPs are ecosystems | Interview 1, 2 and 8 |
| Add if DaaPs use raw or manipulated data | Be more specific in the type of data can be used in a DaaP | Interview 1 |
| Add if DaaPs are user and use case specific | Unclear if DaaPs have a specific focus on a single use case or audience | Interview 1 and 2 |
| Change "sharing" to "monetizing" as "primary purpose" | Differentiate more from Open Data Platforms | Interview 2 |
| Remove the technological platform specification | Superfluous to specify after stating it is a data platform | Interview 2 |
| Add illustration to clarify definition | Confusing phrase "comprises a set of businesses" | Interview 5 |
| Remove the technological platform specification | Confusing as it means something else than a data platform | Interview 7 |

TABLE 7.2: *Feedback on the Definition*

7.4 FINAL DEFINITION

This feedback has led to several final changes for the final definition. As already discussed, after the first interview the "potentially commercial" part was removed from the definition. This resulted in the following definition:

"A Data as a Platform ecosystem comprises a set of organizations utilizing a central data platform, a technological platform, which primary purpose is to share data for use by third parties. This central data platform is supported by an autonomous business model."

The second point of feedback in table 7.2 was regarding the ecosystem perspective taken in the definition. This feedback was independently given by three interviewees. This feedback goes directly against the sub-question for this research, which specifies the creation of a definition for a DaaP ecosystem. Nevertheless, the arguments made by the interviewees, that not all DaaPs are necessarily ecosystems, requires a change of perspective in the definition. "Data as a Platform is a data platform, a technological platform, on which one or more organizations centralize their data for use by third parties. This central data platform is supported by an autonomous business model."

The third and the fourth point of feedback are not taken over in the definition. Both suggestions are concerned with characteristics that the interviewees deem defining for DaaPs. However, as discussed in the characteristics model in the next chapter, for both these categories there is not a single characteristic that specifically applies to DaaPs. To be more exact, for the first suggestion, DaaPs share both raw and manipulated data and, for the second suggestion, DaaPs can have a focus on a single use case or be use case agnostic. Therefore, the definition should not specify which of these is relevant for DaaPs.

The fifth point of feedback is not directly translatable to the newest version of the definition, as the word sharing is removed from this version. Nevertheless, the argument that it is key to specify that companies share their data with the goal of earning money is still applicable to this definition. Therefore, the word "centralize" can still be updated to monetize. This clarifies that fact that the "one or more organiza-

tions” that make their data available do this with the goal of earning money from their data. Moreover, the word centralize may give the idea that a DaaP has to be operated by a third party, this problem is also resolved with this update. Moreover, the word sharing is added between data and platform, to explicitly state that data is shared by the platform, as the change to the word monetize may suppress this aspect of the definition. The new definition becomes:

“Data as a Platform is a data sharing platform, a technological platform, on which one or more organizations monetize their data for use by third parties. This central data platform is supported by an autonomous business model.”

Suggestions six and eight suggest the same change to the definition for completely opposite reasons. The suggestion from interview 2 is supported by the argument that it is obvious that a data sharing platform would consist of a technological platform, making the specification superfluous. On the other hand, the suggestion from interview 7 is based on the argument that there is a significant difference between a technological platform and a data sharing platform and that it should, therefore, not be used a specification without any explanation. Despite this contradiction, the suggestion is taken over in the definition, removing the “technological platform” part. This is done mostly in agreement with the comment from interview 2, that it is an unnecessary specification given the numerous usages of platforms in a technological context. This has led to the final definition:

“Data as a Platform is a data sharing platform on which one or more organizations monetize their data for use by third parties. This central data platform is supported by an autonomous business model.”

This definition is further visualized in Figure 7.1

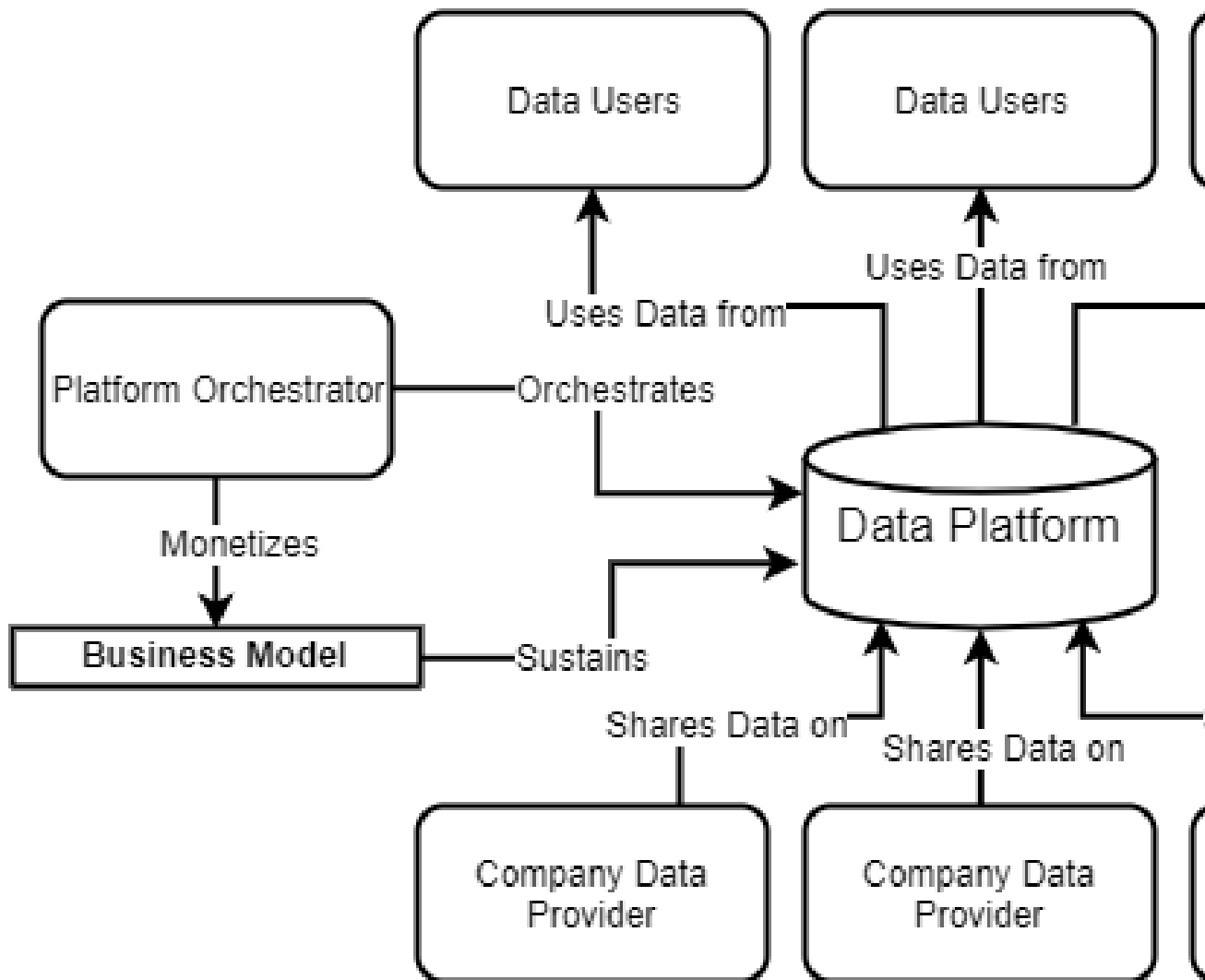


FIGURE 7.1: A visualization of the final definition



CHARACTERISTICS

The second sub-question in this research discusses the characteristics of a DaaP. The goal of this question is the design of a model that illustrates these characteristics. To create this model, input on characteristics of DaaPs or of platforms in related fields is gathered from the literature. Using these characteristics, an initial model is created. This model is, subsequently, tested with experts in eight interviews. During these interviews, the method is updated and validated by the opinions of the interviewees. After the final model is validated by the interviewees, the model is used in a case study, where it is applied to assess an existing DaaP. Based on this case study, a method is described that can be used when an assessment is performed by practitioners.

8.1 LITERATURE REVIEW

Similar to the result of the first sub-question on the definition, the characteristics identified in the literature are scarce. Therefore, the strategy for this research question is to identify the characteristics of the related concepts introduced in the first section of this chapter. Subsequently, some characteristics that do not fit the description are excluded and the remaining characteristics will be validated through expert interviews. Therefore, any identifiable aspect of DaaP ecosystems in literature is noted in this review.

8.1.1 CHARACTERISTICS

The first concept for which characteristics can be found is that of Business Ecosystems, as discussed by Iansiti and Levien [Iansiti and Levien, 2004]. In their expansion on the comparison between business and biological ecosystems, they identify several characteristics that both have in common. First of all, both are characterized by a large number of loosely interconnected participants. Moreover, the healthiness of the overall ecosystem has a significant impact on its participants. When the ecosystem is healthy, its participants can thrive and, in the case of companies, grow, while in an unhealthy ecosystem most if not all participants suffer. Next to that, they name productivity, an ecosystem's ability to transform external input in productive output for the ecosystem, robustness, the ability to adapt to environmental changes and niche creation, the ability to exhibit variety as important indicators for the health of an ecosystem. Finally, they introduce several roles that companies can take in an ecosystem. These roles are keystone, dominator and niche player. The keystone, also known as the orchestrator [Jansen et al., 2009], is the

controller of the ecosystem, they can develop strategies to keep the ecosystem healthy. The dominator aims to control a large portion of the ecosystem directly. They are aggressors and aim to capture as much value as possible themselves in the ecosystem. Finally, the niche player, as the name already suggests, aims to develop a specialized capability that they can leverage to add to and capture some of the value of the ecosystem. These characteristics form the foundation of the ecosystem concept and their applicability for DaaP ecosystems will be tested in the expert interviews.

The second concept discussed is the topic of software ecosystems, as defined by Jansen et al. [Jansen et al., 2009]. First of all, Jansen [Jansen, 2014] more specifically discusses health measures in software ecosystems, specifying metrics for both the project and the network level on the three categories identified by Iansiti and Levien [Iansiti and Levien, 2004]. Mijsters et al. [Mijsters et al., 2018] extend this framework with a sub-ecosystem perspective for ecosystems that rise inside larger ecosystems. While this topic is interesting and relevant also for DaaP ecosystems, these specifications are not used in this research, as ecosystem health assessment is outside of the scope of this research. Furthermore, research into the governance of software ecosystems has introduced several characteristics. The software ecosystem governance analysis framework, as introduced by Baars and Jansen [Baars and Jansen, 2012], names i.e. partnerships, licensing and knowledge sharing as important aspects of the governance of ecosystems. Another characteristic researched in this context is that of the openness of the ecosystem. The openness of an ecosystem determines how freely information streams are accessible and how much control is enforced by the ecosystem orchestrator [Knauss et al., 2014]. These aspects, especially the higher-level characteristics, are relevant, as they describe key dynamics of an ecosystem. Therefore, these will be discussed with experts in the interviews. Finally, data sharing is mentioned more as an important characteristic in ecosystems [Veenstra, 2015]. Various methods for data sharing are discussed that characterize different types of ecosystems. These are also discussed with experts.

A concept not discussed previously is that of IoT ecosystems, which are ecosystems sharing data generated through IoT devices. In this field, Fisher and Davies [Fisher and Davies, 2016] name security, trust, rights, performance and compliance as key concerns that characterize IoT ecosystems. They also note that there are five key roles in these ecosystems. Namely, Information Providers, Application Developers, Analytics Service Providers, Platform Providers and End Users, while noting that, for example, the role of Platform Provider and Information Provider can be performed by the same company or person. Finally, the name data ownership, information as an asset, sensor connectivity, security, scalability and interoperability as key aspects for IoT ecosystems. Apart from sensor connectivity, all these characteristics appear to be relevant for DaaP ecosystems, and therefore, they will be discussed in the interviews. Furthermore, Mattison [Mattison,] names exponential growth, synergy, data liquidity, person-centricity and accordion model for learning as key principles for this ecosystem type. Finally, Ikävalko, Turkama, and Smedlund [Ikävalko et al., 2018] discuss that IoT ecosystems consist of four layers: the device level, the network layer, the service layer and the content layer. Moreover, they find that IoT ecosystems heavily rely on outsourcing, which has led to a lot of value co-creation. Finally, they identify three role archetypes, the ideator, who merely adds knowledge to the ecosystem and gives input for service innovations, the designer, exchange knowledge with the ecosystem to innovate their and develop new services and, the intermediary act as orchestrators facilitating processes. All these characteristics appear to be more related to the specific IoT use case, however, interviewees will discuss their applicability for DaaP ecosystems.

Additionally, Data Collaboratives have been researched to a significant extent. Klievink, van der Voort and Veeneman [Klievink et al., 2018] note that Data Collaboratives are related to the goal-directed collection, sharing and processing of data, that can differ in type, form, specificity and structuredness. Moreover, trust was found to be one of the key considerations that determine a partner's commitment to the collaborative and the effect of the overall rules governing the collaborative was also found to impact the actors. However, these characteristics seem to be more applicable to the collaborative aspect. Susha, Janssen and Verhulst [Susha et al., 2017b] have defined a taxonomy that consists of 14 characteristics, divided into 2 groups, that can be used to classify various types of data collaboratives. In the data sharing and supply group, there is: type of data, content of data, administrative level associated

with data, diversity of data providers, facilitation and degree of access to data. Moreover, on the data use and demand ground there is: target user group, user selection, research or policy problem, incentive to use data, continuity of collaboration, expected outcome of data collaborative, collaboration among data users and purpose of data use. These characteristics can be used to further define the characteristics of DaaP ecosystems. In a later work Susha, Janssen and Verhulst [Susha et al., 2017a] specifically look into the characteristics that define a "Bazaar" type Data Collaborative, something that will be discussed in a later section. They name coordination mechanisms, incentives intensity and control intensity as characteristics that differentiate the "Bazaar" data collaborative type from other types. Susha, Gronlund and van Tulder discuss the philanthropic angle of data collaboratives. They identify security, privacy, timeliness, flexibility, reliability and the ability to combine with other data sources as key characteristics for successful data partnerships.

Gunther, Mehrizi, Huysman and Feldberg [Günther et al., 2017] discuss value creation through the use of Big Data. Through a review of the current literature, they define how governance of information can either be centralized, which eases data governance adoption, or decentralized, which makes governance more difficult. Moreover, they conclude that, as organizations can be reluctant to fully open up their data, organizations tend to control access rights, this can either be done through market mechanisms, such as contracts, or through hierarchies, where one large distributor decides on access policies for smaller parties, this often occurs in supply chain dynamics. They suggest that it may even be possible to create hybrid networks that allow access to both open and controlled data modes. Finally, they name portability, the ability to share data for the use in different context and inter-connectivity the possibility to synthesize data from various sources as key enablers of shared big data and, therefore, important characteristics to develop in such an environment.

Finally, on the topic of Open Data, Buda, Ubacht, Janssen and Sips [Buda et al., 2016] discuss characteristics of Open Data ecosystems. First of all, this work describes several actors. These actors are the private organizations sharing their data, other businesses, that consist of partners, complementors, suppliers, competitors and non-partners, public institutions, according to the authors they play a key role as they used to be on the forefront of open data and, finally, citizens, media, NGOs and academia, who can use the data for their benefit. Furthermore, they describe several business drivers of the operating company that can motivate the open data initiative. These motivations are gaining new revenue streams, community building, internal business improvement and publicity and PR.

8.1.2 TYPES

Next to the characteristics that are directly described in the literature. Several academic works differentiate between different types of data platforms, mentioning key characteristics in the process of describing the difference between the types. These types, like other aspects researched in this literature review, are not defined directly in literature, but different perspectives on creating a typology in related literature fields are discussed, with examples of the types defined for the different fields.

For this topic, the first concept discussed is that of platform ecosystems. Lee, Zhu and Jeffery discuss how different types can be determined based on the purpose of the platform ecosystem. This depends, according to them, on three aspects, namely, platform types, for example, whether it is a content portal or a social network, the business purpose, commercial or non-commercial and the platform strategy. Moreover, Evans [Evans et al., 2011] discuss how different types of market structures can categorize three platform types: coincident, intersecting and monopoly platforms, which are based on competition types. Next to that, Sun, Lou, Li and Wang [?] differentiate between association-oriented and crowdsourcing-based platforms in the context of data collection for health care use cases. Finally, in the context of the data-sharing economy, which stretches the importance of data in the current economic environment, Richter and Slowinski [Richter and Slowinski, 2019] define different modes of sharing data. The first one is a direct data exchange, which occurs between two companies. Secondly, there is data pooling, where several companies share data "about a given service or in general in an industry or an ecosystem". Thirdly, there is a data-sharing platform, which can be a marketplace or a peer-to-peer model.

Moreover, for data collaboratives, numerous ecosystem types have been defined. First of all, Verhulst and Sangokoya [Verhulst and Sangokoya, 2015] define six forms of data collaboratives: trusted research partnerships, prizes and challenges, intermediaries, APIs and corporate data pooling. This taxonomy was formed based on quite anecdotal examples of data collaboratives and it does not prevent overlap between the various types. Susha, Janssen and Verhulst [Susha et al., 2017a], van den Broek and Veenstra [Veenstra, 2015] and [van den Broek and van Veenstra, 2018] apply the inter-organizational modes of governance as originally introduced by Lowndes and Skelcher [Lowndes and Skelcher, 1998] to data collaboratives. These modes are market, bazaar, hierarchy and networks. Markets are characterized by a high level of autonomy for members and are based on agreements between buyers and suppliers, which diminishes the need for trust in the ecosystem. The bazaar mode is a community of actors that cooperate to achieve a common goal, this type of collaboration relies on trust to a significant extent, as there are usually no contracts involved. The hierarchy governance mode is focused on formal relations between individual members, with higher-ranked members having power over lower once. Many supply chains follow this mode. Finally, there is a network which is a hybrid between markets and hierarchies, their governance relies on social contracts between members, moreover, coordination and decision making is done based on consensus between all or most of the participants. An overview of the types with their characteristics in terms of data sharing, coordination mechanisms, control over data and with an example can be found in Figure 8.1 [Veenstra, 2015].

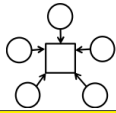

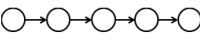
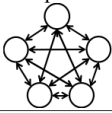
| | Market | Bazaar | Hierarchy | Network |
|---------------------------------|--|---|---|--|
| Type of data sharing | Pooled  | Complex  | Sequential  | Reciprocal  |
| Characteristics of data sharing | Buy and sell data based on (dyadic) transactions | Open up and reuse of data | Data exchange orchestrated by dominant member(s) | Lateral data exchange between individual members |
| Coordination mechanisms | Contracts | Data quality | Power exerted by the dominant member(s) over the others | Trust |
| Control over data | Remains at individual organisations | Open licence means that everyone has access to the data | Determined by the dominant member(s) | Remains at individual organisations |
| Example of data collaboration | Central marketplace for big data | Open data community | Supply chain network | Networked exchange of data |

FIGURE 8.1: *Various Types of Data Collaboratives by van Veenstra et al. [Veenstra, 2015] the yellow elements are used for the characteristics overview*

Finally, in Open Data literature, several taxonomies can be found. First of all, Buda et al. [Buda et al., 2016] define four open data strategies that result in various ecosystem types. These types are based on two variables. First of all, they can either make the data available for free or at a price. Moreover, they can make the ecosystem completely open in terms of access or restrict it. This leads to four strategies over these two dimensions. Data philanthropy, opening up the data for free to everyone, naked corporation, where data is opened for free, but to a limited target audience, monetizing data, where it is open to a limited group and at a cost and, finally, commercial openness, where data is open to everybody, but at a cost. Moreover, Curry and Sheth [Curry, 2018] differentiate ecosystem types by the entities they are formed around. These entities can either be an organization, an activity, a community, a geographical location or within or across industrial sectors. Subsequently, they differentiate types based on the control over data. Directed data ecosystems are centrally controlled and for a specific purpose. Acknowledged data

ecosystems have defined objectives and pool resources, therefore, lacking central control. Collaborative data ecosystems have participants that collaborate voluntarily for a predefined central purpose. Lastly, a virtual data ecosystem has no central management and no agreed-upon purpose. Finally, in the context of Open Government Data platforms, Danneels, Viaene and van den Bergh [Danneels et al., 2017] define two types based on two knowledge epistemologies. The first one is the cognitive view, which says that open data should be organized for ad hoc querying by individual actors and the focus of these ecosystems should be on the one-way traffic of data. However, in the connectionist epistemology, the focus is on connection actors and data, aiming to connect actors that can benefit from further collaboration but would otherwise not have been able to connect directly.

All these types and the specific dimensions that separate them will be considered in this research to determine the various types that can be found in DaaP ecosystems. These types, together with the characteristics found in the previous section will first be further analyzed through the expert interviews. Thereafter, a selection of them will be used to define DaaP ecosystems in more detail.

8.2 INITIAL MODEL

Based on the literature, an initial model is created, this model can be found in table D.1 in Appendix B. The first column of this model indicates the category, that can have several characteristics. The subsequent columns then indicate the potential characteristics that are available for that category. These characteristics have to be implementable by orchestrators of data platforms, at least at the start of the platform, making the options different choices a company can make. This model is used as input for the interviews.

8.3 INTERVIEWS

Due to the limited time available during the interviews, the DaaP characteristics were only discussed during interview 1. This interview was started with a very early version of this model, displayed in table D.1. The goal of this model is to help organizations wanting to start a DaaP by deciding how to start. It describes specific choices companies can make when starting.

A first major point made by the interviewee about the model was that it required definitions. He was uncertain about the meaning of many of the categories, as well as the individual characteristics. Stating that: "I think you need explanations around the topics. I have no idea what you mean with 'orchestrator activities', 'collecting and sharing activities' or how should I look at this?" Next to this, the purpose of the model was quite unclear. He asked: "Does that mean that you need to choose between these options [when starting a DaaP]?" Moreover, as a comment on the overall model, he notes that there is quite some overlap between all the categories. For example, comparing the **Type of data sharing** with the **Data pricing**, which both have elements concerned with paid versus free data.

Next to the major feedback on the model, interviewee 1 commented on many of the specific categories present in the model. The first aspect he disagreed with was the **Research and policy problem** category. He states that it could also be targeted towards a societal or a health problem, asking if these shouldn't be included. Additionally, he questions the compatibility between this category and a commercial data platform. Furthermore, this interviewee is confused by the difference between users and providers, questioning if, for example, **collaboration between data users** could not also involve data providers. Additionally, he says that the *data quality* characteristic in **coordination mechanism** is quite difficult to understand. Stating that data quality is not a way to coordinate users.

Finally, after all this feedback, it was decided that the model needs significant rework. The feedback showed that a more strict review of the selected literature is required with more strict definitions for what constitutes a characteristic, as the interviewee asked many questions on the purpose of the overview. For example, pointing at several characteristics of which it was surely not possible to "make choices" as

they are out of the control of the orchestrating organization. For this reason, it was determined that the definition used to select characteristics needed to be more specific. Using this new definition, an updated model was designed, combining the feedback from interview 1 and a new review of the literature to update the model.

8.4 FINAL MODEL

For this new model, which simultaneously will be the final model after the interviews, the selected articles were analyzed for a second time for characteristics. During this analysis, the requirement that the characteristics are supposed to be choices within a certain category was enforced more strongly. This yielded a new list of characteristics based on the literature found. The table with the input for the new model can be found in Appendix B.2 in Table D.2.

This second review has led to a final model that can be found in Table D.3 in Appendix B.3. This table incorporates the new characteristics selected from the revised literature review, combined with the feedback from interviewee 1. The model consists of 15 categories that have between 2 and 5 options each. Finally, a list of definitions is added for the various categories. These definitions can be found in Table D.4 in Appendix B.4. These definitions can help when applying the model by clarifying the various categories. Moreover, they are key to keep the results consistent when the model is applied by multiple people for the same platform separately.

8.5 CASE STUDY

Finally, the final model with the characteristics is discussed in the case study. As described in chapter 6, the focus of the case study was more on the method and usability of the materials and did not specifically focus on the correctness of the model. An assessment of the characteristics for Platform F was done by the three interviewees. Next to that they were asked about the motivations, situation and method for using the characteristics. Finally, they were asked about the usefulness and understandability of the overview.

8.5.1 CHARACTERISTICS ASSESSMENT

During the case study interviews, an assessment of the characteristics of Platform F is made. During this part of the case study interview, the characteristics were discussed from top to bottom. The results of the assessment can be found in Figure 8.2 This section will briefly describe the direct result of the assessment.

The first topic in the characteristics overview is **Governance**. On the first two characteristics, **Co-ordination Mechanism** and **Control over Data**, there is some misunderstanding over the different options. Especially Persons I and J have a hard time picking one of the options as they find them all somewhat correct. The set-up where the platform is governed by a collaborative of four organizations, but then operated by one is not specifically mentioned and, therefore, no real decision can be made. Person K does select an option for both, but with a similar nuance that sometimes more options are correct. For the remaining two characteristics on this topic, the interviewees do agree. About the **User Selection** topic is no doubt at all and *open* is the unanimous pick. On the **level of centralization** characteristics the response is that it is a balance between the two, data providers are not left completely free, but do have some control.

The second topic is the **Platform**. On this topic nearly all characteristics are picked unanimously. The **Specificity** is determined as *non-specific*. The **Openness** is *Open*. The **Diversity of data providers** *Several from the same industry*. The **Location of Open Data** is *Platform*. The **Continuity of Data Sharing** is *Continuous* and the **Multiplicity** *Many-to-many*. The only topic where no full agreement is on is the **Degree of integration**. Person I and K find it *Inter-industry*. Person J does see a big preference from two main industries, being the geographical services and the real estate industry and, therefore, was not completely sold on either option.

| | Categories | Values | | | | |
|--------------|--------------------------------------|---|--|---|--|---|
| Governance | Coordination Mechanism | Dominant member set standard, through contract or protocol K | Only coordination is that all data is shared via the platform. Only technical maintenance is done collaboratively. | Collaborative analysis by several users for a specific target | | |
| | Control over Data | Hierarchy: - Determined by dominant organization, which tightly controls K | Bazaar: - No control K | Network: - Individual organizations retain control | Control over data in the hand of the buyer | |
| | User selection | On agreement basis (Based on a partnership agreement) | On application basis (Selected for specific application of the data) | Open IJK | | |
| | Level of centralization | Decentralized: Division of Authority, Responsibility and Decision Making IK | Centralization: Authority, Responsibility and Decision Making at central party IK | | | |
| Platform | Specificity | Specific to a single Use Case IJK | Non-specific IJK | | | |
| | Openness | Open IJK | Restricted | | | |
| | Diversity of data providers | One | Several from the same industry IJK | Cross-industry | | |
| | Location of open data | Company website | Platform IJK | Selling on a marketplace | | |
| | Continuity of Data Sharing | On demand | Event-based | Continuous IJK | | |
| | Degree of integration | Intra-industry | Inter-industry IK | | | |
| | Multiplicity | One-to-many data sharing | Many-to-many data sharing IJK | | | |
| Orchestrator | Platform ownership | Set up by sharing company IJK | Created by third party | | | |
| | Ground for data sharing | For a single specific purpose | Based on a strong generic purpose (i.e. curing a disease any way possible) K | No specific purpose I | | |
| | Motivation for starting the platform | Get a seat at the table I | Explore new business models/Commercial purposes | Cut costs IJK | Collaborative innovation/ Research and innovation IJ | Gain legitimacy engaging other stakeholders I |
| Data | Type of data | Processed data IJK | Direct access to a copy of raw data IJK | Modified/enriched data IJK | | |
| | Personal Data | Without personal data IJK | With personal data | | | |

I Person I **J** Person J **K** Person K

CHARACTERISTICS

FIGURE 8.2: The results from the characteristics assessment of Platform F during the case study

The third topic is the role of the **orchestrator** in the ecosystem. For the first characteristic, **Platform Ownership**, the participants agree that the platform is set-up by the sharing company. For the remaining two, more motivational topics, there is some disagreement. Person K found that the **Ground for data sharing** is *Based on a strong generic purpose*, which according to him is promoting open data in general. Person I, however, thinks there is no specific purpose for the data. Finally, there is no agreement on the **Motivation for starting the platform**. Partly, it is found that the list may not be exhaustive, as a major reason for this platform comes from legislation. Next to that, many of the motivations are deemed somewhat relevant.

The final topic is **data**. This topic only consists of two characteristics and both are unanimous. All interviewees find that all **types of data** are being shared. However, that different datasets can contain different types. Finally, they also agreed that in terms of **Personal Data** no *no personal data* is being shared.

Several things can be concluded from these results. First of all, there are many points on which there is agreement between the participants. As stated by Person J, some of the characteristics are more concrete while others are a bit more vague. In general, this seems to align with characteristics on which there was agreement. These things are elements that are observable by just looking at or knowing the platform. While these could be seen as obvious for that reason, the interviewees indicate that it can be useful to consider these aspects when starting out. Organizations might not be familiar with all the options and this overview will help them consider these. A second point brought up primarily

8.5.2 METHOD

In terms of the method that is used during the assessment a very straightforward approach is taken. First of all, interviewees are shown the model and, subsequently, asked if they were to use it how they would go about it. This is a challenging question with limited knowledge from the model and all agree on the approach that they would start on the first row and work their way down. This is also the approach that was planned for this artefact and, therefore, this approach was taken during the assessment. After the assessment, the case study participants were again asked for how they would use the overview if they were to use it again. All agreed that, considering the compact size of the overview, they would again just go top to bottom and that no specific method was necessary.

8.5.3 MOTIVATION

The case study participants were also asked when and why they would use the characteristics. We call this the motivation and situation for using the overview. Person J answered this question in the context of Platform F. This interviewee indicated that most of these options are considered in the early phases of the platform. However, he describes how not all the options can be foreseen when starting off and that some decisions also change as the platform grows. He discusses how the overview would have definitely helped at that point, as some options described by the overview were not considered at that time and were, but could have been relevant for the decisions made at the time. Next to that, he indicates that a reassessment could have helped when making important decisions concerning the platform. Giving an example of discussions on what to do when users were to abuse the platform, he says that the model could have helped consider their initial stance on creating a somewhat restricted platform. At a later point, when that was not deemed necessary anymore it was useful for considering if it should be made completely open.

Person I also focuses on its usefulness when starting out. He describes how the overview can help better consider decisions that may have a long term impact on the success of the platform. He thinks that some of the options described in the model may be missed when starting out and this overview will prevent that. Next to that, it asks question that will help the consideration of future situations and how to deal with them. He recommends platforms to go through the overview before starting off.

| | Usefulness | Understandability |
|----------|------------|-------------------|
| Person I | 5 | 2 |
| Person J | 5 | 2 |
| Person K | 4 | 4 |
| Average | 4,6667 | 2,6667 |

TABLE 8.1: *The performance scores for the characteristics overview*

Person K describes how it can help when starting a platform combined with, for example, a business model canvas assessment. This overview can help make decisions on how to make the platform a successful business opportunity. He says that it can help see what you can do, in terms of decisions you can make and how that can help work towards a value proposition.

8.5.4 PERFORMANCE

Finally, the usefulness and understandability are graded on a Likert scale of 1 to 5. An overview of the results is displayed in Table 8.1. For the *usefulness*, Person I notes that he thinks it is a 5 for the overview, especially for governmental parties working on this. On the other hand, he thinks the likeliness that they will actually use it is closer to a 1. Person J also gives a 5 for the usefulness of the overview. He describes how it is immensely useful at the start of a platform and that organizations should also consider the points at important decision points in their lifecycle. Person K gives a 4 with little explanations.

For the *understandability*, Person I is less positive. He describes how he really needs the author's additional descriptions of the topics to make them understandable. It should be noted that the additional definitions created for the concepts were not provided in the case study interview. He also thinks that many of the concepts are ambiguous and interpretable in multiple ways. However, he also questions how bad that actually is, as the main goal of the model should be to spark a discussion and make a decision based on that, not the exact meaning of the concepts. Person J is also less positive about the understandability. He initially did not understand that the point was to make choices and that it is possible to select multiple. Additionally, he notes that the addition of an "other" option. Finally, he notes that the overview is extendable and that suggestions for extensions should be possible. Person J indicates that, for him, as subject-matter expert, the understandability is a 4, but that improvements can be made to make it more universally understandable.



DAAP CAPABILITIES

This chapter discusses the third and fourth research question regarding the capabilities of a DaaP. The final goal of these questions is to create a research method that can help companies assess and form an improvement plan to mature their DaaP efforts. These research questions are: "What capabilities are relevant for operating or utilizing DaaP ecosystems?" and "How can an organization develop the Capabilities to get them to a level that fits their objectives?" This chapter first discusses the input collected for the model during the literature review. Then it discusses the creation of the first model based on this input. Subsequently it describes the feedback and evolution of the model during the interview phase. Thereafter, the final model is described. Finally, the application of the model in practice during the case study is laid out.

9.1 SYSTEMATIC LITERATURE REVIEW

In the context of Data Governance of platform ecosystems, Lee, Zhu and Jeffery [Lee et al., 2019] note that platform ecosystems can be characterized as either centralized or decentralized. They also note that roles, revenue sharing, trust and control are key concepts for ecosystem governance. Moreover, in the context of data management, they define data ownership, data usage as characteristics on which platform ecosystems can differ. In a different work, Lee, Zhu and Jeffery [Lee et al., 2018] name seven data governance decision domains on which platform ecosystems can differ: Data Ownership, Regulatory Environment, Contribution Measurement, the data use case, conformance, monitoring and data provenance.

Furthermore, Lee, Zhu and Jeffery [Lee et al., 2017b] create a contingency model for data governance in platform ecosystems to assist decision making. They describe seven contingency factors and, subsequently, define them for both centralized and decentralized platform ecosystems. Additionally, for each contingency factor, they describe several data governance factors that influence them. Finally, they describe the papers from which they retrieved the various contingency and data governance factors. The model they created is illustrated in figure 9.1. This model and especially the data governance factors can be useful as potential capabilities relevant for DaaP ecosystems.

All these capabilities found in the literature are displayed in table E.1 in Appendix C. This table consist of a grouping of capabilities with the same or a very similar meaning. Moreover, this list consists of both practices and focus areas and no initial separation is made between the two. These capabilities are

| Contingency Factor | | Definition (centralized <-> decentralized) | Data Governance Factor | Reference |
|-----------------------------|-----------------------|---|--|---|
| Performance strategy | | Performance objective of a platform ecosystem (profit or growth <-> quality) | Data ownership/access definition, data use case, conformance, monitoring | (Weill & Ross 2005; Weber et al. 2009; Hagiu 2014; Evans 2011) |
| Multi-homing strategy | | Degree of affiliation in a platform ecosystem (single-home <-> multi-home) | Data ownership/access definition, data use case | (Weber et al. 2009; Parker & Van Alstyne 2013 and 2014) |
| Degree of market regulation | | Degree of regulation regarding the use of data in a platform ecosystems (highly regulated <-> no regulation) | regulatory environment, conformance | (Weber et al. 2009; Kaisler et al. 2012 and 2013; Khatri & Brown 2010; Ghazawneh & Henfridsson 2010) |
| Governance Configuration | Governance mechanisms | Type of governance of a platform ecosystem (authority-based <- contract-based -> trust-based governance mechanisms) | Contribution measurement, conformance, monitoring, data provenance | (Tiwana 2013; Manner et al. 2013; De Reuver & Bouwman 2011; Ouchi 1979) |
| | Control mechanisms | Type of control of a platform ecosystem (formal <-> informal control) | Data ownership/access definition, conformance, monitoring, data provenance | |
| Platform market structures | | Type of platform market structures based on competition (coincident <- intersecting -> monopoly platforms) | Data ownership/access definition, data use case, monitoring, data provenance | (Evans 2011; Parker & Van Alstyne 2009 and 2014) |
| Open strategy | | Level of openness of a platform ecosystem (open <-> close) | Data ownership/access definition, data use case, monitoring, data provenance | (Gawer & Henderson 2007; Parker & Van Alstyne 2014; Schrieck et al. 2016; Hein et al. 2016; Choudary et al. 2016) |
| Platform maturity | | Level of maturity of a platform ecosystem (immature <-> mature platforms) | Contribution measurement, monitoring, data provenance | (Cusumano 2010; Parker & Van Alstyne 2014; Schrieck et al. 2016; Hein et al. 2016; Choudary 2013; Choudary et al. 2016) |

FIGURE 9.1: Contingency Factors for Data Governance in Platform Ecosystems by Lee, Zhu, Ross [Lee et al., 2017b]

later used to create an initial model. These capabilities will be analyzed, through the expert interviews, for use in the maturity model. Moreover, experts will have the opportunity to supplement the list of capabilities with those that come from their expertise. Moreover, more literature on data governance and data management may be reviewed for further input on data capabilities that could be relevant for a platform, as these results are quite limited. As all capabilities are validated by experts their relevance in the literature to ecosystems and data sharing is not essential. Experts can make their own decisions on the relevance of capabilities on DaaP platforms.

9.2 FOCUS AREA MATURITY MODELS

The Focus Area Maturity Model is designed based on the Development Method for Focus Area Maturity Models as introduced by [van Steenbergen et al., 2010]. This is already briefly discussed in chapter 3 Figure 3.3. This type of maturity model has several benefits over traditional fixed-level maturity models that are specifically applicable for this research. First of all, it allows users to select specific focus areas to focus on. Traditional maturity models often state that a next maturity level is only achieved when all areas are on that level, this is not the case for this type of maturity model. This type allows users to pick the focus areas they find relevant and focus on their improvement. In the context of maturing a DaaP platform this is a major benefit. For data sharing in a platform some focus areas may be more important than others, dependent on the context and use case of the ecosystem. For example, some platforms may require the real-time sharing of data, requiring a high data actuality, while for other DaaP platforms this may be irrelevant. Moreover, a focus area maturity model allows for the modelling of dependencies, both within a focus area and between focus areas. This can also be relevant in the context of DaaP maturity, as some practices may require others before they can mature. For example, data documentation that includes version history, first requires proper version management within the API.

A Focus Area Maturity Model consists of three concepts on different levels of granularity. On the highest level there are **focus areas**, these are general topics discussed on the model and can be seen as categories. Within a focus area fall several capabilities. A **capability** is a single element of a DaaP

of which a maturity level can be determined. This maturity level can be found based on the practices in the capability. A **practice** is a single implementable measure for a capability and falls on a certain maturity level. The current maturity level of a capability is the level before the first unimplemented level. A visualization of these levels, accompanied by an example from the final model, can be found in figure 9.2.

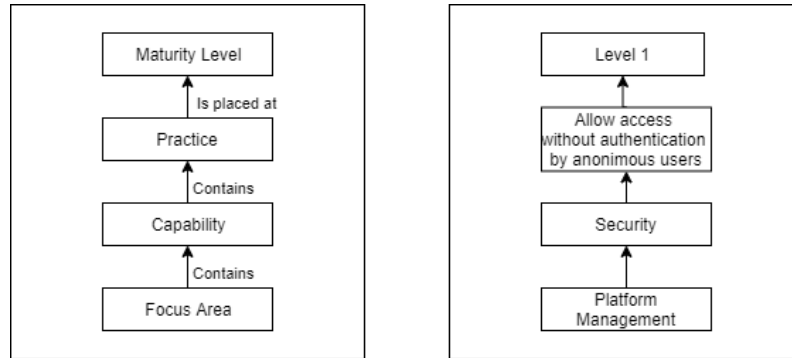


FIGURE 9.2: A meta-model for Focus Area Maturity Models with example

9.3 INITIAL MODEL

Based on all the results from the SLR, an initial Maturity Model is created to be used as discussion material during the interview phase. As discussed the list of capabilities described in table E.1 is quite unorganized. It consists of a lot of capabilities that can either be focus areas or practices and they currently do not indicate maturity.

As a first step towards creating the initial model, the capabilities with a similar topic are put together on the same row. If this row includes a capability that describes the overarching topic, then this capability is made the focus area name. When this is not the case, a focus area name is determined based on the various practices.

As a next step, the practices are put in sequencing that corresponds to their maturity. This is done somewhat pragmatically, as not all capabilities have a description that specifies if they are more mature than other capabilities, however, as the model is later validated by experts, any mistakes can be rectified during the interview phase.

Subsequently, the levels are determined, practices from different focus areas that belong to a similar level of maturity are put in the same level, for this division within levels, the "intra-focus area" dependencies of practices are also considered. Again, this is done pragmatically and with some subjective input from the author. Furthermore, descriptions are created for the various levels, that help determines the maturity level a practice should be in. These descriptions can be found in table 9.1.

Thereafter, practices are renamed to form a consistent practice description for the maturity model. Each practice is put in the imperative form and starts with a verb. Finally, an overarching categorisation is created for the various focus areas, this is again done based on a combination of input from literature and subjective insights by the authors.

| Level | Title | Description |
|---------|--|---|
| Level 1 | Basic data sharing partnership between companies | The sharing company just started of sharing data and they are testing the waters by sharing an insignificant amount of data to a small group of partners. Most likely the revenue from this effort is nullifiable with the additional earnings of the company. |
| Level 2 | Robust Data Platform | The number of functionalities for the data platform is growing, moreover, a limited number of users are actively requesting access to the data. |
| Level 3 | Profitable Data Platform | Monetization from the platform is growing, as more users request access in an automated way. The platform is finding its niche and use cases for the data become more clear. |
| Level 4 | Platform with major company backing | The platform is becoming a major focus from the orchestrating company as a reliable source of revenue. A significant percentage of people in the company are actively working on improving it. |
| Level 5 | Industry standard data platform | The platform has become an industry standard and has been widely adopted by related companies. This has improved the reputation of the orchestrating company significantly. |
| Level 6 | Data platform as grand profit source for operating company | The data platform has become the main source of revenue for the orchestrating company and has therefore become the main focus by top-management and many of the employees working for the company. |
| Level 7 | Expanding data platform | The platform is actively expanding to different industries and use cases, acquiring data sets to achieve that goal and establishing itself as a general source for valuable data. |
| Level 8 | Reigning data platform | The platform is actively processing requests from other companies to share its data and can set the price for that themselves, maintaining a data platform for general use by a wide range of clients has become the main source of revenue for the company and they are seen as a global leader in that front. |

TABLE 9.1: *The Maturity Level Descriptions*

| Focus Area | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Level 6 | Level 7 | Level 8 | Level 9 | Level 10 |
|------------|---------|------------------------------|---------------------------------------|---------|--------------------|---------------------------------------|---------|---------|-------------------------------------|----------|
| Security | | Allow users to give security | Implements watermarking for providers | | Automated security | Provide security levels for providers | | | Use a trusted supplier for security | |

TABLE 9.2: *The First Version of the Security Capability*

9.4 INTERVIEW

9.4.1 DAA P MATURITY MODEL

The main topic of discussion during all interviews was the DaaP maturity model. During these interviews, the interviewees were asked about the correctness of the model, based on their opinion and expertise. They were asked to judge if elements were correct, valid for DaaP and in the correct place in the model. Moreover, they were asked if any elements were missing in the model. The interviews started with a maturity model that was 37 Focus Areas long and 10 levels wide. The complete maturity model used during the interviews can be found in appendix C.

The total amount of feedback given by the interviewees is too extensive to describe in detail in this subsection. Therefore, this subsection gives a detailed description of the feedback received on one of the focus areas, the security focus area. The goal of this description is to exemplify how the interviews were used to update the model. An overview of the interview results on the security topic can be found in table 9.4.

The first interviewee to look at the maturity model and, therefore, at the **security** focus area, was interviewee 2. This interviewee was shown the **security** focus area as displayed in table 9.2. This interviewee started by stretching the importance of security for a DaaP. He states: "**Security** is essential, also when you to prevent that users start cloning your data." Subsequently, even before looking at the specific practices he notes: "I already see that you have **security** here, but also **authentication** and **data access** in other places, when I think of a maturity model I expect a much smaller model, 5 levels and maybe 10 topics it discusses", implying that **authentication**, **data access** and **security** should be merged. To achieve this, some practices should be added: "I expect an identity access management system, some authorization and various types of authentication here." This comment resulted in addition of the practices: *Allow access without authentication by anonymous users*, *Perform authentication through username password / API key*, *Perform user-based API authorization*, *Implement use case-specific best-practice authentication methods* and *Connect Identity Access Management and Authorization system to the security focus area* for the next iteration. Furthermore, interviewee 2 noted that: "*Use a trusted supplier for security* is placed at a high maturity, but I am not sure if that is mature or immature, if you have the expertise, it could be much better to create your security system." This remark initiated the change of *Use a trusted supplier for platform security* from level 8 to level 3 and it introduced *Implement a use case-specific state-of-the-art security system*, as the highest maturity practice for this focus area.

Interviewee 3 had little to add to the **security** focus area. Mainly noting that: "Look at **security**, that one makes sense, with each practice the platform becomes increasingly secure." Furthermore, this interviewee had little to add about the specific practices, stating that: "I am sorry I cannot give feedback on every individual aspect, but I would focus on clustering all the information in the model." For this reason, it was decided that this interviewee agreed with the presence of the **security** focus area, but that he was unable to react to the specific practices and, therefore, his reaction to the security practices was

| Focus Area | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Level 6 | Level 7 | Level 8 |
|-----------------|--|--|--|--------------------------------------|--|---------|---|--|
| Security | Allow access without authentication by anonymous users | Perform authentication through username password / API key | Use a trusted supplier for platform security | Perform user-based API authorization | Implement use case specific best-practice authentication methods | | Connect Identity Access Management and Authorization system | Implement a use case specific state-of-the-art security system |

TABLE 9.3: The Final Version of the Security Capability

not scored.

During interview 4 the security was only discussed briefly. The interviewees in this interview mentioned that their platform, Platform A, was already quite advanced in terms of security and that this was, therefore, not a priority in their current project. Nevertheless, looking through the practices one interviewee did note that the practice *Implement automated security* was rather vague. This statement corroborated a similar statement from interview 2. Therefore, with two interviewees disagreeing and none agreeing, this practice was removed after interview 4. Moreover, these interviewees stated that having multiple practices in a single level, which happened after the feedback from interview 2 was processed, was quite confusing, even for the interview phase of the model. For this reason, the security practices that interviewee 2 had objected to were removed, resulting in a maximum of one practice per maturity level.

During interview 5 the security of the platform was not discussed in detail. This, however, was not because this interviewee had little experience with the security, but because this interviewee would only mention aspects of the model he disagreed with. As there was no substantial feedback on the security focus area, this was considered to be correct in his opinion.

Interviewee 6 was more strict about the **security** practices mentioned in the model. He stated, "how do you plan to run a commercial data platform without having identity access management." This comment argued a sentiment more often repeated by this interviewee, arguing that "I would not consider sharing data when I am not at least near the end of your model." This argument specifically focused on the fact that a DaaP should be a commercial data platform and that the risk that a platform would run into security issues was too high if it did not have a high maturity in this regard. These security threads would make a low maturity data platform, in his opinion, not commercially viable. Therefore, all practices below *connect identity access management and authorization system* were scored as against.

Finally, during interview 8, the resulting **security** focus area was confirmed by the interviewee. He stated that "security I don't know that much about, this looks logical, but we have not yet worked on this", referring to Platform A in the final part of the sentence. This feedback led to the scoring displayed in table 9.4 and the final version of the **security** focus area displayed in table 9.3.

9.5 FINAL MODEL

Each other capability was updated based on the interview results in a similar way. A complete overview of the changes made and the feedback from the interviewee that inspired it can be found in appendix F F.12. Moreover, not only did interviewees suggest changes to the capabilities and the practices, but they also provided feedback on the focus areas inside the model. Based on that feedback, final focus areas were created, the changes to the focus areas are also discussed in the table in Appendix D. Next a table

| Focus Area/Practice | Introduced by | Level | In favour | Against | Total Score | Result |
|---|--------------------------------|-------|-----------|---------|-------------|-------------------------|
| Security | LR | FA | 6 | 0 | 100% | Retained |
| Allow access without authentication by anonymous users | I2 | 1 | 4 | 1 | 80% | Retained |
| Allow users to give security | LR | 2 | 0 | 2 | 0% | Removed after I4 |
| Perform authentication through username password / API key | I2 | 2 | 4 | 1 | 80% | Retained |
| Implement watermarking for providers | LR | 3 | 0 | 2 | 0% | Removed after I4 |
| Use a trusted supplier for security | LR (Added from lvl 8 after I2) | 3 | 4 | 0 | 100% | Retained |
| Perform user-based API authorization | I2 | 4 | 4 | 0 | 100% | Retained |
| Implement automated security | LR | 5 | 0 | 2 | 0% | Removed after I4 |
| Implement use case specific best-practice authentication methods (e.g. Oauth or OpenID) | I2 | 5 | 4 | 0 | 100% | Retained |
| Provide security levels for providers | LR | 6 | 0 | 2 | 0% | Removed after I4 |
| Connect Identity Access Management and Authorization system | I2 | 7 | 5 | 0 | 100% | Retained |
| Use a trusted supplier for security | LR | 8 | 0 | 2 | 0% | Moved to lvl 3 after I4 |
| Implement use case specific security system | I2 | 8 | 4 | 0 | 100% | Retained |

TABLE 9.4: The interview results for the Security Focus Area

with all changes made, a second model is created with the acceptance rate for each of the focus areas and practices that have been in the model at some point. This table follows a similar layout as that of table 9.4 and can be found in appendix F in Table F.1 and onwards.

This has resulted in a final model, which is displayed in Figure E.1 in the next chapter. This model is the result of all the feedback from the interviewees applied to the initial model. Furthermore, in colour, it displays the dependency modelling that shows which practices are dependent on each other. This dependency modelling, which can also be applied across several focus areas, can be used by practitioners to determine if they have all requirements in place to work towards a new practice within a certain focus area.

| Focus Area | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Level 6 | Level 7 | Level 8 | Level 9 | Level 10 |
|------------------------|------------------------------------|---|--|---|--|--|--|--|--|--|
| Platform Management | Security | Allow users to give security | Automated security | Provide security levels for providers | Set strict guidelines for data use and | Set strict guidelines for data use and | Set strict guidelines for data use and | Set strict guidelines for data use and | Set strict guidelines for data use and | Set strict guidelines for data use and |
| | Public Data Goals | Establish research goal | Use mutual alignment to match data | Check behavior in line with goals | Develop organizational models | Develop organizational models | Develop organizational models | Develop organizational models | Develop organizational models | Develop organizational models |
| | Institutional Design | Provide metadata | List types of datasets | Create advanced API materials | Support structural metadata | Support structural metadata | Support structural metadata | Support structural metadata | Support structural metadata | Support structural metadata |
| | Platform maintenance | Maintain original platform setup | Add additional data sets | Model dependencies in data | Create transparent roadmap | Create transparent roadmap | Create transparent roadmap | Create transparent roadmap | Create transparent roadmap | Create transparent roadmap |
| | Decision Making Process | Establish decision making structures within an organization | Create informal process chart for establishing decision rights | Maintain centralized control strategy | Decentralize platform control | Decentralize platform control | Decentralize platform control | Decentralize platform control | Decentralize platform control | Decentralize platform control |
| | Performance Strategy | Multi-homing strategy | Measure contributions | Prohibit multi-homing | Create a contribution measurement model | Create a contribution measurement model | Create a contribution measurement model | Create a contribution measurement model | Create a contribution measurement model | Create a contribution measurement model |
| | Contribution estimation | Provide possible opportunities to stakeholders for monitoring | Identify API metrics | Set usage quotas | Check for imbalances in data | Check for imbalances in data | Check for imbalances in data | Check for imbalances in data | Check for imbalances in data | Check for imbalances in data |
| | Monitoring | Reliability vs. cost trade-off | Keep data up to date | Define and link the use cases and the relevant stakeholders of the data | Support data integration with related datasets | Support data integration with related datasets | Support data integration with related datasets | Support data integration with related datasets | Support data integration with related datasets | Support data integration with related datasets |
| | Platform Measurements | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Data Quality | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| Data Related | Flexibility | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Data use case | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Data Processing | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Partner relationship | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Institutional Design | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Coordination | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Reward System | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Revenue sharing | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Investment | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Compliance and External Regulation | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| Compliance | Internal Compliance | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Policy | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Type of data sharing | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Data Access | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Governance | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Control over Data | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Data Sharing Technology | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Data Processing | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Data provenance | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Authentication | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| Platform Functionality | User Contributions | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Search functionality | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Pricing Mechanism | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Authentication | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | User Contributions | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Search functionality | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Pricing Mechanism | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Authentication | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | User Contributions | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |
| | Search functionality | Let IT determine data structures | Share user data | No relationship with users | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship | Partner relationship |

FIGURE 9.3: The First Version of the DaAP Maturity Model



THE DATA AS A PLATFORM FOCUS AREA MATURITY MODEL

This chapter presents the main result of this research: The DaaP Focus Area Maturity Model (DaaPFAMM) and discusses its contents. Subsequently, it describes a method that can be used by practitioners when the DaaPFAMM is applied to their platform. The final version of the DaaPFAMM is displayed in Figure E.1. A searchable version, that has not been resized to fit a single page can also be found in appendix C.

10.1 THE DAA P FOCUS AREA MATURITY MODEL

This section will go through the contents of the model and describe all the focus areas, capabilities and practices that are in them. The DaaPFAMM is build up out of 6 focus areas. These focus areas are: Platform Management, Data Management, Partner Management, Compliance, Ecosystem Management and Additional Services.

10.1.1 PLATFORM MANAGEMENT

The first focus area is Platform Management. This focus area is the largest in the model, containing 8 capabilities. Platform Management contains the capabilities that the orchestrator of a DaaP needs to manage the platform. This can be to assist users, but also to protect the platform from, for example, users trying to abuse access to the data.

The first capability is **security**. This capability was already discussed extensively in the previous chapter, but it involves several aspects that are important for security. These are authentication, authorization and protection of the platform. For each of these aspects, several practices are determined and placed in the model.

The second capability is **documentation**. Documentation is concerned with describing the contents of the platform and how the platform itself can be used. Another aspect of the documentation is the meta-data that is available for the data. As this capability becomes more mature it also becomes possible to

generate documentation based on the code and contents of the APIs. Moreover, documentation assisting API usage is extended to assistance tools that can be download to ease integration of the platform.

The third capability is **platform maintenance**. This capability is concerned with how actively the platform is maintained and updated by the orchestrator. These updates are concerned with both the API and the data itself. At lower maturity levels the platform has just been created and little is done after that. At higher levels the API is constantly updated with the newest best practices and data is updated when it becomes outdated.

The fourth capability is **Decision Making Process**. This capability has to do how decisions are made about the platform and how transparent an organization is in making them. In the initial levels, decisions are just made by the orchestrator and there is no real transparency behind how this is done. As the levels go up, the transparency in sharing how decisions are made or what the next steps are for the platform increases. Then in the final levels, users are being increasingly involved in decisions. In one of the most mature levels decision rights are handed to users to have a voting say in decisions that are made.

The fifth capability is **monitoring data**. This has to do with how the data itself is monitored and maintained by the orchestrator. On early levels, again, little attention is paid to the relevance of data as it gets older and there are just some datasets available. On later levels, data is regularly updated with changes and users can give feedback on the correctness of the data, for example. Finally, in the last level, a supply chain of the data is displayed indicating all the changes that happened before it was made available on the platform.

The sixth capability is the **Data Use Case**. In this capability, the orchestrator looks at how the data can and must be used. This capability is more present at higher levels, as it is not a must for starting a platform. At the lower levels, it is merely concerned with looking at how data is being used and advertising that to other current and potential users. At later levels, a platform can decide that there are specific use cases for which the data should or should not be used. At the highest level, misuse of this usage goal is punished.

The seventh capability in this focus area is the **data sharing technology**. This has to do with how the data is made available. On the lowest level, there is nothing more than an excel or .csv document that is shared with the users. On higher levels, data is made available, for example, on a company's website or even in a full-blown platform. The final levels of this capability are concerned with the uploading of datasets by third parties, for example by providing functionality where users can share their datasets.

The final capability in this focus area is **platform measurements**. This capability is concerned with the usage of the platform by its users. Often DaaPs are based on fair use agreements that specify how often users are allowed to make calls to the platform. In the lower levels, these agreements are absent and, therefore, little measurements are done on the platform. As this capability becomes more mature, the usage by individual users is actively being tracked. Users are, for example, blocked when they do not comply with their fair use agreement. In the final level, fair use agreements are made flexible, so that users that require many uses can do this in times when the platform is less busy or can pay more for this.

10.1.2 DATA MANAGEMENT

The Data Management focus area is concerned with the data on the platform. It describes the capabilities a platform needs to make its data as useful as possible for its users. This has to do both with the quality of the data itself, as the formats in which it is made available.

The first capability is the **Data Quality**. This is a combination of various factors that are important for the quality of the data. At early levels, it is concerned with keeping the data up to date and making trade-offs between, for example, the cost and reliability of the data. In later levels, validation of the data takes place. This can be done through validation mechanisms, such as checking if a zip code has the correct formatting. Another technique is reconciliation, which is concerned with checking different versions of datasets, for example before and after it is updated, and checking if the row counts and location of values are still logical.

The second capability is the **Data Standardization**. This capability is about the formatting of the data when extracted from the platform. In the earliest levels, the data is available in the formatting that

is most convenient for the platform, usually the format in which it was collected. In later levels, more consistency is created in formatting between datasets and users are asked for input on the format. Finally, flexible formatting is used, where users can utilize their preferences.

The third capability is **Data Processing**. This capability is concerned with how the data is handled before it is made available on the platform. In a very immature platform, it could be possible for data to be shared directly from the database, for example in the form of a snapshot or even raw data. As the platform matures, the data is processed somewhat before sharing to make it more usable for users. In later stages, the data can be turned into linked data, allowing for advanced queries between different datasets. Finally, the data can be cleaned or turned into human-readable data, that eases analytics even further.

The fourth capability is another monitoring capability the **platform monitoring**. The monitoring of the platform is concerned with the availability of the platform. On the initial levels, measures are taken to provide users with ways to determine if the platform is available, for example through a heartbeat test. At latter levels, users are actively informed when APIs or the entire is down. Finally, advanced back-up and post-mortum practices are implemented to prevent the platform from going down.

10.1.3 PARTNER MANAGEMENT

The Partner Management focus area is concerned with how the platform supports important partners. These partners can both be users of the data or the suppliers of data. In contrast to the ecosystem management focus area, this focus area is concerned with the handling of individual partners and the points of more direct communication.

The first capability in this focus area is the **Partner Relationship**. This capability is about the relationship that is being maintained with individual partners. In the early levels, any relationship is mostly absent or limited to some initial screening or the possibility for partners to give feedback. On later levels, both communication between partners as a more intense collaboration with partners is promoted.

The second capability is **incentivisation**. Incentivisation is about promoting data usage for new and active users. The first few maturity levels are, therefore, also concerned with ways to attract new users, such as non-monetary subsidies for early users (e.g. technical assistance). In later levels, incentives are available for current users, such as performance-based rewards for users that use data in new and productive ways that attract more users. On the final level, revenue sharing schemes are suggested between the users and the orchestrator of the platform.

The third capability is **coordination**. This is concerned with coordinating the different partners and their collaboration on the platform. In early levels, orchestrators can take a "take it or leave it" stance towards how partners use the data and if they collaborate. However, as the platform becomes more mature, it can provide some functionalities that will help parties collaborate more effectively. This could, for example, be done by connecting various supply and demand parties when multiple data providers are active on the platform. Moreover, it can share information on the most used use cases of the data to inspire new products. Finally, it can set-up processes such as conflict resolution schemes that help with conflicts between users.

10.1.4 COMPLIANCE

The fourth focus area is concerned with the **compliance** of the platform. This is a smaller focus area that looks both at how the platform itself can be a complaint as well as how it can keep its users in check through internal regulations.

The first capability in this focus area is the **compliance with external regulation**. It should be noted that this is focused on compliance with all regulations except for privacy, as that category received its capability. In general, compliance is essential for a platform, as the fines for non-compliance are regularly hefty enough to take a platform out of business. Therefore, in the lower maturity levels, a good survey of regulations applicable to the platform is crucial. When the platform becomes more mature it can

improve in this aspect by actively chasing certification and creating more processes to be over-compliant, such as responsible disclosure.

The second capability is the **internal regulation** of the platform. This is capability only has a limited number of practices. Much of the internal regulation needs to be covered in the terms of service and code of conduct. This practice is placed at level 1, as this was deemed crucial for platforms that try to generate revenue through their data. On latter levels certain aspects of the terms of service could become more flexible and, for example, negotiable. Moreover, punishments could be created to punish breaches of the Terms of Service.

The third capability is the **privacy**. This practice was isolated from other compliance requirements, as it was a particularly popular topic in literature. Of course, at the lowest level, compliance with GDPR is crucial. However, as a platform becomes more mature it can implement more informed consent functionalities, broadening the availability of personal data when users agree and it can create more privacy ensuring features.

10.1.5 ECOSYSTEM MANAGEMENT

The penultimate focus area is ecosystem management. This focus area is concerned with managing the ecosystem in general. Looking, for example, how to onboard new users to the platform or who controls the data on the platform after it is shared. In contrast to partner management, this is focused on more general management activities that target more than just a single partner.

The first capability in this focus area is **governance**. As the name suggests, governance has to do with how the platform is regulated and what rules are made available by the orchestrators. In the early levels, little control is exerted and, for example, only technical maintenance is done. However, in later levels, more rules can be created to shape the ecosystem. Finally, on the last level, the ecosystem should be able to self-regulate. However, as some interviewees indicated, this is almost a utopia.

The second capability is the **control over data**. This capability is concerned with the control the orchestrator or data providers keep after their data is made available on the platform. In the immature levels, this control is very limited or not present at all. As maturity grows, an initial focus is put on communication, for example, by creating charts showing how data ownership is regulated in different scenarios.

The third capability in this focus area is **onboarding**. Onboarding is all about the experience of new users on the platform. In less mature platforms, it is mostly crucial to connect new users to the platform and little additional practices are present to enrich this experience. At some point, a platform can provide a sandbox mode where users can get a preview of some of the data to test the platform or create a test implementation on their platform. In subsequent levels, more materials are created to inform and advertise the platform to new users. For example, a data playbook can be created showcasing functionalities with the data. Finally, a scalable onboarding functionality can be created with self-service capabilities connecting users to the platform without interruption.

Finally, there is **type of data sharing**. This capability is all about the method with which agreements are made on the platform. On the first level, the platform is just available and there are no real contracts in place to give a legal basis to its usage. These contracts become increasingly important as the maturity grows, where, at some point, premium licensing can be applied to regulate the monetization of the platform. Finally, in an utopic scenario, similar to the self-regulating ecosystem, Open Licensing is applied where there is significant incentive to monetize the platform that no legal methods for monetization is necessary.

10.1.6 ADDITIONAL SERVICES

The final focus area is about any additional services the platform offers. This focus area is the smallest with only 2 capabilities.

The first is the **search functionality**. By many interviewees, this functionality is deemed crucial for a data platform. At lower levels, this functionality is complete absent or very limited, for example, only allowing search between datasets on a meta-data level. A more mature search also provides search on content level in the data and allows for filtering of results. On the final level a pathfinder can be added that connects data to data from a similar entity in another dataset.

The second is the **pricing mechanism**. This is concerned with how the data is priced on the platform and how these prices can be paid. At early levels, the pricing is very simple, for example, through flat pricing or a subscription. On slightly higher levels of maturity an ideal business model is implemented for the data. This can be very platform specific, as in some scenarios a subscription is sensible, while other platforms are better off pricing each API call. Finally, some methods are implemented to come to an ideal price, for example, creating a spot market or implementing congestion pricing, for when the platform is overcrowded.

10.2 DAAPFAMM METHOD

During the Case Study the DaaPFAMM is supported by a method for using its contents. The method consist of five steps and can differ slightly depending on the goal for which the model is applied. It can be used by practitioners to apply the model in a structured way that is likely to result in more useful and correct results than a direct assessment going through the model top to bottom. The method has limited validation in the case study.

10.2.1 THE FIVE DAAPFAMM ASSESSMENT STEPS

Step 1

Before even looking at the model, it is important to determine which of the goals is relevant for an assessment. The method description considers three goals for using the model:

- Assess the current situation of the platform
- Create a starting plan for a new platform
- Create an improvement plan for an existing platform

For this, the method distinguishes three types of maturity level assessments:

- A current maturity assessment, looking at the current situation in the platform
- A target maturity assessment, looking at the desired maturity level of the platform
- A maturity gap assessment, looking at the difference between the current and desired maturity level

A "current situation assessment" exclusively looks at the current situation and during this time only a current maturity assessment is made of the platform. For a starting plan analysis for a new platform, only a target maturity assessment is relevant, as there is no current platform to assess. Finally, for an improvement plan, both the current and the target maturity assessment is relevant for looking at the differences between the current and desired situation. Moreover, for such an assessment a gap analysis looking at the main differences between the two can be relevant.

Moreover, the participants and set-up for the assessment should be determined. Depending on the scale of the assessment there are different options for selecting participants to an assessment. An assessment can be done by as single person working on the platform or a large mixed group of employees and users involved in the platform. The main requirement for participants is, however, that they have significant knowledge on the platform. A more complete group will give a more complete assessment, but will

also be more time consuming. Moreover, involving users in the process can give additional insights in how they view the platform. On the other hand, involving users in an assessment that involves a target maturity assessment, can give users false expectations of what the platform could offer and may result in unrealistic expectations or undesirable promises by the orchestrating organization. All these factors should be considered and balanced when deciding on the participants of an assessment.

Depending on the composition of the group, different applications of the method can be chosen for conducting the assessment. Organizations can, for example, decide to put all the participants in one room for a workshop assessment. During this assessment the participants should come to an agreement on each decision. This, however, can be a difficult and frustrating endeavor, especially when there are widely differing views on the platform. Moreover, it can be difficult to organize such a meeting when users are involved with different schedules. Another option is, therefore, to do individual assessments and to combine or average the results. This ensures that the input from every participant is considered. However, this can be very time-consuming for the people performing the assessment as they have to do individual assessments. This can then be resolved by collecting input through a survey. However, this may cause misunderstandings as no further explanation can be requested for the concepts. Next to this, when the difference between results are widely different, this method does not give many insights in why this is the case. Finally, these methods can be combined by starting off with an individual assessment and concluding with a group discussion when the results are summarized. This reduces the load for participants, as the group discussion can be less extensive, but increases the load for coordinators, as they also have to facilitate this final meeting. Similar to the selection of participants, these aspects should be considered and balanced before an assessment is done.

Finally, the decisions made in this step should be clearly communicated to all participants. Especially when the assessment involves multiple people doing an asynchronous assessment it is important that the type of assessment is clear. Next to that, in that situation it is also key to prepare materials for all participants to ensure a level of continuity between the assessments. The steps 2, 3 and 4, should be performed individually by each of the participants, when this set-up is selected.

Step 2

In the second step a broad determination is made of the current or desired maturity level. Descriptions are created for each of the maturity levels, describing what sort of platform fits each of the maturity levels. These should be used by the organizations assessing their platform to understand the context of each of the maturity levels and to get an initial idea of where they stand, if they do a current maturity assessment, or want to be, when doing a target maturity assessment.

The determined levels do not necessarily need to be used when actually assessing a platform, as the practices for each individual capability is leading. However, a major reason for adding this step is making organizations aware of what maturity is realistic for their platform. It can be easy for organizations to think that they should target the highest maturity level, however, this is not always realistic or even desirable. Therefore, this step is all about aligning and informing an organization before starting off with the actual assessment.

Step 3

The next step is for determining the relevant capabilities for the platform assessed. Not all capabilities are necessarily relevant for every platform. Therefore, it can save time to, before looking at each capability in detail, determine whether or not there are any capabilities that are not at all applicable to the platform. An example of a capability that is irrelevant is the **privacy** capability. Not every DaaP shares personal data and for platform that do not, assessing the privacy measures the platform has taken is not relevant.

This step can be performed by doing a quick read-through of the model. During this read-through it is important to utilize the definitions of the capabilities, to ensure that they are understood correctly. Moreover, if a capability is still unclear after reading the description, it can be useful to look at the practices to determine its meaning. Moreover, it is important to not only consider relevance for the platform, but also for the specific goal of the assessment. If the assessment is done to look at where users can be

supported more actively, then some of the compliance capabilities may be applicable to the platform, but not relevant for the assessment. Usually, many capabilities can quickly be seen as relevant and do not need an extensive discussion. However, for some it can be good to determine if they are useful for an assessment. A discussion about this can further align the goals of an assessment between participants.

As an optional extra step, the strategic value of the capabilities could be determined. Each of the participants can indicate which capabilities they think are strategic, core or foundational. The strategic capabilities are the most important, as they are capabilities in which an organization can achieve a competitive advantage. The core capabilities are less crucial, but are an important key to creating a balanced platform, while the foundational points also build towards a better platform, but are more 'nice to have'. The results from this question can help prioritize improvements when a target maturity assessment is done.

Step 4

After step 3, the capabilities that should be considered in the assessment are known. Now the actual assessment of the platform can start. Capabilities should be discussed per focus area. The reason for this is that for capabilities inside a focus area a similar understanding is necessary. The list of definitions of the capabilities and practices should be consulted when any of them are unclear or a discussion rises on the exact meaning of a concept.

Step 5

In the fifth step of the method the results are determined. In general, this is done by calculating a score for each of the focus areas. This score is a percentage indicating to what extent a focus area is covered by the platform. The score for each focus area can be determined using the following formula:

$$\text{Score} = \text{Total Maturity of Capabilities in Focus Area} / (\text{Number of Capabilities in Focus Area} * 8) * 100$$

This will result in a score for the overall maturity of the platform. This is mostly relevant for current maturity assessments. The score gives a broad overview of the current maturity and is easily comparable between various platforms.

For a target maturity assessment the score is less relevant. For such assessments, the main result comes from looking at the individual practices that should be implemented to reach the target. In these types of assessments the previously mentioned strategic value of the capabilities can help with forming an improvement plan for the platform, where the strategic capabilities are targeted first. Moreover, when both a current and target maturity assessment is done, a gap analysis can provide a productive result. In this analysis the gap between the current and target maturity level of each capability is analyzed. This can be used to create an improvement plan targeting the most lacking capabilities first.

10.3 CASE STUDY

Finally, the maturity model is applied during the case study. Similar to the case study discussion of the characteristics, the case study primarily focuses on the method for applying the model and the motivations and situation for using it. Next to that, the performance of the model is discussed.

10.3.1 MATURITY MODEL ASSESSMENT

During the case study, a maturity assessment is performed for Platform F. The maturity model is discussed with all three participants and their view of the maturity of Platform F is noted. This is done by first selecting the relevant capabilities and, subsequently, picking the maturity of these capabilities. The results from the assessments are created by averaging the results for each practice. Practices that were found to be present by at least 50% of the participants are selected for the final assessment. The resulting assessment model is displayed in Figure 10.2. A more complete overview with the results from each individual participant is displayed in Appendix E Table G.1.

The assessment follows the steps from the Five Step DaaPFAMM Assessment method. The set-up selection described in the first step is determined by the authors. The main reason for this is to reduce the burden on the case study organization, as it removes the need for setting up meetings and finding participants. The set-up selected is that in total three persons are individually interviewed. Two of these are employees at the case study company, while the third is a user of the data. The type of assessment that is selected is the current maturity level assessment, as there is no direct need for a new improvement plan at the case study organization. Finally, after the interviews are conducted, the results are discussed in a joint meeting with the participants from the case study company.

During the second step the levels are considered. However, due to time limitations during the interviews the levels are not actively discussed with the interviewees. During the third step, the relevant capabilities are selected. This results in a unanimous agreement among the interviewees on relevant and irrelevant capabilities. Three capabilities are considered irrelevant for the platform. The first is **Incentivisation**. This capability is considered irrelevant, as the platform does not take any measures to incentivize usage. The main reason for this is that the data is already offered free of charge. Next to that, as indicated by Person I, while they want to encourage everybody that can use the data to use it, their costs increase, but their income stays level when usage increases. This gives them little incentive to actively incentivize the use of the data beyond for cases that are already deemed useful. The second is **Privacy**. The reason to disregard this capability is that the platform does not deal with any personal data. Therefore, they do not have to take privacy measures. This is agreed upon by all interviewees. The third irrelevant capability is the **Pricing Mechanism**. This capability is ignored, because the platform does not monetize their data and, therefore, have no mechanism for pricing their data through their platform.

The fourth step is the actual assessment of the platform on a practice level. During this step, the current level of each relevant capability is selected. A detailed overview of the practice that were selected for each capability is displayed in Appendix E Table G.1. This table shows the presence for every practice in the platform according to the interviewees. Based on this a current maturity level for each of the capabilities can be determined, which is displayed in Table 10.1.

Several interesting situations concerning the model came to light during this step. First of all, the relevance of practices is not considered in the assessment method of the model. While it is determined that capabilities can be irrelevant, it is also possible for practices to be irrelevant for the platform. An example of this for Platform F is in the security capability. In this capability, all interviewees agree that the first practice is present, the second absent, but the third present again. However, omitting the second option, which describes that users can access the platform using an API Key or username and password, is a conscious choice, as the data is freely accessible and that protection is not necessary. Therefore, stating that the maturity level is 1, because this practice is not met, thereby ignoring the third level, which is present, is not an accurate description of the reality. Therefore, it was decided that interviewees could indicate that practices are irrelevant for their specific context and that these practices are ignored. This decision should, however, only be made sparingly and when clear argumentation is present. A second example, which describes a slightly different situation where this occurs is the seventh and eighth practice of the **monitoring platform** capability. Here the seventh level describes a post-mortum, running a recent snapshot of the data, while the eighth level contains a full second back-up system. When this final level is present, the seventh can be considered irrelevant, as an active back-up is superior to a snapshot. Therefore, this seventh level can be considered irrelevant, when the eighth option is present.

A second point of discussion is that potentially not all practices are known by the practitioners. In this case study this is specifically the case for the external interviewee, Person K, who does not know all the internal dealings at Company F. An example of this is the *Create data stewardship practices* practice in the **monitoring data** capability. Person K does not know if these are present. For this reason, this practice and those with a similar problem are also ignored, in a similar way as the irrelevant ones in the previous paragraph.

The final step is the calculation of the score. This step is performed separate from the participants. The results can be found in Table 10.1. As can be seen in the table, the views of the participants differed

significantly for most focus areas. In general, Person I is the most positive about the platform, followed by Person J and, finally Person K. Some of the differences can be explained by the reactions during the interview. Person K looks at more strictly at some of the practices. An example of this is his reaction on the **linked data** practice. Person I and J agree that this practice is implemented in the platform. Person K, on the other hand, argues that, while there are some tests being run with this, it is not omnipresent on the platform and, therefore, he finds that the practice should not be considered implemented. Furthermore, the differences between Person I and J are surprising, as both are from Company F. Partly, this difference can potentially be explained by the different roles of Person I and J. Person I is in his current role less involved with the platform, while Person J is more actively involved. This may have caused Person I to have a less clear view of the platform, filling in some gaps that may not actually have been filled in the platform. Additionally, during this interview, there was little time left for the assessment, this may have resulted in Person I missing some of the mid-level practices that may be absent, while latter ones are present.

10.3.2 METHOD

The method that is used is already described in the previous subsection. Feedback on this method is not given by Person I, as there was no time left to discuss this during the interview. Person J indicated that this method is fitting for the model, as it makes the sizeable model more manageable. Person K also did not have much to add about the method. During the joint meeting the method was further discussed. Both participants did agree with the method as presented in this chapter. However, they did note that some concepts and aspects of the model are relatively hard to understand and that the model should focus on explaining the various aspects of an assessment thoroughly.

10.3.3 MOTIVATION AND SITUATION

The question concerning the motivation and situation for using the model is answered by Person J and K. Person J discusses how the two models could potentially be combined into one. He says that all aspects discussed in both models are relevant, pointing at the view elements that are deemed irrelevant as proof. As feedback he mentions that he is uncertain why the division between the two models are made. After some discussion he agrees that the maturity model is indeed more relevant for creating an improvement plan, while the characteristics are more useful when starting off. Finally, he discusses that the model could be turned into a survey covering the key aspects. This could specifically help when starting out, as, based on the survey, an initial platform architecture could be created.

Person K describes that it helps to determine the maturity of the platform, compared to the broader market. He identifies that level eight should not always be the goal for every platform, because there are some choices you can make that prevent you from reaching level eight. Furthermore, Person K discusses how there can be some contrast or even conflict between some practices on the same level, for example, between pay per use and a self-organizing ecosystem. Finally, he names determining the position in the market or the analysis of competition as concrete use cases for the model.

10.3.4 PERFORMANCE

Finally, the case study participants express their views on the usefulness and understandability of the model. These categories are graded on a Likert scale from 1 to 5. Due to time constraints Person I has not answered these questions. The results are displayed in Table 10.2.

For the *usefulness* Person J describes that it could be useful to connect certain practices across capabilities when they influence each other. As an example he describes that a platform with anonymous access is often coupled with an open license and a self-regulating ecosystem. He describes how it could be relevant to link capabilities and turn them into archetypes. Person K suggests that the model should be extended and should consider the service that is being delivered. He describes that capabilities are

| | Person I | Person J | Person K | Average |
|-------------------------------------|-------------|------------|------------|--------------|
| Platform Maintenance | 86% | 61% | 54% | 71% |
| Security | 4 | 2 | 3 | 3 |
| Documentation | 7 | 6 | 4 | 6 |
| Platform Maintenance | 8 | 4 | 8 | 8 |
| Decision Making Process | 6 | 2 | 4 | 4 |
| Monitoring Data | 7 | 8 | 5 | 7 |
| Monitoring Platform | 8 | 8 | 3 | 8 |
| Platform Maintenance | 8 | 4 | 3 | 4 |
| Data Related | 96% | 46% | 33% | 46% |
| Data Quality | 8 | 3 | 3 | 3 |
| Data Standardization | 8 | 8 | 5 | 8 |
| Data Use Case | 7 | 0 | 0 | 0 |
| Partner Related | 54% | 45% | 45% | 50% |
| Partner Relationship | 1 | 5 | 5 | 5 |
| Incentivisation | N/A | N/A | N/A | N/A |
| Coordination | 4 | 0 | 1 | 1 |
| Onboarding | 8 | 6 | 5 | 6 |
| Compliance | 88% | 38% | 63% | 62,5% |
| Compliance and External Regulations | 6 | 2 | 6 | 6 |
| Internal Regulation | 8 | 4 | 4 | 4 |
| Privacy | N/A | N/A | N/A | N/A |
| Platform Related | 85% | 88% | 53% | 80% |
| Type of Data Sharing | 8 | 8 | 8 | 8 |
| Governance | 8 | 6 | 3 | 6 |
| Control over Data | 3 | 6 | 3 | 3 |
| Data Sharing Technology | 7 | 7 | 5 | 7 |
| Data Processing | 8 | 8 | 2 | 8 |
| Additional Services | 100% | 50% | 50% | 50% |
| Search Functionality | 8 | 4 | 4 | 4 |
| Pricing Mechanism | N/A | N/A | N/A | N/A |

TABLE 10.1: Results Maturity Assessment Case Study

| | Usefulness | Understandability |
|----------|------------|-------------------|
| Person J | 4 | 4 |
| Person K | 3 | 3 |
| Average | 3,5 | 3,5 |

TABLE 10.2: *The performance scores for the Maturity Model*

more concerned with dissemination of data, while others are more focused on working with APIs etc. These differences could be extended upon more, however, he acknowledges that doing this risks creating a model that is extremely large.

For the *understandability*, Person J indicates that sometimes the levels do not represent the quality of the platform. Decisions to stay at the lower levels now provoke the feeling that the platform is limited. However, this is not necessarily the case, as this can be a conscious decision. Person K grades this a 3, without giving any context.

| Focus Area | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Level 6 | Level 7 | Level 8 |
|----------------------|--|---|--|--|---|---|---|---|
| Security | Allow access without authentication by anonymous users | Perform authentication through username password / API key | Use a trusted supplier for platform security | Perform user-based API authorization | Implement use case specific, best-practice authentication methods | Deploy API data through Swagger or similar platforms | Connect Identity Access Management and Authorization system | Implement a use case specific state-of-the-art security system |
| | Provide no documentation | Provide document with basic API description and metadata | List types of datasets | Support structural metadata | Develop Platform with documentation and API descriptions | Swagger or similar platforms | Create assistance tools for API | Provide living documentation |
| | Maintain original platform setup | Establish decision making structures within an organization | Add additional datasets | Manage versioning of the API | Model dependencies in data | Install data mapping tables for data types | Manage versioning of the API | Establish grace period for updates |
| | | Provide opportunities to stakeholders for monitoring data | Create internal process chart for establishing decision rights | Create transparent roadmap | Create standards, conventions and guidelines for platform development | Create alignment processes for widespread involvement from users | Create a governance group to guide product development | Partition decision rights between owner and users |
| Platform Management | | Implement heartbeat test for platform users | Measure platform uptime | Create data stewardship practices | Offer the push of updates on data mutations that shows issues | Create strict monitoring processes, including reporting by users | Create a post-mortem that can replace the system with a snapshot when the service is down | Establish a visible supply chain for the data |
| | | Identify API metrics | Set usage quotas | Warnings and eventually blocking for quota breaking | Throttle abuse of data quotas and make a fair use assessment | | Flexible fair use agreement based on user/datetime | |
| | | Keep data up to date | Make reliability vs. costs trade-off | Check and indicate missing data | Promote data standard usability costs | | Install data validation mechanisms | Perform reconciliation |
| | | Let IT determine data structures | Allow provider to determine data attributes | | | Support data integration with related datasets | Format data according to users' needs | Create flexible data structures |
| Data Management | | Audit every data use case and its processes | Adapt platform to specific use case | Set usage goal for data | Define and link the use cases and the relevant stakeholders of the data | Discuss data usage goal in terms of service | Create new advanced products with data | Monitor if the data use case is met by users and punish misuse |
| | | Perform user screening | Create feedback loops | Ensure continuous collaboration with users | Create working groups and review committees to support users | | Create new avenues and means for cooperation for stakeholders | Create review systems for data suppliers |
| | | Offer non-monetary subsidies to early users | Provide seeding (input to marketplace) | Facilitate knowledge sharing on supply and needs | Create incentive system for users | | Create performance-based rewards schemes for users | Set up revenue sharing or other incentivisation scheme for data supplier partners |
| | | Create a mutual coordination plan for connecting supply and demand | Create communication guidelines between collaborators | Regulate governance sharing capabilities | Regulate collaborations | Set up conflict resolution process | Determine level of gamification | Partition decision rights for data suppliers |
| Partner Related | | Carefully consider communication of open data strategy to potential customers | Create sandbox mode | Create self-service capabilities | Create a showcase section | Offer data playbooks, tutorials and describe data sharing technique | Create scalable onboarding for new users | Allow piloting |
| | | Ensure compliance with applicable regulation | Appoint a data steward | Regulate intellectual property | | Get ISO certified | Set-up use processes for data sharing | Create Responsible Disclosure processes |
| | | Create terms of service and code of conduct | Educate end users on privacy permissions | | Determine punishments for breaches | | Negotiate exchange terms with users | Allow different forms of consent establishment for personal data sharing |
| | | Establish a data exchange | Create contracts for data use | Promote governance standards | Localize licensing conditions | Use premium licensing | Enable data pooling | Use open licensing |
| Compliance | | Do only technical maintenance | Create governance framework | Describe data ownership in platform | | Regulate behaviour through reputational measures | Allow the ecosystem to self-regulate | |
| | | Make individual organization retain control over data | Make individual organization retain control over data | Share data via an open data marketplace | Share data via an actual platform | Support context-based ownership | | |
| | | Share data via excel or csv | Share data on company website | Share data via an actual platform | Implement a review board for control | | | Allow uploading of datasets by users |
| | | Share Snapshot of data | Process data before sharing | Encourage high-quality resources and repress poor-quality resources of platform ecosystems | Share linked data | | Share humanreadable data | |
| Ecosystem Management | | Provide search on metadata level | Allow search on content level | Recommend datasets | | | Assist connecting database by implementing pathfinding | |
| | | Enable basic payment for access | Determine ideal business model for platform context | | | | Create congestion pricing for popular data | Create spot market for short transactions |
| | | | | | | | | |
| | | | | | | | | |
| Additional Services | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |
| | | | | | | | | |

FIGURE 10.1: The Final Version of the DaAP Maturity Model

| Focus Area | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Level 6 | Level 7 | Level 8 |
|----------------------|------------------------------------|---|--|---|---|--|---|---|
| Platform Management | Security | Allow access without authentication by anonymous users | Use a trusted supplier for platform security | Perform user-based API authorization | Implement use case specific best-practice authentication methods | Deploy API data through Swagger or similar platforms | Connect Identity Access Management and authorization system | Implement a use case specific state-of-the-art security system |
| | Documentation | Provide no documentation | List types of datasets | Support structural metadata | Implement a complete Developer Platform with documentation and API descriptions | | Create assistance tools for API | Provide living documentation |
| | Platform Maintenance | Maintain original platform setup | Add additional datasets | Manage versioning of the API | Model dependencies in data | | Manage versioning of the data | Establish grace period for updates |
| | Decision Making Process | Establish decision making structures | Create internal process chart for establishing decision rights | Create transparent roadmap | Create standards, conventions and guidelines | | Create a governance group to guide product development | Partition decision rights between owner and users |
| | Monitoring Data | Provide opportunities to stakeholders for monitoring data | Measure platform uptime | Create data stewardship practices | Offer the push of updates on data mutations | | Create alignment processes for widespread product development | Establish a visible supply chain for the data |
| | Monitoring Platform | Implement heartbeat test for platform users | Set usage quotas | Create alerts for users when the platform is down | Create a status website that shows issues | | Create a post-mortem that can replace the system with a snapshot when the service is down | Run two versions of the platform, where one takes over when the other crashes |
| | Platform Measurements | Identify API metrics | Warnings and eventually blocking for quota breaking | Throttle abuse of data quotas and make a fair use assessment | | | Flexible fair use agreement based on user/datetime | |
| | Data Quality | Keep data up to date | Check and indicate missing data | Make trade-offs between amount of data and usability costs | | | Install data validation mechanisms | Perform reconciliation |
| | Data Standardization | Let IT determine data structures | Allow provider to determine data attributes | Promote data standard governance | | | Format data according to users' needs | Create flexible data structures |
| | Data Use Case | Audit every data use case and its processes | Adapt platform to specific use case attributes | Define and link the use cases and the relevant stakeholders of the data | | | Create new advanced products with data | Monitor if the data use case is met by users and punish misuse |
| Data Management | Partner Relationship | Maintain no relationship with users | Create feedback loops | Ensure continuous collaboration with users | Create working groups and review committees to support users | | Create review systems for data suppliers | |
| | Incentivisation | Offer non-monetary subsidies to early users | Provide seeding (input to one side of the marketplace) | | Create incentive system for users | | Create behavioral rewards for users i.e. through gamification | Set up revenue sharing or other incentivisation scheme for data supplier partners |
| Partner Related | Coordination | Take 'take it or leave it' stance towards connecting providers and users | Create communication guidelines between collaborators | Facilitate knowledge sharing on supply and needs | Regulate collaborations | | Determine level of centralization of the platform | |
| | Onboarding | Carefully consider communication of open data strategy to potential customers | Create sandbox mode | Create self-service capabilities | Create a showcase section | | Create scalable onboarding for new users | Allow piloting |
| Compliance | Compliance and External Regulation | Ensure compliance with applicable regulation | Appoint a data steward | Regulate intellectual property | | | Set-up due processes for data sharing | Create Responsible Disclosure processes |
| | Internal Regulation | Create terms of service and code of conduct | | | | | Negotiate exchange terms with users | |
| Ecosystem Management | Privacy | Ensure complete GDPR compliance | Educate end users on privacy permissions | | Determine punishments for breaches | | Create guardrails to support privacy protection | Allow different forms of consent establishment for personal data sharing |
| | Type of Data Sharing Governance | Establish a data exchange | Create contracts for data use | Promote governance standards | Localize licensing conditions | | Enable data pooling | Use open licensing |
| Additional Services | Control over Data | Exert no control over data | Make individual organization retain control over data | Describe data ownership in various scenarios in the platform | | | Regulate behaviour through reputational measures | Allow the ecosystem to self-regulate |
| | Data Sharing Technology | Share data via excel or csv | Share data on company website | Implement a review board for control | Share linked data | | Support context-based ownership | Allow uploading of datasets by users |
| Additional Services | Data Processing | Share snapshot of data | Process data before sharing | Implement a review board for control | Recommend datasets | | Share human-readable data | |
| | Search Functionality | Provide search on metadata level | Allow search on content level | Allow for filtering search results | | | Assist connecting database by implementing pathfinding | Create spot market for short transactions |
| Additional Services | Pricing Mechanism | Enable basic payment for access | Determine ideal business model for platform context | | | | Create congestion pricing for popular data | |

FIGURE 10.2: The resulting maturity model from the case study assessment. The current level for the ca-



DISCUSSION

In this chapter the findings from this research are discussed. This discussion describes how the sub-questions are answered using the various results gathered in this research. Furthermore, it describes how, using the answers to the sub-questions, the main question of this research can be answered. Moreover, it lays out the limitations of this research and some potential threats to its validity. Finally, some steps for future research on this topic are discussed.

11.1 RESULTS

This section discusses all the results from this research. Firstly, it describes how the sub-questions are answered throughout this research. Secondly, the main research question is answered by combining the answers of the sub-questions.

11.1.1 DEFINITION

The first sub-question of this research looks into a definition for the DaaP topic. As the concept of DaaPs is newly introduced to academic literature in this work, a definition for the topic can give researchers a starting point for the topic to base future research on. Moreover, it is used to strictly scope the topic to data platforms that do and do not fall into the DaaP topic.

The definition is created in two phases. The first phase is the SLR phase. During this phase, input on a definition is gathered in academic literature and, subsequently, used to create an initial definition for the DaaP concept. In the SLR, related definitions are identified and illustrated in a table. Subsequently, aspects relevant to DaaP are reused in the DaaP definition and their design and phrasing is used as inspiration for the DaaP definition. Finally, any missing aspects that are key to the DaaP concept are added to the definition, based on the opinion from the author. Thereafter, the second phase is entered, where the definition is validated in interviews. Here, the interviewees are asked to give their feedback on the definition. This feedback is again illustrated in a table. Where several interviewees had feedback on the same point, the feedback was applied on the definition. Where a point of feedback was only given by a single interviewee, the feedback was assessed by the author, to ensure that it corresponded to the DaaP topic as envisioned.

This process has led to a final definition for DaaP: *”Data as a Platform is a data sharing platform on which one or more organizations monetize their data for use by third parties. This central data platform is supported by an autonomous business model.”* This definition covers the key aspects of a DaaP. First of all, it introduces DaaP as a data sharing platform. Specifically, it is a description of a subset of all data platforms, with some specific requirements.

These requirements are discussed in the remaining part of the definition. The first requirement is that it is a commercial platform. Therefore, it is specified that organizations monetize their data on the platform. This specification requires that organizations should receive some monetary benefits from sharing their data on the platform, detailing a commercial element for the platform. The commercial requirement is reiterated in the second sentence, but now focusing on the platform itself. This sentence specifies that the platform has to have an “autonomous” business model. This requires that not only the providers of the data should have some monetary benefit, but that the platform itself should also be able to support itself with its business model. Of course, this can go hand in hand, when the data provider and platform operator are the same organization, but this is not necessarily the case. Furthermore, there is the “for use by third parties” part. As interviewee 1 rightfully pointed out, in theory you could have an internal data platform at a company that still has its own business model, for example when usage of the platform is billed to the department that uses it. The specification that the data is shared “for use by third parties” details that the data is shared for use outside of the organization that shares it. It, however, does not specify the type of user. It could be that the data is used by other organizations, but it can also be used by consumers, for example.

Finally, there is a final element of the definition, which is not particularly a requirement. That is the “one or more organization” part, that mentions that data on the platform can come both from a big group of data providers and from a single company. This does not particularly exclude many data platforms, but it does elaborate that platforms do not exclusively have to belong to a single sharing and operating company, something that may not be directly apparent from reading a definition without this clarification.

This process has resulted in a validated definition, that is based on both scientific and practical insights. The method used here has a lot in common with a grounded theory approach to defining a concept. First of all, an inductive approach is taken during the literature review, which is characteristic for grounded theory. Wolfswinkel, Furtmueller and Wilderom [Wolfswinkel et al., 2013] describe a grounded theory approach to a systematic literature review. This method takes a similar approach for reviewing literature as used in this research. It suggests selecting papers based on general selection criteria. Subsequently, the selected papers should be coded to collect the insights. This step is also taken in this research, with the alteration that the papers are not coded in a fully ‘open’ way, but rather with some starting definition to speed up the process. The results are subsequently synthesized with as little bias as possible to create the initial definition.

This final definition is used for the rest of this research. Moreover, it is used during the Case Study to explain the topic of DaaP to the participants. Furthermore, it is the answer to the first sub-question, given that the perspective for the definition is changed from an ecosystem perspective to a focus on the platform itself. The definition can also be used by researchers in future work to expand research in this new field, by researching more platforms that fall in this definition.

11.1.2 CHARACTERISTICS

The second sub-question looks into the characteristics a DaaP can have. Characteristics in this context specifically refers to characteristics in which a DaaP can differ and make a choice. To reiterate the definition used during the SLR, a DaaP characteristic is: *”An aspect of a Data as a Platform platform, or related data sharing technology, that describes a key part of that technology. A characteristic belongs to a category that can have several variations that differentiate various implementations of these technologies.”* The result envisioned from this sub-question and definition is an overview of characteristic topics, or categories, which have several variations, the characteristics. A platform should, at least when starting out, be able to consider which of the characteristics they want their platform to have.

This overview is created and validated in three steps. The first step is the SLR, where a collection of characteristics is coded in the selected papers. These characteristics are then synthesized in an initial model. This initial model is the starting point for step 2. During this step, the model is discussed in interviews. Due to the limited time during the interviews, however, the overview was only discussed during a single interview. Based on this feedback and a second review of literature, a final overview of characteristics is created. This overview is used as input for the third phase, where it is applied as a part of an assessment at a Case Study company. During this Case Study, the model is applied at this company and the method for using it is determined. The content around this overview is finalized with a final description of the overview and the method that can be used by practitioners wanting to use the characteristics.

This has led to the model that is displayed in table D.3. This model still has significant overlap with the initial model. However, it is mostly reduced in size, as the requirements for a category or characteristics are toughened. The overview can be divided in four main topics. The first is Governance. Governance is about control and how this is regulated in the platform. This topic starts off with the **Coordination Mechanism** category. This category is about how the platform is coordinated. According to literature, there are several parties that could coordinate the platform. This can be done by a single dominant party, this is then often the orchestrator. Secondly, it can be done with minimal coordination or it can be done in collaboration between all or a group of participants. It is mostly on the orchestrator of the platform to determine the coordination. As described by van der Broek and Veenstra [?], less coordination can lead to more innovation, as participants are more free in trying different use cases and activities. On the other hand, a more coordinated platform can be more profitable, both for the orchestrators and the users, and it can allow the orchestrator more insights in how the platform is used, which can be crucial when data is still partly protected. The second element is the direct control over the data. This element mostly refers to who controls the data after the user has accessed it. Van den Broek and Veenstra [?] describe three scenarios. They correspond with the **Coordination Mechanism**, as they also mention: an orchestrator taking full control, no controlling of the data or individual organizations retaining control, something that is relevant when multiple organizations share data on the same platform. A fourth point is added to this from the interviews and that is that the buyer controls the data as soon as it is accessed and that any rights on the data immediately end when it is purchased. The third category is the selection of users. Literature discusses that this can be done in three ways [Susha et al., 2017b]. Firstly, based on applications, where interested parties apply to get access. Secondly, on agreement basis, where a partnership has to be agreed on before access is granted. Thirdly, open, in which case anybody can access the data directly. Each method, again, gives different levels of control to the orchestrator. An application allows orchestrators to pick their users, who, after that, get access within a certain contract. An agreement, on the other hand, gives orchestrators room to have more contractual control over how the data is used and what they expect in return. The final governance element is **Level of Centralization**. In this category, a platform can either be centralized or decentralized. This category is more of a scale than a direct decision and some balance is important, however, platform mostly lean one way or the other [Fisher and Davies, 2016]. In a decentralized platform the control for the platform lays at the data provider. They, for example, control the documentation and maybe even some marketing or communication for the dataset. In a centralized platform, these responsibilities are for the platform provider.

The second topic is the platform itself. This starts off with the **specificity** of the platform. This describes the use cases of the data on the platform. It gives two options, either the data can be used universally for any purpose, or there is a specific purpose for which the data is to be used. This topic returns later in the maturity model, where it is further worked out in the "Use Case" Focus Area. The second category is the **Openness**. This level, in practice, is somewhere between the governance and the platform topic. It describes how open the platform is for new users. Whether it is restricted and only a specific group of users is allowed to access the platform, or that it is open and can be accessed by anybody. A further description of openness in software ecosystems and its potential consequences is given by Knauss et al. [Knauss et al., 2014]. The third category is **Diversity of Data Providers**, these

characteristic describes the range of data sources on the platform. This aspect was introduced by Susha et al. [Susha et al., 2017b], who name three options for an orchestrator of a platform. These options are, only sharing their own data, finding data providers within the same industry or field and, thirdly, having a wide range of data providers without any necessary overlapping factor. A more specific platform targeted to a industry can have the benefit in targeting users, while an open platform has a bigger pool to attract users from and, therefore, more potential to grow larger. The fourth category is the **location of the data**. The data can of course be stored directly on the company's website, often accompanied by a tool that gives access. It can, however, also be placed on a full-blown platform, with, for example, documentation and registration functionalities. Next to that, it is also possible that the data is stored on a marketplace, where direct access is sold in a webshop environment. As discussed by Case Study participant 2, in practice, a platform is preferable, especially for developers using the data, but for some use cases one of the other options could be more fitting. Then comes the **Continuity of Data Sharing**, which describes when data is shared. Susha et al. describe that, while usually data is shared continuously on a platform, it could be useful to only share data when certain events happen or only on demand, as it gives more control to the operating company. Subsequently, the **degree of integration** is discussed. This refers to how integrated a data platform is within a specific industry. It can either be very industry specific, which is named *intra-industry* or it can be useful across industries, referred to as *inter-industry*. Finally, in this topic, there is multiplicity, which refers to the way the data is shared, this can be one-to-many, with a single provider, or many-to-many, with several data providers.

The next topic is related to the orchestrator. The first element here is the **platform ownership**. This discusses the creator and owner mechanic of the platform. It can either be created and owned by a single party or be orchestrated by one party, while the technical execution lays at another party. Again, the first option allows for more control, but not every company has the capabilities to achieve this. A second category is the **Ground for data sharing**. This describes what the purpose is of sharing the data for the orchestrator. This can be a very specific single purpose, a more generic purpose or no purpose at all. Depending on is, the data is likely to be useful for a larger group of users [?]. Finally, on this topic, there is the **Motivation for starting the platform**. This category, based on the research by Susha et al. [Susha and Janssen, 2017], focuses on the motivation of the orchestrator for sharing data and gives several reasons for sharing data.

The final topic is the smallest, and it is concerned with the data itself. The first category in this topic is the type of data. This topic was deemed important by interviewee 1. It has several options, the data can be *processed*, meaning that, compared to the database it is stored in, it is made user friendly for its users before it is accessed. A second option is that is shared raw, meaning it is shared in a similar way as it is stored. This option was deemed very immature and ineffective by interviewee 2. A final option is that it is enriched, meaning that is combined already with several data sources to make it as useful as possible for users. The final category is privacy related and is concerned with **Personal Data**. It very simply divides platforms where personal data is shared and platforms where no personal data is shared.

This overview has some similarity with, for example, that of [Susha and Janssen, 2017], whose taxonomy is focused on data collaboratives and Cross-Sector Partnerships. Their characteristics are, however, exclusively focused on data sharing for a social goal, which significantly differentiates it from this research. Regardless, some of the categories and characteristics remain relevant and are, therefore, re-used in this overview. Moreover, the overview lacks validation. It is mostly based on literature, as it is only discussed by one interviewee and a literature review is still performed after that interview. This means that there is little to no confirmation from practitioners that these are actual crucial factors for DaaPs. This, however, is also the crucial point. Most categories and characteristics come from research papers that did involve some Case Study or interview phase. However, these phases were usually not on the DaaP topic, for example the work by Susha et al. [Susha and Janssen, 2017] was on Data Collaboratives. Therefore, the validity problem of this result is mostly concerned with its external validity, as the results cannot be generalized to DaaPs. The overall correctness of most categories has been tested in their respective works, their applicability to DaaPs is however not sufficiently confirmed. The subsequent Case Study, where the

overview is applied and used by the participants, does give some validity to the work. However, this part of the research was not specifically shaped to confirm the validity of the model. The Object of Study in the Case Study is more of an Open Governmental Data Platform and not a DaaP and this study, therefore, still cannot confirm the applicability of the characteristics in this context. This problem runs into the representative sampling external validity threat, as described by Wieringa [Wieringa, 2014].

11.1.3 MATURITY MODEL

The third and fourth research sub-questions are answered with the maturity model. The third question looks at general capabilities relevant for DaaPs. Question 4, thereafter, places these capabilities in the maturity model. This Maturity Model is the main contribution from this research. It can be used broadly by practitioners to create, assess and improve the DaaP. Moreover, researchers can use it to identify topics within the DaaP research field for future work. In contrast to the previous sections, a discussion on the model itself and the corresponding method can already be found in chapter 10.2.

11.1.4 ANSWERING THE MAIN QUESTION

Throughout these questions DaaPs are introduced and further formalized in academic literature. This subsection looks to reflect on how the main question is answered with these results. The main question in this research is: *"How can DaaP platforms be utilized by organizations to unlock the value of their data?"* Partly, this question is immediately answered in the definition. DaaPs are defined as a data platform on which companies can monetize their data. Therefore, unlocking monetary value from the data. The remaining research questions focus on how this value can be generated effectively by such a platform. This has resulted in the characteristics and maturity model. In the basis, these models that there is no universal solution for capturing value through a DaaP. However, these models offers handles on how to start and develop a platform. Through iterative development an operating company of a DaaP can use the models to determine how enable more value through the DaaP platform.

As noted by the case study participants, the characteristics can help organizations when building a DaaP. The overview should be assessed to determine how the platform should operate and this could potentially be done in conversation with both data providers and users. Next to that, the Data as a Platform Focus Area Maturity Model can be used to then develop to maturity in those capabilities that are deemed crucial by the users of the platform. Together these materials form a practical answer to the main questions. As the materials when applied to the context of a data platform can help determine what is required for a data platform on which companies can monetize their data, unlocking additional value from it.

11.2 VALIDITY

As any other research, this research is not free from threats to its validity. As a foundation for the analysis of this research the Experimentation in Software Engineering book by Wohlin et al. is used [Wohlin et al., 2012]. In short, this book describes four types of validity: construct, internal, external and conclusion validity. Each type has its threats that could harm the validity of the research. This section describes the validity threats that are relevant for this research. An overview of each validity threat and in what part of the research is displayed in figure 11.1. The specifics of each type of threat is discussed in the following paragraphs in this section.

11.2.1 INTERNAL VALIDITY

The internal validity of the research is concerned with whether the results from the research are credible to the participants in the research. Trochim [Trochim, 2006] also calls this the credibility of the research.

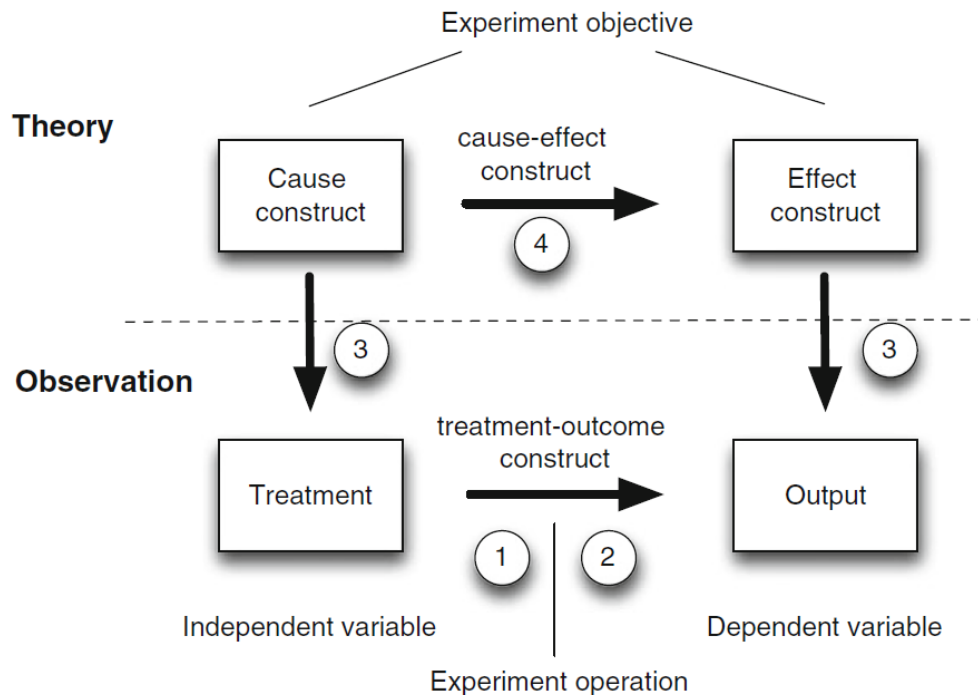


FIGURE 11.1: *Validity Threats according to Wohlin et al. [Wohlin et al., 2012]*

The goal of a qualitative research, like this one, is to create results that are aligned with the view of the participants of the research.

In this research there are some clear threats to this type of validity. First of all, the interviewees are asked to give input on the characteristics and Maturity Model that were already prepared in the SLR. This can significantly have hindered the extend to which the final models reflected the combined opinions of the participants. Asking feedback on the materials forced the participants to focus on the aspects that were already found and reduced their opportunities to give an overall description of their view on the topic. Nevertheless, there were several reasons for taking this approach. First of all, 8 interviews were conducted in the context of this research, this is a relatively limited number, which can be explained by the limited availability of DaaPs in practice. This, however, would mean that a model that was created from scratch through interviews would only have input from 8 participants, who likely would not be able to cover every aspect of the topic. This could either be because of the limited length of each interview or a common lack of expertise on some aspect of DaaPs. Moreover, for the Maturity Model, it is nearly impossible for interviewees to know every practice in a focus area from the top of their heads, risking a lack of completeness of the model. A second argument to involve the literature in an initial model is the large amount of literature that is available. A lot of research is already available on related fields, which can give a lot of additional insights next to the interview results.

While a good argument can be made to accept this threat, several steps are still taken to mitigate this risk. First of all, the requirement is added that each focus area and practice needs to be confirmed by at least 2 interviewees, before it is added to model. This ensures that there is at least agreement from practitioners on the contents. Secondly, the interviewees are asked for additional input throughout the discussion of the model on the topics they appear to have specific insights in. This ensures that the interviewees do not get caught in checking the current model, but also give input on their expertise enriching the model with their knowledge.

A second threat to the internal validity is a lack of completeness in the SLR. As described by Zhou et al. [Zhou et al., 2017], one of the most common validity threats during literature review are incomplete or inappropriate search terms. This threat is present in this research to. As the author is not an expert on the topic and the literature review was conducted in the initial phases of the research, there is some threat that the list of search terms was incomplete or that some items were incorrect. Therefore, the literature review, while conducted systematically, may not be exhaustive. Attempts are made to mitigate these threats by doing a probing search. During this search, initial papers on the topic are read to get a grasp of the terminology. This helps with forming correct search terms for the actual literature review. Moreover, search terms were updated during the review, especially when a search term yielded limited results. Finally, the search terms were validated with several more experienced researchers to validate their correctness.

11.2.2 EXTERNAL VALIDITY

The second validity threat described by Wohlin et al. [Wohlin et al., 2012] is the external validity. This threat is concerned specifically with the generalizability of the research. For example, having an interview sample group that does not represent the overall population. For each of the phases of this research external validity threats can be identified.

Wieringa [Wieringa, 2014] describes several topics that could introduce external validity threats. The first is Object of Study itself. As described by Wieringa, the Objects of Study should fit the "population predicates". This point is not directly applicable to this research, as the Object of Study is not directly observed, but questions are asked to experts on these Objects of Studies. Regardless, the background from the interviewees brings in some risk for this type of external validity threats. First of all, the interviewees are not all working with DaaPs, some should be considered in the category of Open Governmental Data and others, namely interviews 5 and 7, can not yet be seen as a data platform. This threat is, however, reduced by explaining the topic extensively before the interview and asking them to apply their expertise to the DaaP topic. Moreover, for the Maturity Model, it is unlikely that an interviewee was an expert in all the different Focus Areas identified for this topic. In general, this threat was mitigated by the interviewees themselves, as they indicated which topics they were less familiar with and these were mostly skipped. Moreover, as any suggestion had to be confirmed by multiple interviewees, multiple participants had to wrongly agree with something for this to become a problem.

A second external validity threat comes during the Case Study. The platform around which the Case Study was framed cannot be considered a classical DaaP. This platform is a clear example of an Open Governmental Data platform, as the sharing of their data is ordered through law. However, they do have a business model supporting the platform. Revenue in this business model is created by opening up the data from other parties that have similar responsibilities for opening up their data, instead of from the data. This is fairly unique situation that is not excluded by the DaaP definition. However, it cannot be seen as a straightforward business model for many organizations creating a DaaP. Therefore, it cannot be seen as a perfect representation of the larger population. In general this is a problem with Case Studies, as their small sample size often prevent them from representing an entire population. However, this is worsened when a case is selected that is not representative for a big part of the population. Partly, however, this risk had to be accepted as the limited availability of DaaPs made it hard to find a more fitting case. Moreover, attempts are made to mitigate it, by discussing the DaaP topic beforehand with the interviewees. They are asked to indicate where they think that traditional DaaPs may deviate from their situation. This information is used when processing the results.

11.2.3 CONSTRUCT VALIDITY

A third validity threat identified by Wohlin et al. [Wohlin et al., 2012] is the construct validity. This type has to do with the validity of the constructs of the research. In short, this discusses how well defined various aspects of the research are. Construct validity threats can harm the repeatability of the research,

as unclear definitions can lead to different outcomes when applied by different researchers. Moreover, it harms the ability to generalize the results of a single part of the research to the underlying artefact. In this research, for example, if an interviewee misunderstood a focus area in the maturity model, this person's feedback should not be considered correct input for the model.

First of all, during the SLR, as papers were coded based on an initial definition, there is a potential for construct validity threats, as described by Wieringa [Wieringa, 2014]. When constructs are not adequately defined, there is a threat to the validity, as it leaves room for personal interpretation and, therefore, reduces the repeatability of the research. In this research, what definitions are looked for in literature was defined before the coding process as: "A definition of a Data as a Platform platform, or related concept, or a definition from a related data sharing technology concept." However, this definition does not completely take away all threats of bias in the coding for the definition. First of all, as the DaaP concept was not defined in this part of the research, this part of the definition was vague at that point in time. This validity problem is mostly minimized, however, as there was already a relatively concrete idea what this entailed described in the introduction and background section of this research. This input was used to determine what entailed a DaaP platform or related concept. Moreover, the phrases "DaaP related concept" and "related data sharing technology" are quite freely interpretable. This was, however, inevitable, as the goal was to find a broad range of definitions. The severity of this threat was further reduced by the final detailing of "data sharing technology", which gives more direction to the type of definitions being surveyed. A similar argument can be made for the other definitions used in the SLR.

Moreover, during the interview similar problems occurred. As not all interviewees were familiar with Maturity Models, participants needed to be informed to a differing extent. During some interviews, this yielded situations where interviewees were asking details about the model, while parts were already discussed. This misunderstanding may have influenced earlier decisions feedback. Attempts were made to mitigate these risks. Before the model was analyzed by the interviewees, a description was given about how it worked, why the model was made and how it had evolved to this point. This explanation was further shaped by the experience with maturity models, as described by the interviewees. However, as Focus Area Maturity Models are relatively unknown under practitioners, several core aspects were discussed every interview.

Next to this, the maturity model itself too consisted of many constructs that needed a clear definition for them to be understandable by the interviewees. The need for these definitions was further reiterated by interviewee 1, as they were not present at that time. This feedback was later considered and definitions were created for all characteristics, focus areas and practices in their respective materials. These definitions were afterwards used to mitigate the threat of misunderstanding constructs. However, this could not be done completely. Not all constructs in the materials had definitions available that could be used, both the ones in literature, as the ones identified during the interviews did not always have clear definitions. This meant that definitions for these concepts were created by the authors. As the input for these definitions was often limited, this could have resulted in incorrect or incomplete definitions, potentially influencing the subsequent feedback from the interviewee. This threat mostly had to be accepted, due to the limited availability. In a rare instance where a description of a construct by an interviewee sounded incorrect his opinion could have been considered influenced and therefore his feedback could have been ignored. However, usually interviewees were open to indicate this themselves and appropriate steps could be taken from there.

11.2.4 CONCLUSION VALIDITY

The final validity threat discussed by Wohlin et al. [Wohlin et al., 2012] is the conclusion validity. This validity threat has more relevance for quantitative research. It is concerned with the correct application of statistics and the correctness of the results that are concluded from them. During this research no statistics was used and, therefore, a proper conclusion validity, in a quantitative sense, cannot be achieved. Regardless the method described for updating the Maturity Model was designed to give some conclusion validity to the results. The method, where focus areas and practices were only accepted after being

confirmed twice and by more than 50% of the interviewees, mitigated some of this threat.

However, giving statistical significance to these results was not feasible for several reasons. First of all, the research was exploratory in nature, introducing a new field in the form of DaaPs. In general, exploratory research does not lend itself for statistical analysis. Moreover, the materials introduced in this research were unvalidated and incomplete and one of the main goals of the interviews was to extend them with insights from the interviews. Quantitative research does not lend itself for these type of evolutions either. Finally, there are only a limited number of DaaPs available. This, in general, makes it hard to perform research that can be seen as statistically significant, as the sample size is always going to be small.

11.3 LIMITATION

Naturally, some limitations can be identified for this research. These limitations do not necessarily reduce the quality of the research. However, they are elements that, especially in hindsight, could have been done better or more efficient.

The first limitation is one that is relevant for many researches. Preferably, more actual DaaP platform organizations and experts would have participated in this research. In this research, four interviews can be considered to be with actual experts on DaaPs. The other interviews were with experts on related topics. As one of the targets of the interviews was to validate the relevance of some of the topics and concepts found in related literature for the field of DaaPs, a limited number of interviewees limits the reliability of this validation. Next to that, the limited number of interviewees also limited the exposure of different topics relevant for DaaPs. In the maturity model, many concepts have been described that were found relevant for DaaPs by the interviewees, however, having more experts on this topic may have brought to light more relevant topics. However, due to the limited presence of DaaPs in general and the limited time available in this research, more participants could not be interviewed.

A second limitation is that, for the case study, it would have helped to apply the materials to a more "middle of the line" DaaP or even multiple DaaPs. This would have increased the validation of the method and may have given additional insights on the motivations for using the materials. The platform that was used for the case study can be considered a DaaP, however, they use a business model that is not available to many organizations. Therefore, the generalization of the results of the case study to the broader population of DaaPs can only be seen as limited. Having a DaaP with a more traditional business model would have given additional reliability to the case study. Next to that, multiple DaaPs would have given even more validity to the case study. As that would have allowed the comparison between the results that could be made with an assessment and how they can give additional insights.

A third and final limitation is the limited exposure for the actual "value capturing" element, that could be seen as relevant for DaaPs. The specific research question for this work looks at unlocking value and does not specify for whom the value should be unlocked. However, the definition points at the fact that data providers should be able to 'monetize' their data. Directly pointing at a way to capture monetary value. This specific aspect has been discussed with interviewees, but these conversations did not result in concrete recommendations for monetizing data. Naturally, this is a very context-specific topic, as it is dependent on, for example, the market and the data available. Nevertheless, creating a business model was deemed an defining characteristic and this research only to a limited extend discusses this topic.

11.4 FUTURE RESEARCH

Finally, this research gives some interesting insights and opens ways for new research. This section starts off by describing how this research could be continued in new works. This work has some limitations and opens some interesting continuations that other researchers could look at. Thereafter, it describes some more long-term research directions this work could lead to.

As a first step in future research, the direct work of this research could be extended. First of all, the materials that have been created can be extended. As indicated by one of the case study participants, the maturity model does lack detail in some topics. For example, the security topic can be divided in several topics, such as authentication, authorization and platform security. Future work could look at missing topics and extend the model. Next to that, several interviewees suggested other specification in the models. For the maturity model, more overarching topics could be defined that could be a focus point for organizations. An example of this could be flexibility, maturing some capabilities will increase the flexibility of the platform, however, this is not indicated in the model. Finding these cross-capability topics can significantly increase the usefulness of the model. Furthermore, more research could be done into data platform archetypes. Within DaaPs, it is likely that various types can be identified. Finding these types and applying the materials to their specific environments can help specify the models and increase the ease of use. This can, for example, be done by identifying the capabilities that are crucial for the different DaaP types that are found.

Next to the materials created in this research, there are more perspectives that could be applied to DaaPs. A starting point could be looking at the commercial aspect of DaaPs. A research could be done into the business models that can be used by DaaPs. For example, a decision making framework can be drafted that helps organizations determine the best business model for their context. Next to that, the ecosystem perspective of DaaPs can be an interesting area for future research. Topics such as the openness, ecosystem health and governance has been researched for other types of ecosystems, but their impact on DaaPs is still unknown. Future work could, for example, create a ecosystem health model similar to the OSEHO created by Jansen [Jansen, 2014]. Finally, research can be done looking at how data influences DaaPs. In the data field, interesting has been done into data quality or data management and how these impact DaaPs can form an interesting source for future research.

For the long term future, several interviewees suggest that they foresee a future where companies are becoming increasingly open to sharing their data with other companies if this can result in increased value. A major reason for opening up data can already be seen in interview 7, where a data platform is created so that more companies can get access to the software market this company made use of. In this industry, there are only a few organizations creating software for the entire industry. Companies in this industry are looking for individual software solutions, because the data is proprietary and companies don't want to share any information with competitors. The platform created in that industry unifies the formatting of the data and publicizes this format. This is setting up for a next step in the industry where non-proprietary data is being shared between organizations. Future research can look at how organizations can share data in a mutually beneficial way. As the value of data is relative to the use case for which it is used, the value of data can differ for companies. Research can be done in how companies can identify data they own that could bring additional value to other organizations and then help determine what this value is and if it is worth sharing and monetizing this data to that other organization. In this way, mutually beneficial DaaPs can be created where one organization receives monetary value by selling the data and the other can increase its products, services or internal processes using the data.



CONCLUSION

This research has resulted a number of findings regarding DaaPs. In this chapter, these findings are summarized per research question. Moreover, the contribution of these results to academic research is discussed. Finally, suggestions are given for future work.

12.1 CONCLUSION

The goal of this research was to answer the main question:

How can DaaP platforms be utilized by organizations to unlock the value of their data?

While this question is quite broad, this research has tried to answer it by creating several artifacts describing DaaPs. These artifacts were introduced in the sub-questions. Input for these artifacts has been gathered through an extensive systematic literature review, after which initial versions have been created. Subsequently, these artifacts are discussed during eight interviews, where experts on the topic of DaaP and from related field give feedback on the materials. These interviews have resulted in final validated artifacts. These final artifacts are, finally, used in a case study for the assessment of an existing DaaP.

12.1.1 DAAP DEFINITION

The first sub-question is concerned with defining a DaaP. As the topic is new to academic literature, a definition would help future researchers and practitioners understand the field better. Moreover, it would help scope the field, giving some clear guidelines of what is and is not DaaP. The final definition is: "Data as a Platform is a data sharing platform on which one or more organizations monetize their data for use by third parties. This central data platform is supported by an autonomous business model." This definition was created through a literature review and validated by the interviewees.

This answer is the first piece of the puzzle for answering the main question. The main contribution from the definition is that it frames the main question and the artefacts that help answer it. Moreover, it emphasizes some of the important points made in the main research question. It focuses on "monetizing"

data, which could be seen as a specification of the "unlock value from data" that is used in the main question. Finally, combining the definition and the main question, the goal of this research becomes about how data platforms can be used by organizations to monetize their data by sharing it with other parties. The rest of the research questions, therefore, aim to answer this question.

12.1.2 DAAP CHARACTERISTICS

The second sub-question is concerned with the characteristics of a DaaP. The goal of this sub-question is twofold. First of all, a study into the characteristics of such a platform is used gain an increased understanding of the topic. The characteristics can be seen as an extension of the definition and an overview that further describe the aspects that are relevant for DaaPs, but that do not make up key elements for its definition. The second goal is to use the characteristics to provide an initial step towards the answer of the main question. The characteristics can be used to make decisions related to DaaPs a specific implementation of a DaaP with the right characteristics that fit the context of the platform. This can help a platform achieve the goal stated in the main question, unlocking value from data through a DaaP.

The result of this research question is a Characteristics model, extensively discussed in chapter 8 and displayed in table D.3. These characteristics detail the different elements that make up a DaaP and describe different variations of how these elements could be implemented in a DaaP. The description of these characteristics can be combined with knowledge on the specific context of the data that is shared to determine the best set-up of the platform to unlock value from the data.

12.1.3 DAAP MATURITY MODEL

The third and fourth sub-question provide the core of the answer to the main research question. A maturity model has been created that can be used by practitioners to determine how to develop a DaaP, how to mature it and how to grow it into a monetizable platform. This maturity model has been created through the literature review, which resulted in an initial version. The version has been updated, synthesized and validated using the opinions of experts in this field. Finally, it has been extended with a method for how it can be used, which is validated through a case study in which the model is applied in practice to assess a platform.

This model provides an answer to the main question in the following ways. First of all, it shows how DaaP platforms can be created and how they can be improved. The main question asks how DaaPs can be used to unlock value from data. This research concludes, based on the input from both literature and interviewees, that to unlock value from data with DaaPs, a mature platform should be created. The model that has been created in this research can help companies do that. Secondly, accompanied with the model, a method is created that can further assist organizations to mature their platform in different context. The model considers that in a varying context, different elements are required to create a successful DaaP. Therefore, a method is created in that supports organizations in determining how their DaaP can be shaped in their context to enable the unlocking of value in their situation.

12.1.4 FUTURE WORK

Future work could in the look to extend the materials created in this research. An assessment of topics across capabilities or priorities for different DaaP archetypes can significantly increase the usefulness and ease of use of the created materials. Moreover, the ecosystem perspective form DaaPs can be extended upon. Research can , for example, be done in how the openness of the ecosystem impacts the ability to create new products or services in a DaaP or how different data quality concepts influence the ability to capture value from the data. Next to that, the commercial perspective crucial in the DaaP definition can be further developed. Research can be done into business models and new value propositions that can be enabled through the use of DaaPs. Finally, research into the value estimation of data can help increase the presence of DaaPs. As discussed in the previous chapter, the value of data is relative to its use case

and as perceived by its user. Therefore, research into an estimation of the value of data for other parties can help companies select data that is worth monetizing, creating a mutually beneficial platform.



PAPERS LITERATURE REVIEW

| Title | Reference |
|--|--|
| Determinants of information quality in dyadic supply chain relationships | [Myrelid and Jonsson, 2019] |
| Governance of big data collaborations: How to balance regulatory compliance and disruptive innovation | [van den Broek and van Veenstra, 2018] |
| Threats and benefits of power discrepancies between organisations: a supply chain perspective | [Huo et al., 2016] |
| Modes of governance in inter-organisational data collaborations | [Veenstra, 2015] |
| Industry-wide Inter-organizational Systems and Data Quality: Exploratory findings of the use of GS1 standards in the Dutch retail market | [Dalmolen et al., 2015] |
| Creating value through data collaboratives: Balancing innovation and control | [Klievink et al., 2018] |
| Data collaboratives as “bazaars”? A review of coordination problems and mechanisms to match demand for data with supply | [Susha et al., 2017a] |
| Data driven social partnerships: Exploring an emergent trend in search of research challenges and questions | [Susha, 2019] |
| Accessibility and Flexibility: Two Organizing Principles for Big Data Collaboration | [Hemphill and Jackson, 2017] |
| The Data Sharing Economy: On the Emergence of New Intermediaries | [Richter and Slowinski, 2019] |
| Data sharing and open banking | [Brodsky and Oakes, 2017] |
| Challenges of governing inter-organizational relationships: Insights from a case study | [Choi and Kroschel, 2015] |
| Debating big data: A literature review on realizing value from big data | [Günther et al., 2017] |
| Modeling Support for Strategic API Planning and Analysis | [Horkoff et al., 2018] |
| Management and governance of external developer platforms at the example of Akamai, Inc. and Uber Technologies, Inc. | [Umbach and Umbach, 2017] |
| Open data platforms: Discussing alternative knowledge epistemologies | [Danneels et al., 2017] |
| Investigating open data portals automatically: A methodology and some illustrations | [Correa et al., 2018] |
| Towards intelligent open data platforms: Discovering relatedness in datasets | [Sennaike et al., 2018] |
| IoT innovation pulse | [Mattison,] |
| Connecting Digital Cities: Return of Experience on the Development of a Data Platform for Multimodal Journey Planning | [Ouoba et al., 2016] |
| Designing wearable device-based product and service ecosystem | [Sun et al., 2015] |
| Creating internet of things ecosystems | [Fisher and Davies, 2016] |
| The Digital Platform: A Research Agenda | [De Reuver et al., 2018] |
| Design Choices for Data Governance in Platform Ecosystems: A Contingency Model | [Lee et al., 2017b] |
| Design and governance of platform ecosystems–key concepts and issues for future research | [Schrieck et al., 2016] |
| Data Governance for Platform Ecosystems: Critical Factors and the State of Practice | [Lee et al., 2017a] |
| A Data Governance Framework for Platform Ecosystem Process Management | [Lee et al., 2018] |
| Data Governance Decisions for Platform Ecosystems | [Lee et al., 2019] |
| Value creation in the Internet of Things: Mapping business models and ecosystem roles | [Ikävalko et al., 2018] |
| Industrial Open Data: Case Studies of Early Open Data Entrepreneurs | [Lindman et al., 2014] |
| Integrating Open Data Reuse into the Business Models of German Companies | [Stamova, 2016] |
| Exploring the factors that influence adoption of open government data for commercial service innovation in cities | [Maccani, 2016] |
| Governance as a Key Success Factor for Big Data Solutions in Mobility | [Veeneman et al., 2018] |
| Data Collaboratives as a New Frontier of Cross-Sector Partnerships in the Age of Open Data: Taxonomy Development | [Susha and Janssen, 2017] |



INTERVIEWS

B.1 INTERVIEW PROTOCOL

Interview Protocol

Background

The recent trend towards Big Data has exponentially increased the amount of data that is gathered by companies. However, companies are struggling to actually capture value from that data. One potential use case of the data is sharing it in through a data platform. These data platforms, often implemented with an API, allow for the development of new business models related to sharing data in an ecosystem. In this research, such a commercial platform where companies share their data in an ecosystem is coined as Data as a Platform (DaaP). This research looks into both the technical aspects of such a platform, considering the data capabilities required to make a platform successful and the business aspect, looking into potential business models that can be utilized.

Specifically, this research aims to create a maturity model that discusses data capabilities required for these platform and how to develop and mature them. Moreover, it models potential business models through the use of the business model canvas.

Input on these concepts has been gathered through a literature review. In these interviews you will first be asked for your expertise and then you can give feedback on the previous findings.

Goals

The goals of this interview are:

1. Identify and evaluate a definition and characteristics of Data as a Platform
2. Identify and evaluate Data Capabilities required in a Data as a Platform platform
3. Identify and evaluate Business Models that can be used for DaaP platforms

Respondents

There are several types of interviewees that I would like to get input from.

- Data Platform experts

This group of interviewees has some expertise on data sharing in a data platform. They are most of all expert on the data capabilities required for DaaP Platforms specified in the second interview goal. Moreover, they have interesting insights on potential business models.

- Platform operating organization

This group of interviewees is an expert on data platforms and DaaP platforms they have in depth knowledge on key characteristics, data capabilities or at least those important in their context and they have deep insights in at least some of the business models.

- Ecosystem partner organization

This group of interviewees has experience in collaborating with DaaP platforms. In the context of their organization, they know what Data Capabilities are required. Moreover, they have insights in the Business Models that could be used by DaaP platforms.

Type of Platforms

Several types of platforms are researched in this work. As there's no sufficient amount of DaaP organizations and interviewees, several experts on related platform types are also interviewed.

- Data as a Platform interviewees

These are interviewees that have an expertise on the very context of this research. They are familiar with DaaP ecosystems either as a direct part of the organization or from a consultant perspective. Moreover, they can provide input on all goals stated in this research.

- Data Marketplaces interviewees

These interviewees are experts on Data Marketplaces. Data Marketplaces are closely related to DaaP platforms, however, they are not focused on building an ecosystem supporting third party software, but are more focused on selling their data directly for a wide range of use cases. These experts can have insights on Data Capabilities and potentially in some of the business models, as they may have considered related business models before deciding to start a Data Marketplace

- Open (Government) Data interviewees

These interviewees are experts on Open (Government) Data. They are familiar with datasets that are usually made available for free and without business models. Often these are also not made available by companies, but more often by (semi-) governmental organizations or NGOs. These experts have insights in what Data Capabilities are required for data ecosystems and what partners need, but are less familiar with business models.

Questions

Introduction

Permission

Before we start with the interview, I would like to ask you, on the record, for your permission for this interview and to get it recorded. Moreover, before including this interview in this research I would like to make sure that you agree with all your statement and that nothing is included that you do not want discussed. Therefore, I will send you the transcript for you to look over and I would like to get your agreement on the final product. I would like to make sure that you agree with these terms.

Additional Information

The results of this interview are used exclusively for my Thesis. All company and person names you mention will be anonymized from the first transcription of this interview and onward for the rest of this research. This interview will take approximately one hour.

Introducing questions

- Could you introduce yourself?
 - Name
 - Company
 - Tenure at company
 - Function
 - Task description of current role with key expertise

- As mentioned, this research is one the topic of DaaP platforms. Could you tell me about your experience with this or related concepts?
 - Projects
 - Companies
 - Topics
 - Years of experience
- Can you give me more insights on this platform?
 - Type
 - Your role
 - Who started it
 - Size (number of partners)
 - Business Model (overall)

DaaP Definition

As the concept of DaaP has not yet been introduced in literature, the first step of this research is to create a definition for this topic. Based on existing literature on related topics I have created the following definitions:

- From your experience with this topic, which of the definition do you think fits the topic best, as I described it? Do you have any suggestions for alterations?
- Does it have all the key characteristics that differentiate the topic from related topics?
- How would you say the ecosystem you work on relates to this topic and this definition specifically?

DaaP Characteristics

Through the literature review I have identified several key characteristics that describe DaaP ecosystems in more detail. Moreover, these characteristics have various implementations that differentiate DaaP ecosystems. For example, a DaaP ecosystem can be open to anybody or have specific partner selection.

- I have found these characteristics, do you agree with them?
- Based on this, are you missing some still?
- What are the characteristics that define your platform?

DaaP Data Capabilities

For the next step in this research I am interested in looking at some of the data capabilities required for a DaaP platform. These data capabilities are the key capabilities such a platform must have to be useful for customers, but also to allow the monetization of the platform.

- What do you think of these, Focus Areas as the key Focus Areas for DaaP platforms?
- What do you think of the practices that are identified for each of the Focus Areas?
- Based on this, do you know any more that are still missing?
- What are important data capabilities for your platform and why?

DaaP Business Models

Finally, for the final research question, I would like to discuss Business Models that can be used with DaaP platforms.

- I have modeled several business models from literature in the Business Model Canvas, what do you think of these business models?
- Based on this, would you like to add any more business models?
- Finally, what business model does your platform use?

| DaaP Definition |
|--|
| <p>“ Data as a Platform ecosystem comprises a set of organizations utilizing a central data platform, a technological platform, which primary purpose is to share data for use by third parties. This central data platform is supported by an autonomous business model.”</p> |

DaaP Definition

“A Data as a Platform ecosystem is a set of businesses interacting with a shared data market, together with the relationships among them. These relationships are underpinned by a common data platform, which is a technological platform that as primary purpose has the sharing of data. The central data platform in a DaaP ecosystem is supported by a business model.”

“A Data as a Platform ecosystem comprises a set of organizations utilizing a central data platform, a technological platform, which primary purpose is to share data, for, potentially commercial, use by third parties. This central data platform is supported by an autonomous business model.”

B.2 INTERVIEW LETTER

Dear,

For my Master Thesis of the Business Informatics Master's program at the Utrecht University and in collaboration with Deloitte I am researching commercial data platforms. To be more exact, I am researching the functionalities and capabilities of such a platform and I am looking at Business Models that could be used with the platforms. For this, I am looking for interviewees to discuss this topic with. I am looking for someone familiar with either of these situations:

1. Worked on or considered creating a data platform, where an organization's data is shared with third parties for their use cases. This platform was created with a business model in mind.
2. Worked on or considered a product or service, in which data from such a platform would play a prominent role.

Research description

Much work has been done on the advantages of opening up data as an organization, mostly focusing on the advantages of Open Government Data. These advantages include increased transparency and increased innovation. However, there are also advantages to opening up data as a commercial company, such as creating an ecosystem for other parties to add complementary products to a company's main offering strengthening its own product in the process. This research looks at the capabilities a commercial company needs for opening up their data and making it a valid business model.

Research goals

To get an increased clarity on this topic I am looking for subject matter experts to interview for my research. I am doing these interviews with the following goals:

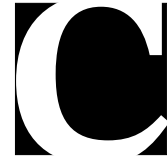
1. Exploring the required capabilities for such a data platform.
2. Determining and validating the potential business models that companies can use to make such a platform profitable.
3. Collecting insights from subject matter experts to illustrate the topic with concrete examples.

The interview will take around 60 minutes and can take place at the desired location of the interviewee or through Skype. The research itself is in all likelihood done at the end of November and all interviewees will get access to the results. Also, the interviews will be fully anonymized and the transcript will, before being processed in the research, be sent to you for verification.

I hope I have supplied you with sufficient information on my research, but if you have any more questions please let me know. Your participation in this research is highly appreciated.

Kind regards,

Yannick



CASE STUDY

C.1 CASE STUDY PROTOCOL

The Case Study Plan

Goal

The main goal of this case study is to validate the materials created in this research and identify how they are best use in practice. These models consist of:

1. A DaaP Focus Area Maturity Model
2. A DaaP Characteristics Framework

There are three goals for performing the case study:

1. Understand how companies want to use the materials
2. Understand the motivations for using such materials
3. Validate that the materials in their current form are useful and easy to use

Why and How

The goals can be achieved in the following ways:

Understand how companies want to use the materials

Why: Understanding how companies are planning on using the materials can give great insights in how the materials can be best framed in the scientific work and outside of that to show the usefulness. Companies, especially those familiar with Maturity Models, likely have their own ways of using such models. Figuring out how they want to use it can inform a method description of how the models can be best used in practice. This method can then be a combination between the practical insights from the application at companies and the theoretical implementation preferred in a scientific context.

How: This information will be retrieved by handing the materials to companies and asking them to use it without giving any suggestions. A prepared method could still be applied at a later point, simultaneously collecting their preferred method and validating a predefined method.

Understand the motivations for using the materials

Why: In order to further optimize the method description for using the materials, it is useful to understand the motivation for using the materials. Understanding why organizations want to use the materials informs a reshaping of the model that fits these motivations.

How: This information can be gathered by directly asking for motivations.

Validate the materials in their current form are useful and easy to use

Why: This is done as a final validation of the model and to determine where future work related to the models should aim towards. Asking about the quality of the material on several axes results in a final validation if the model is actually applicable for practical situations. Moreover, any limitations found by the Case Study participants can either still be improved in this work or input for future research on this topic.

How: This can be done by assessing the results of the assessment in a feedback form given to the case study participants after the assessment has been finished.

Case Study Design

Introduction

This Case Study is performed to validate materials in relation to research into Data as a Platform (DaaP) a concept newly introduced in academic literature by the creator of these materials. Data as a Platform is defined in the following way: *“Data as a Platform (DaaP) comprises a set of organizations utilizing a central data platform, a technological platform, which primary purpose is to share data for use by third parties. This central data platform is supported by an autonomous business model.”*

First, I would like to ask some introductory questions to describe the participants of this Case Study.

| |
|---|
| <ul style="list-style-type: none">- Could you introduce yourself?<ul style="list-style-type: none">○ Name○ Company○ Tenure at company |
| |
| <ul style="list-style-type: none">- Can you describe your job?<ul style="list-style-type: none">○ Function○ Task description of current role with key expertise |
| |
| <ul style="list-style-type: none">- As mentioned, this research is one the topic of DaaP platforms. Could you tell me about your experience with this or related concepts?<ul style="list-style-type: none">○ Projects○ Companies○ Topics○ Years of experience |
| |
| <ul style="list-style-type: none">- Can you give me more insights on this platform?<ul style="list-style-type: none">○ Type○ Your role○ Who started it○ Size (number of partners)○ Business Model (overall) |
| |

Maturity Assessment Method

I would like to continue by starting the maturity assessment. For my research I have created several materials that allow for the assessment of data platforms. The first one is a model that describes characteristics and that can be used to define a Data as a Platform platform in more detail by identifying its characteristics.

Before looking at the model in detail, I do have a question. What do you think would be a motivation for you, or another stakeholder of your Data Platform for using the model I just described?

Now we dive into the assessment of your Data Platform using the model, but before we start the assessment. How would you go about using this model when it was presented to you? What do you think is the best way? Why?

Now let's first access the characteristics relevant for your platform.

And now let's look at the chosen characteristics and select the relevant elements.

What did you think of this method for using the model, compared to what you suggested?

How would you say your view of the use cases and your motivation for using this model changed after seeing it?

How would you grade the usefulness of this model?

1 ()

2 ()

3 ()

4 ()

5 ()

Why?

How would you grade the understandability of this model?

1 ()

2 ()

3 ()

4 ()

5 ()

Why?

How would you grade the Ease of Use of this model?

1 ()

2 ()

3 ()

4 ()

5 ()

Why?

The second model is a Maturity Model that allows the assessment of a Maturity of such a Data as a Platform platform. This Maturity Model consists of Focus Areas. Focus Areas are, subsequently, split up in

Before looking at the model in detail, I do have a question. What do you think would be a motivation for you, or another stakeholder of your Data Platform for using the model I just described?

Now we dive into the assessment of your Data Platform using the model, but before we start the assessment. How would you go about using this model when it was presented to you? What do you think is the best way? Why?

Now let's start the assessment. What do you think is the current maturity of your platform?

Is there a desired level you would like to be at?

And now let's look at the Focus Areas and select the relevant ones.

Now let's assess each of the relevant Focus Areas. What level are you at now?

Considering each relevant Focus Area, what practice would you like to have implemented?

What did you think of this method for using the model, compared to what you suggested?

How would you say your view of the use cases and your motivation for using this model changed after seeing it?

How would you grade the usefulness of this model?

1 ()

2 ()

3 ()

4 ()

5 ()

Why?

How would you grade the understandability of this model?

1 ()

2 ()

3 ()

4 ()

5 ()

Why?

How would you grade the Ease of Use of this model?

1 ()

2 ()

3 ()



4 ()

5 ()

Why?

Business Model Assessment

Could you describe the business model for your Data Platform in detail for me?

| | | | | |
|--|--|--|--|---|
| Key Partners  | Key Activities  | Value Propositions  | Customer Relationships  | Customer Segments  |
| | Key Resources  | | Channels  | |
| Cost Structure  | | Revenue Streams  | | |

C.2 INTERVIEW LETTER

Case Study Data as a Platform Research

Waar gaat mijn onderzoek over?

Mijn onderzoek gaat over commerciële data platformen. Platformen waarop bedrijven hun data delen met andere partijen om producten of services te verbeteren, dit delen gebeurt alleen met een bijpassend business model. Deze platformen geef ik in mijn onderzoek de naam Data as a Platform (DaaP). Neem bijvoorbeeld de KVK, die door middel van een API, waarvoor een abonnement afgesloten moet worden, data deelt met banken zodat zij bedrijven kunnen controleren voordat ze leningen afsluiten.

In mijn vooronderzoek heb ik een aantal artefacten gemaakt, die samen het eindresultaat voor mijn onderzoek vormen. Het gaat hierbij om:

- lijst met karakteristieken voor DaaP ecosystemen
- een Maturity Model die de belangrijkste capabilities uitstippelt
- een verzameling Business Modellen die bedrijven hierbij kunnen gebruiken

Wat ik zoek?

Ik zoek 4 mensen om te interviewen voor mijn Case Study. Bestaande uit een combinatie van:

- mensen binnen het bedrijf dat de data deelt
- klanten die buitenaf gebruik maken van de data
- andere nauw betrokken partijen

Met deze mensen zou ik met de artefacten in een interview van ongeveer 1 tot 1.5 uur een assessment doen van het data platform door middel van de genoemde artefacten. Hierna wil ik ook feedback verzamelen over hoe zinvol en correct de artefacten werden ingeschat.



DAAP CHARACTERISTICS

D.1 FIRST CHARACTERISTICS MODEL

| Data | Alternatives | | | |
|---|--|-------------------------------------|---|------------------|
| Specificity | Specific to a single Use Case | Non-specific | | |
| Structuredness | Structured | Unstructured | | |
| Content of Data | | | | |
| Administrative level associated with data | High | Low | | |
| Ecosystem | | | | |
| Diversity of data providers | Single data provider | Small partnership of data providers | Open platform | |
| Target user group | | | | |
| Research or policy problem | Specific research/policy goal | No specific research/policy goal | | |
| Collaboration among data users | High level of additional collaborations between data users | Little to no collaboration | | |
| Coordination mechanisms | Centralized | Decentralized | | |
| Incentives intensity | Incentives for data use | Incentives for data provision | No incentives | |
| Market Structure | Coincident | Intersecting | Monopoly | |
| Mode of data sharing | Data Pooling | Data Platform Peer-to-Peer | Data Platform Marketplace | |
| Type of data sharing | Buy and sell | Open up | Data exchange orchestrated by dominant member | |
| Openness | Open | Restricted | | |
| Governance | | | | |
| Interoperability | High level of interoperability with different systems | Little to no interoperability | | |
| Data Ownership | Remains at individual organization | Open license | Determined by dominant member | |
| Data Pricing | Paid | Free | | |
| Orchestrator | | | | |
| Activities | Collection | Sharing | Processing | |
| Motivation | Gaining new revenue streams | Community building | Internal business improvement | Publicity and PR |
| Coordination method for Data Sharing | Contracts | Data quality | Power exerted by dominant member | |

TABLE D.1: *First model with Characteristics*

D.2 INPUT SECOND SLR ITERATION

| Aspects | Source |
|--|-------------------------|
| Equal consideration of technical and social aspects | [Tran et al., 2016] |
| Governance Model, Market: - Pool data in a central/pooled marketplace. - Little coordination needed. - Contractual transactions. | [Veenstra, 2015] |
| Governance Model, Market: - Marketplace and contracts between individual organisations - Control over data at individual organisation - Control over data in the hand of the buyer | [Veenstra, 2015] |
| Governance Model, Bazaar: - Data open - Supply and Demand determined by Data Quality - Everyone has access to all data | [Veenstra, 2015] |
| Governance Model, Hierarchy: - Data exchange based on the needs of the dominant member(s). - Dominant member exerts power | [Veenstra, 2015] |
| Governance Model, Network: - Lateral data exchange retaining control - Trust relations form the basis | [Veenstra, 2015] |
| One-to-one data sharing One-to-many data sharing Many-to-many data sharing | [Veenstra, 2015] |
| Hierarchical: - Dominant member set standard, through contract or protocol Bazaar: - No coordination except for technical maintenance - Quality of data determines reuse Network - All types coexist depending on the organizations involved - Only coordination is that all data is shared via the platform | [Veenstra, 2015] |
| Control over Data Hierarchy: - Determined by dominant organization, which tightly controls Bazaar: - No control Network: - Individual organizations retain control | [Veenstra, 2015] |
| Without Personal Data: - Free and open data sharing With Personal Data: - Strict control on data sharing with a hierarchical nature - Even when there's only a small threat of re-identification | [Veenstra, 2015] |
| Ground for data sharing: - For a specific single purpose - Based on a strong generic purpose (say a scientific problem) - No specific purpose | [Veenstra, 2015] |
| Characteristics: - Goal-directed - Various types of data sources (people or phenomena etc.) - Different forms (words, transactions etc.) - Specific or not specific - Structured or unstructured | [Klievink et al., 2018] |
| Data Platform goals: - Public vs. Commercial - Related to Products vs. Services Motivation: - Get a seat at the table - Explore new business models - Cust cost - Collaborative innovation - Gain legitimacy by engaging other stakeholders | [Klievink et al., 2018] |
| Balancing Control and Generativity is a key challenge. Not giving away secrets that help other companies compete, while enabling collaborative innovation | [Klievink et al., 2018] |
| Strict control: - Contracts - Strict governance This hinders the collaborative process. The more the data relates to the primary product of the data sharing company the more control they want to exert. | [Klievink et al., 2018] |

| | |
|---|-------------------------------|
| Governance Arrangement, Market: - Dyadic contractual agreements - Degree of control over the collaboration is high - Primary motivation is competition, companies try to advance their market position by collaborating. - Relationships short-term - High transaction costs - Little coordination required: Data shared in a central place and the rest is done contractually Bazaar: - Community of actors share for a common goal - Reputation is the main motivation - Minor importance for intellectual property - Members waive ownership | [Broek et al., 2018] |
| Hierarchy: - Emphasize formal relations between individual members - Higher ranked members have control Overall motivations: - Research and innovation - Commercial purposes Sharing: - One-to-one - One-to-many - Many-to-many Type of innovation: - Organizational - Social - Both | [Broek et al., 2018] |
| Coordination Mechanisms: - Dominant organization exerts tight control via a contract or protocol - No coordination at all, Quality and Usefulness determine re-use - Only coordination comes from the fact that data is shared via the platform Control over data: - Opt-in and opt-out of use of transaction data by consumers - Determined by the dominant organization - Individual organizations maintain control over the data | [Broek et al., 2018] |
| Four facets of information sharing: - Content, message of communication - Modality, the method of communication and formal vs. informal - Frequency, amount and intensity of communication: routinized or spontaneous - Direction: movement of communication: unidirectional or bidirectional and initiated by the upstream or downstream party | [Myrelid and Jonsson, 2019] |
| Data related: - Type - Content - Administrative level of data - Degree of access - Diversity of data providers - Facilitation mode Use of data: - Policy or research problem - Purpose of use - Target users and user selection - Incentives for use - Expected outcome of collaborative - Continuity of collaboration | [Susha and Janssen, 2017] |
| Platform ownership: Platforms set up by sharing company vs. Platforms created by third parties. | [Richter and Slowinski, 2019] |
| Degree of openness: -Closed, limited to certain cooperating partners. Allow for much control. - Pooling, Data shared in reference to a given service/industry/ecosystem - Open: Data marketplaces | [Richter and Slowinski, 2019] |
| Degree of integration: intra- or interindustry intermediary | [Richter and Slowinski, 2019] |
| Additional Functionalities: Yes or No | [Richter and Slowinski, 2019] |
| Paid access to platform: Yes vs no Open access motivation: - Strong interest in data re-use - Serve public interest | [Richter and Slowinski, 2019] |
| Some networks could work both in open and controlled modes. | [Choi and Kroschel, 2015] |
| Organizations vs. Governmental data | [Choi and Kroschel, 2015] |
| Cognivist: - Platforms are neutral tools for dissemination - Organized for ad hoc querying by individual actors - Limited and one-way interactions - Limited governance, only openness is guaranteed but re-use is not stimulated | [Danneels et al., 2017] |

| | |
|---|---------------------------|
| Connectionist: - Data sharing used to foster connections - Result of the collaboration not visible on the platform - Promoted through i.e. hackatons | [Danneels et al., 2017] |
| Autopoietic: - Platform as living organism - Data from various sources - Includes tools and services for re-use | [Danneels et al., 2017] |
| Crowdsourcing: As many participants as possible | [Sun et al., 2015] |
| Stakeholder roles: - Information providers - App developers - Analytics service providers - Platform providers - Users of information and applications | [Fisher and Davies, 2016] |
| Centralized vs Decentralized: Division of: - Authority - Responsibility - Decision making | [Fisher and Davies, 2016] |
| Check table 1 | [Fisher and Davies, 2016] |
| Control of data: - Participant interdependence: Is interdependence between participants required? Yes requires much coordination - Directed data: Centralized ecosystems for a specific purpose. Participants are independent | [Curry, 2018] |
| Continued: - Acknowledged data ecosystems: Pooled dedicated resources from more sources, changes based on collaboration - Collaborative data ecosystem: Participants interact voluntarily for a central purpose - Virtual data ecosystems: No central management authority and no purpose | [Curry, 2018] |
| Platform type Business purpose: commercial vs not | |
| Three types of data: Raw data, linked data or humanreadable data | [Lindman et al., 2014] |
| Participant profiles: Data analyzer, user experience provider, open data publisher and support service and consultation | [Lindman et al., 2014] |
| Current taxonomies: - How data is collected and opened - Format of data - Data use case - Content and meaning - Differentiate between tracking activities | [Susha and Janssen, 2017] |
| Type of data: personal, natural phenomena etc. Personal data types: - Observed, without explicit knowledge - User-generated - Volunteered | [Susha and Janssen, 2017] |
| Content of data: Words, locations, behavior, ransactions, natur. | [Susha and Janssen, 2017] |
| Administrative level associated with data: Specific vs. unspecific to one purpose | [Susha and Janssen, 2017] |
| Diversity of data providers: One vs several from the same industry vs. Cross industry collaboration | [Susha and Janssen, 2017] |
| Target user group: Academic vs Commercial vs Governmental vs non-profit vs citizens | [Susha and Janssen, 2017] |
| User selection: On agreement basis cs application basis vs open | [Susha and Janssen, 2017] |
| Research or policy purpose: Specified vs unspecified | [Susha and Janssen, 2017] |
| Incentives for data use: tangible vs intangible | [Susha and Janssen, 2017] |
| Continuity of collaboration: on demand vs. event-based vs continuous | [Susha and Janssen, 2017] |

| | |
|---|-------------------------------------|
| Outcome of collaborative: Intervention vs. Data analytics vs innovation. | [Susha and Janssen, 2017] |
| Collaboration among data users: One user vs. selfselected analysis by several users vs collaborative analysis by several users (coordinated) | [Susha and Janssen, 2017] |
| Facilitation: Self-facilitated vs intermediary | [Susha and Janssen, 2017] |
| Degree of access: Real-time vs Direct access to a copy of raw data vs modified/enriched data vs processed data vs open data. | [Susha and Janssen, 2017] |
| Purpose of use: Primary (purpose for which it was collected) vs secondary (similar to the purpose for which it was collected) vs tertiary (different) vs. end-use processed and used by the end-user | [Susha and Janssen, 2017] |
| How is data generated: User-created vs. produced by private organization Data classification: Business data Format of the data Machine readable Accuracy Sensitive Maintenance: Availability must be guaranteed for continuity | [Buda et al., 2016] |
| Technical change: Is a technical change required? Location of open data: - Sharing data on the company website - Via a platform - Sharing/selling directly on a marketplace | [Buda et al., 2016] |
| Costs required for access: Free vs. paid Access: Closed vs Restricted vs Open | [Buda et al., 2016] |
| Data Philanthropy - Opening up data to the whole ecosystem - Great societal impact - Limited impact on the company | [Buda et al., 2016] |
| Naked corporation: - Data freely opened - Opened to a specific target audience - Impact on the company higher | [Buda et al., 2016] |
| Monetizing data: - Data available to a smaller group - Data opened at a cost - Either selling directly or with services | [Buda et al., 2016] |
| Commercial openness: - Company not pre-selecting end users - Available at a fee - Allows for coepetition, as competitors can have access | [Buda et al., 2016] |
| Direct consumption: - API direct to customers - Access to specific data sources - Similar to a brokerage firm, getting revenue through commission | [Malgonde and Bhat-tacherjee, 2014] |
| Wallet companies: Promote applications to economically disadvantaged customers | [Malgonde and Bhat-tacherjee, 2014] |
| Indirect or intangible monetization: - APIs free of charge to gain insights or enhance market presence - Chargebacks, rate limits or quotas Transactional monetization: - Capture revenue through direct usage Freemium Product-based monetization Premium features | [Malgonde and Bhat-tacherjee, 2014] |

TABLE D.2: *Input Characteristics from Literature*

D.3 FINAL CHARACTERISTICS MODEL

D.4 DEFINITIONS OF CHARACTERISTICS

| Categories | Source | Values | | | | | | |
|--------------------------------------|---|---|--|---|---|--|-----------------|-----------------------------|
| Coordination Mechanism | [?] [Broek et al., 2018] | Dominant member set standard, through contract or protocol | Only coordination is that all data is shared via the platform. Only technical maintenance is done collaboratively. Quality of data determines if data is re-used | Collaborative analysis by several users for a specific target | | | | |
| Control over Data | [?] | Hierarchy: - Determined by dominant organization, which tightly controls | Bazaar: - No control | Network: - Individual organizations retain control | Control over data in the hand of the buyer | | | |
| Multiplicity | [?] [Broek et al., 2018] | One-to-many data sharing | Many-to-many data sharing | | | | | |
| Personal Data | [?] | Without personal data | With personal data | | | | | |
| Ground for data sharing | [?] [Susha et al., 2017b] [Susha and Janssen, 2017] | For a single specific purpose | Based on a strong generic purpose (i.e. curing a disease any way possible) | No specific purpose | | | | |
| Motivation for starting the platform | [?] [Klievink et al., 2018] [Broek et al., 2018] | Get a seat at the table | Explore new business models/Commercial purposes | Cut costs | Collaborative innovation/ Research and innovation | | Gain legitimacy | engaging other stakeholders |
| Type of data | [?] [Susha et al., 2017a] [Susha and Janssen, 2017] | Real-time | Direct access to a copy of raw data | Modified/enriched data | Processed data | | | |
| Diversity of data providers | [?] [Susha et al., 2017a] [Susha and Janssen, 2017] | One | Several from the same industry | Cross-industry | | | | |
| User selection | [?] [Susha et al., 2017a] [Susha and Janssen, 2017] | On agreement basis (Based on a partnership agreement) | On application basis (Selected for specific application of the data) | Open | | | | |
| Continuity of Data Sharing | [?] [Susha et al., 2017a] [Susha and Janssen, 2017] | On demand | Event-based | Continuous | | | | |
| Degree of integration | [?] [Umbach and Umbach, 2017] | Intra-industry | Inter-industry | | | | | |
| Level of centralization | | Decentralized: Division of Authority, Responsibility and Decision Making | Centralization: Authority, Responsibility and Decision Making at central party | | | | | |
| Location of open data | [?] [Broek et al., 2018] [Buda et al., 2016] | Company website | Platform | Selling on a marketplace | Pool data in a central/pooled marketplace | | | |
| Specificity | [?] [Susha and Janssen, 2017] | Specific to a single use case | Non-specific | | | | | |
| Openness | [?] [Susha et al., 2017b] | Open | Closed | | | | | |

TABLE D.3: Characteristics found in the Literature Review

| Name | Definition |
|--------------------------------------|--|
| Coordination Mechanism | How is the platform coordinated? |
| Control over Data | Who has control over the data in the platform? |
| Multiplicity | What are the relations of the data sharing? |
| Personal Data | How does the platform handle personal data? |
| Ground for data sharing | What is the purpose of data sharing? |
| Motivation for starting the platform | What is the motivation for starting the platform for the orchestrator? |
| Type of data | What type of data is shared? |
| Diversity of data providers | What data providers are sharing data? |
| User selection | How are users selected for the platform? |
| Continuity of Data Sharing | How continuous is the data shared? |
| Platform ownership | Who owns the platform? |
| Degree of integration | How is the platform integrated in the industry? |
| Level of centralization | How centralized is the platform? |
| Location of open data | Where is the data of the platform located? |
| Specificity | How specific is the platform targeted at one use case? |
| Openness | How open is the platform to new parties? |

TABLE D.4: *Definition for the Characteristics*



DAAP FOCUS AREA

| Name | Source | Elements | | | | |
|---------------------------|------------------------------|--|-------------------------------|--|------------------------------|---|
| Type of Data sharing | [Tran et al., 2016] | Type of information sharing | [?] | standard for data sharing | | |
| Amount of data sharing | [Tran et al., 2016] | Amount of information sharing | [Myreid and Jonsson, 2019] | Appropriate amount refers to the volume of information and should be neither too much nor too little in relation to information receivers' needs (Lee et al., 2002). | | |
| Frequency of data sharing | [Tran et al., 2016] | Frequency of information sharing | | | | |
| Security | [Tran et al., 2016] | Security of IS | [Tran et al., 2016] | Security | [Tiwana et al., 2010] | Security of IS |
| Automated security | [Tran et al., 2016] | Automated ICT-oriented security solutions | | | | |
| Access control | [Tran et al., 2016] | Authorized and authenticated access control | [Kolluru and Meredith, 2001] | Authentication | [Klievink et al., 2018] | Who should access the platform |
| Access control (cont.) | [Myreid and Jonsson, 2019] | Accessibility refers to the ease of locating and obtaining information when needed by information receivers | [Susha and Gil-Garcia, 2019] | Create data access mechanism | [Hemphill and Jackson, 2017] | Accessibility |
| Access control (cont.) | [Lee et al., 2017a] | Data ownership and access definition: presents who owns and uses the data in platform ecosystems | [Fisher and Davies, 2016] | Create licences | [Reuver et al., 2017] | Establish entrance and access rules |
| Data confidentiality | [Kolluru and Meredith, 2001] | Data confidentiality | [Tran et al., 2016] | Confidentiality refers to accidental or intentional disclosure of information | | |
| Data integrity | [Kolluru and Meredith, 2001] | Data integrity | [Tran et al., 2016] | whereas data integrity requires that information be secured against unauthorized modification | | |
| Trust | [Kolluru and Meredith, 2001] | Trust management | [Tran et al., 2016] | Trust (Figure 1) | [Klievink et al., 2018] | Trust relations within data collaboratives |
| Delegation of credentials | [Kolluru and Meredith, 2001] | Delegation of credentials across multiple tiers | [Myreid and Jonsson, 2019] | where accessibility, understandability, ease of operation and appropriate amount are ease of use-related and explain the format and availability of information | | |
| Data availability | [Tran et al., 2016] | The objective of availability is assurance of accessible and uninterrupted information at every node within the supply chain since the implications of interruptions due to degraded ICT system availability can be a dramatic reduction in market responsiveness for the focal firm and wider supply chain. | [Klievink et al., 2018] | that they require continuity to counter uncertainty | | |
| Frequent communication | [Tran et al., 2016] | frequent communication | [Richter and Slowinski, 2019] | Establish open and ongoing channels of communication with stakeholders | [Umbach and Umbach, 2017] | Create formal communications that refer to stakeholders about changes |

| | | | | | | |
|--|---------------------------|---|-------------------------------|--|--|---|
| Frequent communication (cont.) | [Umbach and Umbach, 2017] | Establish direct customer communication with customers | [Umbach and Umbach, 2017] | Open communication with developers | | |
| Partner selection | [Tran et al., 2016] | partner selection | [Richter and Slowinski, 2019] | Partner selection | | |
| Honest and open transaction | [Tran et al., 2016] | honest and open transaction | | | | |
| Create contracts for data use | [Tran et al., 2016] | Formal contract | [?] | contract | [Klievink et al., 2018] | contractual arrangements for data use |
| Create contracts for data use (cont.) | [Choi and Kroschel, 2015] | Determine ownership through formal contracting and selling | [Umbach and Umbach, 2017] | Create individual agreements with users | [Umbach and Umbach, 2017] | Create external contracts with users |
| Create contracts for data use (cont.) | [Umbach and Umbach, 2017] | Create legal agreements for API use | [Lee et al., 2017a] | Create contract-based governance structure (determine everything by contract) | [Broek et al., 2018] | Create a contract to control data sharing and usage |
| Continuous collaboration | [Tran et al., 2016] | Ongoing collaboration | | | | |
| Create personal relationships | [Tran et al., 2016] | Personal relationship | [Lee et al., 2019] | Allow personalized customer contact | | |
| Establish data protocols | [?] | Establish data protocols | [Susha and Gil-Garcia, 2019] | Create standardized protocols | [Broek et al., 2018] | create a protocol for controlled data sharing and usage |
| Establish a data exchange | [?] | Establish data exchange | [Danneels et al., 2017] | Share data one-way | [Susha et al., 2017b] | Manage transfer |
| Establish reporting mechanisms | [?] | Establish reporting mechanisms | | | | |
| Governance | [?] | Inter-organisational governance consists of the arranged institutions and structures to ensure that individuals behave in line with the collective goals, conflicts between individuals are prevented or resolved, and the effective and fair use of collective resources within the inter-organisational collaboration (Provan and Kenis, 2008). | [Klievink et al., 2018] | Governance | [Umbach and Umbach, 2017] | Governance |
| Check behaviour in line with goals | [?] | check if individuals behave in line with the collective goals | | | | |
| Conflict resolution | [?] | , conflicts between individuals are prevented or resolved | [?] | dispute resolution procedures | [Susha et al., 2017b] | Add a separate activity to remedy conflicts |
| Check for fair resource use | [?] | effective and fair use of collective resources | | | | |
| Create authority system | [?] | command structures and authority systems | | | | |
| Create incentive system | [?] | incentive systems | [Klievink et al., 2018] | Incentivisation | [Klievink et al., 2018] | existing relationships, prior collaboration, incentives and pressures (e.g., from peers or government) affect trust |
| Create incentive system (cont.) | [Sennaik et al., 2017] | Encourage creative tools and applications of data | [Sun et al., 2015] | Create contribution benefit schemes | [Klievink et al., 2018] | Create a reward plan |
| Create standard operating procedure | [?] | standard operating procedures | | Create a pricing model | [Fisher and Davies, 2016] | Establish pricing |
| Create non-market pricing system | [?] | non-market pricing systems | [Umbach and Umbach, 2017] | No control exerted, only technical maintenance of the platform | | |
| Do technical maintenance | [?] | no coordination except for (technical) maintenance takes place. | [Lee et al., 2019] | Data ownership and access rights | [Fisher and Davies, 2016] | Respect data ownerships |
| Control over data | [?] | Control over data | [Broek et al., 2018] | Determine license for data | [Parker and Alstynne, 2016] | Openness |
| Control over data (cont.) | [Lee et al., 2017a] | Create authority-based governance structure (ask for permission) | [Buda et al., 2016] | Dominant organization orchestrates and monitors data exchange | [Choi and Kroschel, 2015] | Control data access in a hierarchical fashion through a central commandor |
| Control determined by dominant organization | [?] | determined by the dominant organisation, which tightly controls what happens with the data | [Broek et al., 2018] | No control | | |
| No control over data is exerted | [?] | no control over the data is exerted | [Broek et al., 2018] | | | |
| Individual organization retain control over data | [?] | individual organisations retain control over their data | | Information quality | | |
| Data quality | [Dalmolen et al., 2015] | Data Quality | [Myrelid and Jonsson, 2019] | [25] | | |
| Completeness of data | [Dalmolen et al., 2015] | Completeness of the product information e.g. weight, sizes, shelf life, use by data, pallet stacking, tolerances, flexible packaging | | | | |
| Data validation mechanisms | [Dalmolen et al., 2015] | validation mechanism to identify whether mandatory fields indeed contain information. | | "Credibility refers to how believable and trustworthy information is to its receivers (Lee et al., 2002) | Gustavsson and WŠnström, 2009) and relates to the reputation of data and data sources (Wang and Strong, 1996 | Lee et al., 2002." |
| Data correctness | [Dalmolen et al., 2015] | Correctness is another aspect. Fields may be filled in, however that does not assure that the data is filled in correctly. | [Myrelid and Jonsson, 2019] | | | |

| | | | | | | |
|--|--|--|------------------------------|--|------------------------------|---------------------------------------|
| Logical data control mechanism | [Dalmolen et al., 2015] | logical control mechanism e.g. given the size of the item, the number of item in case or pallet stacking can be calculated and controlled for. | [Dalmolen et al., 2015] | Reduce uncertainty | [Myreliid and Jonsson, 2019] | Prevent inconsistency in data |
| Data consistency "Gustavsson and WŠnstrŠm (2009) also relate understandability to the ease of using information, which here is ease of operation. Ease of operation refers to the ease of processing, aggregating and manipulating information to meet the needs of information receivers (Lee et al., 2002 | [Dalmolen et al., 2015] Manecke and Schoensleben, 2004.)" | Consistency of data remains an issue. In the past not all data was up-to-date | [Klievink et al., 2018] | "Understandability refers to how comprehensive, readable and clear information is for information receivers (Wang and Strong, 1996 | Lee et al., 2002)" | [Myreliid and Jonsson, 2019] |
| Data Interprability | [Dalmolen et al., 2015] | Interpretation is a large challenge for product information in FMCG. | [Myreliid and Jonsson, 2019] | create a useful dataset (Zuiderwijk, 2015). | [Myreliid and Jonsson, 2019] | Relevance of data fields. |
| Data Interprability(cont.) | [Myreliid and Jonsson, 2019] | relevance refers to the extent to which information is value-adding and useful for information receivers as well as the appropriateness of the level of detail | [Susha et al., 2017b] | Useful and usable data for user | [Susha et al., 2017b] | Ensure usability of resource |
| Data relevance | [Dalmolen et al., 2015] | Relevance of data fields. | [Klievink et al., 2018] | Timeliness | [Huo et al., 2016] | Timeliness of data sharing |
| Data timeliness | [Dalmolen et al., 2015] | Timeliness of the data | [Myreliid and Jonsson, 2019] | Process feedback from partners | [Danneels et al., 2017] | Process feedback from users |
| Create feedback loops | [Dalmolen et al., 2015] | Feedback Loops | [Susha et al., 2017b] | Add meta data | [Correa et al., 2018] | Establish metadata |
| Provide metadata | [Klievink et al., 2018] | provide the necessary meta-data | | | | |
| Ensure diversity of data providers | [Klievink et al., 2018] | a significant diversity in data providers, which collaboratives must take into account, given the inherent multiplicity of interests and incentives to join and share data within a data collaborative | [Susha et al., 2017b] | Standardising data formats | [Susha and Gil-Garcia, 2019] | Standardize data collection |
| Ensure diversity of data providers (cont.) | [Choi and Kroschel, 2015] | Standardize data formats | [Choi and Kroschel, 2015] | Develop standards | | |
| Data standization | [Klievink et al., 2018] | data standardisation or processing | [Klievink et al., 2018] | Standardize data types | | |
| Data standization | [Ouoba et al., 2016] | Enable access from service in a structured format | [Veeneman et al., 2018] | | | |
| Piloting | [Klievink et al., 2018] | a first attempt to set up a collaborative, for example, piloting | | | | |
| Institutional design | [Klievink et al., 2018] | Institutional design | | Share data in real-time | [Ouoba et al., 2016] | Create consistent real-time data |
| Real-time data sharing | [Klievink et al., 2018] | Real-time sharing of raw data | [Huo et al., 2016] | | | |
| Create governance rules | [Klievink et al., 2018] | Create rules for governing | | | | |
| Design arrangement for storage personal data | [Broek et al., 2018] | Design and implement arrangements that formulate storage of personal data | | | | |
| Communicate arrangements for storage of personal data | [Broek et al., 2018] | Communicate arrangements for storage of personal data | | | | |
| Assess policies for storage of personal data | [Broek et al., 2018] | Assess policies for storage of personal data | | | | |
| Compliance | [Broek et al., 2018] | Compliance | | | | |
| Data Pooling | [Broek et al., 2018] | Pool data from multiple companies in a central place | [Broek et al., 2018] | Open up data completely | [Danneels et al., 2017] | Establish an Open License |
| Establish open governance structure | [Lee et al., 2017a] | | | | | |
| Open licensing | [Broek et al., 2018] | Make data available with an open license | [Broek et al., 2018] | | | |
| Reputation regulation | [Broek et al., 2018] | Regulate data use through reputation and transparency | | | | |
| Regulate intellectual property | [Broek et al., 2018] | Regulate intellectual property | | | | |
| create design rights | [Broek et al., 2018] | create design rights | | | | |
| Set accountabilities | [Broek et al., 2018] | Set accountabilities | | | | |
| Assess type of data shared for governance | [Broek et al., 2018] | Assess type of data shared for governance | | | | |
| Create contractual data transactions | [Broek et al., 2018] | Create contractual data transactions | | | | |
| Determine data desired by dominant party | [Broek et al., 2018] | Determine data desired by dominant party | [Broek et al., 2018] | Establish informed consent | [Choi and Kroschel, 2015] | Request user consent for data sharing |
| Get informed consent | [Broek et al., 2018] | Get informed consent for personal data use | [Brodsky and Oakes, 2017] | | | |
| Opt-in and opt-out of transaction data by customers | [Broek et al., 2018] | Opt-in and opt-out of transaction data by customers | [Broek et al., 2018] | Detect and prevent unauthorized data use | | |
| Set strict guidelines for data use and punish misuse | [Broek et al., 2018] | Set strict guidelines for data use and punish misuse | [Lee et al., 2019] | | | |

| | | | | | | |
|---|-------------------------------|--|-------------------------------|--|-----------------------|--|
| Objectivity | [Myrelid and Jonsson, 2019] | Finally, objectivity refers to the objectivity and unbiasedness of information (Wang and Strong, 1996) in both its collection and presentation (Lee et al., 2002). | | | | |
| Create involvement | [Myrelid and Jonsson, 2019] | Create involvement | | | | |
| Seek partners with organisational similarity | [Myrelid and Jonsson, 2019] | Seek partners with organisational similarity | | | | |
| Establish top management support | [Myrelid and Jonsson, 2019] | Establish top management support | | | | |
| Informal contracts | [Myrelid and Jonsson, 2019] | use informal contracts | | | | |
| Creating new avenues and means for cooperation for stakeholders | [Susha et al., 2017b] | Creating new avenues and means for cooperation for stakeholders | | | | |
| Promote governance standards | [Susha et al., 2017b] | Promote governance standards | | | | |
| Promote access governance | [Susha et al., 2017b] | Promote access governance | [Susha et al., 2017b] | Create metadata standards for new data | | |
| Promote data standard governance | [Susha et al., 2017b] | Promote data standard governance | [Sennaikie et al., 2017] | | | |
| promote documentation governance | [Susha et al., 2017b] | promote documentation governance | [Susha et al., 2017b] | A widely accepted definition of coordination is Other managing of dependences between entities | | |
| Coordination | [Susha et al., 2017b] | Coordination is a broad concept occurring at various levels (Comfort, Dunn et al., | [Susha et al., 2017b] | Match supply and demand between suppliers and users | | |
| Matching potential data providers and data users. | [Susha et al., 2017b] | Matching potential data providers and data users. | [Richter and Slowinski, 2019] | | | |
| Regulate collaboration | [Susha et al., 2017b] | here is a need for rules on how organizations collaborate. | | | | |
| Attract intermediary for connecting data supply and demand | [Susha et al., 2017b] | Attract intermediary for connecting data supply and demand | [Susha et al., 2017b] | No control over data sharing (take it or leave it) | | |
| take 'take it or leave it' stance towards connecting providers and users | [Susha et al., 2017b] | take 'take it or leave it' stance towards connecting providers and users | [Susha et al., 2017b] | | | |
| Create a mutual coordination plan for connecting supply and demand | [Susha et al., 2017b] | Create a mutual coordination plan for connecting supply and demand | [Susha et al., 2017b] | Survey users for needs | [Lee et al., 2018] | Consider needs of participating group |
| Facilitate knowledge sharing on supply and needs | [Susha et al., 2017b] | Facilitate knowledge sharing on supply and needs | [Fisher and Davies, 2016] | | | |
| Maintain control over data after data sharing | [Susha et al., 2017b] | Maintain control over data after data sharing | | | | |
| Create privacy agreements | [Susha et al., 2017b] | Create privacy agreements | | | | |
| Anonimize and aggregate privacy sensitive data | [Susha et al., 2017b] | Anonimize and aggregate privacy sensitive data | | | | |
| Keep control over data use by using a secure digital environment for data sharing | [Susha et al., 2017b] | Keep control over data use by using a secure digital environment for data sharing | | | | |
| Create norms on data use | [Susha et al., 2017b] | Create norms on data use | [Susha et al., 2017b] | Defining an explicit data owner should be implemented according to the identified regulations | | |
| Appoint a data steward | [Susha et al., 2017b] | Appoint a data steward | [Lee et al., 2017a] | | | |
| Match research problem with data characteristic | [Susha et al., 2017b] | Match research problem with data characteristic | | | | |
| Use mutual adjustment to match data and problem | [Susha et al., 2017b] | Use mutual adjustment to match data and problem | | | | |
| Allow provider to determine data attributes | [Susha et al., 2017b] | Allow provider to determine data attributes | | | | |
| Offer data playbooks, tutorials and describe data sharing technique | [Susha et al., 2017b] | Offer data playbooks, tutorials and describe data sharing technique | | | | |
| align incentives for data sharing with data users | [Susha et al., 2017b] | align incentives for data sharing with data users | [Susha et al., 2017b] | Create value for ecosystem members | | |
| Formulate value proposition for all participating sides of the platform | [Susha et al., 2017b] | Formulate value proposition for all participating sides of the platform | [Danneels et al., 2017] | | | |
| Data integration | [Susha and Gil-Garcia, 2019] | Data integration | | | | |
| Exchange additional resources | [Susha and Gil-Garcia, 2019] | Exchange data and/or resources | | | | |
| Allow different forms of consent establishment for personal data sharing | [Susha and Gil-Garcia, 2019] | Allow different forms of consent establishment for personal data sharing | [Susha et al., 2017b] | Flexibility | | |
| Flexibility | [Richter and Slowinski, 2019] | Flexibility | [Choi and Kroschel, 2015] | | | |
| Create transparency of tools | [Richter and Slowinski, 2019] | Create transparency of tools | [Susha et al., 2017b] | Establish data transmission according to user needs | | |
| Format data according to users | [Richter and Slowinski, 2019] | Format data according to users | [Choi and Kroschel, 2015] | | | |
| Create agile platform management | [Richter and Slowinski, 2019] | Create agile platform management | [Richter and Slowinski, 2019] | Create as much information about the data available | [Tiwana et al., 2010] | open API documentation and accessibility |
| Create rigorous documentation | [Richter and Slowinski, 2019] | Create rigorous documentation | [Richter and Slowinski, 2019] | Allow data from various structures | | |

| | | | | | | |
|--|-------------------------------|--|-------------------------------|---|---------------------------|--|
| Create flexible data structures | [Richter and Slowinski, 2019] | Create flexible data structures | [Ouoba et al., 2016] | | | |
| Create communication guidelines between collaborators | [Richter and Slowinski, 2019] | Create communication guidelines between collaborators | | | | |
| Create review systems for data suppliers | [Richter and Slowinski, 2019] | Create review systems for data suppliers | | | | |
| Perform partner screening | [Richter and Slowinski, 2019] | Perform partner screening | | | | |
| Create certification for providers | [Richter and Slowinski, 2019] | Create certification for providers | | | | |
| Provide security levels for providers | [Richter and Slowinski, 2019] | Provide security levels for providers | | | | |
| Implements watermarking for providers | [Richter and Slowinski, 2019] | Implements watermarking for providers | [Richter and Slowinski, 2019] | Process payments | [Umbach and Umbach, 2017] | Facilitate transactions through the platform |
| Process transactions | [Richter and Slowinski, 2019] | Process transactions | [Richter and Slowinski, 2019] | | | |
| create codes of conducts | [Richter and Slowinski, 2019] | create codes of conducts | | | | |
| Create API | [Brodsky and Oakes, 2017] | Create API | | | | |
| Streamline data access | [Brodsky and Oakes, 2017] | Streamline data access | | | | |
| Create guardrails to support privacy protection | [Brodsky and Oakes, 2017] | Create guardrails to support privacy protection | | | | |
| Create data category based security protocols | [Brodsky and Oakes, 2017] | Create data category based security protocols | | | | |
| Educate end users on privacy permissions | [Brodsky and Oakes, 2017] | Educate end users on privacy permissions | | | | |
| Keep data up to date | [Choi and Kroschel, 2015] | Keep data up to date | | | | |
| Install data mapping tables for data types | [Choi and Kroschel, 2015] | Install data mapping tables for data types | [Choi and Kroschel, 2015] | Clean data before sharing | | |
| Clean data before sharing | [Choi and Kroschel, 2015] | Clean data before sharing | [Buda et al., 2016] | | | |
| Create spot market for short transactions | [Choi and Kroschel, 2015] | Create spot market for short transactions | | | | |
| Negotiate exchange terms with users | [Choi and Kroschel, 2015] | Negotiate exchange terms with users | | | | |
| Create practices of data disclosure and screening | [Choi and Kroschel, 2015] | Create practices of data disclosure and screening | | | | |
| Implement a review board for control | [Choi and Kroschel, 2015] | Implement a review board for control | | | | |
| Let IT determine data structures | [Choi and Kroschel, 2015] | Let IT determine data structures | | | | |
| Implement risk management | [Choi and Kroschel, 2015] | Implement risk management | | | | |
| Advocate transparency | [Choi and Kroschel, 2015] | Advocate transparency | | | | |
| Offer users control over data | [Choi and Kroschel, 2015] | Offer users control over data | | | | |
| Create portability for the data | [Choi and Kroschel, 2015] | Create portability for the data | | | | |
| Create interconnectivity for the data | [Choi and Kroschel, 2015] | Create interconnectivity for the data | | | | |
| Develop organizational models | [Choi and Kroschel, 2015] | Develop organizational models | | | | |
| Develop principles | [Choi and Kroschel, 2015] | Develop principles | | | | |
| Develop policies | [Choi and Kroschel, 2015] | Develop policies | | | | |
| Realign work practices to realize stakeholder value | [Choi and Kroschel, 2015] | Realign work practices to realize stakeholder value | [Lee et al., 2017a] | impose certain regulations on the user participation to reap the benefits of ecosystem growth | [Fisher and Davies, 2016] | Create terms and conditions |
| Create terms of service | [Umbach and Umbach, 2017] | Create terms of service | [Umbach and Umbach, 2017] | Penalize regulation breaking | | |
| Determine punishments for breaches | [Umbach and Umbach, 2017] | Determine punishments for breaches | [Parker and Alstyne, 2016] | | | |
| Create legal agreements for API use | [Umbach and Umbach, 2017] | Create legal agreements for API use | | | | |
| Enforce legal agreement | [Umbach and Umbach, 2017] | Enforce legal agreement | | | | |
| Monitor resource use by customers | [Umbach and Umbach, 2017] | Monitor resource use by customers | | | | |
| Assess risk of access policy | [Umbach and Umbach, 2017] | Assess risk of access policy | | | | |
| Create governance framework | [Umbach and Umbach, 2017] | Create governance framework | | | | |
| Align accountability with business objectives | [Umbach and Umbach, 2017] | Align accountability with business objectives | | | | |
| Establish decision making structures within an organization | [Umbach and Umbach, 2017] | Establish decision making structures within an organization | | | | |
| Create alignment processes for widespread involvement from users | [Umbach and Umbach, 2017] | Create alignment processes for widespread involvement from users | | Create a set of tools and rules | | |
| Create developer products | [Umbach and Umbach, 2017] | Create developer products | [Umbach and Umbach, 2017] | Create sandboxes | | |
| Create sandboxes | [Umbach and Umbach, 2017] | Create sandboxes | [Tiwana et al., 2010] | | | |
| Create performance-based rewards schemes | [Umbach and Umbach, 2017] | Create performance-based rewards schemes | | | | |
| Establish decision rights | [Umbach and Umbach, 2017] | Establish decision rights | | | | |
| Create internal process chart for establishing decision rights | [Umbach and Umbach, 2017] | Create internal process chart for establishing decision rights | | | | |
| Create transparent roadmap | [Umbach and Umbach, 2017] | Create transparent roadmap | | | | |
| Create a governance group to guide product development | [Umbach and Umbach, 2017] | Create a governance group to guide product development | | | | |

| | | | | | |
|---|---------------------------|---|---------------------------|---|--|
| Create standards conventions and guidelines for API development | [Umbach and Umbach, 2017] | Create standards conventions and guidelines for API development | | | |
| Create a showcase section | [Umbach and Umbach, 2017] | Create a showcase section | | | |
| Create working groups and review committees to support partners | [Umbach and Umbach, 2017] | Create working groups and review committees to support partners | | | |
| Create behavioral rewards | [Umbach and Umbach, 2017] | Create behavioral rewards | | | |
| Limit access to steer new application development | [Umbach and Umbach, 2017] | Limit access to steer new application development | | | |
| List types of datasets | [Danneels et al., 2017] | List types of datasets | | | |
| Foster the health of the ecosystem | [Danneels et al., 2017] | Foster the health of the ecosystem | | | |
| Add additional data sets | [Danneels et al., 2017] | Add additional data sets | | | |
| Maintain original platform setup | [Danneels et al., 2017] | Maintain original platform setup | | | |
| Share data according to a one-size fits all approach | [Danneels et al., 2017] | Share data according to a one-size fits all approach | | | |
| Adapt platform to specific use case | [Danneels et al., 2017] | Adapt platform to specific use case | | Adaptivity to help ecosystem partners | |
| Adapt to environmental changes | [Danneels et al., 2017] | Adapt to environmental changes | [Ouoba et al., 2016] | | |
| Allow actors to alter or add to the platform | [Danneels et al., 2017] | Allow actors to alter or add to the platform | | | |
| Govern alterations from actors | [Danneels et al., 2017] | Govern alterations from actors | | | |
| No relationship with users | [Danneels et al., 2017] | No relationship with users | | | |
| Name intermediaries to foster connection between re-use actors | [Danneels et al., 2017] | Name intermediaries to foster connection between re-use actors | | Define clear roles and responsibilities for users | |
| Consider various roles of users | [Danneels et al., 2017] | Consider various roles of users | [Lee et al., 2018] | | |
| Fill empty roles | [Danneels et al., 2017] | Fill empty roles | | | |
| Focus on two-way interactions of data users by establishing offline connections | [Danneels et al., 2017] | Focus on two-way interactions of data users by establishing offline connections | | | |
| Allow the ecosystem to self-regulate | [Danneels et al., 2017] | Allow the ecosystem to self-regulate | | | |
| Support structural metadata | [Sennaik et al., 2017] | Support structural metadata | | | |
| Allow for filtering search results | [Sennaik et al., 2017] | Allow for filtering search results | | | |
| Recommend datasets | [Sennaik et al., 2017] | Recommend datasets | | | |
| Create search | [Sennaik et al., 2017] | Create search | | | |
| Allow uploading of datasets | [Sennaik et al., 2017] | Allow uploading of datasets | | | |
| Provide search on metadata level | [Sennaik et al., 2017] | Provide search on metadata level | | | |
| Allow search on content level | [Sennaik et al., 2017] | Allow search on content level | | | |
| Enable social media sharing of data | [Sennaik et al., 2017] | Enable social media sharing of data | | | |
| Enable federation between datasets | [Sennaik et al., 2017] | Enable federation between datasets | | | |
| Provide analysis tools | [Sennaik et al., 2017] | Provide analysis tools | | | |
| Allow extensions to core features | [Sennaik et al., 2017] | Allow extensions to core features | | | |
| Support personalization | [Sennaik et al., 2017] | Support personalization | | | |
| Support programmatic access to data resources | [Sennaik et al., 2017] | Support programmatic access to data resources | | | |
| Make trade-offs between amount and usability costs | [Ouoba et al., 2016] | Make trade-offs between amount and usability costs | | | |
| Reliability vs. costs trade-off | [Ouoba et al., 2016] | Reliability vs. costs trade-off | | | |
| Create advanced API materials | [Ouoba et al., 2016] | Create advanced API materials | | offer flexible business support systems to developers | |
| Create assistance tools for API | [Ouoba et al., 2016] | Create assistance tools for API | [Fisher and Davies, 2016] | | |
| Associate data types from various parties | [Sun et al., 2015] | Associate data types from various parties | | | |
| Audit commercial rights | [Fisher and Davies, 2016] | Audit commercial rights | | | |
| Audit legal compliance | [Fisher and Davies, 2016] | Audit legal compliance | | | |
| Model dependencies in data | [Fisher and Davies, 2016] | Model dependencies in data | | | |
| Make sure datasets are predictable | [Fisher and Davies, 2016] | Make sure datasets are predictable | | | |
| Trust | [Fisher and Davies, 2016] | Trust | | | |
| Enable usage-based billing | [Fisher and Davies, 2016] | Enable usage-based billing | | | |
| Create SDKs | [Reuver et al., 2017] | Create SDKs | | | |
| Partition decision rights between owner and users | [Lee et al., 2017a] | Partition decision rights between owner and users | | | |
| Data Governance | [Lee et al., 2017a] | Data Governance | | | |
| Create performance strategy | [Lee et al., 2017a] | Create performance strategy | [Lee et al., 2017a] | Centralize key decisions | |
| Maintain centralized control strategy | [Lee et al., 2017a] | Maintain centralized control strategy | [Parker and Alstyn, 2016] | Share platform control with other parties | |
| Decentralize platform control | [Lee et al., 2017a] | Decentralize platform control | [Parker and Alstyn, 2016] | | |

| | | | | | | |
|---|---------------------|---|-----------------------------|---|--------------------|---------------------------------|
| Introduce formal data quality control mechanisms | [Lee et al., 2017a] | Introduce formal data quality control mechanisms | [Lee et al., 2017a] | Determine multi-homing rules | | |
| Establish multi-homing strategy | [Lee et al., 2017a] | Establish multi-homing strategy | [Parker and Alstynne, 2016] | | | |
| Allow users to multi-home | [Lee et al., 2017a] | Allow users to multi-home | | | | |
| Prohibit multi-homing | [Lee et al., 2017a] | Prohibit multi-homing | | | | |
| Configure governance structure | [Lee et al., 2017a] | Configure governance structure | [Lee et al., 2017a] | Establish revenue sharing | [Lee et al., 2019] | Enable revenue sharing |
| Revenue sharing | [Lee et al., 2017a] | Revenue sharing | [Lee et al., 2018] | | | |
| . The definition for all types of data should be clarified to support revenue sharing and to keep control of data flow in platform ecosystems | [Lee et al., 2017a] | . The definition for all types of data should be clarified to support revenue sharing and to keep control of data flow in platform ecosystems | | | | |
| . Responsible, Accountable, Consulted, and Informed (RACI) approach is applicable to define data ownership, but it can be expanded to include Accessible for the definition of access rights for platform ecosystems. | [Lee et al., 2017a] | . Responsible, Accountable, Consulted, and Informed (RACI) approach is applicable to define data ownership, but it can be expanded to include Accessible for the definition of access rights for platform ecosystems. | | | | |
| Definition criteria: refers to the regulatory environment that could affect the ownership and use of the data in platform ecosystems | [Lee et al., 2017a] | Definition criteria: refers to the regulatory environment that could affect the ownership and use of the data in platform ecosystems | [Lee et al., 2019] | Audit compliance | [Lee et al., 2018] | Survey regulations |
| Therefore, identifying relevant regulations and laws regarding all types of data in platform ecosystems is necessary. | [Lee et al., 2017a] | Therefore, identifying relevant regulations and laws regarding all types of data in platform ecosystems is necessary. | [Lee et al., 2018] | | | |
| A decision model based on legal aspects, and a mechanism to track and notify the compliance of the rules should be applied in data governance for platform ecosystems. | [Lee et al., 2017a] | A decision model based on legal aspects, and a mechanism to track and notify the compliance of the rules should be applied in data governance for platform ecosystems. | [Lee et al., 2017a] | Create a contribution measurement model | | |
| Contribution estimation: is a mechanism to measure user contribution against value creation by providing data. | [Lee et al., 2017a] | Contribution estimation: is a mechanism to measure user contribution against value creation by providing data. | [Lee et al., 2019] | | | |
| support revenue sharing | [Lee et al., 2017a] | support revenue sharing | | | | |
| encourage high-quality resources and to repress poor-quality resources of platform ecosystems | [Lee et al., 2017a] | encourage high-quality resources and to repress poor-quality resources of platform ecosystems | | | | |
| A contribution measurement model should be adopted for accurately identifying, acknowledging and rewarding the contributors based on the platforms incentive strategy | [Lee et al., 2017a] | A contribution measurement model should be adopted for accurately identifying, acknowledging and rewarding the contributors based on the platforms incentive strategy | [Lee et al., 2019] | Classify all types of data with use cases | [Lee et al., 2018] | Define categories and use cases |
| Data use case: is related to how to use the data in platform ecosystems | [Lee et al., 2017a] | Data use case: is related to how to use the data in platform ecosystems | [Lee et al., 2018] | Provide data categories and use cases | [Lee et al., 2019] | Classify data |
| Data categories of platform ecosystems should include various sources of data (not only user content): i.e. data from users, systems and business processes | [Lee et al., 2017a] | Data categories of platform ecosystems should include various sources of data (not only user content): i.e. data from users, systems and business processes | [Lee et al., 2018] | Define data use cases | | |
| the use cases and the relevant stakeholders of the data (including individual use case for each data category) should be clearly defined and linked. | [Lee et al., 2017a] | the use cases and the relevant stakeholders of the data (including individual use case for each data category) should be clearly defined and linked. | [Lee et al., 2018] | | | |
| consistency and integrity of the cases must be considered over the lifecycle of the data to support a visible data supply chain. This factor supports both data ownership and data usage governance domains. | [Lee et al., 2017a] | consistency and integrity of the cases must be considered over the lifecycle of the data to support a visible data supply chain. This factor supports both data ownership and data usage governance domains. | [Lee et al., 2017a] | Conformance | [Lee et al., 2019] | Identify compliance |
| Conformance: is defined as an audit for compliance based on strict processes and rules | [Lee et al., 2017a] | Conformance: is defined as an audit for compliance based on strict processes and rules | [Lee et al., 2018] | | | |
| minimizing illegal use of identifiable personal data and focusing on data due processes are required | [Lee et al., 2017a] | minimizing illegal use of identifiable personal data and focusing on data due processes are required | | | | |
| review process to assess every application and content against their regulations | [Lee et al., 2017a] | review process to assess every application and content against their regulations | [Tiwana et al., 2010] | Track API consumption | [Lee et al., 2019] | Monitoring |
| Monitoring: Data flow should be monitored | [Lee et al., 2017a] | Monitoring: Data flow should be monitored | [Lee et al., 2018] | | | |
| data stewardship practices and a visible supply chain are necessary | [Lee et al., 2017a] | data stewardship practices and a visible supply chain are necessary | | | | |

| | | | | | | |
|--|------------------------|--|------------------------|----------------------------------|------------------------|--|
| take into account platform users needs for traceability of who/what has accessed or modified data in a data supply chain | [Lee et al., 2017a] | take into account platform users needs for traceability of who/what has accessed or modified data in a data supply chain | | | | |
| strict monitoring processes, including reporting by users, should be considered to detect and notify all activities | [Lee et al., 2017a] | strict monitoring processes, including reporting by users, should be considered to detect and notify all activities | [Lee et al., 2019] | Track data usage | [Lee et al., 2019] | Data provenance |
| Data provenance: means to trace the derivation history of the data transparently for all participating group | [Lee et al., 2017a] | Data provenance: means to trace the derivation history of the data transparently for all participating group | [Lee et al., 2018] | | | |
| Metadata is a foundational element to provide transparency and visibility of the use of data | [Lee et al., 2017a] | Metadata is a foundational element to provide transparency and visibility of the use of data | | | | |
| Standardizing metadata should be considered for tracking information of the data in platform ecosystems | [Lee et al., 2017a] | Standardizing metadata should be considered for tracking information of the data in platform ecosystems | | | | |
| shared data or derived data through transformation or analysis has to be considered for the original data owner | [Lee et al., 2017a] | shared data or derived data through transformation or analysis has to be considered for the original data owner | [Lindman et al., 2014] | Recognize privacy sensitive data | [Lindman et al., 2014] | Maintain relationship with developer community |
| Continues review processes to determine personal identifiable information | [Lee et al., 2018] | Continues review processes to determine personal identifiable information | [Lee et al., 2018] | | | |
| Use, share and sell data without losing ownership | [Lee et al., 2018] | Use, share and sell data without losing ownership | | | | |
| Allow deletion of data | [Lee et al., 2018] | Allow deletion of data | | | | |
| Create decision process | [Lee et al., 2018] | Create decision process | | | | |
| Align with platform governance concepts | [Lee et al., 2018] | Align with platform governance concepts | | | | |
| Address all data types when making decisions | [Lee et al., 2018] | Address all data types when making decisions | | | | |
| Consider platform context | [Lee et al., 2018] | Consider platform context | | | | |
| Develop a decision model for data ownership | [Lee et al., 2018] | Develop a decision model for data ownership | | | | |
| Consider extensive regulations for non-user content | [Lee et al., 2018] | Consider extensive regulations for non-user content | | | | |
| Contribution management | [Lee et al., 2018] | Contribution management | | | | |
| Consider types of contributions for contribution management | [Lee et al., 2018] | Consider types of contributions for contribution management | | | | |
| Measure contributions | [Lee et al., 2018] | Measure contributions | | | | |
| Involve various participating groups for conformance | [Lee et al., 2018] | Involve various participating groups for conformance | | | | |
| Audit every data use case and its processes | [Lee et al., 2018] | Audit every data use case and its processes | | | | |
| Provide possible opportunities to stakeholders for monitoring | [Lee et al., 2018] | Provide possible opportunities to stakeholders for monitoring | | | | |
| Make data supply chain visible for monitoring | [Lee et al., 2018] | Make data supply chain visible for monitoring | | | | |
| Enable data owners to trace history of data use | [Lee et al., 2018] | Enable data owners to trace history of data use | | | | |
| Record all use | [Lee et al., 2018] | Record all use | | | | |
| Establish principles and policies | [Lee et al., 2018] | Establish principles and policies | | | | |
| Establish ownership and privacy policies | [Lee et al., 2018] | Establish ownership and privacy policies | | | | |
| Charactize data by ownership | [Lee et al., 2018] | Charactize data by ownership | | | | |
| Develop ownership definition model | [Lee et al., 2018] | Develop ownership definition model | | | | |
| Allow users to store analysis results | [Lee et al., 2018] | Allow users to store analysis results | | | | |
| Audit data and processes | [Lee et al., 2018] | Audit data and processes | | | | |
| Delete data and use history | [Lee et al., 2018] | Delete data and use history | | | | |
| Handle data owner provenance reports | [Lee et al., 2018] | Handle data owner provenance reports | | | | |
| Collaborate with other domains | [Lee et al., 2019] | Collaborate with other domains | | | | |
| Allow users to give security | [Lee et al., 2019] | Allow users to give security | | | | |
| Implement due process | [Lee et al., 2019] | Implement due process | | | | |
| Track data changes | [Lee et al., 2019] | Track data changes | | | | |
| Determine reward types (exposure/subsidy or reputation) | [Lee et al., 2019] | Determine reward types (exposure/subsidy or reputation) | | | | |
| Share data via excel or csv | [Lindman et al., 2014] | Share data via excel or csv | | | | |
| Process data before sharing | [Lindman et al., 2014] | Process data before sharing | | | | |
| Share raw data | [Lindman et al., 2014] | Share raw data | | | | |
| Share linked data | [Lindman et al., 2014] | Share linked data | | | | |
| Share humanreadable data | [Lindman et al., 2014] | Share humanreadable data | | | | |

| | | | | | |
|---|---------------------------|---|---------------------------|----------------------------------|--|
| Versioning of the API | [Lindman et al., 2014] | Versioning of the API | | | |
| Localize licensing conditions | [Veeneman et al., 2018] | Localize licensing conditions | | | |
| Create a single point of entry for the platform | [Veeneman et al., 2018] | Create a single point of entry for the platform | | | |
| Carefully consider communication of open data strategy to potential customers | [Buda et al., 2016] | Carefully consider communication of open data strategy to potential customers | | | |
| Allow users to fix errors in data | [Buda et al., 2016] | Allow users to fix errors in data | | | |
| Share data on company website | [Buda et al., 2016] | Share data on company website | | | |
| Share data via an actual platform | [Buda et al., 2016] | Share data via an actual platform | | | |
| Share data via an actual platform | [Buda et al., 2016] | Share data via an actual platform | [Parker and Alstyn, 2016] | Offer price subsidies as rewards | |
| Offer subsidies monetary subsidies for early users | [Parker and Alstyn, 2016] | Offer subsidies monetary subsidies for early users | | | |
| Offer non-monetary subsidies to early users (tech support i.e.) | [Parker and Alstyn, 2016] | Offer non-monetary subsidies to early users (tech support i.e.) | | | |
| Allow seeding (provide input to one side of the marketplace) | [Parker and Alstyn, 2016] | Allow seeding (provide input to one side of the marketplace) | | | |
| Focus on small connected community | [Parker and Alstyn, 2016] | Focus on small connected community | | | |
| Launch by pickybacking other network | [Parker and Alstyn, 2016] | Launch by pickybacking other network | | | |
| Create rules to maintain platform coherence | [Parker and Alstyn, 2016] | Create rules to maintain platform coherence | | | |
| Create congestion pricing for popular data | [Parker and Alstyn, 2016] | Create congestion pricing for popular data | | | |
| Create quality review for data | [Parker and Alstyn, 2016] | Create quality review for data | | | |
| Create scalable onboarding for new users | [Tiwana et al., 2010] | Create scalable onboarding for new users | | | |
| Identify API metrics | [Tiwana et al., 2010] | Identify API metrics | | | |
| Set usage quotas | [Tiwana et al., 2010] | Set usage quotas | | | |
| Set throttling limits for excessive use | [Tiwana et al., 2010] | Set throttling limits for excessive use | | | |
| Employ Oauth 2 and SAML for authentication | [Tiwana et al., 2010] | Employ Oauth 2 and SAML for authentication | | | |
| Do OpenID connect for auth | [Tiwana et al., 2010] | Do OpenID connect for auth | | | |
| Allow incorporation of various development environments | [Tiwana et al., 2010] | Allow incorporation of various development environments | | | |
| Developer sites | [Tiwana et al., 2010] | Developer sites | | | |
| Use a trusted supplier for security | [Tiwana et al., 2010] | Use a trusted supplier for security | | | |
| Create self-service capabilities | [Tiwana et al., 2010] | Create self-service capabilities | | | |

TABLE E.1: *Capabilities for Maturity Model found in Literature*

E.1 FIRST VERSION DAAPFAMM

| Focus Area | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | Level 6 | Level 7 | Level 8 | Level 9 | Level 10 |
|---------------------------|---------|---|--|---|---|--|---|--|--|--|
| Security | | Allow users to give security | Implements watermarking for providers | | Automated security | Provide security levels for providers | | | Use a trusted supplier for security | |
| Public Data Goals | | Establish research goal | | Match research problem with data characteristic | Use mutual adjustment to match data and problem | Check behaviour in line with goals | | Set strict guidelines for data use and punish misuse | | |
| Institutional Design | | | Develop principles for data platform | | | Develop organizational models | | Realign work practices to consider stakeholder value | | |
| Documentation | | Provide metadata | List types of datasets | Create advanced API materials | Support structural metadata | Create assistance tools for API | | Associate data types from various parties | Real-time data sharing | |
| Platform maintenance | | Maintain original platform setup | Add additional data sets | Model dependencies in data | Create transparent roadmap | Install data mapping suites for data types | | Versioning of the API | | |
| Decision Making Processes | | Establish decision making structures within an organization | Create internal process chart for establishing decision rights | Create transparent roadmap | Create standards and conventions and guidelines for API development | Create alignment processes for widespread involvement from users | | Establish decision rights | Partition decision rights between owner and users | Create a governance group to guide product development |
| Performance Strategy | | | Maintain centralized control strategy | | Decentralize platform control | | | Introduce formal data quality control mechanisms | | |
| Multi-homing strategy | | Allow users to multi-home | | | Prohibit multi-homing | | | | | |
| Contribution estimation | | | Measure contributions | | Create a contribution measurement model | | Consider types of contributions for contribution management | | | |
| Monitoring | | Provide possible opportunities to stakeholders for monitoring | | Create data stewardship practices | Allow deletion of data and use history | Create strict monitoring processes, including reporting by users | | a visible supply chain are necessary | Take into account platform users' needs for traceability | |
| Platform Measurements | | Identify API metrics | | Set usage quotas | | Set throttling limits for excessive use | Install data validation mechanisms | | Allow incorporation of various development environments | |

TABLE E.2: Initial version of DaapFAMM 1/3

| Area | Issue | Impact | Strategic Alignment | Key Objectives | Key Activities | Key Resources | Key Risks | Key Deliverables | Key Metrics | Key Stakeholders | Key Dependencies | Key Enablers | Key Constraints |
|-----------------|--|--|---|---|---|-------------------------------|------------------------------------|-------------------------------------|-------------|------------------|------------------|--------------|-----------------|
| Data Related | Data Quality | Reliability vs. costs trade-off | Keep data up to date | Check and indicate missing data | Make trade-offs between amount and quality of data | Create personal relationships | Create certification for providers | Adapt platform to specific use case | | | | | |
| | Data Standardization | Let IT determine data structures | Allow providers to determine data attributes | Support data integration with related datasets | Promote data standard governance | | | | | | | | |
| | Trustability | Create interconnectedity for the data | Support data integration with related datasets | Support data integration with related datasets | Consider consistency and integrity of the data cases | | | | | | | | |
| | Data use case | Generate data by ownership | Reflect on link between use cases and the data | Identify sources of data for data cases | Identify sources of data for data cases | | | | | | | | |
| | Data Processing | Process data before sharing | Show linked data | Show linked data | Show human-readable data | Share behavioral feedback | | | | | | | |
| | Partner Relationship | No relationship with users | Create feedback loops | Ensure diversity of data providers | Align incentives for participating with data users | | | | | | | | |
| | Partner Selection | Create review systems for data suppliers | Create incentive system | Create performance-based rewards | Perform partner screening | | | | | | | | |
| | Recruitment | Determine reward types (expense/subsidies) | Create a mutual coordination plan for co-creation | Facilitate knowledge sharing on supply and demand | Facilitate knowledge sharing on supply and demand | | | | | | | | |
| | Coordination | Take time to have if success/coordination providers and demand | Focus on small community | Launch by kickstart together network | Allow scaling (provide input to case side of the marketplace) | | | | | | | | |
| | Reward System | Offer non-monetary subsidies to early users (tech support) | Support revenue sharing | Determine revenue sharing based on data types | | | | | | | | | |
| Revenue sharing | Support revenue sharing | Enable social media sharing of data | Enable social media sharing of data | Create showcase section | | | | | | | | | |
| Involve ment | Carefully consider communication of opportunities to potential customers | Stock partners with opportunities/visibility | | | | | | | | | | | |

TABLE E.3: Initial version of DaaPFAMM 2/3

| Focus Area | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 | |
|---------------------|--|---|--|---|---|--|
| Platform Management | Security | Allow users to give security | Implements watermarking for providers | Match research problem with data characteristic | Automated security Use mutual adjustment to match data and problem | |
| | Public Data Goals | Establish research goal | | | | |
| | Institutional Design | | Develop principles for data platform | | | |
| | Documentation | Provide metadata | List types of datasets | Create advanced API materials | Support structural metadata | |
| | Platform maintenance | Maintain original platform setup | Add additional data sets | Model dependencies in data | Create transparent roadmap | |
| | Decision Making Process | Establish decision making structures within an organization | Create internal process chart for establishing decision rights | Create transparent roadmap | Create standards conventions and guidelines for API development | |
| | Performance Strategy | | | | | |
| | Multi-homing strategy | Allow users to multi-home | Maintain centralized control strategy | | Decentralize platform control Prohibit multi-homing Create a contribution measurement model accurately identifying, acknowledging and rewarding the contributors based on the platforms' incentive strategy | |
| | Contribution estimation | | Measure contributions | | | |
| | Monitoring | Provide possible opportunities to stakeholders for monitoring | | | Create data stewardship practices | Allow deletion of data and use history |
| Data Related | Platform Measurements Data integrity / Data Completeness / Data correctness / Data consistency / Data relevance / Data timeliness | Identify API metrics Reliability vs. costs trade-off | Keep data up to date Allow provider to determine data attributes | Set usage quotas Check and indicate missing data | Make trade-offs between amount and usability costs Promote data standard governance | |
| | Data Standardization | Let IT determine data structures | | | | |
| | Flexibility | Create interconnectivity for the data | the use cases and the relevant stakeholders of the data (including individual use case for each data category) should be clearly defined and linked. | Support data integration with related datasets | Data categories of platform ecosystems should include various sources of data (not only user content); i.e. data from users, systems and business processes | |
| | Data use case | Characterize data by ownership | | | | |
| | Data Processing | Share raw data | Process data before sharing | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

FIGURE E.1: Focus Area and Practices found in the Literature Review Top Left

| Focus Area | Level 6 | Level 7 | Level 8 | Level 9 | Level 10 |
|---------------------|--|--|---|--|---|
| Platform Management | Security | | | Use a trusted supplier for security | |
| | Public Data Goals | Provide security levels for providers | | Set strict guidelines for data use and punish misuse | |
| | Institutional Design | Check behaviour in line with goals | | Realign work practices to consider stakeholder value | |
| | Documentation | Develop organizational models | | Associate data types from various parties | Real-time data sharing |
| | Platform maintenance | Create assistance tools for API | | Versioning of the API | |
| | Decision Making Process | Install data mapping tables for data types | | Establish decision rights | Partition decision rights between owner and users |
| | Performance Strategy | Create alignment processes for widespread involvement from users | | Introduce formal data quality control mechanisms | Create a governance group to guide product development |
| | Multi-homing strategy | | | | |
| | Contribution estimation | | Consider types of contributions for contribution management | | |
| | Monitoring | strict monitoring processes, including reporting by users, should be considered to detect and notify all activities | | | take into account platform users: needs for traceability of who/what has accessed or modified data in a data supply chain |
| Data Related | Platform Measurements Data integrity/ Data completeness / Data consistency / Data relevance / Data timeliness | Set throttling limits for excessive use | | a visible supply chain are necessary | Allow incorporation of various development environments |
| | Data Standardization | Create portability for the data | | Create personal relationships | Create certification for providers |
| | Flexibility | consistency and integrity of the cases must be considered over the lifecycle of the data to support a visible data supply chain. This factor supports both data ownership and data usage governance domains. | | | Adapt platform to specific use case |
| | Data use case | Share humanreadable data | | | Audit every data use case and its processes |
| | Data Processing | Continuous collaboration | | Create behavioural rewards | Create working groups and review committees to support partners |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |

FIGURE E.2: Focus Area and Practices found in the Literature Review Top Right

| Focus Area | Level 1 | Level 2 | Level 3 | Level 4 | Level 5 |
|--------------------------|------------------------------------|---|--|---|--|
| Partner Relationship | Partner Relationship | No relationship with users | Create feedback loops | | Frequent communication |
| | Partner Selection | | | Ensure diversity of data providers | Perform partner screening |
| | Incentivisation | Determine reward types (exposure/subsidy or reputation) | Create incentive system | Create performance-based rewards schemes | |
| | Coordination | Take 'take it or leave it' stance towards connecting providers and users | Create a mutual coordination plan for connecting supply and demand | Facilitate knowledge sharing on supply and needs | Regulate collaboration |
| | Reward System | Offer non-monetary subsidies to early users (tech support i.e.) | Focus on small connected community | Launch by pickybacking other network | Allow seeding (provide input to one side of the marketplace) |
| Compliance | Revenue sharing | Support revenue sharing | | Determine revenue sharing based on data types | |
| | Involvement | Carefully consider communication of open data strategy to potential customers | Seek partners with organisational similarity | Enable social media sharing of data | Create a showcase section |
| | Compliance and External Regulation | | | Regulate intellectual property | Determine punishments for breaches |
| | Internal Compliance | Survey regulations | | create codes of conducts | Educate end users on privacy permissions |
| Privacy | | Create terms of service | Anonymize and aggregate privacy sensitive data | Design arrangement for storage personal data | |
| | | Create privacy agreements | | | |
| Platform Data Management | Type of Data Sharing | Establish a data exchange | | Informal contracts | Create contracts for data use |
| | Data Access | Create self-service capabilities | Support programmatic access to data resources | Streamline data access | |
| | Governance Control over Data | Allow the ecosystem to self-regulate No control over data is exerted | Do only technical maintenance | Promote governance standards Control determined by dominant marketplace | Configure governance structure |
| | Data Sharing Technology | Share data via excel or csv | Share data on company website | encourage high-quality resources and to repress poor-quality resources of platform ecosystems | Share data via an actual platform |
| | Data Processing | | Track data changes | Create practices of data disclosure and screening | |
| Platform Functionalities | Data provenance | | Employ OAuth 2 and SAML for authentication | Promote access governance | Record all use |
| | Authentication | Assess risk of access policy | | | |
| | Documentation | Provide metadata | List types of datasets | Create advanced API materials | Support structural metadata |
| | User Contributions | Provide search on metadata level | Allow uploading of datasets | Allow for filtering search results | Allow users to fix errors in data |
| | Search functionality | Provide search on metadata level | Allow search on content level | Allow for filtering search results | |
| Pricing Mechanism | Process transactions | Create pricing system | | | Enable usage-based billing |

FIGURE E.3: Focus Area and Practices found in the Literature Review Bottom Left

| Focus Area | Level 6 | Level 7 | Level 8 | Level 9 | Level 10 | |
|--------------------------|------------------------------|---|---|---|---|---|
| Partner Related | Partner Relationship | | | | | |
| | Partner Selection | Align incentives for data sharing with data users | Creating new avenues and means for cooperation for stakeholders | | | |
| | Incentivisation | Set up conflict resolution process | | Offer subsidies monetary subsidies for early users | | |
| | Coordination | | | | | |
| | Reward System | | | | | |
| | Revenue sharing | | | Set up revenue sharing according to Responsible, Accountable, Consulted, and Informed (RACI) approach | | |
| | Involvement | Offer data playbooks, tutorials and describe data sharing technique | | Allow Piloting | Create sandboxes | |
| | Compliance | Compliance and External Regulation | Set-up due processes for data sharing | Negotiate exchange terms with users | Appoint a data steward | Implement risk management Create legal agreements for API use Enforce legal agreement |
| | | Internal Compliance | | | | |
| | | Privacy | Get informed consent for personal data | Create guardrails to support privacy protection | | Allow different forms of consent establishment for personal data sharing |
| Platform Data Management | Type of Data Sharing | | Enable Data Pooling | | Use Open licensing | |
| | Data Access | | | | | |
| | | Create scalable onboarding for new users | | | | |
| | Governance Control over Data | | | | | |
| | | Individual organization retain | Regulate behaviour through reputational measures | Align with platform governance concepts Offer users control over data | | Localize licensing conditions |
| | Data Sharing Technology | Create SDKs | | | | |
| | Data Processing | Clean data before sharing | | Implement a review board for control | | |
| | Data provenance | | Enable data owners to trace history of data use | | Handle data owner provenance reports | Opt-in and opt-out of data use collection |
| | Authentication | | | Enable delegation of credentials | Do OpenID connect for auth | |
| | Platform Functionalities | Documentation | | | | |
| User Contributions | | Create assistance tools for API | | Associate data types from various parties | | |
| Search Functionality | | Govern alterations from actors Recommend datasets | | | | |
| Pricing Mechanism | | Recommend datasets | Create congestion pricing for popular data | | Create spot market for short transactions | |

FIGURE E.4: Focus Area and Practices found in the Literature Review Bottom Left



INTERVIEW UPDATES MATURITY MODEL

F.1 INTERVIEW AGREEMENT ON MATURITY MODEL

F.2 CHANGE LOG MATURITY MODEL

| Focus Area | Practice | From | To | Reason | Inter- vie- wees |
|------------------|---------------------------------------|------|----|--|------------------------|
| Security | Allow users to give security | 2 | X | Unclear | 2 |
| Security | Implements watermarking for providers | 3 | X | Not really security related | 2 |
| Security | Implement automated security | 5 | X | Too vague | 2 |
| Security | Provide security levels for providers | 6 | X | Too vague | 2 |
| Security | Use a trusted supplier for security | 9 | 3 | Should be done way earlier | 2 |
| Public Data Goal | X | FA | X | Merged with Data Use Case | - |
| Public Data Goal | Check behaviour in line with goals | 6 | X | Merge with 9 into new 6 | 2,3 |
| Documentation | Provide metadata | 2 | X | merge with PDF/Word into new 2 | 5,8 |
| Documentation | Support structural metadata | 5 | 4 | Here the model can become slimmer, not a big difference with level 3 | 5 |

| | | | | | |
|-------------------------|---|----|---|--|-------|
| Documentation | Implement a complete Developer Platform with documentation and API descriptions | 6 | 5 | Here the model can become slimmer, not a big difference with level 5 | 5 |
| Documentation | Deploy API data online through Swagger or similar platforms | 4 | 6 | This is really a nice to have and can be put later | 5 |
| Documentation | Create assistance tools for API | 6 | 7 | Also really a nice to have | 5 |
| Documentation | Associate data types from various parties | 8 | X | Unclear | - |
| Platform Maintenance | Maintain original platform setup | 2 | 1 | This is the 'do nothing' step, so at level 1 | 2 |
| Performance Strategy | X | FA | X | Integrate in coordination | 3,5 |
| Multi Homing Strategy | X | FA | X | Integrate in internal regulations | 3,5 |
| Contribution estimation | X | FA | X | Integrate in incentivisation | 2,3,5 |
| Monitoring platform | Create alerts when the platform is down for users | - | 4 | Just measuring doesn't do anything alerting helps you acknowledge problems | 5 |
| Platform Measurements | Set usage quotas | 4 | 3 | This you really should do earlier | 2,5 |
| Platform Measurements | Warnings and eventually blocking for quota breaking | 5 | 4 | This you really should do earlier | 2,5 |
| Platform Measurements | Throttle abuse of data quotas and make a fair use assessment | 6 | 5 | This you really should do earlier | 2,5 |
| Platform Measurements | Flexible fair use agreement based on user/datetime | 8 | 7 | This you really should do earlier | 2,5 |
| Data Quality | Make reliability vs. costs trade-off | 2 | 3 | Can't have two things in the same place | - |
| Data Quality | Keep data up to date | 3 | 2 | This you should do really early on, nobody is paying for useless data | 5 |
| Data Standardization | Create new advanced products with data | 10 | 9 | From GraphQL, just an example | 8 |
| Flexibility | X | FA | X | Flexibility is something to integrate in several Fas | 2,3,5 |
| Data use case | Create data category based security protocols | 7 | X | This one is very unclear | 2 |
| Data use case | Characterize data by ownership | 2 | X | This one is very unclear | 2 |

| | | | | | |
|----------------------|--|----|---|--|-----|
| Data use case | Adapt platform to specific use case | 10 | 3 | A platform should be well aware of the use cases of its customers | 2,5 |
| Data use case | Audit every data use case and its processes | 9 | 2 | A platform should be well aware of the use cases of its customers | 2,5 |
| Data use case | Define and link the use cases and the relevant stakeholders of the data (including individual use case for each data category) | 3 | 5 | Can't have two things in the same place | - |
| Data use case | Include various sources of data for data categories of platform ecosystems (not only user content): i.e. data from users, systems and business processes | 5 | X | Very unclear | 2 |
| Data use case | Establish research goal | 2 | 4 | Merged from Public data goal | 2 |
| Data use case | Match research problem with data characteristic | 4 | 6 | Merged from Public data goal | 2 |
| Data use case | Use mutual adjustment to match data and problem | 5 | 8 | Merged from Public data goal | 2 |
| Data Processing | Share raw data | 1 | X | You really can't monetize raw data | 2 |
| Partner Relationship | Frequent communication | 5 | X | Practice under Partner relationship | - |
| Partner Relationship | Ensure continuous collaboration with users | 7 | 4 | Can't have two things in the same place | - |
| Partner Relationship | Create working groups and review committees to support users | 9 | 5 | Je users meenemen in een soort board kan veel eerder al | 2 |
| Partner selection | X | FA | X | Merged with Partner Relationship | 2 |
| Incentivisation | Determine reward type (exposure/subsidy or reputation) | 2 | X | This is barely something you can implement | 5 |
| Incentivisation | Create Incentive System | 3 | 5 | Very advanced | 2 |
| Incentivisation | Create Performance-based rewards scheme | 4 | 6 | Very advanced | 2 |
| Incentivisation | Create behavioral rewards | 9 | 8 | More incentive than relationship related | - |
| Incentivisation | Set up revenue sharing or other incentivisation scheme for data supplier partners | 5 | X | From Revenue sharing, other methods of incentivisation can be used too | 5 |

| | | | | | |
|------------------------------------|---|--------|----|---|-------|
| Coordination | Determine level of centralization of the platform | X | 9 | From the Performance Strategy FA | - |
| Reward System | X | FA | X | Merged | 2,5,8 |
| Revenue sharing | X | FA | X | Merged | 2,5,8 |
| Onboarding | Seek partners with organisational similarity | 3 | X | No double entries | - |
| Onboarding | Enable social media sharing of data | 5 | X | Not relevant for onboarding | 5 |
| Compliance and External Regulation | Ensure compliance with applicable regulation | X | 1 | From all the GDPR points, you have to be GDPR compliant, no matter the maturity | 5 |
| Compliance and External Regulation | Appoint a data steward | 5 | 3 | This is very important for data platforms | 5 |
| Internal Regulation | Create terms of service and code of conduct | 2 or 5 | 1 | Can't share data without a code of conduct or terms of service | 2,5 |
| Internal Regulation | Create legal agreements for API use | 9 | X | Falls under different thing and is essential at level 1 | 2 |
| Internal Regulation | Enforce legal agreement | 10 | X | Falls under different thing and is essential at level 1 | 2 |
| Privacy | Create privacy agreements | 3 | 1 | Merged into Ensure GDPR compliance | 5 |
| Privacy | Anonimize and aggregate privacy sensitive data | 4 | 1 | Merged into Ensure GDPR compliance | 5 |
| Privacy | Design arrangement for storage personal data | 5 | 1 | Merged into Ensure GDPR compliance | 5 |
| Privacy | Educate end users on privacy permissions | 3 | 2 | Almost directly needed | 5 |
| Type of Data Sharing | Informal contracts | 4 | X | Not a thing, contracts need to be formal | 2,5 |
| Data Access | X | FA | X | Merged in incentives | - |
| Governance | Allow the ecosystem to self-regulate | 1 | 10 | This is dream scenario | 2,5 |
| Governance | Configure governance structure | 5 | X | No double entries | - |
| Governance | Localize licensing conditions | 9 | 4 | Localizing is only translations, that should be done earlier | 2 |
| Data processing | Create practices of data disclosure and screening | 4 | X | X | - |

| | | | | | |
|----------------------|--|----|---|---|-----|
| Data processing | Implement a review board for control | 8 | 4 | X | 2 |
| Data provenance | X | FA | X | Not sure what this is | 2,8 |
| Control over data | Make individual organization retain control over data | 5 | 2 | This is how this usually happens | 2,8 |
| Control over data | Offer users control over data | 5 | X | Not done | 2,8 |
| Control over data | Develop a decision model for data ownership | 3 | X | Unclear | 2 |
| Control over data | Describe data ownership in various scenarios in the platform | 4 | Ownership depends on the exact way the data is acquired | 2,8 | |
| User contributions | Allow users to fix errors in data | 5 | X | This is not done, feedback is fine but a fix on the database is undesirable | 2,8 |
| User contributions | Allow uploading of datasets | 3 | 8 | Very advanced | 2 |
| Search Functionality | Provide search on metadata level | 2 | 1 | This is the absolute minimum, finding a dataset | 2,5 |
| Search Functionality | Allow search on content level | 3 | 2 | This is the almost the minimum, finding a dataset | 2,5 |
| Search Functionality | Assist connecting database by implementing pathfinding | X | 8 | X | 5 |
| Pricing mechanisms | Determine ideal business model for platform context | X | 3 | X | 5 |
| Pricing mechanisms | Enable basic payment for access | X | 1 | X | 5 |
| Monitoring data | Allow deletion of historic data | 5 | X | Not done | 2,5 |
| Data Standardization | Support data integration with related datasets | 7 | 6 | Slim the model | - |
| Data Standardization | Format data according to users | 8 | 7 | Slim the model | - |
| Data Standardization | Create new advanced products with data | 9 | 8 | Slim the model | - |

| | | | | | |
|------------------------------------|---|----|---|---|-----|
| Data Standardization | Create flexible data structures | 10 | 9 | Slim the model | - |
| Onboarding | Create sandboxes | 10 | X | 2 was found the right level | 2 |
| Governance | Allow the ecosystem to self-regulate | 1 | 9 | Is a utopy | 2,5 |
| Security | Implement use case specific state-of-the-art security system | 9 | 8 | Slim the model | - |
| Decision Making Proces | Establish decision rights | 7 | X | Combine with Partition decision rights | 5 |
| Decision Making Proces | Create a governance group to guide product development | 9 | 7 | This is not more advanced than partition rights | 2 |
| Monitoring Data | Establish a visible supply chain for the data | 8 | 7 | Slim the model | - |
| Monitoring Data | Take into account platform users' needs for traceability of who/what has accessed or modified data in a data supply chain | 9 | 8 | Slim the model | - |
| Data Standardization | Create flexible data structures | 9 | 8 | Slim the model | - |
| Incentivisation | Create behavioral rewards for users | 8 | 7 | Slim the model | - |
| Incentivisation | Set up revenue sharing or other incentivisation scheme for data supplier partners | 9 | 8 | Slim the model | - |
| Governance | Determine level of centralization of the platform | 9 | 7 | Slim the model | - |
| Compliance and External Regulation | Implement risk management | 8 | X | Vague | 2 |
| Compliance and External Regulation | Create Responsible Disclosure processes | 9 | 8 | Slim the model | - |
| Privacy | Allow different forms of consent establishment for personal data sharing | 9 | 8 | Slim the model | - |
| Type of Data Sharing | Use Open licensing | 9 | 8 | Slim the model | - |
| Governance | Align with platform governance concepts | 8 | X | Vague | 2 |
| Governance | Allow the ecosystem to self-regulate | 9 | 8 | Slim the model | - |

| | | | | | |
|-------------------------|---|----|---|---|-----|
| Pricing Mechanism | Create spot market for short transactions | 9 | 8 | Slim the model | - |
| Monitoring Data | Take into account platform users' needs for traceability of who/what has accessed or modified data in a data supply chain | X | 8 | Same as establish a data supply chain | - |
| Monitoring Data | Establish a visible supply chain for the data | 7 | 8 | Take position of the other supply chain element | - |
| Governance | Allow the ecosystem to self-regulate | 1 | X | Belongs on the end | 2,5 |
| Data Sharing Technology | Sell data on an open data marketplace | 4 | 3 | X | - |
| Data Sharing Technology | Share data via an actual platform | 5 | 4 | X | - |
| Data Sharing Technology | Create API | 3 | X | Merged with platform | - |
| Monitoring Platform | Create a post-mortum that can replace the system with a snapshot when the service is down | 8 | 7 | X | 5,6 |
| Monitoring Platform | Run two versions of the platform, where one takes over when the other crashes | X | 8 | X | 5,6 |
| User Contributions | Allow feedback on data | 7 | X | Already in "Monitoring the Data" | - |
| User Contributions | Govern alterations from actors | 6 | X | "Suggestions okay, but this is not done" | 2,5 |
| User Contributions | FA | FA | X | It's empty | - |

TABLE F.12: *Change Log Maturity Model during Interviews*

| Capability | Practice | Level | Source | Description | I2 | I3 | I4 | I5 | I6 | I7 | I8 | In favour | Against | Score |
|---------------|---|-------|--------|--|----|----|----|----|----|----|----|-----------|---------|-------|
| Security | Allow access without authentication by anonymous users | 1 | I2 | Allow direct access to the platform without any authentication practice in place. | 1 | 1 | 1 | 1 | 0 | | | 4 | 1 | 80% |
| Security | Perform authentication through username password / API key | 2 | I2 | Ensure that users first have to make a Username and Password account or have to request an API key before they can access the platform. | 1 | 1 | 1 | 1 | 0 | | | 4 | 1 | 80% |
| Security | Use a trusted supplier for platform security | 3 | [31] | Install security measures for the platform, as provided by a trusted supplier specialized in security. | 1 | 1 | 1 | 1 | 0 | | | 4 | 1 | 80% |
| Security | Perform user-based API authorization | 4 | I2 | Authorization mechanisms are in place to give users access to specific parts or datasets on the platform | 1 | 1 | 1 | 1 | 0 | | | 4 | 1 | 80% |
| Security | Implement use case specific best-practice authentication methods | 5 | I2 | Figure out what the best-practice is for user authentication in the industry or for the use-case of the platform and implement that best-practice. | 1 | 1 | 1 | 1 | 0 | | | 4 | 1 | 80% |
| Security | Connect Identity Access Management and Authorization system | 7 | I2 | Centralize the management of User authentication and authorization to the platform in a single system | 1 | 1 | 1 | 1 | | | | 4 | 0 | 100% |
| Security | Implement a use case specific state-of-the-art security system | 8 | I2 | Hire security experts to implement a state-of-the-art security system fitting to the use-case and industry standard of the platform | 1 | 1 | 1 | 1 | | | | 4 | 0 | 100% |
| Documentation | Provide no documentation | 1 | I2 | There is no documentation in the platform. | 1 | 1 | 1 | 1 | 1 | | 1 | 6 | 0 | 100% |
| Documentation | Provide document with basic API description and metadata | 2 | I2 | A PDF or Word documentation is created and maintained with information on the data in the data platform. | 1 | 1 | 1 | 1 | 1 | | 1 | 6 | 0 | 100% |
| Documentation | List types of datasets | 3 | [16] | Describe the various types of datasets available on the platform. | | 1 | 1 | 1 | 1 | | 1 | 5 | 0 | 100% |
| Documentation | Support structural metadata | 4 | [18] | Add structural metadata to the platform to support developers' data access. | 1 | 1 | 1 | 1 | 0 | | 1 | 5 | 1 | 83% |
| Documentation | Implement a complete Developer Platform with documentation and API descriptions | 5 | I2 | Add a developer platform to the platform detailing documentation, registration processes and other important information for the platform. | 1 | 1 | 1 | 1 | 1 | | 1 | 6 | 0 | 100% |
| Documentation | Deploy API data through Swagger or similar platforms | 6 | I2 | Use Swagger to create automated and standardized API materials | 1 | 1 | 1 | 1 | 1 | | 1 | 6 | 0 | 100% |
| Documentation | Create assistance tools for API | 7 | [19] | Create assistance tools that help developers with using the API and checking the results of API calls. | | 1 | 1 | 1 | 0 | | 1 | 4 | 1 | 80% |
| Documentation | Provide living documentation | 8 | I2 | Provide living documentation that automatically updates after API developments | 1 | 1 | 1 | 1 | 1 | | 1 | 6 | 0 | 100% |

TABLE F.1: Interviewees in favour on capabilities Security and Documentation

| Capability | Practice | Level | Source | Description | I2 | I3 | I4 | I5 | I6 | I7 | I8 | In favour | Against | Score |
|-------------------------|---|-------|--------|---|----|----|----|----|----|----|----|-----------|---------|-------|
| Platform Maintenance | Maintain original platform setup | 1 | [16] | No maintenance is done on the platform it just stayed as published initially. | 1 | 1 | 1 | 1 | | | 1 | 5 | 0 | 100 |
| Platform Maintenance | Add additional datasets | 3 | [16] | Additional datasets are added to keep the data up to date. | 1 | 1 | 1 | 1 | | | 1 | 5 | 0 | 100 |
| Platform Maintenance | Manage versioning of the API | 4 | [26] | The API is versioned for each new version to show users that they need to update. | 1 | 1 | 1 | 1 | | | 1 | 5 | 0 | 100 |
| Platform Maintenance | Model dependencies in data | 5 | [21] | Dependencies between data is modelled to ensure that up to date data and aged data aren't miscombined. | 1 | 1 | 1 | 1 | | | 1 | 5 | 0 | 100 |
| Platform Maintenance | Install data mapping tables for data types | 6 | [14] | Data mapping tables are created to map various data types of different datasets. | 1 | 1 | 1 | 1 | | | 1 | 5 | 0 | 100 |
| Platform Maintenance | Manage versioning of the data | 7 | [26] | Manage the versioning of the data, detailing the last time it was updated and its correctness. | 0 | 1 | 1 | 1 | | | 1 | 4 | 1 | 80 |
| Platform Maintenance | Establish grace period for updates | 8 | I2 | The API is versioned for each new version to show users that they need to update. | 1 | 1 | 1 | 1 | | | 1 | 5 | 0 | 100 |
| Decision Making Process | Establish decision making structures within an organization | 2 | [15] | Create a basic process for decision making in the organization. | 1 | 1 | 1 | 1 | 1 | | 1 | 6 | 0 | 100 |
| Decision Making Process | Create internal process chart for establishing decision rights | 3 | [15] | Create a decision process chart that establishes decision rights from various roles inside the company. | 1 | 1 | 1 | 1 | 1 | | 1 | 6 | 0 | 100 |
| Decision Making Process | Create transparent roadmap | 4 | [15] | A transparent roadmap for new functionalities is created on the platform. | 1 | 0 | 1 | 1 | 1 | | 1 | 5 | 1 | 83 |
| Decision Making Process | Create standards, conventions and guidelines for platform development | 5 | [15] | Focus on using best-practices and industry-standards when making decisions | 1 | 0 | 1 | 1 | 1 | | 1 | 5 | 1 | 83 |
| Decision Making Process | Create alignment processes for widespread involvement from users | 6 | [15] | Align decision with users to ensure widespread involvement in the platform. | 1 | 0 | 1 | 1 | 1 | | 1 | 5 | 1 | 83 |
| Decision Making Process | Create a governance group to guide product development | 7 | [15] | Create a group inside the company that guides both users as the orchestrator in their decision making and that helps understand the other side. | 1 | 0 | 1 | 1 | 1 | | 1 | 5 | 1 | 83 |
| Decision Making Process | Partition decision rights between owner and users | 8 | [23] | Make a fair partition for decision rights between users and the orchestrator of the platform. | 1 | 1 | 1 | 1 | 1 | | 1 | 6 | 0 | 100 |

TABLE F.2: Interviewees in favour on capabilities Platform Maintenance and Decision Making Process

| Capability | Practice | Level | Source | Description | I2 | I3 | I4 | I5 | I6 | I7 | I8 | In favour | Against | Score |
|---------------------|---|-------|--------|--|----|----|----|----|----|----|----|-----------|---------|-------|
| Monitoring Data | Provide opportunities to stakeholders for monitoring data | 2 | [24] | Allow stakeholders of the platform to monitor their part of the platform. | | | 1 | 1 | 1 | | 1 | 4 | 0 | 100 |
| Monitoring Data | Create data stewardship practices | 4 | [23] | Advance the platform with data stewardship practices from the main company. | | | 1 | 0 | 1 | | 1 | 3 | 1 | 75 |
| Monitoring Data | Create strict monitoring processes, including reporting by users | 6 | [13] | Allow users to report abuse of the platform and strictly monitor every action on the platform to corroborate this. | | | 1 | 1 | 1 | | 1 | 4 | 0 | 100 |
| Monitoring Data | Establish a visible supply chain for the data | 8 | [24] | Create a visible supply chain for the data showing where it is coming from. | | | 1 | 1 | 1 | | 1 | 4 | 0 | 100 |
| Monitoring Platform | Implement heartbeat test for platform users | 2 | I2 | Allow users to easily test if the platform is up. | 1 | | 1 | 1 | 1 | | 1 | 5 | 0 | 100 |
| Monitoring Platform | Measure platform uptime | 3 | I2 | Measure the uptime of the platform to ensure SLAs. | 1 | | 1 | 1 | 1 | | 1 | 5 | 0 | 100 |
| Monitoring Platform | Create alerts for users when the platform is down | 4 | I4 | Proactively alert users when functionalities of the platform are down to manage downtime. | | | 1 | 1 | 1 | | 1 | 4 | 0 | 100 |
| Monitoring Platform | Create a status website that shows issues | 5 | I2 | Create a website that gives an overview of the datasets that are up. | 1 | | 1 | 1 | 1 | | 1 | 5 | 0 | 100 |
| Monitoring Platform | Create a post-mortum that can replace the system with a snapshot when the service is down | 7 | I2 | Create an advanced post mortum system that takes over when the platform is down. | 1 | | 1 | 1 | 1 | | 1 | 5 | 0 | 100 |
| Monitoring Platform | Run two versions of the platform, where one takes over when the other crashes | 8 | I4 | Run synchronized versions of the platform where a second can take over when the first one is down. | | | 1 | 1 | 1 | | 1 | 4 | 0 | 100 |

TABLE F.3: Interviewees in favour on capabilities Monitoring Data and Monitoring Platform

| Capability | Practice | Level | Source | Description | I2 | I3 | I4 | I5 | I6 | I7 | I8 | In favour | Against | Score |
|-----------------------|--|-------|--------|---|----|----|----|----|----|----|----|-----------|---------|-------|
| Platform Measurements | Identify API metrics | 2 | [31] | Come up with API metrics that quantify the performance of the API | 1 | | 1 | 0 | | | 1 | 3 | 1 | 75 |
| Platform Measurements | Set usage quotas | 3 | [31] | Set a quota for the amount of data users are allowed to consume in a period of time. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Platform Measurements | Warnings and eventually blocking for quota breaking | 4 | I2 | Enforce the quotas by warning and eventually blocking irregular use. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Platform Measurements | Throttle abuse of data quotas and make a fair use assessment | 5 | I2 | Throttle usage that closes in on quotas to prevent excessive use. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Platform Measurements | Flexible fair use agreement based on user/datetime | 7 | [31] | Create flexible SLAs that allow for throttling based on overall platform capabilities. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Data Quality | Keep data up to date | 2 | [14] | Update data when it becomes out of data or when new data is available. | | 1 | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Data Quality | Make reliability vs. costs trade-off | 3 | [19] | Trade off reliability of the data with the costs required to ensure it. | | 1 | 1 | 0 | | | 1 | 3 | 0 | 75 |
| Data Quality | Check and indicate missing data | 4 | [14] | Give indicators for missing values in the dataset to make sure users handle them correctly. | | 1 | 1 | 0 | | | 1 | 3 | 0 | 75 |
| Data Quality | Make trade-offs between amount of data and usability costs | 5 | [19] | Closely consider the amount of data that is shared and the usability of the data. | | 1 | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Data Quality | Install data validation mechanisms | 7 | [4] | Create or install systems that validate the data in the system and that warns for irregular values. | | 1 | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Data Quality | Perform reconciliation | 8 | I8 | Set-up a reconciliation mechanisms to continuously check the correctness of stored data | | | | | | | 1 | 1 | 0 | 100 |

TABLE F.4: Interviewees in favour on capabilities Platform Measurement and Data Quality

| Capability | Practice | Level | Source | Description | I2 | I3 | I4 | I5 | I6 | I7 | I8 | In favour | Against | Score |
|----------------------|---|-------|--------|--|----|----|----|----|----|----|----|-----------|---------|-------|
| Data Standardization | Let IT determine data structures | 1 | [14] | Let the IT infrastructure determine the type of data, what is the easiest. | 1 | | | 1 | | | 1 | 3 | 0 | 100 |
| Data Standardization | Allow provider to determine data attributes | 3 | [9] | Let the provider of the data determine the data types and attributes. | 1 | | | 1 | | | 1 | 3 | 0 | 100 |
| Data Standardization | Promote data standard governance | 5 | [9] | Create a data standard and promote and govern its use. | | | | 1 | | | 1 | 2 | 0 | 100 |
| Data Standardization | Support data integration with related datasets | 6 | [10] | Compare data with related datasets and support data type integration with those sets. | | | | 1 | | | 1 | 2 | 0 | 100 |
| Data Standardization | Format data according to users' needs | 7 | [12] | Allow users to determine how the data is structured. | | | | 1 | | | 1 | 2 | 0 | 100 |
| Data Standardization | Create flexible data structures | 8 | [12] | Create flexible data structures that allow users to get access to various types of the same data. | 1 | | | 1 | | | 0 | 2 | 1 | 67 |
| Data Use Case | Audit every data use case and its processes | 2 | [24] | Very specifically audit how data is used to achieve the use cases at users in order to promote it. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Data Use Case | Adapt platform to specific use case | 3 | [16] | Focus the platform on a specific use case. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Data Use Case | Set usage goal for data | 4 | I2 | Set the goal with which the data is shared in the platform, so participating companies can conform to it. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Data Use Case | Define and link the use cases and the relevant stakeholders of the data | 5 | [23] | Identify use cases of the data and link them to relevant stakeholders to promote the data. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Data Use Case | Discuss data usage goal in terms of service | 6 | I2 | Set the goal with which the data is shared in the platform, so participating companies can conform to it. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Data Use Case | Create new advanced products with data | 7 | I8 | Create data dissemination products to facilitate specific users by tailoring the data output to their needs. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Data Use Case | Monitor if the data use case is met by users and punish misuse | 8 | I2 | Monitor the use of data by the users continuously and check if it is in line with the data use case. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |

TABLE F.5: Interviewees in favour on capabilities Data Standardization and Data Use Case

| Capability | Practice | Level | Source | Description | I2 | I3 | I4 | I5 | I6 | I7 | I8 | In favour | Against | Score |
|----------------------|---|-------|--------|--|----|----|----|----|----|----|----|-----------|---------|-------|
| Partner Relationship | Maintain no relationship with users | 1 | [16] | No relationship with users. | | | 1 | 0 | | | 1 | 2 | 1 | 67 |
| Partner Relationship | Perform user screening | 2 | [12] | Screening is performed on partners to ensure that only trusted partners are involved. | 1 | | 1 | 0 | | | 1 | 3 | 1 | 75 |
| Partner Relationship | Create feedback loops | 3 | [4] | Feedback loops are created that allow users to give input on the platform. | | | 1 | 0 | | | 1 | 2 | 1 | 67 |
| Partner Relationship | Ensure continuous collaboration with users | 4 | [1] | There is a continuous joined collaboration with the users to improve the platform. | | | 1 | 0 | | | 1 | 2 | 1 | 67 |
| Partner Relationship | Create working groups and review committees to support users | 5 | [15] | Groups are created that support and promote the user perspective from within the orchestrator company. | 1 | | 1 | 0 | | | 1 | 3 | 1 | 75 |
| Partner Relationship | Create new avenues and means for cooperation for stakeholders | 7 | [9] | New or additional ways of collaborating with the platform users are explored. | | | 1 | 0 | | | 1 | 2 | 1 | 67 |
| Partner Relationship | Create review systems for data suppliers | 8 | [12] | Data suppliers are reviewed through a system before their data is shared. | 1 | | 1 | 0 | | | 1 | 3 | 1 | 75 |
| Incentivisation | Offer non-monetary subsidies to early users | 2 | [30] | Offer early adoptors of the platform additional non-monetary benefits. | 1 | | 1 | 1 | | | | 3 | 0 | 100 |
| Incentivisation | Provide seeding (input to one side of the marketplace) | 3 | [30] | Provide input on the demand side to promote new data supplier to join. | 1 | | 1 | 1 | | | | 3 | 0 | 100 |
| Incentivisation | Create incentive system for users | 5 | [9] | Create a system that incentivizes data use by the users. | 1 | | 1 | 1 | | | | 3 | 0 | 100 |
| Incentivisation | Create performance-based rewards schemes for users | 6 | [15] | Create advanced transparent reward schemes that promote good performance. | 1 | | 1 | 1 | | | | 3 | 0 | 100 |
| Incentivisation | Create behavioral rewards for users i.e. through gamification | 7 | [15] | Behavioral rewards are created to reward supporting platform users. | 1 | | 1 | 1 | | | | 3 | 0 | 100 |
| Incentivisation | Set up revenue sharing or other incentivisation scheme for data supplier partners | 8 | [23] | Create a scheme for sharing revenue between users. | 1 | | 1 | | | | 0 | 2 | 1 | 67 |

TABLE F.6: Interviewees in favour on capabilities Partner Relationship and Incentivisation

| Capability | Practice | Level | Source | Description | I2 | I3 | I4 | I5 | I6 | I7 | I8 | In favour | Against | Score |
|--------------|---|-------|--------|--|----|----|----|----|----|----|----|-----------|---------|-------|
| Coordination | Take 'take it or leave it' stance towards connecting providers and users | 1 | [9] | Merely make the data available. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Coordination | Create a mutual coordination plan for connecting supply and demand | 2 | [9] | Create a coordination plan to ensure the demanded data can be become available. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Coordination | Create communication guidelines between collaborators | 3 | [12] | Create communication guidelines to protect conversation between providers of data. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Coordination | Facilitate knowledge sharing on supply and needs | 4 | [9] | Facilitate knowledge sharing within the platform on the data needs. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Coordination | Regulate collaborations | 5 | [9] | Create rules to comply with to support fair collaborations between platform users. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Coordination | Set up conflict resolution process | 6 | [3] | A conflict resolution process is created to manage conflicts between various users or users and the platform. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Coordination | Determine level of centralization of the platform | 7 | [23] | Determine whether a centralized strategy, where all control of the platform is in the hand of the platform orchestrators, or a decentralized strategy, where the platform is managed by each data provider, is preferable. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Onboarding | Carefully consider communication of open data strategy to potential customers | 2 | [29] | Communicate the strategy the platform users for opening up data to its potential users. | | | 1 | 0 | | | 1 | 2 | 1 | 67 |
| Onboarding | Create sandbox mode | 3 | [15] | Create a sandbox mode that allows users to test the platform. | 1 | | 1 | 0 | | | 1 | 3 | 1 | 75 |
| Onboarding | Create self-service capabilities | 4 | [31] | Allow users to onboard themselves through the platform. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Onboarding | Create a showcase section | 5 | [15] | Create a showcase section displaying the creations of users. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Onboarding | Offer data playbooks, tutorials and describe data sharing technique | 6 | [9] | Create playbooks and tutorials to support new users. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Onboarding | Create scalable onboarding for new users | 7 | [31] | Make onboarding scalable differentiating between different types of users with different wishes. | 0 | | 1 | 1 | | | 1 | 3 | 1 | 75 |
| Onboarding | Allow piloting | 8 | [5] | Allow users to test the platform through a piloting scheme. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |

TABLE F.7: Interviewees in favour on capabilities Coordination and Onboarding

| Capability | Practice | Level | Source | Description | I2 | I3 | I4 | I5 | I6 | I7 | I8 | In favour | Against | Score |
|------------------------------------|--|-------|--------|---|----|----|----|----|----|----|----|-----------|---------|-------|
| Compliance and External Regulation | Ensure compliance with applicable regulation | 1 | I2 | Survey the regulations relevant for your platform and ensure compliance with them | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Compliance and External Regulation | Appoint a data steward | 3 | [9] | Appoint a person to check the compliance of all data actively. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Compliance and External Regulation | Regulate intellectual property | 4 | [6] | Check and regulate intellectual property on the platform. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Compliance and External Regulation | Get ISO certified | 6 | I2 | Get an ISO certification to show the quality of the data platform. | 1 | | | | | | | 1 | 0 | 100 |
| Compliance and External Regulation | Set-up due processes for data sharing | 7 | [23] | Set up a due-process for data sharing that checks the specific compliance of every transaction. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Compliance and External Regulation | Create Responsible Disclosure processes | 8 | I2 | Create responsible disclosure processes that allow users to disclose any compliance issues with the platform. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Internal Regulation | Create terms of service and code of conduct | 1 | [15] | Create a terms of service for users to accept with basic regulations of the platform. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Internal Regulation | Determine punishments for breaches | 5 | [15] | Punish any breaches of the terms of service or code and actively enforce the rules. | 1 | | 1 | 0 | | | 1 | 3 | 1 | 75 |
| Internal Regulation | Negotiate exchange terms with users | 7 | [14] | Allow specific SLAs to support users whose wishes go beyond the ToS | 1 | | 1 | 0 | | | 1 | 3 | 1 | 75 |
| Privacy | Ensure complete GDPR compliance | 1 | I2 | Ensure complete compliance with the european privacy regulation to prevent privacy infringements. | 0 | | 1 | 1 | | | 1 | 3 | 1 | 75 |
| Privacy | Educate end users on privacy permissions | 2 | [13] | Ensure that end-users understand the privacy measures taken by the platform. | 0 | | 1 | 1 | | | 1 | 3 | 1 | 75 |
| Privacy | Get informed consent for personal data | 6 | [6] | Go to any user that has personal data on the platform and try to get informed consent. | 0 | | 1 | 0 | | | 1 | 2 | 2 | 50 |
| Privacy | Create guardrails to support privacy protection | 7 | [13] | Create guardrails for users to prevent them from making privacy infringements themselves. | 0 | | 1 | 1 | | | 1 | 3 | 1 | 75 |
| Privacy | Allow different forms of consent establishment for personal data sharing | 8 | [13] | Allow users to give specific informed consent to data use cases. | 0 | | 1 | 1 | | | 1 | 3 | 1 | 75 |

TABLE F.8: Interviewees in favour on capabilities Compliance and External Regulation, Internal Regulation and Privacy

| Capability | Practice | Level | Source | Description | I2 | I3 | I4 | I5 | I6 | I7 | I8 | In favour | Against | Score |
|----------------------|--|-------|--------|---|----|----|----|----|----|----|----|-----------|---------|-------|
| Type of Data Sharing | Establish a data exchange | 1 | [3] | A basic data exchange is established that enables data sharing. | 0 | | 1 | 0 | | | 1 | 2 | 2 | 50 |
| Type of Data Sharing | Create contracts for data use | 2 | [1] | Contracts are used to regulate the use of data by users. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Type of Data Sharing | Use premium licensing | 6 | I2 | Premium licensing is used to ensure monetization of the data. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Type of Data Sharing | Enable data pooling | 7 | [6] | Data pooling is enabled that allow various sources to combine the data. | | | 1 | | | | 1 | 2 | 0 | 100 |
| Type of Data Sharing | Use open licensing | 8 | [6] | An open license is used that opens up the data to a wide range of users. | 0 | | 1 | | | | 1 | 2 | 1 | 67 |
| Governance | Do only technical maintenance | 2 | [3] | Only perform technical maintenance on the platform. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Governance | Create governance framework | 3 | [15] | Create a framework to support governance actions on the platform. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Governance | Promote governance standards | 4 | [9] | Promote the governance standards to increase transparency. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Governance | Localize licensing conditions | 5 | [27] | Create licences localized to the languages of each specific user. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Governance | Regulate behaviour through reputational measures | 7 | [6] | Create reputational features on the platform that regulate the usage by users. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Governance | Allow the ecosystem to self-regulate | 8 | [16] | Allow users to regulate how the platform is governed. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Control over Data | Exert no control over data | 1 | [3] | No control is exerted over the data, it can be freely reshared. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Control over Data | Make individual organization retain control over data | 2 | [3] | Individual organizations get control over the data after buying it. | | | 1 | 0 | | | 1 | 2 | 1 | 67 |
| Control over Data | Describe data ownership in various scenarios in the platform | 4 | | Describe scenarios of data sharing and indicate through each step who the owner of the data is to increase understandability. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Control over Data | Support context-based ownership | 7 | I2 | Dependent on the context of the data, ownership regulations can differ. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |

TABLE F.9: Interviewees in favour on capabilities Type of Data Sharing, Governance and Control over Data

| Capability | Practice | Level | Source | Description | I2 | I3 | I4 | I5 | I6 | I7 | I8 | In favour | Against | Score |
|-------------------------|--|-------|--------|---|----|----|----|----|----|----|----|-----------|---------|-------|
| Data Sharing Technology | Share data via excel, csv or SOAP | 1 | [26] | Data is shared through a single excel or csv file. | 0 | | 1 | 1 | | | 1 | 3 | 1 | 75 |
| Data Sharing Technology | Share data on company website | 2 | [29] | Data can be downloaded from the company website. | 0 | | 1 | 0 | | | 1 | 2 | 2 | 50 |
| Data Sharing Technology | Sell data on an open data marketplace | 3 | [29] | Data can be bought and sold on the platform's marketplace. | 1 | | 1 | | | | 1 | 3 | 0 | 100 |
| Data Sharing Technology | Share data via an actual platform | 4 | [29] | An actual platform is created on which the data is shared with additional features. | 1 | | 1 | 0 | | | 1 | 3 | 1 | 75 |
| Data Sharing Technology | Create Developer assist tools | 6 | I2 | A developer assist tool is created to support data usage by users. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Data Sharing Technology | Allow uploading of datasets by users | 8 | [18] | Allow users to participate in the platform by uploading their own dataset. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Data Processing | Share Snapshot of data | 1 | I2 | Process the raw data to create a snapshot of the current state of the database and share that. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Data Processing | Process data before sharing | 2 | [26] | Process the data before sharing to only share the useful parts. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Data Processing | Encourage high-quality resources and repress poor-quality resources of platform ecosystems | 3 | [23] | High quality resources are encouraged, while low quality resources are repressed without actual cleaning. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Data Processing | Implement a review board for control | 4 | [14] | Create a review board that reviews the data before it is shared. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Data Processing | Share linked data | 5 | [26] | Create linked data before sharing it. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Data Processing | Clean data before sharing | 6 | [14] | Completely clean the data before sharing | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Data Processing | Share humanreadable data | 7 | [26] | Translate the data so it becomes humanreadable before sharing it. | 0 | | 1 | 1 | | | 1 | 3 | 1 | 75 |

TABLE F.10: Interviewees in favour on capabilities Data Sharing Technology and Data Processing

| Capability | Practice | Level | Source | Description | I2 | I3 | I4 | I5 | I6 | I7 | I8 | In favour | Against | Score |
|----------------------|--|-------|--------|---|----|----|----|----|----|----|----|-----------|---------|-------|
| Search Functionality | Provide search on metadata level | 1 | [18] | Create a basic search on metadata level to find datasets. | 0 | | 1 | 0 | | | 1 | 2 | 2 | 50 |
| Search Functionality | Allow search on content level | 2 | [18] | Allow users to search on content level to find specific lines of data they require. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Search Functionality | Allow for filtering search results | 3 | [18] | Enable the filtering of search results to increase usability. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Search Functionality | Recommend datasets | 5 | [18] | Recommend related datasets that the user could use to create more value. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Search Functionality | Assist connecting database by implementing pathfinding | 7 | 15 | Implement pathfinding functionalities that shows related content to a specific data item from different datasets. | | | 1 | 1 | | | 1 | 3 | 0 | 100 |
| Pricing Mechanism | Enable basic payment for access | 1 | 12 | Allow users to pay for the data they use. | 1 | | 1 | 1 | | | 1 | 4 | 0 | 100 |
| Pricing Mechanism | Determine ideal business model for platform context | 3 | 12 | Determine the business model that best fits the platform and the needs of the users and the orchestrators. | 1 | | 1 | | | | 1 | 3 | 0 | 100 |
| Pricing Mechanism | Create congestion pricing for popular data | 7 | [30] | Create congestion pricing that makes popular data more expensive at times that data is congested. | | | 1 | | | | 1 | 2 | 0 | 100 |
| Pricing Mechanism | Create spot market for short transactions | 8 | [14] | Create a spot market that allows for short and cheap transactions of low-demand data. | | | 1 | | | | 1 | 2 | 0 | 100 |

TABLE F.11: Interviewees in favour on capabilities Compliance and Search Functionality and Pricing Mechanism



CASE STUDY RESULTS

G.1 CASE STUDY RESULTS

| | | Erwin | Jeroen | Joris | |
|---------------|--|-------|--------|-------|----------|
| Security | Allow access without authentication by anonymous users | 1 | 1 | 1 | 100 |
| Security | Perform authentication through username password / API key | 0 | 0 | 0 | 0 |
| Security | Use a trusted supplier for platform security | 1 | 0 | | 50 |
| Security | Perform user-based API authorization | 1 | 0 | 0 | 33,33333 |
| Security | Implement use case specific best-practice authentication methods | 0 | 0 | 0 | 0 |
| Security | Connect Identity Access Management and Authorization system | 0 | 0 | 0 | 0 |
| Security | Implement a use case specific state-of-the-art security system | 0 | 0 | 0 | 0 |
| | | | | | |
| Documentation | Provide no documentation | 1 | 1 | 1 | 100 |
| Documentation | Provide document with basic API description and metadata | 1 | 1 | 1 | 100 |
| Documentation | List types of datasets | 1 | 1 | 1 | 100 |
| Documentation | Support structural metadata | 1 | 1 | 1 | 100 |

| | | | | | | |
|-------------------------|--|---|---|---|---|----------|
| Documentation | | Implement a complete Developer Platform with documentation and API descriptions | 1 | 1 | 0 | 67 |
| Documentation | | Deploy API data through Swagger or similar platforms | 1 | 1 | 0 | 66,66667 |
| Documentation | | Create assistance tools for API | 1 | 0 | 0 | 33,33333 |
| Documentation | | Provide living documentation | 0 | 0 | 0 | 0 |
| | | | | | | |
| Platform Maintenance | | Maintain original platform setup | 1 | 1 | 1 | 100 |
| Platform Maintenance | | Add additional datasets | 1 | 1 | 1 | 100 |
| Platform Maintenance | | Manage versioning of the API | 1 | 1 | 1 | 100 |
| Platform Maintenance | | Model dependencies in data | 1 | 0 | 1 | 67 |
| Platform Maintenance | | Install data mapping tables for data types | 1 | 0 | 1 | 67 |
| Platform Maintenance | | Manage versioning of the data | 1 | 1 | 1 | 100 |
| Platform Maintenance | | Establish grace period for updates | 1 | 1 | 1 | 100 |
| | | | | | | |
| Decision Making Process | | Establish decision making structures within an organization | 1 | 1 | 1 | 100 |
| Decision Making Process | | Create internal process chart for establishing decision rights | 1 | 0 | 1 | 66,66667 |
| Decision Making Process | | Create transparent roadmap | 1 | 1 | 1 | 100 |
| Decision Making Process | | Create standards, conventions and guidelines for platform development | 1 | 1 | 0 | 66,66667 |
| Decision Making Process | | Create alignment processes for widespread involvement from users | 1 | 1 | 1 | 100 |
| Decision Making Process | | Create a governance group to guide product development | 0 | 0 | 0 | 0 |
| Decision Making Process | | Partition decision rights between owner and users | 0 | 1 | 0 | 33,33333 |
| | | | | | | |
| Monitoring Data | | Provide opportunities to stakeholders for monitoring data | 1 | 1 | 1 | 100 |
| Monitoring Data | | Create data stewardship practices | 1 | 1 | | 100 |

| | | | | | |
|-----------------------|---|---|---|---|-----|
| Monitoring Data | Offer the push updates on data mutations | 1 | 1 | 1 | 100 |
| Monitoring Data | Create strict monitoring processes, including reporting by users | 1 | 1 | 0 | 67 |
| Monitoring Data | Establish a visible supply chain for the data | 0 | 1 | 0 | 33 |
| | | | | | |
| Monitoring Platform | Implement heartbeat test for platform users | 1 | 1 | 1 | 100 |
| Monitoring Platform | Measure platform uptime | 1 | 1 | 1 | 100 |
| Monitoring Platform | Create alerts for users when the platform is down | 1 | 1 | 0 | 67 |
| Monitoring Platform | Create a status website that shows issues | 1 | 1 | 1 | 100 |
| Monitoring Platform | Create a post-mortum that can replace the system with a snapshot when the service is down | 1 | 0 | 0 | 33 |
| Monitoring Platform | Run two versions of the platform, where one takes over when the other crashes | 1 | 1 | 1 | 100 |
| | | | | | |
| Platform Measurements | Identify API metrics | 1 | 1 | 1 | 100 |
| Platform Measurements | Set usage quotas | 1 | 0 | 1 | 67 |
| Platform Measurements | Warnings and eventually blocking for quota breaking | 1 | 1 | 0 | 67 |
| Platform Measurements | Throttle abuse of data quotas and make a fair use assessment | 1 | 0 | 0 | 33 |
| Platform Measurements | Flexible fair use agreement based on user/datetime | 1 | 0 | 0 | 33 |
| | | | | | |
| Data Quality | Keep data up to date | 1 | 1 | 1 | 100 |
| Data Quality | Make reliability vs. costs trade-off | 1 | 0 | 1 | 67 |
| Data Quality | Check and indicate missing data | 1 | 0 | 0 | 33 |
| Data Quality | Make trade-offs between amount of data and usability costs | 1 | 1 | 0 | 67 |
| Data Quality | Install data validation mechanisms | 1 | 0 | 0 | 33 |
| Data Quality | Perform reconciliation | 1 | 1 | 0 | 67 |
| | | | | | |
| Data Standardization | Let IT determine data structures | 1 | 1 | 1 | 100 |

| | | | | | |
|----------------------|---|-----|-----|-----|-----|
| Data Standardization | Allow provider to determine data attributes | 1 | 1 | 1 | 100 |
| Data Standardization | Promote data standard governance | 1 | 1 | 1 | 100 |
| Data Standardization | Support data integration with related datasets | 1 | 1 | 0 | 67 |
| Data Standardization | Format data according to users' needs | 1 | 1 | 0 | 67 |
| Data Standardization | Create flexible data structures | 1 | 1 | 0 | 67 |
| | | | | | |
| Data Use Case | Audit every data use case and its processes | 1 | 0 | 0 | 33 |
| Data Use Case | Adapt platform to specific use case | 1 | 1 | 0 | 67 |
| Data Use Case | Set usage goal for data | 1 | 1 | 0 | 67 |
| Data Use Case | Define and link the use cases and the relevant stakeholders of the data | 1 | 1 | 0 | 67 |
| Data Use Case | Discuss data usage goal in terms of service | 1 | | 0 | 50 |
| Data Use Case | Create new advanced products with data | 1 | | 0 | 50 |
| Data Use Case | Monitor if the data use case is met by users and punish misuse | 0 | 0 | 0 | 0 |
| | | | | | |
| Partner Relationship | Maintain no relationship with users | 1 | 1 | 1 | 100 |
| Partner Relationship | Perform user screening | 0 | 1 | 1 | 67 |
| Partner Relationship | Create feedback loops | | 1 | 1 | 100 |
| Partner Relationship | Ensure continuous collaboration with users | | 1 | 1 | 100 |
| Partner Relationship | Create working groups and review committees to support users | | 1 | 1 | 100 |
| Partner Relationship | Create new avenues and means for cooperation for stakeholders | | 0 | 0 | 0 |
| Partner Relationship | Create review systems for data suppliers | 0 | 0 | 0 | 0 |
| | | | | | |
| Incentivisation | Offer non-monetary subsidies to early users | N/A | N/A | N/A | |
| Incentivisation | Provide seeding (input to one side of the marketplace) | N/A | N/A | N/A | |
| Incentivisation | Create incentive system for users | N/A | N/A | N/A | |
| Incentivisation | Create performance-based rewards schemes for users | N/A | N/A | N/A | |

| | | | | | |
|------------------------------------|---|-----|-----|-----|----------|
| Incentivisation | Create behavioral rewards for users i.e. through gamification | N/A | N/A | N/A | |
| Incentivisation | Set up revenue sharing or other incentivisation scheme for data supplier partners | N/A | N/A | N/A | |
| | | | | | |
| Coordination | Take 'take it or leave it' stance towards connecting providers and users | 1 | 0 | 1 | 66,66667 |
| Coordination | Create a mutual coordination plan for connecting supply and demand | 1 | 0 | 0 | 33,33333 |
| Coordination | Create communication guidelines between collaborators | 1 | 0 | 0 | 33 |
| Coordination | Facilitate knowledge sharing on supply and needs | 1 | 0 | 0 | 33 |
| Coordination | Regulate collaborations | 0 | 0 | 0 | 0 |
| Coordination | Set up conflict resolution proces | 0 | 0 | 0 | 0 |
| Coordination | Determine level of centralization of the platform | 0 | 0 | 0 | 0 |
| | | | | | |
| Onboarding | Carefully consider communication of open data strategy to potential customers | 1 | 1 | 1 | 100 |
| Onboarding | Create sandbox mode | 1 | 0 | 0 | 33 |
| Onboarding | Create self-service capabilities | 1 | 0 | 1 | 67 |
| Onboarding | Create a showcase section | 1 | 1 | 1 | 100 |
| Onboarding | Offer data playbooks, tutorials and describe data sharing technique | 1 | 1 | 0 | 67 |
| Onboarding | Create scalable onboarding for new users | 1 | 0 | 0 | 33 |
| Onboarding | Allow piloting | 1 | 0 | 0 | 33 |
| | | | | | |
| Compliance and External Regulation | Ensure compliance with applicable regulation | 1 | 1 | 1 | 100 |
| Compliance and External Regulation | Appoint a data steward | 1 | 0 | 1 | 67 |
| Compliance and External Regulation | Regulate intellectual property | 1 | 0 | 0 | 33 |
| Compliance and External Regulation | Get ISO certified | 1 | 0 | 1 | 67 |
| Compliance and External Regulation | Set-up due processes for data sharing | 0 | 0 | 0 | 0 |

| | | | | | |
|------------------------------------|--|-----|-----|-----|-----|
| Compliance and External Regulation | Create Responsible Disclosure processes | 0 | 0 | 0 | 0 |
| | | | | | |
| Internal Regulation | Create terms of service and code of conduct | 1 | 1 | 1 | 100 |
| Internal Regulation | Determine punishments for breaches | 1 | 0 | 0 | 33 |
| Internal Regulation | Negotiate exchange terms with users | 1 | 1 | 0 | 67 |
| | | | | | |
| Privacy | Ensure complete GDPR compliance | N/A | N/A | N/A | |
| Privacy | Educate end users on privacy permissions | N/A | N/A | N/A | |
| Privacy | Get informed consent for personal data | N/A | N/A | N/A | |
| Privacy | Create guardrails to support privacy protection | N/A | N/A | N/A | |
| Privacy | Allow different forms of consent establishment for personal data sharing | N/A | N/A | N/A | |
| | | | | | |
| Type of Data Sharing | Establish a data exchange | 1 | 1 | 1 | 100 |
| Type of Data Sharing | Create contracts for data use | 1 | 1 | 1 | 100 |
| Type of Data Sharing | Use premium licensing | 1 | 1 | 1 | 100 |
| Type of Data Sharing | Enable data pooling | 1 | 1 | 1 | 100 |
| Type of Data Sharing | Use open licensing | 1 | 1 | 1 | 100 |
| | | | | | |
| Governance | Do only technical maintenance | 1 | 1 | 1 | 100 |
| Governance | Create governance framework | 1 | 1 | 1 | 100 |
| Governance | Promote governance standards | 1 | 1 | 0 | 67 |
| Governance | Localize licensing conditions | 1 | 1 | 0 | 67 |
| Governance | Regulate behaviour through reputational measures | 1 | 0 | 0 | 33 |
| Governance | Allow the ecosystem to self-regulate | | 1 | 0 | 50 |
| | | | | | |
| Control over Data | Exert no control over data | 1 | 1 | 1 | 100 |
| Control over Data | Make individual organization retain control over data | 1 | 1 | 1 | 100 |
| Control over Data | Describe data ownership in various scenarios in the platform | 0 | 1 | 0 | 33 |
| Control over Data | Support context-based ownership | 0 | 0 | 0 | 0 |
| | | | | | |

| | | | | | |
|-------------------------|--|-----|-----|-----|-----|
| Data Sharing Technology | Share data via excel, csv or SOAP | 1 | 1 | 1 | 100 |
| Data Sharing Technology | Share data on company website | 1 | 1 | 1 | 100 |
| Data Sharing Technology | Sell data on an open data marketplace | 1 | 1 | 1 | 100 |
| Data Sharing Technology | Share data via an actual platform | 1 | 1 | 1 | 100 |
| Data Sharing Technology | Create Developer assist tools | 1 | 1 | 0 | 67 |
| Data Sharing Technology | Allow uploading of datasets by users | 0 | 0 | 0 | 0 |
| | | | | | |
| Data Processing | Share Snapshot of data | 1 | 1 | 1 | 100 |
| Data Processing | Process data before sharing | 1 | 1 | 1 | 100 |
| Data Processing | Encourage high-quality resources and repress poor-quality resources of platform ecosystems | 1 | 1 | 0 | 67 |
| Data Processing | Implement a review board for control | 1 | 1 | 1 | 100 |
| Data Processing | Share linked data | 1 | 1 | 1 | 100 |
| Data Processing | Clean data before sharing | 1 | 1 | 0 | 67 |
| Data Processing | Share humanreadable data | 1 | 1 | 0 | 67 |
| | | | | | |
| Search Functionality | Provide search on metadata level | 1 | 1 | 1 | 100 |
| Search Functionality | Allow search on content level | 1 | 1 | 1 | 100 |
| Search Functionality | Allow for filtering search results | 1 | 1 | 1 | 100 |
| Search Functionality | Recommend datasets | 1 | 0 | 0 | 33 |
| Search Functionality | Assist connecting database by implementing pathfinding | 1 | 0 | 0 | 33 |
| | | | | | |
| Pricing Mechanism | Enable basic payment for access | N/A | N/A | N/A | |
| Pricing Mechanism | Determine ideal business model for platform context | N/A | N/A | N/A | |
| Pricing Mechanism | Create congestion pricing for popular data | N/A | N/A | N/A | |
| Pricing Mechanism | Create spot market for short transactions | N/A | N/A | N/A | |

TABLE G.1: Case Study results Maturity Assessment Platform F

LISTING OF FIGURES

| | | |
|------|--|-----|
| 3.1 | The Focus Area Maturity Model | 10 |
| 3.2 | The Design and Engineering Cycle by Wieringa [Wieringa, 2014] | 11 |
| 3.3 | Development Method for Focus Area Maturity Models based on van Steenbergen et al. [van Steenbergen et al., 2010] | 12 |
| 3.4 | Complete Research Method | 13 |
| 4.1 | The Systematic Guide to Literature Review Development (a yellow line indicates a skipped step) | 15 |
| 4.2 | The filter process of the literature review | 19 |
| 5.1 | The incremental model update method used in this research. | 25 |
| 5.2 | The process used for processing the interviews. | 28 |
| 5.3 | An overview of the codes used for the interview. | 30 |
| 7.1 | A visualization of the final definition | 44 |
| 8.1 | Various Types of Data Collaboratives by van Veenstra et al. [Veenstra, 2015] the yellow elements are used for the characteristics overview | 48 |
| 8.2 | The results from the characteristics assessment of Platform F during the case study | 51 |
| 9.1 | Contingency Factors for Data Governance in Platform Ecosystems by Lee, Zhu, Ross [Lee et al., 2017b] | 55 |
| 9.2 | A meta-model for Focus Area Maturity Models with example | 56 |
| 9.3 | The First Version of the DaaP Maturity Model | 62 |
| 10.1 | The Final Version of the DaaP Maturity Model | 74 |
| 10.2 | The resulting maturity model from the case study assessment. The current level for the capabilities is indicated with with a thick right border. | 75 |
| 11.1 | Validity Threats according to Wohlin et al. [Wohlin et al., 2012] | 81 |
| E.1 | Focus Area and Practices found in the Literature Review Top Left | 128 |
| E.2 | Focus Area and Practices found in the Literature Review Top Right | 129 |
| E.3 | Focus Area and Practices found in the Literature Review Bottom Left | 130 |
| E.4 | Focus Area and Practices found in the Literature Review Bottom Left | 131 |

LISTING OF TABLES

| | | |
|------|---|-----|
| 2.1 | Issues with sharing data found by the European Commission | 6 |
| 2.2 | Factors to improve willingness to share data found by the European Commission | 7 |
| 4.1 | The mapping of the SLR questions on the Research Questions and SLR Goals | 17 |
| 4.2 | Coding Node Definitions | 19 |
| 5.1 | Overview of the Interview Participants | 26 |
| 6.1 | An Overview of the Case Study Participants | 33 |
| 7.1 | Data as a Platform Related Definitions | 39 |
| 7.2 | Feedback on the Definition | 42 |
| 8.1 | The performance scores for the characteristics overview | 53 |
| 9.1 | The Maturity Level Descriptions | 57 |
| 9.2 | The First Version of the Security Capability | 58 |
| 9.3 | The Final Version of the Security Capability | 59 |
| 9.4 | The interview results for the Security Focus Area | 60 |
| 10.1 | Results Maturity Assessment Case Study | 72 |
| 10.2 | The performance scores for the Maturity Model | 73 |
| D.1 | First model with Characteristics | 109 |
| D.2 | Input Characteristics from Literature | 113 |
| D.3 | Characteristics found in the Literature Review | 114 |
| D.4 | Definition for the Characteristics | 115 |
| E.1 | Capabilities for Maturity Model found in Literature | 124 |
| E.2 | Initial version of DaaPFAMM 1/3 | 125 |
| E.3 | Initial version of DaaPFAMM 2/3 | 126 |
| E.4 | Initial version of DaaPFAMM 3/3 | 127 |
| F.12 | Change Log Maturity Model during Interviews | 138 |
| F.1 | Interviewees in favour on capabilities Security and Documentation | 139 |
| F.2 | Interviewees in favour on capabilities Platform Maintenance and Decision Making Process | 140 |
| F.3 | Interviewees in favour on capabilities Monitoring Data and Monitoring Platform | 141 |
| F.4 | Interviewees in favour on capabilities Platform Measurement and Data Quality | 142 |
| F.5 | Interviewees in favour on capabilities Data Standardization and Data Use Case | 143 |
| F.6 | Interviewees in favour on capabilities Partner Relationship and Incentivisation | 144 |

| | | |
|------|--|-----|
| F.7 | Interviewees in favour on capabilities Coordination and Onboarding | 145 |
| F.8 | Interviewees in favour on capabilities Compliance and External Regulation, Internal Regulation and Privacy | 146 |
| F.9 | Interviewees in favour on capabilities Type of Data Sharing, Governance and Control over Data | 147 |
| F.10 | Interviewees in favour on capabilities Data Sharing Technology and Data Processing . . . | 148 |
| F.11 | Interviewees in favour on capabilities Compliance and Search Functionality and Pricing Mechanism | 149 |
| G.1 | Case Study results Maturity Assessment Platform F | 156 |

BIBLIOGRAPHY

- [Arnaut et al., 2018] Arnaut, C., Pont, M., Scaria, E., Berghmans, A., and Leconte, S. (2018). Digital Single Market Study on data sharing between companies in Europe This study was carried out for the European Commission by.
- [Baars and Jansen, 2012] Baars, A. and Jansen, S. (2012). A Framework for Software Ecosystem Governance. pages 168–180. Springer, Berlin, Heidelberg.
- [Ball, 2012] Ball, A. (2012). Review of data management lifecycle models. *IDMRC*.
- [Bertot and Choi, 2013] Bertot, J. C. and Choi, H. (2013). Big data and e-government. In *Proceedings of the 14th Annual International Conference on Digital Government Research - dgo '13*, page 1, New York, New York, USA. ACM Press.
- [Booth et al., 2016] Booth, A., Sutton, A., and Papaioannou, D. (2016). *Systematic approaches to a successful literature review*. Sage.
- [Brodsky and Oakes, 2017] Brodsky, L. and Oakes, L. (2017). Data Sharing and Open Banking. *McKinsey on Payments*, (July):16–23.
- [Broek et al., 2018] Broek, T. V. D., Fleur, A., and Veenstra, V. (2018). Technological Forecasting & Social Change Governance of big data collaborations : How to balance regulatory compliance and disruptive innovation. *Technological Forecasting & Social Change*, 129(November 2017):330–338.
- [Buda et al., 2016] Buda, A., Ubacht, J., Janssen, M., and Sips, R.-j. (2016). Decision Support Framework for Opening Business Data. (January).
- [Choi and Kroschel, 2015] Choi, T. and Kroschel, I. (2015). CHALLENGES OF GOVERNING INTER-ORGANIZATIONAL RELATIONSHIPS : INSIGHTS FROM A CASE STUDY Governance and Value Chains – Fundamentals and Research. pages 1–16.
- [Correa et al.,] Correa, A. S., Paulo, S., and Paulo, S. Investigating open data portals automatically : a methodology and some illustrations.
- [Correa et al., 2018] Correa, A. S., Zander, P.-O., and da Silva, F. S. C. (2018). Investigating open data portals automatically. *Proceedings of the 19th Annual International Conference on Digital Government Research Governance in the Data Age - dgo '18*, pages 1–10.
- [Curry, 2018] Curry, E. (2018). Next-Generation Smart Environments : From System of Systems to Data Ecosystems. *IEEE Intelligent Systems*, 33(June):69–76.
- [Dalmolen et al., 2015] Dalmolen, S., Moonen, H., and van Hillegersberg, J. (2015). Industry-wide Inter-organizational Systems and Data Quality : Exploratory findings of the use of GS1 standards in the Dutch retail market. *Twenty-first Americas Conference on Information Systems*, pages 1–7.

- [Danneels et al., 2017] Danneels, L., Viaene, S., and Bergh, J. V. D. (2017). Open data platforms : Discussing alternative knowledge epistemologies. *Government Information Quarterly*, 34(3):365–378.
- [Davenport et al., 2012] Davenport, T. H., Barth, P., and Bean, R. (2012). How 'Big Data' Is Different. *MIT Sloan Management Review*, 54(1).
- [De Reuver et al., 2018] De Reuver, M., Sørensen, C., and Basole, R. C. (2018). The digital platform: A research agenda. *Journal of Information Technology*, 33(2):124–135.
- [DiCicco-Bloom and Crabtree, 2006] DiCicco-Bloom, B. and Crabtree, B. F. (2006). Making sense of qualitative research: The qualitative research interview. *Medical Education*, 40(4):314–321.
- [Eisenhardt, 1989] Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *Academy of Management Review*, 14(4):532–550.
- [European Commission, 2017] European Commission (2017). Final results of the European Data Market study measuring the size and trends of the EU data economy | Digital Single Market.
- [Evans et al., 2012] Evans, D. S., Epstein, R., Schmalensee, R., Chiang, L., Murphy, J., Joyce, S., Schilt, M., and Tuladhar, N. (2012). Governing Bad Behavior By Users of Multi-Sided Platforms, 27 Berkeley Tech. *Berkeley Technology Law Journal*, 27:1201.
- [Evans et al., 2011] Evans, D. S., Schmalensee, R., Noel, M. D., Chang, H. H., and Garcia-Swartz, D. D. (2011). Platform Economics: Essays on Multi-Sided Businesses.
- [Fink, 2005] Fink, A. (2005). *Conducting research literature reviews: from the Internet to paper*.
- [Fisher and Davies, 2016] Fisher, M. and Davies, J. (2016). CREATING INTERNET OF THINGS ECOSYSTEMS. 9(4):9–14.
- [Ghazawneh and Henfridsson, 2015] Ghazawneh, A. and Henfridsson, O. (2015). A paradigmatic analysis of digital application marketplaces. *Journal of Information Technology*, 30(3):198–208.
- [Günther et al., 2017] Günther, W. A., Mehri, M. H. R., Huysman, M., and Feldberg, F. (2017). Debating big data : A literature review on realizing value from big data. 26:191–209.
- [Harrington, 1998] Harrington, J. L. (1998). *Relational database design and implementation*.
- [Hemphill and Jackson, 2017] Hemphill, L. and Jackson, S. (2017). Accessibility and Flexibility Two Organizing Principles for Big Data Collaboration.
- [Horkoff et al., 2018] Horkoff, J., Linkman, J., Hammouda, I., Knauss, E., Debbiche, J., and Freiholtz, M. (2018). Modeling Support for Strategic API Planning and Analysis.
- [Huo et al., 2016] Huo, B., Wang, Q., Zhao, X., and Schuh, S. (2016). Threats and benefits of power discrepancies between organisations : a supply chain perspective. *International Journal of Production Research*, 7543:0.
- [Iansiti and Levien, 2004] Iansiti, M. and Levien, R. (2004). Strategy as Ecology. *Harvard Business Review*, 82(3).
- [Ikävalko et al., 2018] Ikävalko, H., Turkama, P., and Smedlund, A. (2018). Value Creation in the Internet of Things : Mapping Business Models and Ecosystem Roles. 8(3):5–16.
- [Jansen, 2014] Jansen, S. (2014). Measuring the health of open source software ecosystems: Beyond the scope of project health. *Information and Software Technology*, 56(11):1508–1519.

- [Jansen et al., 2009] Jansen, S., Finkelstein, A., and Brinkkemper, S. (2009). A sense of community: A research agenda for software ecosystems. In *2009 31st International Conference on Software Engineering - Companion Volume*, pages 187–190. IEEE.
- [Jetzek et al., 2014] Jetzek, T., Avital, M., and Bjorn-Andersen, N. (2014). Data-driven innovation through open government data. *Journal of Theoretical and Applied Electronic Commerce Research*, 9(2):100–120.
- [Kassen, 2018] Kassen, M. (2018). Open data and its institutional ecosystems : A comparative cross-jurisdictional analysis of open data platforms. 61(1):109–129.
- [Kitchenham and Charters, 2007] Kitchenham, B. and Charters, S. (2007). issue: EBSE 2007-001. *Technical report*, 2(3).
- [Klievink et al., 2018] Klievink, B., van der Voort, H., and Veeneman, W. (2018). Creating value through data collaboratives. *Information Polity*, 23(4):379–397.
- [Knauss et al., 2014] Knauss, E., Damian, D., Knauss, A., and Borici, A. (2014). Openness and requirements: Opportunities and tradeoffs in software ecosystems. In *2014 IEEE 22nd International Requirements Engineering Conference (RE)*, pages 213–222. IEEE.
- [Kolluru and Meredith, 2001] Kolluru, R. and Meredith, P. H. (2001). Security and trust management in supply chains.
- [Lee et al., 2017a] Lee, S. U., Zhu, L., and Jeffery, R. (2017a). Data Governance for Platform Ecosystems : Critical Factors and the State of Practice. *PACIS 2017 Proceedings*.
- [Lee et al., 2018] Lee, S. U., Zhu, L., and Jeffery, R. (2018). A Data Governance Framework for Platform Ecosystem Process Management.
- [Lee et al., 2019] Lee, S. U., Zhu, L., and Jeffery, R. (2019). Data Governance Decisions for Platform Ecosystems. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 6:6377–6386.
- [Lee et al., 2017b] Lee, S. U., Zhu, L., and Ross, J. (2017b). Design Choices for Data Governance in Platform Ecosystems – A Contingency Model. *arXiv preprint arXiv:1706.07560*, pages 1–10.
- [Lindman et al., 2014] Lindman, J., Kinnari, T., and Rossi, M. (2014). Industrial open data : Case studies of early open data entrepreneurs.
- [Louise Barriball RGN Research Assistant and While RGN RHV Cert Ed, 1994] Louise Barriball RGN Research Assistant, K. B. and While RGN RHV Cert Ed, A. (1994). Collecting data using a semi-structured interview: a discussion paper. Technical report.
- [Lowndes and Skelcher, 1998] Lowndes, V. and Skelcher, C. (1998). The Dynamics of Multi-organizational Partnerships: an Analysis of Changing Modes of Governance. *Public Administration*, 76(2):313–333.
- [Maccani, 2016] Maccani, G. (2016). Exploring the factors that influence adoption of open government data for commercial service innovation in cities. *PQDT - Global*.
- [Malgonde and Bhattacharjee, 2014] Malgonde, O. and Bhattacharjee, A. (2014). Innovating Using Big Data: A Social Capital Perspective Twentieth Americas Conference on Information Systems. (Ibm 2012):1–9.
- [Mattison,] Mattison, J. INTERNET OF THINGS AND DATA ANALYTICS HANDBOOK.

- [Mijsters et al., 2018] Mijsters, Y., Mustafa, A., Mihai, I., and Jansen, S. (2018). On the nature of software sub-ecosystems and their health. pages 25–32.
- [Myrelid and Jonsson, 2019] Myrelid, P. and Jonsson, P. (2019). Determinants of information quality in dyadic supply chain relationships. *International Journal of Logistics Management*, 30(1):356–380.
- [Najjar, 2013] Najjar, Mohammad S Kettinger, W. J. (2013). Data Monetization: Lessons from a Retailer’s Journey. *MIS Quarterly Executive*, 12(4).
- [Ojo and Millard, 2017] Ojo, A. and Millard, J. (2017). *Government 3.0 – Next Generation Government Technology Infrastructure and Services: Roadmaps, Enabling Technologies & Challenges*, volume 32.
- [Ojo et al., 2016] Ojo, A., Porwol, L., Waqar, M., Stasiewicz, A., Osagie, E., Hogan, M., Harney, O., and Zeleti, F. A. (2016). Realizing the Innovation Potentials from Open Data: Stakeholders’ Perspectives on the Desired Affordances of Open Data Environment. pages 48–59. Springer, Cham.
- [Okoli, 2015] Okoli, C. (2015). A Guide to Conducting a Standalone Systematic To cite this version : A Guide to Conducting a Standalone Systematic Literature Review. *Communications of the Association for Information Systems*, 37:879–910.
- [Ouoba et al., 2016] Ouoba, J., Lahti, J., and Ahola, J. (2016). Connecting Digital Cities Return of Experience on the Development of a Data Platform for Multimodal Journey Planning.
- [Parker and Alstyne, 2016] Parker, G. and Alstyne, M. V. (2016). Platform Strategy. (January).
- [Provan and Kenis, 2008] Provan, K. G. and Kenis, P. (2008). Modes of network governance: Structure, management, and effectiveness. *Journal of Public Administration Research and Theory*, 18(2):229–252.
- [Reuver et al., 2017] Reuver, M., Sørensen, C., and Basole, R. C. (2017). The digital platform: a research agenda. *Journal of Information Technology*, (May 2018).
- [Richter and Slowinski, 2019] Richter, H. and Slowinski, P. R. (2019). The Data Sharing Economy : On the Emergence of New Intermediaries. *IIC - International Review of Intellectual Property and Competition Law*, 50(1):4–29.
- [Santhanam and Hartono, 2003] Santhanam, R. and Hartono, E. (2003). Issues in Linking Information Technology Capability to Firm Performance. *MIS Quarterly*, 27(1):125.
- [Scaria et al., 2018] Scaria, E., Berghmans, A., Pont, M., Arnaut, C., and Leconte, S. (2018). Study on data sharing between companies in Europe. Technical report.
- [Schreieck et al., 2016] Schreieck, M., Wiesche, M., and Krcmar, H. (2016). Design and Governance of Platform Ecosystems – Key Concepts and Issues for Future Research. *European Conference on Information Systems 2016 Proceedings*, pages 1–20.
- [Scott, 2017] Scott, A. (2017). What is ‘open data’ and why should we care?
- [Sennaïke et al., 2017] Sennaïke, O. A., Waqar, M., Osagie, E., Hassan, I., Stasiewicz, A., and Ojo, A. (2017). Towards Intelligent Open Data Platforms. (September):414–421.
- [Sennaïke et al., 2018] Sennaïke, O. A., Waqar, M., Osagie, E., Hassan, I., Stasiewicz, A., Porwol, L., and Ojo, A. (2018). Towards intelligent open data platforms: Discovering relatedness in datasets. *2017 Intelligent Systems Conference, IntelliSys 2017, 2018-Janua(September):414–421*.
- [Stamova, 2016] Stamova, T. (2016). Integrating open data reuse into the business models of German companies. pages 1–112.

- [Sun et al., 2015] Sun, X., Lou, Y., Li, T., and Qi, W. (2015). Designing wearable device-based product and service ecosystem.
- [Susha, 2019] Susha, I. (2019). Data driven social partnerships : Exploring an emergent trend in search of research challenges and questions. *Government Information Quarterly*, 36(1):112–128.
- [Susha and Gil-Garcia, 2019] Susha, I. and Gil-Garcia, J. R. (2019). A Collaborative Governance Approach to Partnerships Addressing Public Problems with Private Data. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, pages 2892–2901.
- [Susha and Janssen, 2017] Susha, I. and Janssen, M. (2017). Data Collaboratives as a New Frontier of Cross-Sector Partnerships in the Age of Open Data : Taxonomy Development. pages 2691–2700.
- [Susha et al., 2017a] Susha, I., Janssen, M., and Verhulst, S. (2017a). Data collaboratives as “bazaars”?: A review of coordination problems and mechanisms to match demand for data with supply. *Transforming Government: People, Process and Policy*, 11(1):157–172.
- [Susha et al., 2017b] Susha, I., Janssen, M., Verhulst, S., and Pardo, T. (2017b). Data Collaboratives. pages 604–606.
- [Tilson et al., 2012] Tilson, D., Sørensen, C., and Lyytinen, K. (2012). Change and control paradoxes in mobile infrastructure innovation: The Android and iOS mobile operating systems cases. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, pages 1324–1333. IEEE Computer Society.
- [Tiwana et al., 2010] Tiwana, A., Konsynski, B., and Bush, A. A. (2010). Platform evolution: Coevolution of platform architecture, governance, and environmental dynamics. *Information Systems Research*, 21(4):675–687.
- [Tran et al., 2016] Tran, T. T. H., Childerhouse, P., and Deakins, E. (2016). Supply chain information sharing: Challenges and risk mitigation strategies. *Journal of Manufacturing Technology Management*, 27(8):1102–1126.
- [Trochim, 2006] Trochim, W. (2006). Qualitative validity.
- [Umbach and Umbach, 2017] Umbach, C. and Umbach, C. (2017). Signature redacted Signature redacted Signature redacted.
- [van den Broek and van Veenstra, 2018] van den Broek, T. and van Veenstra, A. F. (2018). Governance of big data collaborations: How to balance regulatory compliance and disruptive innovation. *Technological Forecasting and Social Change*, 129:330–338.
- [van der Aalst, 2014] van der Aalst, W. M. P. (2014). Data Scientist: The Engineer of the Future. In *Enterprise Interoperability VI*, pages 13–26. Springer International Publishing, Cham.
- [van Steenberg et al., 2010] van Steenberg, M., Bos, R., Brinkkemper, S., van de Weerd, I., and Bekkers, W. (2010). The Design of Focus Area Maturity Models. pages 317–332. Springer, Berlin, Heidelberg.
- [Veeneman et al., 2018] Veeneman, W., Voort, H. V. D., Hirschhorn, F., Steenhuisen, B., and Klievink, B. (2018). Research in Transportation Economics PETRA : Governance as a key success factor for big data solutions in mobility. *Research in Transportation Economics*, 69(December 2017):420–429.
- [Veenstra, 2015] Veenstra, A. F. V. (2015). Modes of Governance in Inter-Organizational Data Collaborations. pages 0–12.

- [Verhulst and Sangkoya, 2015] Verhulst, S. and Sangkoya, D. (2015). Data Collaboratives: Exchanging Data to Improve People's Lives. *Medium*.
- [Wang, 2018] Wang, F. (2018). Understanding the dynamic mechanism of interagency government data sharing. *Government Information Quarterly*, 35(4):536–546.
- [Wieringa, 2014] Wieringa, R. J. (2014). The Design Cycle. In *Design Science Methodology for Information Systems and Software Engineering*, pages 27–34. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [Wohlin et al., 2012] Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., and Wesslén, A. (2012). *Experimentation in software engineering*, volume 9783642290442. Springer-Verlag Berlin Heidelberg.
- [Wolfswinkel et al., 2013] Wolfswinkel, J. F., Furtmueller, E., and Wilderom, C. P. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems*, 22(1):45–55.
- [Yin, 2017] Yin, R. (2017). *Case study research and applications: Design and methods*. Sage publications, 6th edition.
- [Zhou et al., 2017] Zhou, X., Jin, Y., Zhang, H., Li, S., and Huang, X. (2017). A map of threats to validity of systematic literature reviews in software engineering. In *Proceedings - Asia-Pacific Software Engineering Conference, APSEC*, pages 153–160. IEEE Computer Society.
- [Zimmermann and Pucihar, 2015] Zimmermann, H.-D. and Pucihar, A. (2015). Open Innovation, Open Data and New Business Models. *SSRN Electronic Journal*.
- [Zuiderwijk et al., 2015] Zuiderwijk, A., Janssen, M., Poulis, K., and van de Kaa, G. (2015). Open data for competitive advantage. pages 79–88.
- [Zuiderwijk et al., 2016] Zuiderwijk, A., Janssen, M., Van De Kaa, G., and Poulis, K. (2016). The wicked problem of commercial value creation in open data ecosystems: Policy guidelines for governments. *Information Polity*, 21(3):223–236.