

This is the author's final, peer-reviewed manuscript as accepted for publication (AAM). The version presented here may differ from the published version, or version of record, available through the publisher's website. This version does not track changes, errata, or withdrawals on the publisher's site.

Identifying the business model dimensions of data sharing: a value-based approach

Alvaro E. Arenas, Jie Mein Goh and Brian Matthews

Published version information

Citation: AE Arenas, JM Goh and B Matthews. "Identifying the business model dimensions of data sharing: a value-based approach." *Journal of the Association for Information Science and Technology*, vol. 70, no. 10 (2019): 1047-1059.

DOI: [10.1002/asi.24180](https://doi.org/10.1002/asi.24180)

This version is made available in accordance with publisher policies. Please cite only the published version using the reference above. This is the citation assigned by the publisher at the time of issuing the AAM. Please check the publisher's website for any updates.

This item was retrieved from **ePubs**, the Open Access archive of the Science and Technology Facilities Council, UK. Please contact epubs@stfc.ac.uk or go to <http://epubs.stfc.ac.uk/> for further information and policies.

Identifying the Business Model Dimensions of Data Sharing: A Value-Based Approach

Alvaro E. Arenas

IE Business School, IE University, Spain

Jie Mein Goh

Beedie School of Business, Simon Fraser University, Canada

Brian Matthews

Scientific Computing Department, Science and Technology Facilities Council, UK

Author Note

Correspondence concerning this article should be addressed to Alvaro Arenas, IE Business School, Maria de Molina 11-13-15, Madrid 28006, Spain. Email: Alvaro.Arenas@ie.edu.

Abstract

This study aims to investigate the underlying business model of organizations that have data sharing at the core of their activities. Previous work has stressed that data-sharing projects need to be sustainable in the long term, and highlighted the need for a deeper understanding of the operation model of existing data-sharing initiatives. To investigate this important issue, we take a qualitative approach to uncover the dynamics of value creation in data sharing. Using a case study method, we examined two data-sharing sites across different areas. We conducted semi-structured interviews with managers from data centers and other stakeholders, and reviewed documents about the technical and managerial practices in order to determine the main characteristics of their business models. In addition, we applied the e3-value modelling methodology to tease out the value flows within each site. Our findings have demonstrated the importance of the value network dimension of a business model, as data sharing relies on a set of actors creating and getting value in the process, and the significance of intangible assets. The main contributions of this study include extending current understanding on data-sharing business models by analyzing key dimensions, and uncovering how value is created and transferred in data sharing.

Keywords: data sharing, business models, value creation

Identifying the Business Model Dimensions of Data Sharing: A Value-Based Approach

Introduction

Data sharing is the practice of making data used for scholarly research available to other investigators. Data sharing usually involves the collection and analysis of data across multiple individuals, institutions and research sites. Sharing data in these cases becomes more than just the exchange of final datasets, as it involves the establishment of complex information and communication infrastructures, as well as organizational issues for managing such collaborations (Birnholtz & Bietz, 2003).

In the last few years, there has been tremendous interest from both private organizations and governments around the world to initiate projects on data sharing. For instance, one prominent project on data sharing undertaken by scientists across different countries was the "Human Genome Project" (HGP), an international research effort to sequence and map all the genes of the human species. The \$3.8 billion invested by the U.S. government in the HGP from 1988 to 2003 contributed to \$796 billion in economic impact and the generation of \$244 billion in total personal income (Tripp & Grueber, 2011). In 2010 alone, the human genome sequencing projects and associated genomics research and industry activity generated \$67 billion in U.S. economic output and supported 310,000 jobs, which produced \$20 billion in personal income. The genomics-enabled industry also provided \$3.7 billion in federal taxes during 2010. Data sharing has been key in stimulating innovation and entrepreneurship in this area.

Data sharing projects need to be sustainable in the long term (e.g. Berman *et al.*, 2010; Kaye & Hawkins, 2014) and more research is necessary for a deeper understanding of the operation model of existing data-sharing initiatives and determining whether a clear value proposition can be articulated for all stakeholders. To investigate this important issue, we examine the underlying business model of organizations where data sharing is central to their activities. The main research question forming the basis of our study is the following: *How is value created through data sharing?*

We take a qualitative approach to answer this question and uncover the dynamics of value creation in data sharing. Using a case study method, we examined two data sharing sites covering two different knowledge areas: chemistry & crystallography, and atmospheric sciences. We conducted semi-structured interviews with managers, end users and other stakeholders at the data centers, and review documents about their technical and managerial practices in order to determine the main characteristics of their business models. We used the

e3-value modelling methodology (Gordijn & Akkermans, 2003), which is widely used for analyzing business models with a value-based lens, to tease out the value flows within each site.

This study makes several important contributions to the literature on information science and technology. First, our study extends current understanding on data-sharing business models by analyzing key dimensions such as value proposition, value architecture, value network and finance in the context of data sharing. Second, our study advances existing data and information sharing research by uncovering how value is created and transferred in data sharing. Third, previous research on data sharing and data preservation has focused on tangible assets and their costs. Our results show the importance of intangible assets when pursuing data sharing.

The paper is structured as follows. In the next section, we define important terms such as business models and value creation, and discuss the most important findings in the recent literature. We then present the research design followed by the main findings. We close the paper with a discussion about the results and the main conclusions.

Literature Review

Data Sharing

Several studies have highlighted the importance of data sharing (Dallmeier-Tiessen *et al.*, 2014; Williamson *et al.*, 2015). Data sharing refers to “*the voluntary provision of information from one individual or institution to another for purposes of legitimate scientific research* (Boruch, 1985). Historically, data sharing has been considered a hallmark of scientific practice, allowing the confirmation of research findings, especially through the replication of results, or the reuse of data in new ways to advance knowledge.

Technology and large datasets in Science have opened up the prospects of sharing data, making data accessible to research communities around the world. Cyberinfrastructure (Atkins, 2003), e-science (Andronico *et al.*, 2011), and e-research (Anandarajan, 2010) initiatives in the United States, the European Union and other countries look at ways to allow scientists to collaborate on the creation and use of large data sets, exploiting information and telecommunication technologies.

Previous research on data sharing has included studies on barriers and enablers of data sharing (Williamson *et al.*, 2015; Dallmeier-Tiessen *et al.*, 2014; Tenopir *et al.*, 2011). For instance, Tenopir *et al.* (2011) found that barriers to effective data sharing are rooted in the practice and culture of the research process as well as the researchers themselves. Many

organizations do not provide support for their researchers for data management, but if conditions such as formal citation and sharing reprints are met, researchers would be willing to share their data, which indicates that recognition may be an incentive in data sharing. Dallmeier-Tiessen *et al.* (2014) found societal, academic and research benefits coming from data sharing; for example, among the academic benefits highlighted is the contribution to the public understanding of science, and among the research benefits emphasized are the validation of scientific results and the reuse of data. However, there is a lack of studies on the operational model of data-sharing organizations, and, in particular, on their relation to the sustainability of the data centers.

The case for data sharing as a public good has gained much traction in recent years, advocated by leaders in the scientific community such as Boulton *et al.* (2012), and policy statements from national and international bodies, for example (Pilat & Fukasaku, 2007; Castro & Korte, 2015; Douglass *et al.*, 2014). However, while it is well acknowledged that the sustainability of digital preservation is a challenge (e.g. Berman *et al.*, 2010), the value proposition for data sharing is much less well understood. There have been some well-known approaches to discussing the costs of data preservation, for instance (Beagrie *et al.*, 2008; Rosenthal *et al.*, 2012), and some approaches to categorizing the value associated with data holdings (Beagrie *et al.*, 2012, Caruso *et al.*, 2013). However, understanding the business models used in data sharing still requires systematic investigation.

Business Models

A business model (BM) generally refers to the way firms structure their activities in order to create and capture value (Zott *et al.*, 2011). The management literature presents several perspectives on the business model. For instance, Magretta (2002) introduces business models as a holistic description of how a firm does business; by contrast, Teece (2010) follows a resource-based perspective in which a business model articulates how an organization converts resources and capabilities into economic value. Osterwalder and Pigneur (2010) propose a consolidated view of a business model as a composition of several elements such as value proposition, value-creation and delivery activities, cost structure and revenue model. In all those perspectives, the notion of value takes a central position.

In the information systems community, business models are seen as a tool for depicting, innovating and evaluating business logics in organizations, especially in IT-enabled and digital industries (Veit *et al.*, 2014). The literature presents several conceptualizations of business models. A fruitful line of research is the development of

business models categories, identifying main characteristics of each category. Rappa (2003) proposed categories for online businesses such as advertising, brokering, subscription and utility, among others. In similar lines, Clemons (2009) developed several online business models not based on advertising, starting from two basic categories: models that sell some product, experience, content or service; and models that provide access to consumers and charge for access.

In order to clarify the business model concept, Al-Debei and Avison (2010) proposed a unified framework consisting of four fundamental aspects: dimensions, functions, reach and modelling principles. The *BM dimensions* comprise the primary constructs of the BM, which includes value proposition, value architecture, value network and value finance. In relation to *BM functions*, a BM can be used as a conceptual tool for aligning business strategy and technology, mediating between technology and the attainment of goals, and supporting decision-making functions to generate knowledge capital. The *BM reach* component indicates that the BM is an intermediate layer between business strategy and business processes. Finally, the principles that guide the *modelling of BMs* see a BM as “a conceptual coherent framework that provides a holistic but abstract understanding of the underlying business logic of an organization” (p. 365).

For our study, we employ Al-Debei and Avison’s BM dimensions as the main theoretical lens to examine our research question, due to their emphasis on different viewpoints in which value is created and managed within organizations. The *value proposition* dimension denotes “a way that demonstrates the business logic of creating value for customers and/or to each party involved through offering products and services that satisfy the needs of their target segments” (p. 366). The *value architecture* dimension denotes “an architecture for the organization including the technological architecture and organizational infrastructure that allows the provisioning of products and services in addition to information flow” (p. 366). The *value network* represents “a way in which an organization enables transactions through coordination and collaboration among parties and multiple companies” (p. 366). The *value finance* dimension indicates “a way in which organizations manage issues related to costing, pricing, and revenue breakdown to sustain and improve the creation of revenue” (p. 366). As shown in previous work, different industries may place a different emphasis on those four dimensions. For instance, for telecommunication providers, the value architecture has a key prominence, since it is the primary enabler of value propositions (Olla & Patel, 2002).

Research Design

This research consists of a qualitative multiple case study of two organizations that have data sharing at the core of their activities. The case study method is a well-established approach for conducting information science research (Yin, 2013). A multiple-case study was preferred to a single case study since the objective was to analyze the operation of the organizations in order to establish the main characteristics of their business model, rather than obtaining deep knowledge about one single organization (Eisenhardt, 1989). While the case study approach has limitations such as generalizability, many information science researchers have applied this method (Chu, 2015), especially for questions that are exploratory as ours.

The Cases

Our sample comprises two organizations covering data-sharing practices in two academic disciplines: material sciences, represented by the Cambridge Crystallographic Data Centre (CCDC¹); and environmental sciences, represented by the British Atmospheric Data Centre (BADDC²). They were purposefully selected as they are recognized leading organizations in their respective fields, and present a variability in areas such as operating budgets, target audiences and their size, as well as the diversity of data collections and access methods, offering a wide range of data-sharing practices (see Table 1).

The Cambridge Crystallographic Data Centre (CCDC). The CCDC provides structural databases and software for pharmaceutical discovery, materials development, research and education. The crystallographic community is often used as an “exemplar of the benefits and practices of sharing data” (Bruno & Groom, 2014) and one of the most well-known repository for this community belongs to CCDC. The CCDC compiles and distributes the Cambridge Structural Database (CSD) (Groom et al., 2016), the world's repository of experimentally determined organic and metal-organic crystal structures, and related applications software. The CSD started in 1965 with the goal of facilitating knowledge discovery as a community, and it has evolved to become a multi-user platform allowing both the storage and requests for crystallographic data from the academic community and commercial users (e.g. pharmaceuticals) all over the world. It is still experiencing huge growth and users all over the world, especially those from the crystallographic community,

¹ <http://www.ccdc.cam.ac.uk/>

² <http://www.badc.ac.uk/>

are depositing structures into its consolidated repository that now contains over 850,000 structures.

The British Atmospheric Data Centre (BADC). The BADC is the Natural Environment Research Council's (NERC) designated data center for the atmospheric sciences. The role of the BADC is to assist UK researchers to locate, access and interpret atmospheric data, and to ensure the long-term integrity of the data produced by NERC projects. BADC has also become the de-facto point of contact for UK researchers to access the meteorological products of third party data providers including the UK Met Office and the European Centre for Medium Range Weather Forecasting.

The BADC came into existence in 1994 in response to NERC's requirement for a dedicated UK data center for atmospheric research. Currently BADC has a user community with more than 22,500 users and includes more than 1 Petabyte of accessible online data. The data held at the BADC are of two types. First, datasets produced by NERC-funded projects; these datasets are of high priority, since NERC has designated BADC to be the primary long-term archive of these data. Second, third party datasets that are required by the UK atmospheric research community are most efficiently made available through one location.

Table 1. Key characteristics of the organizations

	The Cambridge Crystallographic Data Centre (CCDC)	The British Atmospheric Data Centre (BADC)
Description	Facilitate knowledge discovery for the crystallography community. Spinoff from Cambridge university.	"Assist UK atmospheric researchers to locate, access and interpret atmospheric data and to ensure the long-term integrity of atmospheric data produced by the NERC projects"
Operational History	1965 to present	1994 to present
Users	Academic researchers mainly in crystallography community, Commercial users (e.g. pharmaceuticals, agrochemical companies)	Atmospheric researchers
Data shared	Crystallographic structures	Atmospheric data

Data Collection

We triangulated data from multiple information sources. The data sources include semi-structured interviews with senior managers and key stakeholders in the two organizations as well as written and electronic documents; and interviews with experts on data sharing from other organizations. Scholars working on data sharing have previously identified main stakeholder groups (Dallmeier-Tiessen *et al.*, 2014), as illustrated in Table 2. We use that proposed stakeholder classification to collect data in our interviews, as presented in Table 3.

Table 2. Stakeholders and roles in the data sharing process.

Adapted from (Dallmeier Tiessen *et al.*, 2014)

Stakeholder group	Roles
Funders	Research funders Infrastructure funders
Researchers	Data producers Data consumers
Research and education organizations	Research planners and managers Librarians
Data management and infrastructure service providers	Data center managers and staff Other infrastructure providers
Publishers	Publishers
Policy-makers	National policy makers Regional policy makers
Business	
General public	

Table 3. Interviews per stakeholder group

Organization	Stakeholder Group	Number of Interviews	Medium
CCDC	Manager	2	Videoconference
			Videoconference
	Researcher	1	Videoconference
	Librarian	1	Videoconference
BADC	Manager	2	Person
			Videoconference

	Researcher	1	Telephone
	Research Funder	1	Videoconference
	Facility Provider	1	Telephone
Experts from Other Organizations	Data Center Manager	1	Telephone
	Publisher	1	Videoconference
	Facility Provider	1	Person

Interview questions were generated for several stakeholder groups based on Al-Debei and Avison's BM Dimensions model (see Appendix A). We interviewed four key stakeholders at CCDC and five at BADC. We also interviewed three data-sharing experts from other organizations who are in senior positions in projects and standardization organizations on data sharing. Their input was used to obtain a better insight about data sharing and to contrast their views with those of the interviewees from the centers. Among the 12 interviews, seven interviews were conducted via videoconference, three interviews were conducted via telephone, and two interviews were conducted face-to-face. We took into consideration the variety of mediums used for the interviews, and followed guidelines published elsewhere (Sturges & Hanrahan, 2004; Hanna, 2012; Farooq & de Villiers, 2017). Interviews lasted 46 minutes on average. Nine interviews were recorded and later transcribed for analysis purposes, and for the interviews which were not recorded, extensive notes were taken. Transcriptions made a total of 135 pages. To supplement the findings, we obtained documents from the organizations directly and from publicly available sources (e.g. websites, annual reports, etc.), totaling 536 pages.

To model how value flows between stakeholders group we applied the e3-value model (Gordijn & Akkermans, 2001), a methodology providing a conceptual model of the value transfers in a business network. [The eE3-value model](#) includes constructs such as actors (e.g. buyers, sellers), value objects (e.g. payment, goods), and value transfers to exchange value objects, which can [be](#) graphically represented.

Data Analysis

Data was analyzed in two steps. In the first step, we conducted a within-case analysis to review the unique patterns of each organization with the objective of identifying the

elements of the business model and value flow within each site. A team of two researchers independently conducted coding of all the transcripts. We followed a two-step coding process. First, we conducted thematic analysis (Boyatzis, 1998) where we coded our data based on Al-Debei and Avison's BM dimensions. In this phase, we coded the entire dataset of interviews identifying all the relevant extracts that correspond to the main themes: value proposition, value architecture, value network and value finance. After reviewing the themes, we checked the themes against secondary data to see if they were found in other data sources. In the second step, a cross-case analysis was conducted in search for common patterns and differences in the business models of the organizations.

The two cases are non-profit organizations. Traditionally, the literature on business models has concentrated on for-profit organizations, where the companies look for ways to create value for a customer, and the main source of revenue is the customer paying for the value (Amit & Zott, 2001). With a few exceptions, that is not true in the non-profit sector. When a non-profit finds a way to create value for a beneficiary/customer (e.g. managing the research data produced by a scientist), it is not required to produce an economic return for such an activity. Here, we leverage on previous work in social entrepreneurship that calls for having a plausible value proposition for each stakeholder group (Guclu *et al.*, 2002).

Within-Case Analysis

In this section, we analyze the business model dimensions of the organizations. For each site, we start by describing its main stakeholder groups, and then describe its BM dimensions.

The Cambridge Crystallographic Data Centre (CCDC)

Stakeholders. Table 4 describes main stakeholders for CCDC. We highlight two stakeholder groups: data producers, which are mainly researchers, and data consumers, which include the academic community and industrial users.

Table 4. Stakeholders group at CCDC

Stakeholder group / Role	Description
Data Producer / Researchers	Main data producers are researchers, who deposited data; in the majority of the cases, this is motivated by the potential of publications.
Data Consumer / Academic community	Main data consumers are scientists interested in analyzing data for their research studies.

Data Consumer / Business (Industrial users)	Main industrial users come from the pharmaceutical industry, using the data for research activities such as drug discovery.
Academic Publishers	CCDC has agreements with publishers; crystal structures associated to publications are deposited in CCDC.
Data-sharing organizations	Collaboration with the Protein Data Bank, PubChem or ChemSpider for sharing knowledge, for example.
Software vendors	Companies that provide software to solve crystal structures.

Value network. The value network represents a way in which transactions are enabled through coordination and collaboration among parties, multiple companies and the stakeholders. Here we follow the view that the value network is a multiparty stakeholder network describing how value is exchanged among stakeholders. The e3-value methodology is used to describe and analyze the value network. Figure 1 illustrates the flow of value in the case of CCDC. Stakeholders are represented as actors, and we are modelling the interchange of resources, both tangible and intangible ones. Main data producers for CCDC are researchers, who tend to be members of the academic community who deposit structures usually associated to an academic publication. The advantage of depositing data into CCDC is to get data management for the data deposited, which provides a unique identifier (DOI) and helps in data preservation. In addition, data creators can enhance their recognition and reputation, since data can be reused by scientists worldwide either as an individual structure or through knowledge derived from the collection of structures.

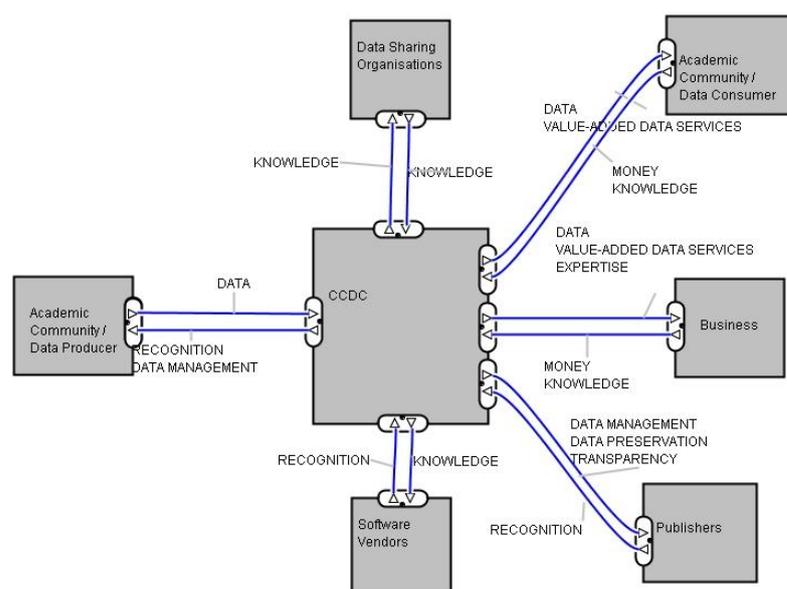


Figure 1. Value flow at CCDC.

Value Proposition. The value flow analysis serves as a base for identifying value propositions for CCDC stakeholders. Several value propositions emerged for the stakeholder groups, as presented in Table 5.

Table 5. Value propositions at CCDC.

Stakeholder role	Value proposition	Example from interviews
Data producers	Data management	<p>“We [CCDC] provide unique identifiers for the datasets so that they can be cited in papers”.</p> <p>“I think that [the] curation and sustainability they [CCDC] provide is very important”.</p>
	Recognition of their contribution	<p>“Sometime the person who did all the work doesn’t get all the credit and arguably a crystal structure determination can stand in its own right as a citeable object”.</p> <p>“[Having your structure in CCDC database] is important for citations, that gives you more visibility, and more visibility gives you more impact and that helps with funding”.</p>
Data consumers	Access to datasets produced by other researchers	<p>“Datasets are deposited and ... accessible through [CCDC] web interfaces”.</p> <p>“The real benefit is that it makes finding crystallographic data so easily, and much easier than in other disciplines”.</p>
	Access to value-added data services	<p>“We [CCDC] provide value added services [database and software] on top of the raw data that’s deposited with us”.</p> <p>“It allows you ... to visualize the structure for statistical analysis. They [CCDC] allow you to place your results very quickly into the context of everything else”.</p>
	Generation of new research	<p>“We’ve done other research where you’re not collecting any new data at all, and you’re using the data in CCDC ... You may want to know how close molecules can navigate, ... and how far you can push molecules. We have answered questions like that by using data mining on Cambridge database”.</p>
Publisher	Transparency in the research process	<p>“Providing [publication] referees with access to the data pre-publication if they want to look at that as part of the evaluation process”.</p> <p>“When someone comes to review your article, they can easily get the data, look over it, see what they need to see and to make a recommendation for the paper to publish it or not”.</p> <p>“[Data as] supplementary materials of publications are important if you want to go back and check anything”.</p>

	Preservation of data linked to published articles	“The majority of structures are associated with a publication, we need to make it as easy as possible for people to submit that data along the publication ... we provide unique identifiers for the datasets so that they can be cited in papers”.
Business	Specialized data services	“Building on top of our expertise and our familiarity with the data, there are tools that we are providing specifically to help people who are working in pharmaceutical development, which is a stage after drug discovery”.
	Expertise	“We tend to take a global view to partnership. That’s why we’re trying to establish partnerships with those organizations [pharmaceutical companies, for example]. We are providing them more than software, we are providing them with our expertise”.
Software vendors	Knowledge interchange	“[Software vendors] interested in taking advantage of CSD data to optimize some of their methodologies or just to provide a very easy means of access”.
Data-sharing organizations	Knowledge interchange	“We’ve got various collaborations in place [with the Protein Data Bank] ... that helps protein crystallographers in terms of the workflows that the Protein Data Bank have in place”.

Value architecture. The value architecture denotes the technological and organizational infrastructure required to generate value to all stakeholders. According to Al-Debei and Avison (2010), the foundation of the value architecture is the resource-based view of the firm, in which an organization corresponds to a collection of specific resources and the manager’s ability in combining the resources to exploit market opportunities is central to the performance of the organization. CCDC can be seen as a customer-centric organization. The technological infrastructure has been tailored to ease the sharing of data among researchers; on top of that, scientific tools for visualization and 3D search and analysis of structures, among other tools, have been developed.

“The internal informatics system has been built around SharePoint and Microsoft CRM. ... That system manages the depositions, it manages the process of making those public and it also takes care of making the datasets really available to the global community. It also then interacts with scientific tools that we have developed in CCDC to take care of going from the deposited data through to an entry in the database that is at the core of our value added products”.

Value finance. Value finance refers to the way organizations manage issues related to costing, pricing and revenue breakdown to sustain and improve its creation of revenue. CCDC follows a “freemium model” (Teece, 2010), where some of the services are offered free to the community, and additional value-added services are offered with a premium. As a non-profit organization, all income is invested in the management and improvement of the system.

“We have a model where we, develop software around searching the data base and using the research tools. We sell that for a license, for a fee, it’s an annual fee. Its different license for different amounts of usage, different types of business”.

The British Atmospheric Data Centre (BADC)

Stakeholders. Table 6 describes the stakeholders identified in the operation of BADC.

Table 6. Stakeholders at BADC

Stakeholder group / Role	Description
Data Producer / Instruments	Main data producers include instruments such as satellites and microscopes, capturing environmental observations and measures.
Data Producer / Collaborating Organizations	BADC stores data provided by collaborating organizations such as UK Meteorology Office and NASA.
Data Consumer / Researchers	Main data consumers are scientist interested in analyzing data for their research studies.
Data Consumer / General Public	Access the data for general interest or for verifying claims made by researchers.
Funders / Research funders	Main research funder for BADC is NERC via research grants
Funders / Infrastructure funders	Main research facility provider is STFC, who holds the center in one of its laboratories, and manages large-scale research facilities such as telescopes.

Value network. Figure 2 illustrates the flow of value in the case of BADC. Main data producers for BADC are scientific instruments like telescopes and satellites, rather than scientists. The organizations managing those instruments produce data as part of their normal operation, and BADC offers to those organizations data-management planning, in the way of data curation and license management; in addition, the data producers obtain recognition, such as being referenced in publications by researchers using the data. Another group of

important data producers comprises collaborating organizations such as NASA and the UK Meteorology Office, which also provides some of their data to BADC as part of their wide-dissemination programs, and looks for the data to be used by data consumers to obtain reputational gain in return. Both data providers may provide the data with restrictions in the license of use. Main data consumers are research scientists who access the data for their research activities. For BADC, it is crucial to maintain statistical information about the data it stores, which includes information such as number of downloads and how the data is used. Another important data consumer is the general public.

Two important intangible assets are recognition and transparency. Recognition is a value received by both organizations holding instruments, and collaborating organizations that provide data to BADC, represented in the form of reference to the organizations by researchers using the data. Transparency in the research process is an essential component of the scientific method; it is usually associated with the disclosure of hypothesis, procedure and methods, study material, data analytic plan, and raw data and metadata (Campbell et al., 2014). In the interviews, transparency was consistently highlighted as a key value for both research funders and the general public.

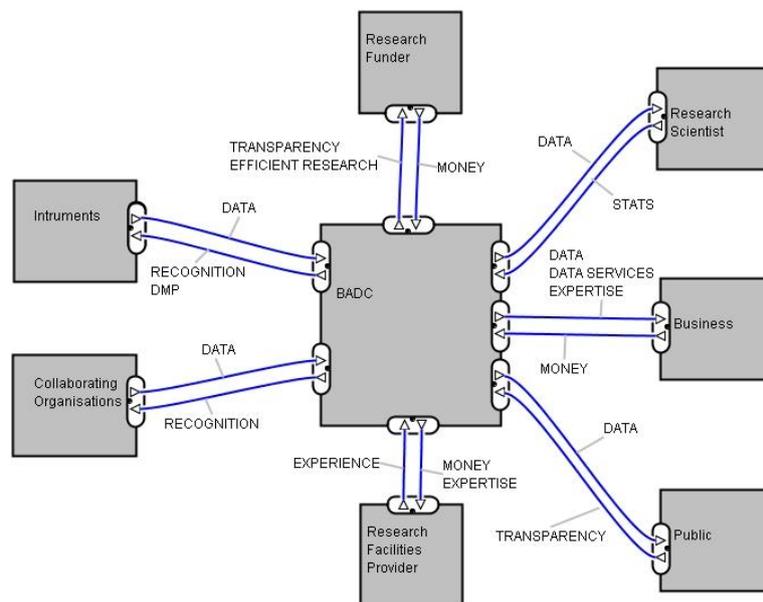


Figure 2. Value flow at BADC.

Value Proposition. Some value propositions were identified for the stakeholders groups, which are presented in Table 7.

Table 7. Value propositions at BADC.

Stakeholder role	Value proposition	Example from interviews
Data producers	Data management	<p>“We take data from the research community and ensure that is preserved in the long term and can be reused”</p> <p>“One of the things we provide is ... data management plan ... advising on the type of license ... keeping statistics of data use ...”.</p> <p>“All we [data producers] need to do is the data and then they [BADC] take care of preserving the data”.</p>
	Recognition of their contribution	<p>“People are getting credit for their research by getting [their] data reused ... they are getting a DOI ... and can be cited and published”.</p> <p>“For the researchers who are producing data, the benefit ... is the impact of their work”.</p>
Data consumers	Access to new datasets	<p>“Individual scientists can do better work because they have more datasets they can use to apply to a problem”.</p>
	Access to new data services	<p>“With Jasmin [a data-intensive cluster facility] we are sharing our data just as a simple file system ... That's a huge opportunity for people, because people are coming and processing all of it at once.”.</p> <p>“Jasmin has become increasingly important over the last few years ... it has proof to be very useful”.</p>
	Generation of new research	<p>“People are able to go back and look at all the data from different perspectives”.</p> <p>“The more the dataset is known, the more you can start to look for or compare it with other datasets”.</p>
Research funders	Efficiency in the research process	<p>“We provide data curation and make the data available to other researchers, so that they can generate new research from the data.”.</p>
Research Facilities	Gaining expertise	<p>“[BADC] has very demanding requirements, we gain expertise with them that we can apply to other projects”.</p>
Business	Expertise	<p>“We have projects for business where they ask us to produce some new data from existing data, and some kind of climate model”.</p> <p>“We work with businesses collaboratively on projects funded by that party”.</p>
General public	Transparency in the research process	<p>“During the climate change debate, we have a peak in downloading climate data, we guess people [are] corroborating other results”.</p>

Value architecture. All BADC data are available on-line through their website or via a file-transfer service. In addition, software is provided to assist in the manipulation of the data and information is provided on the data collection procedures, formats, data quality, contact names and references to journal papers associated with the data. Furthermore, BADC has developed value-added data products such as averaged and gridded data and the generation of video sequencing of data for viewing large datasets.

Value finance. In the case of BADC, approximate two thirds of their funding comes from UK research councils and then one third by other means, through European Commission, space agencies, direct government departments or commercial activity. Data consumers are not charged by the use of the data.

Cross-Case Analysis

All interviewees highlighted the importance of data sharing within their organizations. For the CCDC, “*data sharing, data dissemination is right at the core of what we exist to do ... we’re primarily an organization that’s set up to share the data that’s deposited with us or we have*”. According to its mission, the role of the BADC is “*to assist UK atmospheric researchers to locate, access and interpret atmospheric data and to ensure the long-term integrity of atmospheric data produced by NERC projects*”.

Data management and access to new data services were mentioned by the two organizations as key value propositions for data producers and data consumers respectively. Data management is at the core of data sharing activities; it includes making data available for reuse, data curation and preservation, and providing quality assurance (Mayernik, 2015). Access to new data services is also valued by data consumers. There is a variety of services offered by each organization. The term “value added services” was emphasized during the interviews in both organizations, covering several key services such as online portals for rapid search and data visualization. BADC also utilized new services and infrastructure for exploiting their data; the example that was mentioned, JASMIN³, was a data analysis environment that allows

³ <http://www.jasmin.ac.uk/>

scientists to collaborate in self-managing group workspaces, enabling models and algorithms to be evaluated alongside curated archive data.

A central dimension is the value-network dimension. The organizations emphasize on partnership and collaboration as central to their data sharing activities. For instance, CCDC mentioned “*we partner with various industries to help them do the research and so most of the software tools we have developed that are based in our collection of data have come from research problems in the real world*”. On the other hand, BADC mentioned “*we have collaborations and interactions with key partners*”, partners such as their host organization –STFC, their key funding organization –NERC, the UK Met Office and the European Space Agency, among others; they also mentioned “*we work with business sometimes collaboratively*”.

The value architectures provide evidence of a focus on customer centricity. Our analysis unveiled that there was a constant evolution of the technological architecture with updates and additions. Several of these updates were done to improve usability of the software and technological platforms. A usable interface to the software or platforms helped customers derive value easily and in the process increase the stickiness of the platform (i.e. lock in). Further, the organizations have dedicated customer support serviced either in person or through an online forum. Similar to the improvements to the technological architecture, having the skills and competences related to customer support enabled the organizations to respond to customers’ needs quickly and increase lock in. In addition, the data centricity was implemented via an organizational structure that was clearly aligned with their value propositions.

Discussion and Conclusions

In this study, we examined the business model dimensions of data sharing, following a value-based approach by applying Al-Debei and Avison (2010) framework. Two major insights emerged from our findings in relation to the BM dimensions. First, the value-network dimension arose as a key dimension for data sharing, where all stakeholders contribute to the success of data-sharing activities generating value from themselves and the network. Second, there is a range of value propositions for the

stakeholders participating in data sharing, and activities in the value network aim to achieve such value propositions for all participants.

According to Allee (2000), “*a value network generates economic value through complex dynamic exchanges between one or more enterprises, its customers, suppliers, strategic partners, and the community*”. In our cases much of the value generated corresponds to intangible value, value that goes beyond the actual service and that is not accounted for in traditional financial measures.

Analyzing the value networks at the two centers, the main asset interchanged is data, but most of interchanged assets correspond to intangible assets. Main intangible assets identified in the management literature encompass reputation, brand image, product quality, employee know-how, the level of trust between the people or organizations forming the relationships, etc. (Teece, 2015). In our study, main intangibles encompass data management, data preservation, [and](#) recognition in the use of data, knowledge and expertise. For instance, some of the value created for data producers include intangible assets such as data management and recognition in the use of the produced data. For data consumers, value created includes access to data and to data-management services, as well as the expertise provided by personnel from the data-sharing organizations. On the other hand, value for funder organizations includes improving transparency in the research process, and recognition in the research process by reuse of the data generated from funded projects. Science can only progress if there is corroboration among colleagues by reproducing research results (Munafò *et al.*, 2017), and data sharing contributes to this endeavor by making data from previous studies available. Finally, in addition to data, value for data-sharing organizations includes intangible such as knowledge and recognition of the work done.

The several assets generated in data sharing can be seen as the result of a range of value propositions that satisfies the needs of the stakeholders, as shown in Tables 5 and 7. As mentioned before, social entrepreneurship is characterized by having value propositions for each stakeholder group (Guclu *et al.*, 2002), and this has been a strong aspect in our cases. Each stakeholder group has plausible value propositions and the role of the data centers has consisted in coordinating the delivery of such value propositions, i.e. how to co-create value among all participants.

Value Co-Creation in Data Sharing

Value co-creation mechanisms are the means to generate value not only for the focal firm but also for all other stakeholders. In the management literature, there are three main value creation mechanisms: novelty, efficiency, and complementarity (Amit & Zott, 2001). *Novelty* represents the value created through new ways of conducting profitable exchange, which can be achieved by connecting previously unconnected parties, by linking transaction participants in new ways, or by designing new transaction mechanisms. *Efficiency* represents the value created through streamlining transactions and coordination activities. *Complementarity* focuses on identifying and exploiting complementary resources/capabilities among the partners such that together they are a source of value that a partner could not build on its own.

All three mechanisms were present in our findings, with complementarity being a central one and closely related to the BM value-network dimension. For instance, CCDC creates novel, value-added data services such as visualization, on top of the data provided by data producers. Efficiency is a word that emerged several times during the interview with BADC managers. Research funders get efficient research from BADC via data management and preservation from the data resulting from funded projects. The simple case of complementarity appears in the form of *data preservation*, a service provided by the organizations to the data providers, co-creating value for themselves and all their user communities. In addition, the two organizations develop IT-based value-added services on top of the “raw” data, creating value for several stakeholders. For instance, CCDC defines new services such as visualization; BADC offers data analytics services via the JASMIN infrastructure.

Important elements in value co-creation for intangible assets are value conversion and value enhancement (Allee, 2008). *Value conversion* is achieved by converting one type of value input into another kind of value as an output. Above examples of developing IT-based value-added services on top of the “raw” data deposited in the data centers are examples of value conversion. *Value enhancement* is achieved by adding features that make the value output unique. An example of value enhancement is adding metadata to the “raw” data, so that it can be further exploited by other members of the value network. Both value conversion and value enhancement may contribute to the sustainability of the organizations by transforming intangible assets in tangible assets with economic value.

The Emergence of the Community Business Model

Using the categories of business models proposed by Rappa (2003), both CCDC and BADC follow a Community Model, whose viability depends on participation of community members in the data sharing process. The Community Model depends strongly on users' loyalty, with users having a high involvement in the generation of products and services, and revenues can be based on either the sale of ancillary products and services, voluntary contributions, advertising or subscription for premium services (Rappa, 2003). Examples of Community Model include Open Source Software (OSS) and open accessible content developed collaboratively by volunteers like the case of Wikipedia; in the case of OSS, revenue is generated from services like system integration or product support, instead of licensing code for a fee; by comparison, Wikipedia's main source of income comes from contributions of the public.

There are important differences in the way the community model is implemented in both organizations: CCDC's main source of funding comes from subscriptions for premium services, and BADC's main source of funding comes from the UK government via NERC.

In the case of CCDC, subscriptions follow a freemium model, with some services free for the whole community, and access to value-added services available via an annual license fee, with versions for academics and enterprises. Other digital repositories like figshare⁴ and Dryad⁵ follow a Community Model but with slightly similar subscription models. figshare is a multidisciplinary repository for research materials supported by Digital Science of Macmillan Publishers; for researchers, figshare offers a limited amount of free storage space for private use, and an unlimited amount of storage space for publicly shared materials; however, institutions and publishers who want to use figshare should pay a license fee. Dryad is a nonprofit organization that provides long-term access to scientific and medical contents at no cost to researchers, educators or students, but it also charges a fee to institutions and publishers.

On Sustainability and Business Models for Data Sharing

⁴ <https://figshare.com/>

⁵ <https://www.datadryad.org/>

The extant literature recognizes the need for further research on sustainable data sharing. Previous authors have recognized the importance of the business model concept when discussing the sustainability of organizations (Stubbs & Cocklin, 2008; Boons *et al.*, 2013) since a business model highlights the value creation logic of an organization, its effects, and potentially allows for new governance forms such as public private partnerships or social business, among others.

In one of the early works about sustainability and business models, Stubbs and Cocklin (2008) revealed a set of normative principles of organizational development that form an “ideal type” of sustainability-oriented business model. Their ideal type comprised attributes such as providing economic as well as ecological or social value through offering products and services; having performance measurement systems that include financial as well as social and environmental indicators; and adopting a stakeholder view of the firm, understating stakeholder engagement and collaboration as key for sustainability. Some of the attributes of sustainability-oriented business models were present in both CCDC and BADC. The value-network analysis stressed the relevance of stakeholders in the generation of value for both organizations. Our findings showed how the creation of value-added services on top of the data deposited in the centers generated not only economical value but also user engagement. In the case of CCDC, researchers praised their data mining and visualization tools; in the case of BADC, researchers praised tools like JASMIN, making the work much easier.

Our study has shown the importance of intangible assets in data sharing; consequently, another element contributing to the sustainability of business models is related to value conversion and enhancement by transforming some intangible assets into tangible assets that may generate some economical value. CCDC manifested that knowledge acquired via their industrial users had contributed to the improvement of their tools and the creation of new services for the whole user community.

Previous work on sustainable digital information services has highlighted the required balance among economic, social and environmental sustainability (Chowdhury, 2013). In our cases, most identified values contribute to achieving economic and social sustainability. Hence, a pending task in data sharing is the inclusion of values that also lead to environmental sustainability.

Implications for Research and Practice

Our study makes important theoretical contributions to the emerging body of knowledge about business models for data sharing. We showed the importance of the value-network dimension in data sharing, identified value propositions for stakeholders in data sharing, and highlighted the relevance of value co-creation. These results will serve as a basis for further research on sustainable business models in data sharing. Our work paves the way for further research on value conversion and enhancement in data sharing, especially on how converting intangibles assets into tangible ones that contribute to the sustainability of data-sharing organizations.

The results obtained offer practical implications for the data and information sharing communities. Our study is, to our best knowledge, the first to apply the e3-value methodology to analyze value flow in data sharing. Application of this methodology is useful for revealing how value flows among the different stakeholders participating in data-sharing projects, and for identifying intangible assets. We suggest that data-sharing organizations should pay attention to the creation of collaboration networks and take into consideration intangible assets when measuring the effectiveness of their data-sharing programs.

Limitations and Future Research

The BM dimensions constitute one component of the complete framework proposed by Al-Debei and Avison, which also includes BM function, reach and guidelines for modelling. Hence, a natural direction for future research is to investigate the application of the complete framework to data sharing.

One limitation of this study relates to the number of cases. Future research could extend the study to include a larger number of cases, covering a wider number of sectors using data sharing. Another limitation relates to the fact that both centers studied are non-profit organizations and, consequently, an avenue for future research is to extend the study to for-profit organizations. It would be worth examining if value flows in a similar way in for-profit data sharing and whether stakeholder engagement – which is key for our case studies- is equally valid.

Finally, the e3-value framework has been used as the base for developing business model architecture and patterns in other domains (Mettler & Eurich, 2012; Glova et al., 2014). Another possible research direction includes developing business model architecture for data sharing.

References

- Al-Debei, M. M., & Avison, D. (2010). Developing a unified framework of the business model concept. *European Journal of Information Systems*, 19(3), 359-376.
- Allee, V. (2000). Reconfiguring the value network. *Journal of Business Strategy*, 21(4), 36-39.
- Allee, V. (2008). Value network analysis and value conversion of tangible and intangible assets. *Journal of Intellectual Capital*, 9(1), 5-24.
- Amit, R., & Zott, C. (2001). Value creation in e-business. *Strategic Management Journal*, 22(6-7), 493-520.
- Anandarajan, M. (Ed.). (2010). *E-research collaboration: Theory, techniques and challenges*. Springer Science & Business Media.
- Andronico, G., Ardizzone, V., Barbera, R., Becker, B., Bruno, R., Calanducci, A., ... & La Rocca, G. (2011). E-Infrastructures for e-science: a global view. *Journal of Grid Computing*, 9(2), 155-184.
- Atkins, D. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure*.
- Beagrie, N., Chruszcz, J., & Lavoie, B. (2008). Keeping research data safe: A cost model and guidance for UK universities. JISC report 2008. Retrieved from <http://www.jisc.ac.uk/media/documents/publications/keepingresearchdatasafe0408.pdf>.
- Beagrie, N., Duke, M., Patel, M., Lyon, L., Hardman, C., Kalra, D., Lavoie, B. and Woollard, M. (2012). The krds benefit analysis toolkit: Development and application. *International Journal of Digital Curation*, 7(2), pp.64-67.
- Berman, F., Lavoie, B., Ayris, P., Cohen, E., Courant, P., Dirks, L., & Friedlander, A. (2010). Sustainable economics for a digital planet: Ensuring long-term access to digital information: Final report of the blue ribbon task force on sustainable digital preservation and access.
- Birnholtz, J. P., & Bietz, M. J. (2003). Data at work: supporting sharing in science and engineering. In *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work* (pp. 339-348). ACM.
- Boons, F., Montalvo, C., Quist, J., & Wagner, M. (2013). Sustainable innovation, business models and economic performance: an overview. *Journal of Cleaner Production*, 45, 1-8.

- Boruch, R. F. (1985). Definitions, products, distinctions in data sharing. *Sharing research data*, 89-122.
- Boulton, G., Campbell, P., Collins, B., Elias, P., Hall, W., Laurie, G., O'Neill, O., Rawlins, M., Thornton, J., Vallance, P. and Walport, M. (2012). Science as an open enterprise. *The Royal Society*.
- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage.
- Bruno, I. J., & Groom, C. R. (2014). A crystallographic perspective on sharing data and knowledge. *Journal of Computer-aided Molecular Design*, 28(10), 1015-1022.
- Campbell, L., Loving, T. J., & Lebel, E. P. (2014). Enhancing transparency of the research process to increase accuracy of findings: A guide for relationship researchers. *Personal Relationships*, 21(4), 531-545.
- Caruso, G., Briguglio, L., Matthews, B., Tona, C., Albani, M. (2013) Modelling Data Value in Digital Preservation. In 10th International Conference on Preservation of Digital Objects, Lisbon, Portugal, 2-6 Sep 2013.
- Castro, D., & Korte, T. (2015). Open Data in the G8: A Review of Progress on the Open Data Charter. 2015-04-25]. <http://www.datainnovation.org/2015/03/open-data-in-the-g8>.
- Chowdhury, G. (2013). Sustainability of digital information services. *Journal of Documentation*, 69(5), 602-622.
- Chu, H. (2015). Research methods in library and information science: A content analysis. *Library & Information Science Research*, 37(1), 36-41.
- Clemons, E. K. (2009). Business models for monetizing internet applications and web sites: Experience, theory, and predictions. *Journal of Management Information Systems*, 26(2), 15-41.
- Dallmeier-Tiessen, S., Darby, R., Gitmans, K., Lambert, S., Matthews, B., Mele, S., & Wilson, M. (2014). Enabling sharing and reuse of scientific data. *New Review of Information Networking*, 19(1), 16-43.
- Douglass, K., Allard, S., Tenopir, C., Wu, L., & Frame, M. (2014). Managing scientific data as public assets: Data sharing practices and policies among full-time government employees. *Journal of the Association for Information Science and Technology*, 65(2), 251-262.
- Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532-550.

- Farooq, M. B., & de Villiers, C. (2017). Telephonic qualitative research interviews: when to consider them and how to do them. *Meditari Accountancy Research*, 25(2), 291-316.
- Glova, J., Sabol, T., & Vajda, V. (2014). Business models for the internet of things environment. *Procedia Economics and Finance*, 15, 1122-1129.
- Gordijn, J., & Akkermans, J. M. (2003). Value-based requirements engineering: exploring innovative e-commerce ideas. *Requirements Engineering*, 8(2), 114-134.
- Groom, C. R., Bruno, I. J., Lightfoot, M. P., & Ward, S. C. (2016). The Cambridge structural database. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, 72(2), 171-179.
- Guclu, A., Dees, J. G., & Anderson, B. B. (2002). The process of social entrepreneurship: Creating opportunities worthy of serious pursuit. *Center for the advancement of Social Entrepreneurship*, 1-15.
- Hanna, P. (2012). Using internet technologies (such as Skype) as a research medium: A research note. *Qualitative Research*, 12(2), 239-242.
- Kaye, J., & Hawkins, N. (2014). Data sharing policy design for consortia: challenges for sustainability. *Genome medicine*, 6(1), 4.
- Magretta, J. (2002). Why business models matter. *Harvard Business Review*, 80(5), 86-92.
- Mayernik, M. S. (2015). Research data and metadata curation as institutional issues. *Journal of the Association for Information Science and Technology*.
- Mettler, T., & Eurich, M. (2012). A “design-pattern”-based approach for analyzing e-health business models. *Health Policy and Technology*, 1(2), 77-85.
- Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., du Sert, N. P., ... & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1, 0021.
- Olla, P., & Patel, N. V. (2002). A value chain model for mobile data service providers. *Telecommunications Policy*, 26(9-10), 551-571.
- Osterwalder, A., & Pigneur, Y. (2010). *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. John Wiley and Sons, New Jersey.
- Pilat, D., & Fukasaku, Y. (2007). OECD principles and guidelines for access to research data from public funding. *Data Science Journal*, 6, OD4-OD11.
- Rappa, M. (2003). Business models on the web. Available at *Managing the Digital Enterprise website: <http://digitalenterprise.org>*.

- Rosenthal, D. S., Rosenthal, D. C., Miller, E. L., Adams, I. F., Storer, M. W., & Zadok, E. (2012). The economics of long-term digital storage. *Memory of the World in the Digital Age, Vancouver, BC*.
- Stubbs, W., & Cocklin, C. (2008). Conceptualizing a “sustainability business model”. *Organization & Environment, 21*(2), 103-127.
- Sturges, J. E., & Hanrahan, K. J. (2004). Comparing telephone and face-to-face qualitative interviewing: a research note. *Qualitative Research, 4*(1), 107-118.
- Teece, D. J. (2010). Business models, business strategy and innovation. *Long Range Planning, 43*(2-3), 172-194.
- Teece, D. J. (2015). Intangible assets and a theory of heterogeneous firms. In *Intangibles, Market Failure and Innovation Performance* (pp. 217-239). Springer International Publishing.
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., ... & Frame, M. (2011). Data sharing by scientists: practices and perceptions. *PloS one, 6*(6), e21101.
- Tripp, S., & Grueber, M. (2011). Economic impact of the human genome project. *Battelle Memorial Institute*.
- Veit, D., Clemons, E., Benlian, A., Buxmann, P., Hess, T., Kundisch, D., & Spann, M. (2014). Business models. *Business & Information Systems Engineering, 6*(1), 45-53.
- Williamson, K., Kennan, M. A., Johanson, G., & Weckert, J. (2015). Data sharing for the advancement of science: Overcoming barriers for citizen scientists. *Journal of the Association for Information Science and Technology*.
- Yin, R. K. (2013). *Case study research: Design and methods*. Sage publications.
- Zott, C., Amit, R., & Massa, L. (2011). The business model: recent developments and future research. *Journal of Management, 37*(4), 1019-1042.

Appendix A

Interview Questions

Organization and its Mission

1. How does data sharing relate to the mission of your organization?
2. What is the organizational structure for data sharing? Is there a data-sharing champion?

The BM Value Dimensions of the Organization

Value Proposition

3. Who are the main users/customers in your organizations? What are the benefits users get from data sharing?
4. Do you have an example of new opportunities in your organization from data sharing?

Value Network

5. Which organizations are key partners in your data-sharing activities?
6. What benefits do your partners derive from data sharing?

Value Architecture

7. Can you describe the technological architecture of your organization for data sharing?
8. What skills and competences are needed to manage such an architecture?

Value Finance

9. Which investments did your organization make to enable data sharing?
10. How is data sharing enabled? What is the cost? What are the trends?
11. How many full time employees in your organization are involved in enabling data sharing?
12. Is there a pricing mechanism for or income from data sharing?