ABSTRACT

Transforming Viscous Data into Liquid Data:
How Does Intermediating through Digital Platforms Impact Data?

Jordana Jeanne George, Ph.D.

Mentor: Dorothy E. Leidner, Ph.D.

This study examines how a data platform intermediary enables the evolution of viscous data into liquid data. Viscous, or difficult to use, data is the result of data usage problems that often plague information systems. Data may be viscous because of poor quality, staleness, size issues, unusable formats, missing metadata, unknown history, mysterious provenance, poor access for users, and inability to move data between systems. Viscous data is problematic to use and difficult to incorporate into decision making. On the other hand, liquid data is high quality, formatted to be machine-readable, has provenance and metadata, is easy to move in and out of different systems, is accessible by users, and lends itself well to being used for decision making. Using a longitudinal case study that follows a data platform intermediary startup company from late 2015 to 2018, I break down elements of the platform into data users, data providers, and data intermediaries. Using a lens from the Community of Practice literature, I show how social learning, data wrangling, data complementing, and data liberalizing on a digital data platform transform data from viscous to liquid. This work contributes by providing a different perspective to

data management, a means to address a dearth of data skills, and a way to make data more usable for both individuals and institutions.

Transforming Viscous Data into Liquid Data:
How Does Intermediating through Digital Platforms Impact Data

by

Jordana J. George, B.A., M.B.A., M.F.A.

A Dissertation

Approved by the Department of Information Systems

____

Jonathan Trower, Ph.D., Chairperson

Submitted to the Graduate Faculty of
Baylor University in Partial Fulfillment of the
Requirements for the Degree
of

Doctor of Philosophy

Approved by the Dissertation Committee

____

Dorothy E. Leidner, Ph.D., Chairperson

____

Jonathan Trower, Ph.D.

____

Timothy Kayworth, Ph.D.

____

Debra Burleson, Ph.D.

____

Peter Klein, Ph.D.

Accepted by the Graduate School
August 2019

____

J. Larry Lyon, Ph.D., Dean

Page bearing signatures is kept on file in the Graduate School.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank the co-founders and employees of data.world, without whom I would not have been able to complete this research. I would also like to thank the founder and members of Data for Democracy and the members of Data Journos of DC for their insightful perspectives. Last, I would like to thank my advisor, Dr. Dorothy E. Leidner, for her unfailing support and wisdom.

DEDICATION

To my family, Douglas George, Jeffrey George, and Thomas George,

for their unwavering love and support

CHAPTER ONE

Introduction

Is Data the New Oil? How one startup is rescuing the world's most valuable asset.
— Forbes, 2017

The thought-provoking article in Forbes magazine related how data.world, a young Austin, Texas startup was founded with the mission to democratize and network data in the new data economy where data was the most valuable asset. Data has long been a key element in IS dating back to the first computers of the 1950s designed for data processing (Hirschheim and Klein 2012). Technology is an information processing tool (Orlikowski and Iacono 2001) and information is grounded in data (Alavi and Leidner 2001). However, storing rows and columns of data for monthly sales reports or key performance metrics is a different activity than data work we see today. Organizations are exploiting data through combining data from multiple sources, not just company data. Data exhaust from IoT artifacts may be aggregated with government open data, social media, and visualizations to create new insights. Modern data use succeeds not so much on gathering data as making use of it, which are the cornerstones of the data economy (The Economist, 2017). I define the data economy as an economic environment where those who both have data and can exploit it hold greater economic power and strategic advantage (Chahal, 2014; Elbaz, 2012; Otto & Aier, 2013; Zech, 2017). Having data alone is not enough. It is important for IS scholars to understand the character and implications of how organizations use data today.

In this work I wanted to understand the main challenges of exploiting data and how these challenges could be addressed.

To better understand how data is used and the nature of data itself, I spent two years at data.world conducting a qualitative grounded theory study of the startup, the people involved, the data in question, and the value proposition the firm provided. I gathered information from the company co-founders, employees, users, the press, and external data groups to build a rich picture of where data usage is at present and how data.world, a data platform intermediary, identified and addressed the challenges of exploiting data.

There are several main obstacles to successfully working with data. The primary obstacle is not gathering the data, but the frequent inability to make use of collected data. Although vast quantities of data are produced every minute (Brynjolfsson, Geva, & Reichman, 2016), the people who need the information from the data can rarely get it. The factors that contribute to this state include data in unusable condition, inability to get to the data or move it in or out of other systems and tools, and a dearth of data skills required to work the data (Shah, Horne, & Capellá, 2012; Thomas & Rosenman, 2006; Wladawsky-Berger, 2017). Enormous amounts of data are now available that could solve many global problems from the environment to economics to health to education (Kirkpatrick, 2013). If those who need the data cannot use it, however, it is useless. The term I found to describe difficult-to-use and hard-to-access data is *viscous* (Mallah, 2018; van Schalkwyk, Willmers, & McNaughton, 2016). Viscous data is a contrast to *liquid* data, which is easily accessible, in usable formats, and meets quality standards, among other characteristics (Chui, Manyika, & Kuiken, 2014; Jetzek, 2016). This led me to observe that one of the main obstacles to using data is the difficulty in changing viscous data into liquid data. The inability to use data results in missed opportunities. For example, there are rich insights to be found when data is aggregated from multiple sources (Chiang, Grover, Liang, & Zhang, 2018). I use two examples, Pinterest and IBM/The Weather Company, to illustrate.

Pinterest is a digital scrapbook company, recently valued at $12.7 billion in its 2019 IPO (Griffith, 2019). The Pinterest model allows users to save or "pin" images and articles to virtual bulletin boards, and intersperses content with e-commerce shopping and ads. Pinterest holds enormous amounts of data on users' likes and preferences and these data assets hold significant value for advertisers and retailers (Aslam, 2017; Benner, 2017; D'Onfro, 2014), by enabling highly targeted marketing both on and off the Pinterest platform. The data from the platform aggregates users' interests, the ads they respond to, which content is saved, and what combinations of pins comprise user accounts and user categories within their accounts. For example, user JWI has 21 "boards" (virtual bulletin boards) in her Pinterest account. These might include such varied topics as Restoring Vintage Furniture, Wine Cocktails, Bathroom Remodeling, Homemade Skincare Ideas, Cute Cats, and Instant Pot recipes. The aggregation of 21 varied topics provides insights similar to "market basket" business analytics in that we may find that users that share these particular boards are also interested in Charming Small Towns (travel sales opportunity) or Comfortable and Cute Clothing (clothing sales opportunity). At an even more granular level, which data is aggregated in similar categories may provide insights. For example, boards categorized as Gardening are common on the platform, but the boards vary in what is "pinned" to that board. Some users might pin Easy Maintenance Plants, Too Much Zucchini from Your Vegetable Garden, and Greenhouses Made from Old Windows, while others pin Bug-repellent Plants, Building a Firepit, and Brick Patios. These individual aggregations provide marketing opportunities when the data shows that even a generally understood topic such as Gardening has different facets that may not have been obvious before.

Another example is IBM. In 2015, IBM purchased The Weather Company with a goal to provide hyper-local weather forecasts along with the forecasted impact on logistics and consumer buying behaviors to client companies (Booton, 2016). To illustrate how weather data can be combined with logistics and purchasing habits, I look at the grocery store chain HEB in Texas. When bad weather such as heavy rains, flooding, hurricanes, and tornados occur, customers change their buying patterns. Instead of purchasing fresh produce and fresh meats, customers focus on bottled water, batteries, canned meats, and bread (Morago, 2017). When a store has advance warning of bad weather it can redirect produce and fresh meat shipments to other locales, which reduces waste. It can also redirect bottled water, bread, batteries, and canned tuna to the weather-afflicted stores where such items may experience shortages. These examples demonstrate a few ways how aggregated data adds value. However, when data can't be used, not only are opportunities missed but the cost of storing and securing data becomes an expense with no little payback (Alexander, 2016).

While much of the scholarly research on data management is focused on either the mechanics of data, such as data warehousing or analytics (Sen, Ramamurthy, & Sinha, 2012), or on knowledge management, such as the creation and transfer of knowledge (Szulanski, Ringov, & Jensen, 2016), relatively few works address the strategic value of data or address the problem of why data is not being used to its full potential (Chiang, Grover, Liang, & Zhang, 2018). Chiang et al., (2018) state, “Analysis of data without generating value offers no contribution to an organization, regardless of whether data are big or small.” I identify the main problem as how to get value out of data, and specifically, I focus on the issues of unusable, viscous data and how to enable organizations to get the most out of their data. It is a problem because data could offer great value but rarely reaches its potential for strategic advantage. This is further clouded by current data fads where managers are instituting data programs without fully understanding the costs, value, or how to use the data (Chen, Chiang, & Storey, 2012). The general objective of this paper is to identify the roadblocks that prevent strategic use of data and how those challenges may be overcome in a particular setting. Thinking along these lines, I decided to conduct a longitudinal study of Austin-based data.world, which is a data platform that serves as an intermediary for data providers and data users. Factors in my decision to work with data.world included prior acquaintance with several people at the company and everyone’s open and welcoming attitude towards my research. I spent a day onsite at data.world every other Friday for over a year, was provided access to Slack chats ranging from 2015 to 2018, and was given a range of other data from user interviews to usage statistics, although I did not use all of the data because there was too much for the scope of this dissertation. My method was iterative as I asked questions and listened and learned, then reflected upon the new information. This generated new questions and new reflections and refined my thoughts about liquid and viscous data and how platform intermediation changed things. From these concepts and my investigation of data.world I formed my research question: *How does intermediating through digital platforms impact data?*

The data.world experience enriched my knowledge of who created data, who had data access, and how people and organizations used data. Perhaps most importantly, my experience with data.world exposed me to data problems. While the big data and data warehousing literature espoused data volume, velocity, veracity and variety (Chiang, Grover, Liang, & Zhang, 2018), I found other aspects of data management that were equally or even more important. First, the democratization of data has become more popular at both government and organizational levels. Data democratization means allowing data access to a wider audience. First introduced in the open government data movement, democratization has begun to spread to intra-organizational and inter-organizational data sharing. Second, even when data is opened and made available, if people lack the skills to work the data, it has little practical use. Third, data work and education are improved through participation in data communities. Fourth, data platform intermediaries such as data.world address these aspects by providing a cloud platform with easy access, automating tools for moving and transforming data, and enabling sociality through social media features.

In terms of theory, the analysis of my research yielded these mechanisms: social learning, data wrangling, data complementing, data liberation, and data liquidating, which resulted in the democratization of data. In short, liquidating data not only made it easily available but also made it usable by people with minimal data skills, thus broadening both who had access to data and who could use it. When liquid data was made available on the platform, it resulted in the democratization of data.

Theory does not always provide direction for practitioners, however, and sometimes practitioners are quick to jump on the latest managerial fad such as big data

(Abrahamson, 1991). I hope to provide direction for practitioners when considering data projects. There is considerable cost to gathering, storing, and using data and if firms don't realize value from the data, it can be a considerable business expense that provides no payback. On the other hand, when organizations can harness data and utilize it appropriately, they may discover numerous opportunities for growing the business, increasing strategic advantage, or improving the bottom line. One of the biggest problems for practitioners is viscous data. It is expensive to keep around, hard to access, difficult to use, and useless for decision making. However, when firms change viscous data into liquid data, such as through the use of data platform intermediaries, they are better positioned to leverage the value of their data.

The paper continues as follows: First I provide the foundations of my research in a review of the literature in data, data systems, and the data cycle. Next, I provide my methodology and research setting. I then provide my analysis and examine it in the Discussion section. Last, I conclude with implications, limitations, and future research directions.

CHAPTER TWO

Foundations

Overview of the Literature

In this chapter I provide a background on the various topics that inform this study. I begin with the methods used in developing the literature review. After that, I discuss data, including, open data, the semantic web, and aspects and characteristics of data. In the next section, I unpack data systems, which involves data warehousing and platforms. I next investigate data people which I break down into data users, data providers, and data intermediaries. Last, I discuss communities of practice, which I use as theoretical sensitizing devices (Glaser & Holton, 2004).

Literature Review Method

Google Scholar, Web of Science, and general Google Search were primarily used for this literature review, and backward and forward searches employed on particularly relevant articles. Google Search was especially useful to find business journalism on the topic of data. The key words used in the searches included data, platforms, data economy, open data, and data intermediary. Article lists were sorted by relevance and incrementally filtered again by limiting dates to more recent work. The search period began with no limitations, then was filtered to 2000-2018, and finally to 2003-2018. Once the lists were under 1000 articles, individual articles were manually selected from the larger lists by viewing the title and abstract for relevance and rigor. In the literature selection, priority was given to articles that fell into top tier journals, including the IS Senior Scholars' Basket of

Eight, the University of Texas Dallas top business journals list (http://jindal.utdallas.edu/the-utd-top-100-business-school-research-rankings/list-of-journals), and the Financial Times 50 (https://www.ft.com/content/3405a512-5cbb-11e1-8f1f-00144feabdc0).

Due to a lack of top journal articles in the relatively new fields of the data economy and open data, research from conference proceedings, practitioner journals, and lesser known or niche publications were also utilized. To maintain quality, those journals ranked two or greater from the Chartered Association of Business Schools Academic Journal Guide (ABS AJG) were included. There are several excellent practitioner journals ranked at two by the ABS AJG, such as MIS Quarterly Executive; therefore, two seemed to be an appropriate cut-off. The conference proceedings that were searched include the International Conference on Information Systems (ICIS), the Hawaii International Conference on Systems Science (HICSS), the American Conference on Information Systems (AMCIS), the European Conference on Information Systems (ECIS), the Pacific Conference on Information Systems (PACIS), Academy of Management (AOM), and the Institute of Electrical and Electronics Engineers conferences (IEEE). The final filtered list resulted in 366 articles and several books to be examined in more depth for inclusion in the literature review. Of these, 228 were used in the literature review.

In approaching the literature, my goal was to remain true to grounded theory principles while still maintaining rigor, relevance, and staying methodologically coherent (Templier and Paré, 2015). Because the topic of data intermediary platforms is relatively new, I opted for a theoretical literature review that offers "context for identifying, describing, and transforming into a higher order" (Paré, Trudel, Jaana, & Kitsiou, 2015).

Literature reviews are helpful for new topics through increasing visibility and theoretical discussions, not just for mature topics (Webster and Watson, 2002). While the literature informed the research question of how digital data platform intermediaries change viscous data into liquid data, I did not focus on gap spotting. As a new topic, there are far too many gaps for this to provide value to the researcher, therefore I developed an organizing review. Literature reviews fall into four quadrants divided by a vertical axis of research objective moving from synthesis to theorizing, and a horizontal axis of review focus that flows from broad description to a narrow view for trend/gap identification. Figure 1 (from Leidner, 2018) illustrates the framework. As an organizing review, my goal was to synthesize the literature with the case study and provide a foundation for theorizing. Organizing reviews are appropriate for new phenomenon (Leidner, 2018).

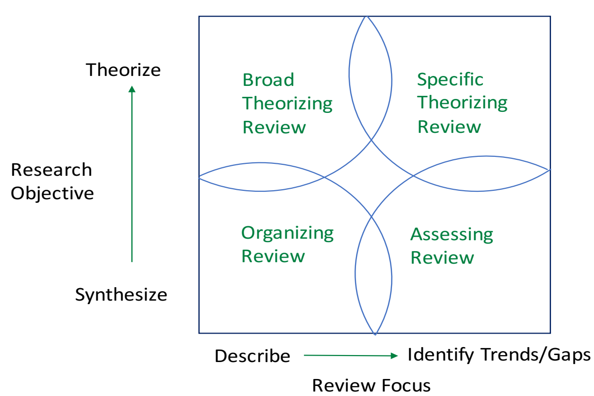

Figure 1. A Polylithic Framework of RTD Papers (from Leidner, 2018)

We now move to the literature, which I break down into four sections that all related to the data.world case: data, data systems, the data cycle, and sensitizing devices, which is a method I use to provide perspective but without *a priori* expectations.

The data section covers aspects of data, meta data and provenance, viscous and liquid data, open data, semantic web, and data workers. Data systems are broken down into data repositories, data warehouses, and data platforms while the data cycle examines data providers, users, and intermediaries. The relationships between these topics are illustrated in Figure 2.

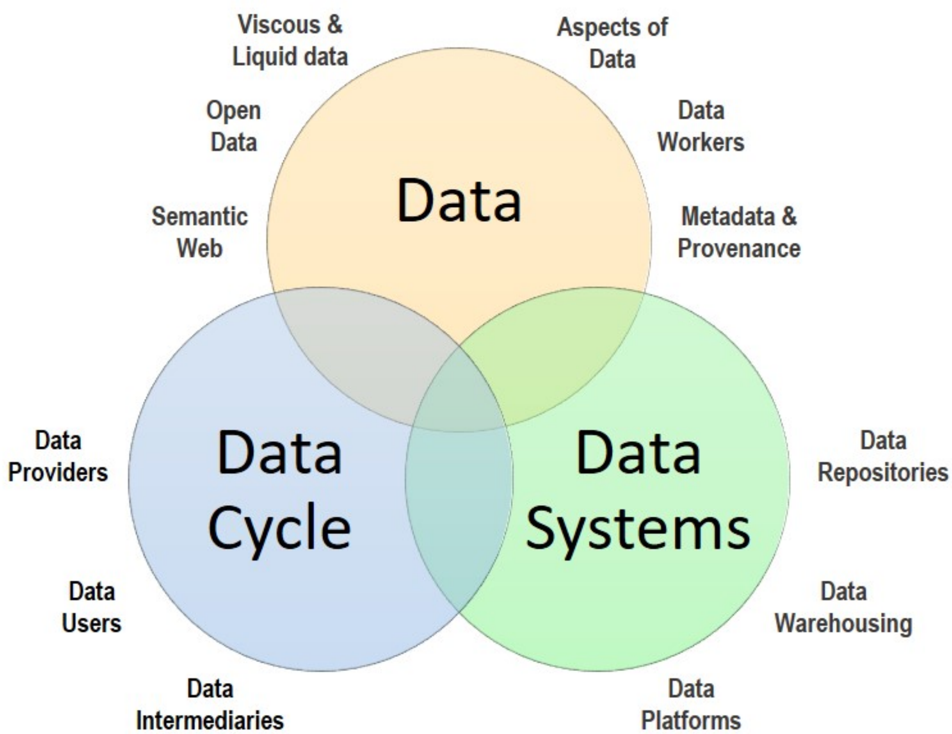

Figure 2. Literature Review Data Topics

Data

Data is a raw material. When data is processed it becomes information. When information is given context, it becomes knowledge (Alavi and Leidner 2001). I approach this research with the mindset that data is the foundation of knowledge, therefore, I explore several facets of data that impact this research. This section is broken into several topics:

data in general, meta data and provenance, viscous and liquid data, open data, semantic web, and data workers.

Because this dissertation explores the structures, people, governance, and economies of data, it is logical to provide some background on the object of this activity: the data itself, with some history and its current state. Data is one of the oldest topics in Information Systems with research originating in database management methodologies from the 1970s and 1980s (Chaudhuri, et al., 2011). Modern business intelligence and analytics are based in the statistical, analytical, and data mining processes developed decades before big data (Chen, et al., 2012). However, the text files of the 1970s bear little resemblance to the broad range of 21st century data that require far more than SQL and OLAP (online analytical processing) (Chaudhuri, et al., 2011). Chen et al., (2012) provide a clear progression of how data management has changed, which they label BI&A 1.0, 2.0, and 3.0 for the three stages of business intelligence and analytics. BI&A 1.0, the earliest form, was based on structured content and database management systems. BI&A 2.0 featured early unstructured content and web services. The last phase, BI&A 3.0, expands to include content from IoT, sensors, and mobile technologies (Chen et al., 2012). As data expanded into data warehousing and big data, it was commonly described in terms of 3 Vs: velocity, volume, and variety (Sagiroglu & Sinanc, 2013; Zikopoulos & Eaton, 2011). This has since been expanded to reflect the greater diversity of modern data to include volume, variety, velocity, value, and complexity (Kaisler, et al., 2013).

- Volume is the double-edged sword of 21st century data. Where data was once scarce it is now growing exponentially and disrupting traditional data warehouses and data management practices (Chen, et al., 2012; Marr, 2016). The business press and data industry vendors have drawn attention to data growth with evermore impressive headlines:

- AnalyticsWeek: More data was created in the past couple of years than in the "entire previous history of the human race" (Kumar, 2017)
- Forbes: 163 Zettabytes of data may be created by 2025 (Cave, 2017)
- IBM: We currently create 2.5 quintillion bytes of data every day (IBM, 2017)

Contrary to logic, the data deluge may exacerbate scarcity because it will become so difficult to find the wheat among the chaff, or the data one is searching for in the vast quantities, formats, and sources available. Combining data from several sources may lead to violations of traditional database principles, which in turn leads to imperfect datasets which result in the inability to find the data you need (Helland, 2011). One way to deal with imperfect data is to satisfice, or decide that the big picture from the data is good enough (Helland, 2011; Mayer-Schönberger and Cukier, 2014).

Data formats today demonstrate enormous variety. Graphics provide a particularly captivating subset of data with still images, video, and even embedded cryptocurrency that can serve as a mechanism for digital image rights (KodakOne, 2017; Roose, 2018). One of the more interesting applications of graphics data combined with facial recognition software is the Google Arts and Culture app for smartphones. The app has a feature that allows users to take a selfie that is compared to images of portraits from 1,500 museums in 70 countries, purporting to find the user's doppelgänger (Chen, 2018; Shu 2018). Even with seemingly innocent entertainment of this sort, a range of data concerns cropped up. When the selfie tool was released, users in Texas could not access it because of the state's privacy laws regarding biometric data (Reitz, 2018). As the selfie feature's international popularity grew, so did concerns about the Euro-centric portrait collections that were prominent in the dataset. Lack of representation offended many people from non-white races and cultures, (Chen, 2018; Shu, 2018). Freely sharing data, or "opening the kimono" theoretically should reduce conflict and increase unity through transparency (Bertot, et al., 2010); however, we may need to go through a period of uncomfortable discovery before we reap the benefits of openness (Berrone, et al., 2016).

There are other characteristics about data that make it different. First, data is non-rivalrous in that it may be used over and over again without diminishment (Opher et al., 2016; The Economist, 2017; Zeleti et al., 2016). Second, the value of data may increase with usage, as we see with social media that depend on network effects to grow (Borgatti & Halgin, 2011; The Economist, 2017; Weber, et al., 2012; Wladawsky-Berger, 2017). Third, data often holds little value in small quantities and but achieves great value in large quantities, even when terminal usage targets individual users (Bechmann, 2013; Brynjolfsson, et al., 2016; Rosenbaum, 2013). In summary, working with data today requires greater understanding of complicated relationships and characteristics.

Metadata and Provenance

It is interesting to note that metadata, or data about the data, was often “a second-class citizen in the world of databases and data warehouses” (Sen, 2004). Yet this has changed both in data warehousing and especially in both the semantic web and in data repositories because data without accompanying information often has little value. Metadata is where search begins (Garshol, 2004). Metadata and provenance -- the origin and history of the data -- add tremendous value. Metadata is important so that the data producers, the data publishers, and the end-users of the data have what they need to use the data. Metadata must be "consistent, comprehensive, and flexible" (Musgrave, 2003). Metadata might include schema, title, dates, or creators among others. It is particularly important for non-text media such as images, video, or audio where text cannot be searched. Provenance, on the other hand, provides a record of where the data is from, who has worked on it, what changes might have been made or which errors were corrected and how. It adds a layer of transparency. Provenance can tell us how missing data was handled or how standardization was accomplished. Provenance is important for trust in the data because it implies accountability (Moreau, 2010; "Provenance Incubator Group," 2009).

Both open data and the semantic web literature have provided a number of new data concepts, terminologies, and ways to improve data access and usage. Liquid and viscous data are terms that describe how easily data flows from provider to user, with value increasing along with liquidity (Manyika, et al. 2013; van Schalkwyk, et al. 2016).

Viscous and Liquid Data

Viscous data is difficult to use, and those difficulties arise from myriad reasons (Mallah, 2018; Wang, 2012). Viscous data may be provided in a PDF format or placed on a web page instead of a machine-readable CSV file. This would require a user to either scrape the web page or copy the PDF text (if the report allowed it) and paste it into a spreadsheet. It could then be saved as a CSV file and uploaded into an analysis tool. Viscous data may have latencies or lag times, rendering it useless for time-sensitive decisions (Vorhies, 2012). It may be too large to be easily moved from a repository to an analytics tool. It might be poor quality, lacking accuracy or missing fields. Data may also be viscous because of missing metadata. When tags are missing, search becomes very difficult. Non-standard field names are a constant problem when trying to consolidate data, as are non-standard file formats. Basically, viscous data has low usability because the data is difficult to work with (Mallah, 2018; Sreenivasan, 2017; Vorhies, 2014; Wang, 2012). Therefore, viscous data holds less value than it might if those problems could be solved.

On the other hand, liquid data is easy to use. It flows through systems with ease, it is high quality, on time, and offers a lot of information about the data, its sources, and its history (Chui, et al., 2014; Jetzek 2016; Manyika, et al. 2013). An example of liquid data would be a dataset that is updated in a machine-readable CSV file in a timely manner, is complete, accurate, and is easy to access and use. The file is accompanied by meta-data, data dictionaries, and a summary. Perhaps it even has an audit trail of who has worked on the file and what changes they made and an API to move it to another system. Data liquidity is accomplished through standardization, common file formats, and ease of access such as APIs or quick downloads of zipped files. Complementary files also make data more liquid, such as data dictionaries and summaries. Data viscosity and liquidity are summarized in Table 1.

The most viscous data is known as dark data. Gartner (2016) defines dark data as "information assets" that companies stockpile and ignore. Such data may include those required for legal reasons and may not even allow access for organizational use. In these cases, dark data is a burden to the organization due to the cost of storage and security with no benefit to offset the expense. On the other hand, where dark data is not legally bound, data analysis vendors such as IBM suggest that dark data may hold goldmines of unexploited data assets (Moreno, 2016). The challenge in converting this data into usable formats remains a daunting task. Identifying the causes of viscous data helps organizations mitigate the problems. The reason we want to convert viscous data into liquid data is to be able to use it.

Table 1. Viscous and Liquid Data

Attribute	Viscous	Liquid
Completeness	Incomplete data is difficult to analyze & provides inaccurate analyses	Complete data gives a more accurate picture.
Machine-readability	Non-machine-readable data must be transformed before use. It is difficult to access & hard to search.	Machine-readable data is easily uploaded and searched.
Coding	Non-standard coding makes it difficult to understand what the data represents.	Standardized coding helps analysts comprehend the meaning behind the data.
No tags or keywords	Lack of tags makes search difficult.	Tagging and keywords make search more efficient.
Timeliness	Untimely data may be no longer accurate or may be too late to be of value.	Timely data reflects what is actually occurring now.
Metadata	Missing metadata may lead to misunderstanding.	Complete metadata helps analysts understand the material.
Provenance	Lack of history may cause incorrect assumptions about data veracity and accuracy, such as who worked on the data or how missing data was handled.	Provenance offers the history of the data so that analysts can draw upon others who have worked on it or understand if/how the data was manipulated.
Accessibility	Data that is hard to access is used by fewer people.	Data that is easy to access is used by more people.
Darkness	Dark data that is intentionally hidden or ignored provides little to no value & may cost to keep up.	Data that is open garners more users and provides greater overall benefit.

Open Data

Open data is a phenomenon that has been around for several decades but has become more prominent in the past decade. It provides valuable information for economic, regulatory, research, and educational development (Zeleti, et al. 2016; Kassen, 2013). I include this section on open data, how it is created and how it is used, to inform the reader about its significance in modern data work and prepare the reader for its role in the data platform data.world. Economic improvements are fueled by increases in efficiency, innovation, transparency, and participation (Bertot, et al., 2010; Jetzek, et al., 2013). Open data is primarily associated with open government data (OGD) offering datasets ranging from educational statistics to farming, weather to censuses, and budgets to contracts (Jaeger & Bertot, 2010; Link et al., 2017; Data.gov, 2017; Data.gov.uk, 2018). It is important for governments to maintain open data initiatives over time so that data assets retain their quality (Zeleti, et al. 2016). Missing data, redacted data, and removed data all negatively impact the worth of open data. If government budgets cut data collection funding for a year, the lost data may not be recoverable. This can result in data gaps that reduce value. Managed properly, however, OGD has the potential to be an important driver of the economy. The European Commission estimates that the returns from OGD could be worth €40 billion a year to the EU economy (European Commission 2017).

Open data is not limited to OGD, however. A few for-profit companies have embraced the concept of open data and freely share both raw data and formatted reports and visualizations as part of their business strategy. As one example, the real estate website Zillow.com, provides a full set of open data resources through Zillow Research. The mission statement of this group states, "Zillow Research aims to be the most open, authoritative source for timely and accurate housing data and unbiased insight. Our goal is to empower consumers, industry professionals, policy makers and researchers to better understand the housing market." The resources provided by Zillow Research include raw data downloads, full reports, and visualizations (Zillow.com/research, 2017). Another source of open data is found on Google where top search trends are not only listed in reports, but email subscriptions and APIs are provided for automated access to the data (Trends.google.com, 2018).

Open data offers a number of benefits to those that can leverage it, impacting efficiencies, innovation, transparency, and participation (Jetzek, et al., 2013). Efficiency improves when gatekeepers are removed and data is easily available, while innovation arises from the vast stores of new information available that can highlight opportunities.

Transparency, such as that in open government data, improves governance and fiduciary operations as it exposes bribery, favoritism, larceny, and other illegal or unethical activities. Last, openness invites participation from a range of parties, including private business, citizen scientists, digital activists, the press, and researchers (Jetzek, et al., 2013; Manyika et al., 2013; Veljković, et al., 2014; Zuiderwijk & Janssen, 2014).

Semantic Web

The semantic web is the brainchild of world wide web creator Tim Berners-Lee and is seen as the tool that will provide linked, semantic data on the web. The semantic web is an extension of the current world wide web. The current web is syntactic or structural, where everything follows rules of syntax. (Berners-Lee, 2009; Berners-Lee, et al., 2001; Bernstein, et al., 2016). However, humans are not structured like this. We just "know" things and that meaning is semantics. We "know" that noontime is during the day or that rain will get you wet. In the semantic web, semantics are the *meanings* of things defined for an application, which allows systems and people to better understand relationships, qualities, and knowledge of the web content. The semantic web offers well-defined meanings to resources, services, and data that enable people and computers to coordinate and collaborate together (Islam, et al., 2010).

The semantic web starts with the "magic of hyperlinks," where "anything can link to anything" (Berners-Lee, et al., 2001). The semantic web is more than just loading content - it is creating links so that people and machines can explore and find more content on the "web of data." When you find one thing and it is linked, it leads you to the next thing. "Properly designed, the Semantic Web can assist the evolution of human knowledge as a whole" (Berners-Lee, 2009; Berners-Lee, et al., 2001; Bernstein, et al., 2016).

Linked data is used for the purpose of sharing knowledge and publishing content so that it can be found. It is based on the premise that more linkages increase the usefulness of data. When things are linked, it is easier to explore and find new things (Berners-Lee, et al., 2001; Bernstein, et al, 2016). The main goal behind using linked data is to easily enable knowledge sharing and publishing. The basic assumption is that the usefulness of linked data will increase the more it is interlinked with other data.

The technology behind the semantic web is based on ontologies. "Ontology is one of the most vital components of semantic web. It is a collection of concepts, attributes, relationships among the concepts and instances of these concepts. It provides the vocabulary of a particular domain such that web contents related to that particular domain can be understood" (Islam, et al., 2010). Ontology languages are XML or non-XML based. XML languages include the two most widely used languages, RDF and OWL, although there are a number of others. One of the challenges in building ontologies is that people don't know what they don't know (Bera et al., 2011). Guided visual ontologies are helpful for people navigating through content (Ibid), but less so for machines. Berners-Lee identified rules for building for the semantic web, and which interestingly also apply to open data. These are also illustrated in Figure 3. Data that followed these rules is known as 5-Star Data.

1. Make the data available on the web: assign URIs to identify things.
2. Make the data machine readable: use HTTP URIs so that looking up these names is easy.
3. Don't use proprietary formats (later added to the original four guidelines as these became a problem).
4. Use publishing standards: when the lookup is done provide useful information using standards like RDF.

5. Link your data: include links to other resources to enable users to discover more things.
(5-Star Open Data, 2010; Berners-Lee, Hendler, & Lassila, 2001)

Figure 3. 5 Stars of Linked Open Data

While the semantic web has not grown at the pace that many have desired, it has gained considerable ground in two decades. Over 2 1/2 billion web pages now have semantic links. Organizations, governments, and enterprises use the semantic web (Sheth, 2003). Private firms include Google, Facebook, Oracle, and Microsoft, but it is also used by non-tech organizations such as news agencies (the New York Times and the BBC), libraries, and museums (Attard, Orlandi, Scerri, & Auer, 2015; Bernstein, Hendler, & Noy, 2016). Semantic web has been embraced in healthcare, as well, such as the World Health Organization. Semantic web has even begun to seep into traditional data warehouses where OLAP is combined with RDF and OWL (Nebot & Berlanga, 2012). Several things impact the adoption of semantic web technologies. Absorptive capacity is a primary issue (Joo, 2011). Other influences include demand pull from new services and search/integration problems with current systems; technology pushes from others such as government, business potential, suppliers; environmental conduciveness; organizational competence; users' over-expectations (which negatively impacted adoption); additional financial investments required; building and pooling ontologies with others, and whether benefits could be demonstrated (Joo, 2011).

An example of how the semantic web is used may be found in the Amsterdam Museum project. In 2010 the museum, filled with thousands of cultural objects relating to Amsterdam and its people, put its collection online using a creative commons license. The collection includes images of a wide variety of artifacts from art to furniture to clothing to books. The museum can only physically display 20% of the collection at any one time due to space constraints, therefore putting it online increases visibility. The museum uses a digital data management system to handle metadata and authority files. The collection is machine-readable and offers an API. The entire collection of 30,000 pieces (approximately 90% of the pieces not on display) can be downloaded or the database can be searched for creator, year, material, title, or type among other criteria. Many European museums try to use the same linked data aggregator, such as Europeana, because they are too small to create their own linked data. The Europeana Linked Data pilot pulls from 200 institutions to create their metadata records, which are then restructured for their data model. In addition to the metadata, the Amsterdam museum also provided a thesaurus which provided 28,000 concepts, terms, and relations. A search page from the Amsterdam Museum website is illustrated in Figure 4.

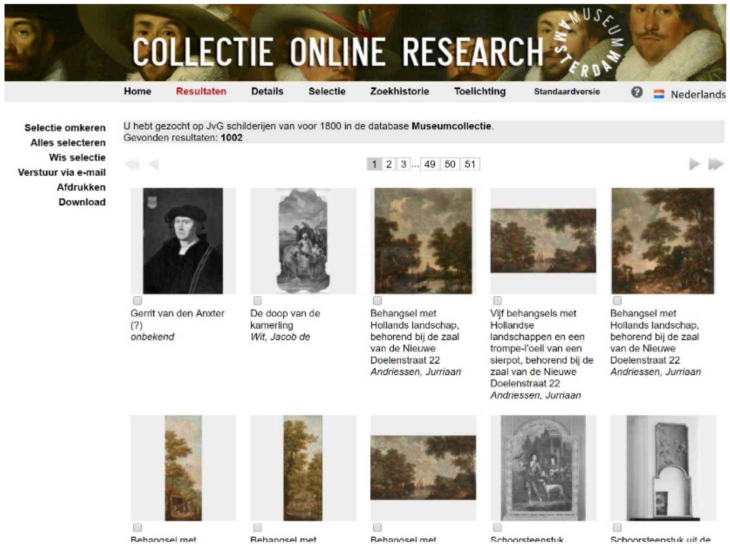

Figure 4. Amsterdam Museum Item Search (Amsterdam Museum, 2015)

The semantic web has its challenges, however, not the least of which is nay-sayers who believe it is doomed to fail because it is simply not feasible (Floridi, 2009) or those who believe it isn't moving fast enough (Anderson & Rainie, 2012). The first challenge is that information types vary tremendously, and what is good information for a machine is not necessarily good information for a human, and vice versa (Berners-Lee, Hendler, & Lassila, 2001). Quality is also a problem because an error at the source can be spread to thousands or millions of content pages (Assaf & Senart, 2012; Fürber, 2016). This is also related to issues with trust on the semantic web (Artz & Gil, 2007) and security (Lee, Upadhyaya, Rao, & Sharman, 2005). Language is an issue, as well. English speaking countries launched the semantic web but now other countries are joining it. Therefore, there is a need to address linguistic issues so that data can be accessed in multiple languages. Multiple languages present a problem with semantic web data. Systems such as *lemon* (an RDF ontology lexicon) can be used for multi-lingual projects and increase translation accuracy. It can include special metadata such as provenance (where it was translated it) and the reliability score of translations (Montiel-Ponsoda, Gracia del Río, Aguado de Cea, & Gómez-Pérez, 2011).

Data Workers

Thirty years ago, organizations relied upon analysts to crunch the data, but these roles are evolving into both more skilled data workers (data scientists) and less skilled data workers such as managers and general employees (Davenport, Barth, & Bean, 2012; Lyytinen, Grover, & Clemson University, 2017). Analysts of the past were fairly isolated. They created reports and ran queries for a number of departments and had little expertise in the business. Data scientists, on the other hand, are creatives that have deep understanding of their business domain and how to get data and work it. Data scientists also tend to be more IT savvy, and many have advanced degrees in computer science or statistics (Davenport, Barth, & Bean, 2012). In the past analysts were hard to find. Today, data scientists are even more so (Chamberlin, 2016; Davenport & Patil, 2012). At the opposite end of the spectrum are employees who are workers or managers in the core disciplines of the organization, such as Finance or Marketing or Operations. More organizations are finding value in data-based decision making, from genetics research to A/B testing to email response metrics (Vassilakopoulou, Skorve, & Aanestad, 2018). If data problems such as too much data and a lack of skills hamper experienced people, the data neophytes attempting to use data for decision making are even more vulnerable to failure, and "the new critical deficiency under which managers operate today is their inability to discover new relevant information" (Lyytinen, Grover, & Clemson University, 2017).

The general lack of data skills is a matter of concern. Training data workers at all levels is a long-term objective and requires educational prerequisites in mathematics, statistics, stochastics, databases, and computer use (Aalst & Damiani, 2015). Many companies around the globe are finding that the available talent pool for data workers is too small and people often lack advanced skills (Bakhashi, et al., 2014). The implications of a data worker shortage include missed strategic opportunities for organizations and place a large constraint on data industry growth (Chamberlin, 2016; European Commission, 2017).

To summarize this section on data, we reviewed the following topics: aspects of data, meta data and provenance, viscous and liquid data, open data, semantic web, and data workers. Data seems to be simultaneously scarce and ubiquitous; it is scarce because it is so difficult to draw information out of it due to too much data, a wide range of format that inhibits search and usage, and enormous volumes. Data also often lacks context such as metadata and provenance, which can result in distrust of the data or misuse. Difficult to use data, or viscous data, has relatively little value and often has great cost to store and protect. Its counterpart, liquid data, is easily usable due to standardization, machine-readable format, and the addition of context. When data is opened, such as government open data, it can stimulate economies or provide for joint solutions to global problems. Linking data, particularly open data, provides semantic meaning and the ability to expand search on the web. Exploiting data is one of the fastest growing activities of the 21st century, where data is an asset that has become simultaneously cheap and expensive, if one can find enough skilled workers to manage it.

Data Systems

In this section I discuss the major information system types that house data: repositories, data warehousing, and data platforms.

Data Repositories

A repository provides online access to a large number of raw data sets (Janssen & Zuiderwijk, 2014). The onus is on the user to do something with the raw data. Open data is often stored in public repositories, but private and semi-private repositories also exist (Janssen & Zuiderwijk, 2014; Lakomaa & Kallberg, 2013). A number of organizations and governments are now requiring grant recipients to publish their data in open repositories in the interest of transparency and trust. The US National Science Foundation grants and the National Science Foundation of China both required data to be published in open repositories (Link, et al., 2017). Repositories by themselves are just collections of data sets, therefore, they require a portal or other means of access and search, along with tools such as decision support systems or analysis (Angst, Agarwal, Sambamurthy, & Kelley, 2010; Janssen & Zuiderwijk, 2014). EMR systems illustrate the use of repositories, where clinical data and patient records are stored (Angst, Agarwal, Sambamurthy, & Kelley, 2010). The key differentiator between repositories and other types of data systems is that they are simply a storage site. The data must be moved elsewhere to be used.

Data Warehousing

"A data warehouse is a subject-oriented, integrated, time-variant, nonvolatile collection of data" (Inmon, 1993). Data warehousing is an organized collection of structured data with text and/or numeric formats and is often based on transactional records, such as sales data. Data warehouses, like repositories, are storage sites, and data must be extracted to employ with decision support systems, analytics, or other tools in order to use it. W. H. Inmon, known as the father of the data warehouse, offers this further description of his main points:

A data warehouse is:

- subject oriented (data are organized around areas of interest, such as customers)
- integrated (data from multiple sources are combined around a common key)
- time-variant (historical data are maintained)
- nonvolatile (users do not update the data)
- collection of data in support of management decision processes. (Inmon, 2005)

Data warehousing changed how data and information was managed in organizations when it was first introduced. It lowered the cost of information, made data available that was previously hidden to decision makers, and it could provide consistent and high quality data at the point in time that it was needed (Inmon 2005; Sen, Ramamurthy, & Sinha, 2012). Sometimes data marts are confused with data warehouses, but they are different. Data marts are typically focused on one area of the business and don't have the complexity of data warehouses. While data warehouses can feed data marts, it is a lot more challenging to accomplish that the other way around (Watson, Ariyachandra, & Matyska, 2001).

Despite its long history and potential for organizational benefits, data warehousing still has its challenges and many initiatives have failed (Sen, Ramamurthy, & Sinha, 2012; Shin, 2003). Data warehousing is expensive, which is the first hurdle. The technologies cost money. People need to be able to do something with the data, as it does no one any good simply sitting in storage. Analysts need access to the data, tools to work with it, and the skills to use the tools. Lack of documentation is a problem for both data sets and for data warehouse operations. There is an absence of audit trails and no provenance is provided. Finally, trust in the data is a concern. This is particularly important when multiple source systems feed the data warehouse. Lack of trust also comes from quality issues that revolve around how recent the data is, accuracy and missing data, format issues, inconsistencies, and duplicates (Sen, Ramamurthy, & Sinha, 2012; Shin, 2003). Data warehouses go through three stages of growth: chartering, a growth period, and a mature period, and each stage has unique needs in terms of technologies, human resources, impacts, and costs. Anticipating these sequential stages can help the organization navigate the pitfalls that commonly befall data warehousing initiatives. However, there are still problems with turning the contents of data warehouses into usable data (Batra, 2017; Russom, 2016). Many organizations are concerned with modernizing their data warehouses in order to deal with historical problems and new ones, such as big data (Russom, 2016). The next evolution that many of these firms are looking for may be data platforms.

Data Platforms

The last type of data home I discuss is the data platform. Platforms in general possess a “layered architecture of digital technology” with that includes a device layer, network layer, service layer, and content layer (Parker, Van Alstyne, & Jiang, 2016; Yoo, Henfridsson, & Lyytinen, 2010). Platform providers furnish services, content, governance, and infrastructure that empower interactions between participants (Constantinides, Henfridsson, & Parker, 2018; Greenberg, 2018). Platforms include users or "complementors" that put users in touch with complementary systems, and "sponsors" that create the platform technologies (also known as platform owners). These roles can be closed or open. Platform owners (sponsors) must consider interoperability with rival platforms, licensing for additional platform providers, and widening sponsor activities (Eisenmann, Parker, & Alstyne, 2009). Users and "complementors," however, look at backward compatibility, exclusive rights, and taking over platform complements. Platforms also change over time and will move towards closedness in regard to sponsorship (platform ownership) and towards openness in regard to users (Eisenmann, et al., 2009; Parker, Van Alstyne, & Jiang, 2016).

One of the paradoxical things about platforms is that they "invert" the traditional model of the firm. Platforms make more money from improving the business processes of other firms than from producing a product (Parker, Van Alstyne, & Jiang, 2016). Platforms are also dependent upon third parties for success. For example, gaming platforms, iOS and Android would be useless without games and apps developed by third parties (Parker, Van Alstyne, & Jiang, 2016); Tiwana, 2015). One way that platforms add value is through standardization. In order to use the platform, all participants (users/complementors, sponsors/platform owners, third party developers) must use the standards put in place by the sponsor, and there is a corresponding increase in performance when participating in platforms (Greenberg, 2018). Yet platform owners need to create boundaries so there is enough openness for third party developers to succeed, while at the same time retaining enough control so that the platform owner maintains authority and coordination (Boudreau, 2017). Also, third parties will perform better when intellectual property controls are in place (Greenberg, 2018).

There are all sorts of platforms, from gaming consoles to social media. Social media is defined by platforms. Offline social networks do not depend on platforms. Social media brings platforms to the forefront, particularly because their design impacts how they are used. New features in social media, such as network visualization and content search, create different outcomes from offline social networks because other users see the network, how people use the platform, and then make their own novel uses. (Kane, Alavi, Labianca, & Borgatti, 2013). Platforms lend themselves well to sociality or social interaction. The main features of social media include digital profiles of the user, relational ties between users, search and privacy capabilities, and network transparency that reveals relational ties (Kane, Alavi, Labianca, & Borgatti, 2013). All of these features increase sociality on platforms. Platforms also change the nature of work and workers, increase speeds of exchange, and provide greater access to previously marginalized groups. Last, platforms provide location independence and automation, elements that are sometimes difficult to achieve with repositories and warehouses (Constantinides, Henfridsson, & Parker, 2018).

Most of the extant literature on platforms is concerned with large firms with many third party developers and partners, such as Microsoft, Apple, Facebook, Nintendo, Google, and the like (Ceccagnoli, Forman, Huang, & Wu, 2012; Eisenmann, Parker, & Alstyne, 2009; Parker, Van Alstyne, & Jiang, 2016; Yoo, Henfridsson, & Lyytinen, 2010). Data platforms, however, are somewhat different. Data platforms are a new generation of data storage and often make use of cloud and internet technologies. They are designed to handle enormous amounts of data cheaply, multiple data types beyond text and numbers, and have strong search capabilities (Davenport, Barth, & Bean, 2012). They use new tools such as Hadoop, JSON, and SPARQL. Even automobiles are now being turned into data platforms (Mandel, 2013). A number of platforms include tools for cleaning, importing and exporting, transforming, and even analysis. Like all data homes, data platforms have their challenges. Despite the advances in handling big data and multiple formats, many of the old problems still exist, such as trust in the data and quality.

To summarize this section, data repositories, data warehousing, and data platforms make up the various information systems that provide the backbone of data management. Repositories are virtual storage sites for data, which must be moved elsewhere to be used, while warehouses are organized collections of structured data such as transactional data. Data platforms, on the other hand, enable participant interactions between themselves and with the platform content, tools, and services. All three types are in use today. Regardless of which home data is housed in, all data is subject to the data cycle, which is discussed next.

The Data Cycle

In this section I unpack how data is created and used at the micro (people), meso (organization), and macro (global/government) levels. Data usage follows a fairly simple process model. It is created by providers and utilized by users. Sometimes it goes through an intermediary, but not necessarily, and relationships are bidirectional. This is illustrated in Figure 5. I break down the data cycle into three components: data users, data providers, and data intermediaries. Although data users, providers, and intermediaries have different roles, the same people or groups may fall into one or both of the other buckets. Data users may provide their own data, as we see in national level governments that use census data for budgetary expenditures. If that same government has an open data initiative, they may serve as an intermediary for external data users such as businesses. The same government platform may also serve as an intermediary for other providers such as local governments and public organizations that have no data platform of their own. I begin with data providers.

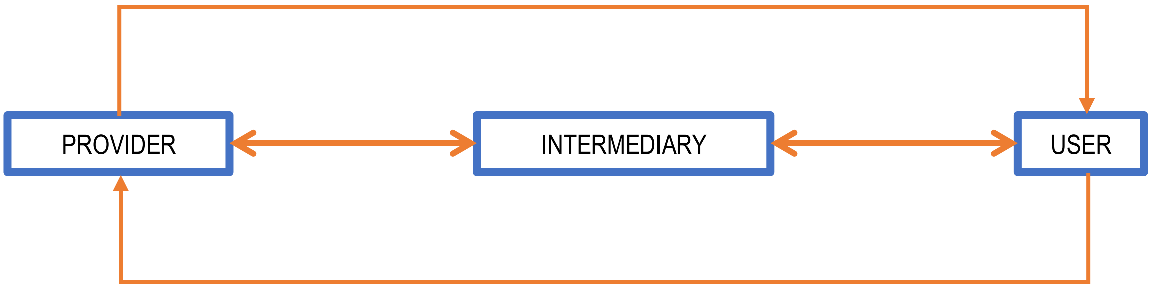

Figure 5. Data Flows

Data Providers

Data providers supply data. To understand data providers, I first separate them into three tiers: Individual, Organizational, and Government/Global. Government and global entities include governmental, inter-governmental (IGO), and non-governmental (NGO) agencies. The Organizational tier includes for-profit business and non-profit regional institutions. The last tier, Individuals, also covers a wide range and provides data for myriad reasons, sometimes deliberately and often unknowingly.

Open data is an important part of data work (Chui, et al., 2014; European Commission 2017; Manyika, et al., 2013). Open government data (OGD) or Public Sector Information (PSI) is the most common type of open data (Janssen, 2011; Janssen, et al., 2012). Although government information is often the first thing that comes to mind when discussing open data, many people tend to think of it as a recent phenomenon (Zhu, 2017). However, public access to government information is centuries old in some countries. In 1766 the Swedish Freedom of the Press Act required open government documents while France provided open access to budget information in the Declaration of the Rights of Man in 1789 (Anderson, 1908; Manninen, 2006). The US Federal Depository Library Program was created in 1895 through the Government Printing Office with the intent to "provide free, ready, and permanent public access to Federal Government information, now and for future generations" ("Federal Depository Libraries" N.D.). Seventy years later the US Freedom of Information Act of 1966 was created to provide additional public access to government records (FOIA.gov, n.d.; Shattuck 1988). It was soon followed by similar laws in Norway, the Philippines, France, Australia, Republic of Korea, Israel, Mexico, Belize, and Ukraine among many others (National Freedom of Information Coalition, 2017). This brief list simply shows how widespread the open information movement has spread. From the beginning of the concept, open data was sought not only for transparency in government but also for reuse in the private sector (Zhu, 2017). Open data and freedom of information are keys to economic success (Benkler, 2006; Janssen et al., 2012). To date, dozens of countries around the world have introduced open data legislation and many continue to do so, as exemplified by the 2016 Kenya Access to Information Act. While proponents of openness laud this progress, there are gaps in implementation and quality standards, even in countries with strong open data initiatives (Bertot, et al., 2012).

Global non-governmental organizations (NGO) are also major providers of open data. The United Nations, the World Bank, the World Trade Organization (WTO), and the International Monetary Fund (IMF) all provide a wealth of high quality open data in standardized formats.

> "The IMF publishes a range of time series data on IMF lending, exchange rates and other economic and financial indicators. Manuals, guides, and other material on statistical practices at the IMF, in member countries, and of the statistical community at large are also available." (International Monetary Fund website, February 2018)

In a more recent phenomenon enabled by the world wide web, a few organizations have opted to share data openly along with governments. Organizational data providers include companies like Google, local and regional public entities such as school districts and utility companies, and higher education. Weather Undergound (wunderground.com) offers free weather data via API, charging a fee only when API calls exceed 10 per minute or 500 per day. A sample of the available data fields includes current weather conditions, 3-day, 10-day and hourly forecasts, satellite thumbnails, radar images, alerts, and tidal information, as well as many others (wunderground.com, 2018). Weather Undergound provides the free data to encourage developers to source weather information from Weather Underground when they build applications. Another common source of open data is school districts in the US. Because of a US law commonly known as No Child Left Behind (NCLB), public schools have accountability requirements and school data such as student performance, facilities, percent of free/reduced school lunches, student demographics, and teacher experience are publicly available (US Department of Education 2017). While early implementations of NCLB presented the data in PDF and report formats, today large metropolitan school districts often post full datasets that may be easily downloaded.

The newest breed of data provider is at the individual level. Technology has enabled ordinary people to participate in data collection, recording, manipulation, analysis, and publication. Individual data providers may include researchers, hobbyists, and citizen scientists, to name a few categories. These are people who create and share data on an eclectic range of topics. The citizen scientist, for example, "is a volunteer who collects and/or processes data as part of a scientific enquiry" (Silvertown, 2009) and as a group, citizen scientists have become a valuable tool for expanding scientific research (Bonney et al., 2009). Citizen scientists have audited natural resources such as lakes and rivers, counted migrating birds, and measured their own intestinal flora, (Conrad & Hilchey, 2011; eBird.com, 2017); Gura, 2013; Sullivan et al., 2009). New tools such as eBird, a wild bird tracking phone app, make citizen science easier than ever before (eBird.com, 2018; Graham, et al., 2011). Yet, why do millions of people participate in citizen science? A wish to contribute to science may be a motivator for this group, along with a strong interest in the subject and a desire for recognition (Raddick et al., 2013; Rotman et al., 2012). However, the motivations of citizen scientists may vary from other individual data providers, such as the dedicated reader who charts every book he ever read, as seen in Andy Sernovitz' 1526 row book list dataset or in @History's 39-file dataset on the history of Scottish witches (data.world, 2017). People have long enjoyed collecting as a hobby (McIntosh & Schmeichel, 2004) and digital data collections may simply be the newest manifestation of this activity.

Outside of citizens scientists, there is minimal research on individuals who share data, however, we may find correlations in the knowledge sharing literature. Knowledge sharers often offer their wisdom without expectation of reciprocity and do so for enjoyment and increased social capital (Wasko & Faraj, 2005). It is logical that individual data providers may collect and open their data for similar reasons. Knowledge sharing motivation varies across cultures, however. In China, for example, tight social networks known as *guanxi* intertwine mutual benefit and reciprocity and have a significant impact on knowledge sharing (Huang, et al., 2011). Cultural influences on knowledge sharing are not only geographical but also organizational. Company culture impacts an individual's likelihood of sharing knowledge (Alavi, et al., 2005). The knowledge sharing literature may provide a foundation for understanding why individuals provide data, which in turn may aid in encouraging, guiding, and managing individual data warriors. From individuals to global institutions, data providers cover a broad scope, but dividing providers into tiers aids our understanding. Table 2 provides a summary of data providers that summarizes this group and its three tiers.

Table 2. Taxonomy of Data Providers

Provider Tier	Entity Examples	Data Examples*
Global/ National/ Governmental	Government open data portals (UK, Brazil, US, Russia, Israel, Singapore, Ghana, etc.)	Economic, agricultural, manufacturing, budgets and expenditures, census and demographic, tax information
Organizational	Google, Zillow.com, Smithsonian American Art Museum, universities, K-12 school districts, survey/marketing firms, libraries	Search term trends, educational performance, real estate, book collections, art collections
Individual	Citizen scientists, hobbyists, researchers, students	Bird counts, police brutality incidents, personal favorites lists, recipes, wine and beer

* Note that data examples are representative and not all entities offer all data forms listed

Data Users

Like data providers, I also divide data users into three tiers: governmental/global, organizational, and individual (Safarov, et al., 2017). Governments and NGOs use data for policy, regulation, initiatives, and legislation. These entities are often the primary users of the data they produce and may combine it with data from other sources (Dawes, et al., 2016; Oriko, 2017). Cross-border data flows have increased over 45 times since 2006 as a demonstration of this sharing (Luxton, 2016; Manyika, et al., 2016) A number of world problems may be studied by macro-view global and governmental entities in such varied topics as health, hunger, famine, climate, refugees, agriculture, and economics (Bertot, et al., 2014). Global organizations use data to respond to disasters, mitigate pandemics, and provide aid. To achieve operational success, however, trade and economic growth depend mightily on cross-border data streams (Mandel, 2014). At the government level, demographic and census data may be used for resource allotment or defining voting districts, while industrial and employment data may influence policy and regulation (Dawes, et al., 2016).

Another large group of data users is organizational, and governments and global entities recognize and encourage this because it fosters economic activity. The value provided to organizations by open data is just beginning to be evidenced in empirical research that looks at business models and methods ((Magalhaes, et al., 2014; Jetzek, et al., 2013). There are many opportunities for profit-seeking and nonprofit organizations to employ data as it becomes more ubiquitous. Nonprofit organizations include entities such as the Sunlight Foundation, which accelerates government transparency through data (Suhaka & Tauberer, 2012; Sunlight Foundation, 2017). Other non-profits include scientific and research enterprises, for example SRI and RAND. These organizations use a wide range of externally acquired data, such as weather, population, or geographic data to complement their own data production (Marsh, et al., 2006). Many types of educational institutions use data, as well (Marsh, et al., 2006; Suhaka & Tauberer, 2012). Universities use data operationally for strategic planning and assessment, as well as for research (Banta, et al., 2009), while primary and secondary schools may depend on data to identify characteristics and patterns in their communities (Feldman & Tung, 2001).

Profit-seeking organizational data users employ data operationally and strategically, as do nonprofits, or may use it for production and marketing activities. Drawing from the open data literature in particular, scholars and practitioners have identified a number of different business models that are applicable for profit-seeking organizational data users (Deloitte UK, 2012; Ferro & Osella, 2013; Suhaka & Tauberer, 2012; Zeleti, et al., 2016).

While some of the extant research articles identify dozens of business models, others condense the landscape to three: data enablers, integrators, and facilitators (Magalhaes, et al., 2014). Based on the literature, which is generally focused on re-use of OGD, I suggest that in data work beyond OGD, organizational users utilize data resources in four ways: a) data is used strategically to create competitive advantage; b) data is used operationally to reduce cost; c) data is treated as raw material that is transformed into a product; or d) data is used as a complementary or promotional product alongside other offerings for sale.

Strategic data users employ data for planning, forecasting, research and development that bolster the firm's main business. (Deloitte UK, 2012; Zeleti, et al., 2016). In contrast, *operational* data users utilize data for cost savings and increasing efficiency (Zeleti, et al., 2016). One method of cost savings uses openness to increase data quality and dissemination (Zeleti, et al., 2016), while adopting an open source policy can allow firms to leverage community members to expand and improve the data for everyone's advantage (Ferro & Osella, 2013; Howard, 2014; Spencer, 2009; Zeleti et al., 2014).

A number of data users create new products with the data they consume. The first type is known as *Premium*, where acquired data is cleaned up and put in an easy-to-search and ready-to-use format, such as Lexis-Nexis (Ferro & Osella, 2013; Howard, 2014; Suhaka & Tauberer, 2012; Zeleti, et al., 2016). Sometimes there is no value added to the data product, as is found in *White-Label Development* where data from source A is simply purchased and relabeled as a new product from seller B (Ferro & Osella, 2013; Howard, 2014). Last, the *Startup* model adds considerable value to acquired data through manipulation and aggregation, creating a brand new product for their market (Suhaka & Tauberer, 2012).

The last group uses data for promotions and sales of other products. In the *Freemium* model, limited data is free while expanding quantities, services, or other features incurs a cost (Ferro & Osella, 2013; Suhaka & Tauberer, 2012; Zeleti, et al., 2016). Following a tried and true strategy, *Infrastructural Razor & Blades* or *Parts of Tools* offers free data but a cost is incurred in order to use it, similar to offering free razors and charging for blades or giving away a mobile phone in order to charge for cell service (Ferro & Osella, 2013; Howard, 2014; Zeleti, et al., 2016). The final promotional business model is labeled *Free as Branded Advertising*. In this model, data is offered as a free product for organizational promotion, as might be seen with Zillow.com's real estate trends datasets (Ferro & Osella, 2013; Howard, 2014; Suhaka & Tauberer, 2012). Table 3 summarizes the literature on profit-seeking organizational data users and how they use data in their business models.

Looking at the four types of data user business models, we can further abstract their focus into just two foci: revenue and expense. The revenue focused business models include groups a (strategic) and d (complementary/promotional), while expense focused models contain b (operational) and c (raw material). Revenue focus looks towards building the business, expanding reach, increasing sales, and dominating the competition. The expense view, on the other hand, prioritizes decreasing costs, improving quality, and increasing efficiencies. Figure 6 illustrates these relationships. It is important to understand how data is used to gain a better perspective of the data.world platform and its value proposition.

Table 3. How Organizational Data Users Employ External Data

Type	Model	Description	Author
	(Legend: a=strategic, b=operational, c=raw material, d=complementary/promotional)		
a	Supports core business/Indirect Benefit to business	Bolsters the business' main competencies and guides planning	Deloitte UK, 2012; Zeleti, et al., 2016
b	Cost Saving	Uses openness to increase data quality and dissemination	Zeleti, et al., 2016
b	Open Source	Similar to open source software, data is offered free for reuse & distribution. Like OSS, Open data is shared with the community who then improve and build upon it for the benefit of everyone	Ferro & Osella, 2013; Howard, 2014; Spencer, 2009; Zeleti et al., 2014
c	Premium	High quality, ready-to-use data is offered for a price	Ferro & Osella, 2013; Howard, 2014; Suhaka & Tauberer, 2012; Zeleti, et al., 2016
c	White-Label Development	Data from source A is purchased and relabeled as a new product from seller B	Ferro & Osella, 2013; Howard, 2014
c	Startup	Use, combine, and manipulate data to create a new service or product	Suhaka & Tauberer, 2012
d	Freemium	Limited data is free while expanding quantities, services, or other features incurs a cost	Ferro & Osella, 2013; Suhaka & Tauberer, 2012; Zeleti, et al., 2016
d	Infrastructural Razor & Blades; Parts of Tools	Data is free and subsequent tools to use it incurs cost, similar to offering free razors and charging for the blades required to use it	Ferro & Osella, 2013; Howard, 2014; Zeleti, et al., 2016
d	Free as Branded Advertising	Data is offered as a free product for organizational promotion	Ferro & Osella, 2013; Howard, 2014; Suhaka & Tauberer, 2012

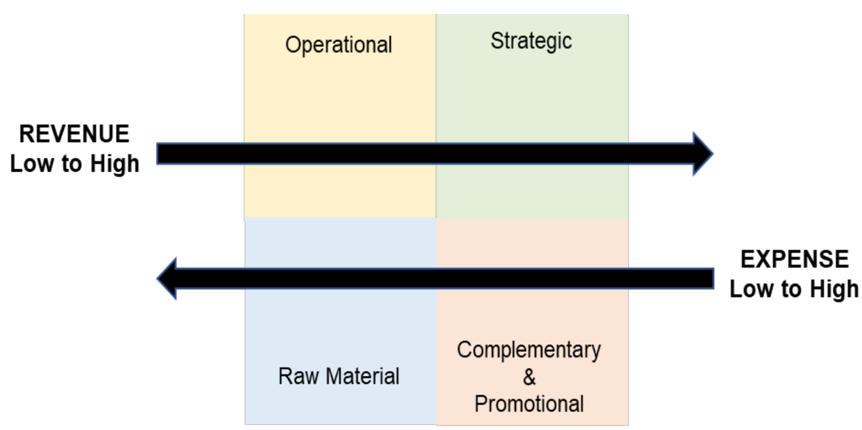

Figure 6. Data User Business Model Framework

While organizations turn data into profits and progress, individuals have also jumped on the data bandwagon, although this remains one of the least studied user groups compared to organizations (Safarov, et al., 2017). The new millennium has seen the rise of data journalists and data hobbyists, along with more traditional researchers and students. Individual data users must overcome the hurdle of acquiring data skills before data can be put to use and this requirement limits individual participation (Magalhaes, et al., 2014; Safarov, et al., 2017). I include background on data journalists because they were an important part of the data.world community.

Data journalists are a rising subgenre in the press corps (Coddington, 2015; Gray, et al., 2012). Data journalism is different from traditional journalism in that it is a combination of using "a nose for news" with the ability to use data to "tell a compelling story" (Gray, et al., 2012). In data journalism, data can be both a tool and a source. Data journalism is important because technology has enabled news to travel at internet speed and be generated by multiple sources (Howard, 2014). News is immediate and is no longer filtered, curated, and disseminated by agencies. However, data can successfully be used to sort out facts, search out deeper analysis, and link phenomenon to provide more meaningful news stories (Coddington, 2015).

Data journalists use a range of tools that they employ at different stages. These activities may be characterized as computer-assisted reporting, data journalism, or computational journalism (Coddington, 2015). Tasks may include investigative information gathering, analyzing connections, and spotting trends. Infographics are also important in data journalism to convey a story through illustration (Gray, et al., 2012). The Guardian's Datablog (https://www.theguardian.com/data) offers excellent examples of data journalism with its multitude of topics and eye-grabbing visualizations that tell a story with few words and great impact. Data journalism doesn't replace traditional journalism, it enhances it, as it can provide news stories that are unavailable from other sources (Coddington, 2015; Gray, et al., 2012; The Guardian 2018). For the data journalist, building stories from data offers "less guessing, less looking for quotes" (Gray, et al., 2012). Data based news stories provide fact-based articles that are more likely to be perceived as accurate. A 2017 poll sponsored by data.world found that 78% of US adults would have more trust in the news if they had access to the data underlying news stories (Globe Newswire, 2017).

However, just because information comes in data form, it still must be vetted. Journalists cannot rely on veracity and accuracy simply because the information comes in data format. "Like any source, it should be treated with skepticism; and like any tool, we should be conscious of how it can shape and restrict the stories that are created with it" (Paul Bradshaw, Birmingham City University, 2017 via databox.com 2017). Not all journalists can attempt data journalism, either. Gaining the expertise required for data journalism can be a challenge, such as learning data science skills, analysis, and visualizations (Howard, 2014).

Learning data skills is a common issue for individual data users, whether they use data professionally, such as data journalists, or for pleasure, such as data hobbyists, researchers and students (Anderson & Rainie, 2012). Using data for hobbies is not new, but it has become more ubiquitous. Horse racing, for example, has a long history of using data for predicting the outcome of races (Snyder, 1978). Today hobbyists use data to in video gaming (Darzentas, et al., 2015), sports predictions and analysis (Haghighat, et al., 2013; McGarry, 2009), and genealogy (Cannon-Albright et al., 2013). Health enthusiasts commonly use self-generated data from wearable technology, such as Fitbit, to track and improve health outcomes (Giddens, et al., 2016; Li, et al., 2011). We also see financial data used by individuals. Banks often provide access to online data tools that illustrate how a customer spends their money or how they can better manage their debt and savings (Thakur, 2014). In the same vein, car enthusiasts are beginning to use the data from digitally enabled vehicles to enhance their hobby (Markovich, 2017). As data collection increases, it is likely that individuals and hobbyists will find new ways to use data in a wide range of activities. However, the state of the data, how viscous or liquid it is, and how available it is, will determine how well individuals will be able to utilize it. I suggest that individuals, as opposed to organizations, will find it more challenging to learn the skills and tools required to exploit data and their success is more dependent upon intermediaries that facilitate data usage. Data users are summarized in Table 4.

Data Intermediaries

Intermediaries add value by making data easier to use. Raw data from providers is often viscous, dirty, and presented in non-machine-readable formats. Viscous data by its nature is difficult to use, hard to search, and inhibits broad utilization by its user-hostile nature, but intermediaries can decrease data viscosity and increase liquidity (Lämmerhirt, et al., 2017; Mallah, 2018; van Schalkwyk, et al., 2016). Increasing liquidity improves accessibility to the data, how it can be used, and who can use it. With data science skills at a premium, anything that reduces skill requirements is likely to increase usage.

Table 4. Data Users

Level	Who-General	Types of Data	Who-Specific	Examples	Outcomes
Global/ National/ Governmental	Governments, IGOs	Tax entities, government agencies	US Federal Reserve Bank, World Bank, CDC	Property values, consumer price index (to calculate inflation rate), census	Provides basis for tax or discount rates according to changes, redrawing district lines for representatives
Organizational	For-profit Organizations	Product development, marketing, sales, sometimes the data is the product	Home Depot	Demographics, consumer preferences, efficacy, health trends, weather trends Home Depot runs a weather center during hurricane season that uses data to track hurricanes and deploy supplies to stores in affected areas before and after a weather event	Helps develop new markets, increase revenues, commercialization, strategic planning for competitive advantage, cost reduction
	Non-profit Organizations	Services development, grants, needs assessment, resource allocation, NGO	Red Cross	Demographics, poverty, education, health, environment, crime, hunger, slavery, homelessness Open data was used extensively to provide services for victims of Hurricane Harvey in Houston 2017	Serves constituents, target funding, strategic planning to achieve a mission
Individual	Researchers	Dedicated labs, higher education	Published authors in a large number of fields	Scientific data, health data, populations, poverty rates	Aids new research, replications, extended research, new health solutions, engineering solutions, weather/geo predictions and tracking
	Journalists	Data Journalism: Providing news through previously unavailable data and compelling infographics, also known as open journalism	The Economist, The Wall Street Journal, BBC News Visual Journalism	Varied topics	Provides new way to present news, previously unseen information brought to light through data tools and automation

Level	Who-General	Types of Data	Who-Specific	Examples	Outcomes
	Individuals	Students, homebuyers/renters, entrepreneurs, armchair auditors	Individuals	Ancestry.com (personal history), Fitbit (personal health data), MyFitnessPal (personal and open nutrition data), Zillow.com (open real estate data), amatueur Armchair auditors use open data to audit government contracts and expenditures O'Leary 2015	Provide a way to seek information for personal benefit (social, economic, psychological, health)

Intermediaries increase data liquidity in several ways, such as manipulation, aggregation, storage, and ease of movement. Manipulation includes data cleaning, altering data into standardized machine-readable formats, and adding meta-data. Musgrove (2003) suggests that meta-data is a broad term for any information about the data that improves analysis, presentation, explanation, and usage. Aggregation combines data, often from varied sources, into clean usable datasets. Storage is another service provided by repositories. There are thousands of data repositories available, many pre-dating the internet age and focused around government entities or narrow interest groups, such as certain scientific fields (Marcial & Hemminger, 2010; Nature, 2017). However, new technologies have opened the door to a wider range of repositories that offer greater access and more added value such as tools and collaboration. One of the most important benefits that intermediaries provide is moving data – getting it in and out. The most feature-rich intermediaries provide integrations and APIs to easily upload and download datasets, some of which can be enormous. Integrations can be critical in data dissemination.

As I have done with data users and providers, I divide data intermediaries into the three tiers of global/government, organizational, and individual. Global and government entities are somewhat unique in that they may simultaneously create data, use data, and serve as an intermediary. Government open data portals and IGOs (inter-governmental organization) exemplify this segment. Some examples include the World Bank and International Monetary Fund (IMF), data.gov.in (India), data.gov.uk (UK), and opendata.go.ke (Kenya). In Europe, an early proponent of open data, 27 countries provide data portals with only four not participating (European Data Portal, 2017). Although government and IGO portal growth is expanding globally, on many continents it is the exception not the rule, and where portals do exist but may be sparsely populated (World Wide Web Foundation, 2015). In contrast, many NGOs have not joined the open data movement although it offers benefits for these organizations (Fernando, 2017; Oriko, 2017).

Organizations are common intermediaries, in both profit-seeking and non-profit guises. Repositories and platforms make up a large portion of this group. As described earlier, repositories serve as virtual storage facilities, but some have evolved through digitization into collaborative and social platforms that encourage data sharing, co-authoring, and discovering like-minded people and groups. The best known non-profit platform is Wikimedia, which includes much more than Wikipedia. With its other properties (Wikimedia Commons, Wiktionary, Wikibooks, and others) it provides open data, media files, and other data assets (Wikimedia, 2017).

Among profit-seeking organizations, GitHub, data.world, and Kaggle exemplify platform based intermediaries. These organizations not only store datasets (and in the case of GitHub, code), but offer community interaction, as well. Tools may be provided to aid in analysis, such as auto-created data schemas or visualizations, and APIs and integrations ease the process of getting data in and out. data.world has a particular focus on the semantic web and its data linking is capabilities are unique and valuable. The world wide web is still in its infancy and is often a "medium of documents for people rather than of information that can be manipulated automatically" (Berners-Lee, et al., 2001; Joo, 2011). Beyond serving as platform intermediaries, organizations offer a range of commercial data services that find, sell, trade, clean, match, aggregate, separate, digitize, and transcribe data assets. Like repositories, data services are not new, but do offer new features to help customers manage data, such as cleaning.

Yet intermediaries do not have to be large organizations to make an impact. Technology has enabled individuals to serve as intermediaries and add significant value to data. An exemplar class in this area is data activism. Falling within the realm of digital activism, data activists provide data services for the greater good (George and Leidner, 2018; Sieber & Johnson, 2015). Example tasks they undertake include saving and archiving at-risk open data, cleaning, tagging, and adding meta-data. Data activists work alone or at organized civic hackathons with groups such as Data for Democracy and DataRescue. Data activism has increased over the past few years and is making progress at local, regional, and national levels (Data for Democracy, 2017; Data Rescue National Neighborhood Indicators Partnership, 2017; Harmon, 2017; Sunlight Foundation, 2017). Data intermediaries are summarized in Table 5.

To summarize, the extant literature tells us that data intermediaries may be governmental/global, organizations, or individuals, or a combination of two or all three categories. Their goal is to make data more usable through labor and/or applied technology. Data intermediary tasks may include data cleaning and manipulation, aggregation, moving data, and storage. These activities result in greater data, such as data rescued from redaction or removal, data democratization, data moved from inaccessible sites to easily accessible sites, dirty data transformed into clean data sets, and the creation of new data from combinations of individual data sets. On the other hand, data intermediaries vary considerably. We have little understanding of which types of intermediaries are most effective and which methods yield the greatest impact, nor how the different types interact with each other. We also don't fully understand the potential negative aspects of data intermediation. While there appears to be some evidence that intermediaries make data more usable, we have little in the way of theory to explain how this occurs and under which circumstances. To address this lack of understanding, I ask the following research question: How does intermediating through digital platforms impact data?

In summary, the three parts of the data cycle are data users, data providers, and data intermediaries. The data cycle is a process model whereby data is created by providers and employed by users. It may or may not go through an intermediary and all relationships are bidirectional. In this next section, I wrap up the literature by reviewing the sensitizing device I used in my research.

Summary of the Literature

To provide background for better understanding my research with data.world, I reviewed the literature for data, data systems, the data cycle, and a sensitizing device. at the end of each section, I summarized the literature into what we know and what we don't know. Putting it all together, we know that data has great potential and more and more people and organizations are becoming wise to new data opportunities and value. We learned that opening data has great value in a number of different scenarios ranging from small business to finding solutions for global warming. We also know that platforms provide an ecosystem that empowers interactions between participants (Constantinides, Henfridsson, & Parker, 2018; Greenberg, 2018), and online communities in particular benefit from platforms.

Table 5. Data Intermediaries

Level	Description	Activities	Who-Specific	Examples	Outcomes
Global/ National/ Governmental	eGov Portals	Promote open government data	city, state, country level	Taiwan, Australia and the UK, 1st and 2nd ranked (tied) countries for openness (Global Open Data Index)	Transparency, economic advantage, reduction in unethical political behavior
	Policy Makers	Develop policy and promote legislation	Federal Trade Commission, European Commission	EU Data Protection Directive of 2012, FTC data broker protection recommendations to Congress	Personal data privacy laws in EU tightened, FTC brought charges against data brokers who sold personal data to illicit buyers
Organizational	Services	Tools, visualizers, code libraries	data.world, Google Data Studio, Google Data Explorer, Code for Science, digital-activism.org, The R Project/CRAN	data.world data visualizer and APIs, Google Data Studio, Google Data Explorer, R and R Studio, Apache Hadoop, Python, SPARKL	Help people use the data because data alone is not very useful unless it can be applied
	Repositories	Free and paid repositories, specialist and generalist, data repository certification	data.world, Re3Data.org, Kaggle, figshare, Dryad Digital Repository, Harvard Dataverse, AWS Public Datasets (Amazon)	Generalist repositories: data.world, Dryad Digital Repository, Harvard Dataverse, figshare Specialist repositories: DNA DataBank of Japan (DDBJ), Integrated Taxonomic Information System (ITIS), ClinicalTrials.gov Certifiers of data repositories: World Data Systems or Data Seal of Approval	Provide a place to store/retrieve the data, those with social aspects (data.world) include collaboration and community aspects
	Data Brokers	Anonomously gather and sell personal data	Acxiom, Corelogic, Datalogix, eBureau, ID Analytics, Intelius, PeekYou, Rapleaf, and Recorded Future	These organizations have sensitive identifying data, such as name, birthdate, and social security number, census data, internet browsing trends, health, social media data, financial data, and information on personal habits.	While most purchasers use the data for fraud detection or other non-criminal activity, data brokers have been found to sell data to illicit buyers.

Level	Description	Activities	Who-Specific	Examples	Outcomes
Individual	Data Activists	Volunteer hacktivists	Data for Democracy, Data Rescue, Data Refuge	Open Data Day/March 2017 - data hackers around the US worked open data research, tracking money flows, open environmental data, and open human rights data.	Social benevolence, increases in open data, publicizes open data, provides additional data for data users
	Education	Free and paid training for data sciences, expression/script suggestions, tutorials, wizards	MOOCs (Coursera, EdS, Udacity, MIT, UT Austin Extension, Harvard University), bootcamps, colleges, IBM, KD Nuggets	Individual courses, bootcamps (Thinkful, NYC Data Science Academy), 16 week courses and 30 week courses, certification (Harvard Extension)	Provides training for people to learn and use data, covers a variety of beginning skill levels and levels of advancement
	Other organizations	Non-profits, NGOs	Open Data Institute, Open Data Watch, Global Open Data Index	Provide aid to data providers, monitor the state of open data, particularly in regards to nation/states	Social benevolence, increases in open data, publicizes open data, provides additional data for data users

Communities of practice provide apprenticeships, mentoring, and guidance within a craft, which allows participants to improve their skills at a quicker pace and with greater quality and have great promise for data communities. Last, we know that data is created by providers, may go through an intermediary, and is ultimately employed by data users. These three functions may be performed by separate entities or one may perform all the functions. The functions exist at individual, organizational, and global/governmental levels.

The literature review has provided a background in the key elements impacting the data.world case, including understanding data and data work and its varied forms and challenges, data systems, and the data cycle. I identified several areas in the literature that still need answers. Despite progress from data warehousing to big data to analytics to visualization, data work continues to experience 1) a lack of workers skilled in using data; 2) limited access to data by those who could use it; 3) difficulty moving data in or out of systems; and 4) inability to leverage data to advantage because of little or no context in the data. Further research could focus on any of these areas to provide greater insight into finding solutions for researchers and practitioners.

There is much we don't know, such as how to break down data silos, get data to the right people, or how to balance data security and privacy with openness. After decades we still know little about how to incent workers to learn analytical skills or make data usable to the vast majority of people. We don't know how to get data to the people who need it, and indeed, we often don't even know who needs it or what they could do with it. Therefore, I proceed with the following research question: *How does intermediating through digital platforms impact data?*

Sensitizing Device

In grounded theory, theory is built from the data. However, one does not go into grounded theory with a *tabula rasa* or blank slate (Dey, 1999; Glaser & Holton, 2004; Glaser and Strauss, 1967; Urquhart, 2013). While grounded theory does not use a priori hypotheses, sensitizing devices may be used to provide perspective and a lens to view the research (Gregor, 2006; Klein & Myers, 1999). For my sensitizing device, I used the literature from communities of practice. I selected this concept because data work, particularly on platforms like data.world, has an element of sociality that is kin to communities of practice.

A community of practice is comprised of a domain, a community, and a practice and the concept is based on the premise that learning is based on social participation (Lave & Wenger, 1991; Wenger, 1999). New participants join a community of practice for the purpose of learning, and they can learn through just being part of the community, which Lave and Wenger call a "legitimate peripheral participant." New participants learn through different modes of action and understanding the different meanings associated with each mode. People learn not just from a master, such as a master-apprentice situation, but from the community surrounding the learning experience because it provides context. Learning is a social activity (Barton & Tusting, 2005; Hildreth & Kimble, 2004; Lave & Wenger, 1991).

Communities of practice are comprised of diverse participants with different goals, skills, and levels of participation.   Besides teaching skills, communities of practice impart relationships between entities in the practice, power, ethics, and the practice's place in the world. Members learn what it means to be in the community - what jokes are appropriate, how to speak the community language. People build their identities through interaction with their communities of practice, and learning is a goal for new members that seek to emulate more advanced members (Brown & Duguid, 2000; Lave & Wenger, 1991). As mentioned earlier, a community of practice must have three things: a domain of shared interest, a community of members that engage, interact, and build relationships, and practice, which is the actual practice of the topic, not just shared interests. Members "develop a shared repertoire of resources: experiences, stories, tools, ways of addressing recurring problems—in short, a shared practice. This takes time and sustained interaction" (Wenger, 1999). Communities of practice have many activities, including problem solving, asking for assistance, seeking experience, reusing assets, coordination and synergy, discussing developments, documentation projects, visits, and mapping knowledge/identifying gaps. Communities of practice may be large or small, online or in-person, local or global. They are not new but have been used by humans for centuries. Those who use communities of practice include organizations, government, education, associations, non-profits, international development, and software development (Wenger, 1999; Wenger 2010). Communities of practice offer ways to promote innovation, develop social capital, spread knowledge within the community and extend tacit knowledge. A community of practice may come about naturally because of the members' common interest in a particular domain or area or it may be created specifically to build knowledge in a specific topic. The communication of the community includes sharing of information and personal experiences that participants learn from other members. This helps them grow their skills and personal development. Hildreth & Kimble (2004) provide the following description of communities of practice: A domain frames the community and is the common topic of interest, while the practice includes those practices and activities shared by the group. Participants utilize a repertoire of shared resources and tools for the community to use. Participants in communities of practice have common ground and often share backgrounds and similar needs and requirements, as well as a common purpose. Communities also evolve over time, as situations, people, relationships, topics, and needs change. Relationships between participants are an important part of the community and storytelling is another way for members to share their experiences while also building social capital and legitimacy. Last, a community may be formal or informal, and may not even realize they are a community of practice. The idea of formality vs. informality is controversial as some scholars believe that individual legitimacy comes from formal recognition in the group while Lave and Wenger see legitimacy as being informally through community consensus (Hildreth & Kimble, 2004).

Yet communities of practice are not a panacea. There are limits to their usefulness and situations where they are not effective (Roberts, 2006). Communities of practices may be used in a wide range of settings, but they may also be co-opted for an inappropriate setting (Roberts, 2006). Things that impact the effectiveness of communities of practice include power in the community - who has it and how it is used, trust, predispositions such as members learning towards particular directions, size and spatial reach, and how quickly change occurs in the community.

The case of data.world relied greatly on the social nature of its platform community, the interaction between members, the social media features of following and liking, and frequent asking and giving of advice for solving data problems. Applying the concept of communities of practice to the context of data.world provides a way to explain how and why platform sociality impacts data and data work. I suggest that it is important to understand the relationship between sociality and data work outcomes in order to expand our comprehension of the processes that convert viscous data into liquid data. In the next section I discuss my research methods and process.

CHAPTER THREE

Method

My research strategy focuses on a single case study of a data intermediary, data.world, because its unique position as a data platform provides perspectives from three areas: data providers, intermediaries, and data users. IS literature has a long tradition of employing single case studies to understand complex phenomena (Dubé & Paré, 2003; Levina & Ross, 2003; Seidel, et al., 2013). Single case studies are also especially useful "when a phenomenon is broad and complex, when a holistic, in-depth investigation is needed, and when a phenomenon cannot be studied outside the context in which it occurs" (Dubé and Paré 2003). Yin (2003) advocates that case studies are beneficial when researching "how," "when," and "why" questions, and suggests that cases are also useful for "contemporary phenomenon within some real-life context" (Yin, 2003).

To conduct this case study, I employed established methods from qualitative researchers, including Myers, 1997, Sarker & Lee, 2002; Schultze & Avital, 2011, Trauth, 2000; and Walsham, 2006. This included semi-structured interviews, personal notes and observations, electronic communications, and attending meetings, workshops, and conferences. This study uses grounded theory methodology (GTM) for analysis and to build theory based on the phenomenon. I draw from Birks, et al., 2012; Glaser & Strauss, 1967; Urquhart, 2013; and Urquhart & Fernández, 2013. Grounded theory is an iterative process that uses the data itself to develop theory (Glaser and Strauss 1967; Myers 1997). GTM provides "continuous interplay between data collection and analysis" (Urquhart, et al., 2010). GTM has a distinct split between the Glaserian and Straussian strands. While Corbin and Strauss use four levels of coding (open, axial, selective, and theoretical), Glaser and Strauss use only three, omitting the often difficult and controversial axial coding (Urquhart, 2013). In this paper I use Glaserian grounded theory method.

As per GTM practice, I did not start with a theory in a positivist fashion, but reviewed literature before, during, and after data collection to inform my thoughts about what I was seeing in the field (Glaser and Strauss, 1965; Urquhart, 2013). The literature review was built slowly during this time, as new topics of relevance became apparent. I observed the subject case and its participants, reflected upon my observations, and then revised my interview questions and approaches in an iterative fashion. I started with questions informed by a broad research thrust (how do intermediaries add value) and the literature. Because data exploitation on such a grand scale is a relatively new topic, I availed myself of a wide range of sources to understand it, including journal and conference papers, blogs, books, and other media for a grand total of 344 reference sources. The source types and quantities of literature are summarized in Table 6. Initial broad questions revolved around the role of each participant, where they saw themselves in the company, their background, how they described the operational aspects of the platform, their view of the value of the platform, and the role of the company in the marketplace. I found themes emerging around linked semantic data, open data, and data socialization where the platform brought people together. In later months my questions altered to reflect changes in the company. My questions now had more external focus and revolved around the value of the platform, its transformational nature for those using the platform, and how it was different than data warehouses. The themes evolved from socialization to the ability of collaboration to provide context and ultimately transform viscous data into liquid data. I also made notes during and after my visits to document my observations about the office environment, culture, company meetings, and growth. In addition to employee interviews, I was able to interview several users of the platform. These individuals were data activists or data journalists. These interviews typically lasted 30 minutes, although a few were as long as an hour. Last, I was given access to 77 anonymous transcribed user interviews that the company had conducted for early stage product development. These users ranged from academics and students to data scientists and mangers. I did not talk to these users directly.

Table 6. Literature Resource Types

Type	Count	Type	Count
Blog	56	Forum post	1
Book or book section	36	Journal article	188
Computer program or dataset	1	Magazine article	5
Conference paper	21	Newspaper article	5
Dictionary/encyclopedia entry	5	Presentation	1
Report	25		
Total		344	

My process was theoretical sampling, where I used an iterative method to constantly compare data, analyze it, and look for new information based on my interviews and other sources over time (Glaser & Holton, 2004; Glaser & Strauss, 1967). Interviewees made suggestions about people to talk to and reports and books to read, even giving me books on the semantic web to add to my personal library. They invited me to public workshops, conferences, and hackathons where I was able to learn more and connect with more people. I also made memos during my time at the company and during my analysis. The various sources of data (interviews conducted by me and by company employees, archival data, press, and social media) allowed me to triangulate the data and provided greater construct validity while helping me to converge my inquiry streams (Yin, 2003).

The iterative process produced some rabbit trails, but eventually led to an emerging theory about the role of data platform intermediaries. Some of the rabbit trails included benefit corporations, data journalism, data philanthropy, data activism, high tech entrepreneurship, and open data. Several of these topics extended into conference and journal papers and all added to the richness of the data.world case. However, in the interest of creating parsimonious and relevant theory for this dissertation, I reigned in the scope after a year with the company and focused on the research question of how digital data platform intermediaries transform viscous data into liquid data. Once I was able to hone in on this research question I was able to reach theoretical saturation in my data collection in 2018 (Glaser and Strauss, 1967).

Site Selection, Data Collection, and Description

This research is based on my experiences with data.world, an Austin, Texas start up that describes itself as “the social network for data people” (data.world, 2017). My choice of data.world (data.world) for a case study was serendipitous. I was acquainted with several of the co-founders and employees from my own previous employment at other Austin firms. Upon learning I was pursuing a doctorate in IS and had an interest in studying the new company, I was invited by the Chief Product Officer to the data.world office to “hang out, use our wifi, and drink our coffee.” I visited data.world every other Friday between August 2016 and April 2017, with additional visits during late winter/spring of 2018 and email communications between May 2017 and January 2018. Data that I was given access to originated in August 2015. I attended meetings, interviewed employees, met users, socialized at a few happy hours, and was generally able to become a proverbial fly on the wall. The Chief Product Officer at one point described me as observing data.world "like Gorillas in the Mist," a 1988 film about Dian Fossey's observations of gorillas in the wild. While some researchers may decide a priori upon a specific type of organization on which to base a case study that best answers their research question (Levina & Ross, 2003), I did not develop the research question for this dissertation until I had been working with data.world for a year.

My data collection efforts span August 2016 through February 2018, and some of the data (specifically Slack transcripts) began in late 2015, providing for a 2+ year longitudinal study with multiple inputs. Through data.world, I was introduced to other groups, such as Data for Democracy, a data activist group, and Datajournos, a group dedicated to data journalism. My data was collected onsite at data.world, at workshops, at conferences, at offsite meetings, and remotely through web conferencing, email, Slack, and on the data.world platform. Dubé & Paré (2003) suggest that case study research should "include better documentation particularly regarding issues related to the data collection and analysis processes." To reflect this guidance, I provide a detailed list of source materials and analysis tools in Table 7 and interview details in Table 8.

Table 7. Data Sources

Data Sources	Data Features	Use in Analysis
Primary data source Interviews	30 interviews from data.world employees 2 interviews with Data for Democracy members 2 interviews with data journalists 77 data.world Alpha user interviews 2 data.world Beta user interviews	Tool: NVivo 11 qualitative analysis These interviews provided the foundations for the research through the thoughts, feelings, motivations, emotions, and efforts of those interviewed.
Primary data source Slack transcripts from data.world and Data for Democracy.	286 Single-spaced printed pages	NVivo 11 qualitative analysis The Slack transcripts were just as helpful as interviews in providing insight into the decisions, motivations, methods, reasoning, and operations of the firm, perhaps even more so as participants were talking to each other, not to an interviewer.
Conferences, meetings, and workshop observations	Attended a variety of events including data.world company meetings, an e-gov open data conference, hackathon, and a data journalist group meeting.	The events helped add another layer of understanding about these organizations. It also permitted triangulation with other sources.
Primary data source: Surveys	Data for Democracy User Participation Survey N=184, 14 questions (11 closed ended, 3 freeform text).	PLS-SEM exploratory analysis on the closed ended questions NVivo 11 qualitative analysis on the 3 text questions The survey data was helpful in learning about users of the data.world platform and understanding data activists and their role.
Primary data source data.world investor, business development, and press presentations	data.world Operating Principles - 19 slides data.world Overview of data.world - 27 slides data.world Benefit Corporation presentation - 6 slides data.world for Professionals - 8 slides data.world Linked Data 101 - 39 slides data.world Q4 Design Deep Dive - 42 slides data.world Rough Screen Shots - 10 slides	These presentations provided a view into how data.world wants to be perceived. This material presents the company's vision. It is a different perspective than that provided to users as the firm attempts to educate potential investors and partners about the data economy, the value of intermediaries, and the role of linked data.

Data Sources	Data Features	Use in Analysis
Primary data source Scripts for use by data.world employees	data.world Data Collaboration Workshop Script - 4 pgs data.world Press Tour Demo Script - 2 pgs data.world Script for Data Scientist Case Study Research Design - 5 pgs	This material offers a glimpse into data.world operations and how employees interact with users to increase usage.
Primary data source Notes, Documentation, User communications on public message threads, and General Information	Investigator's Notes - 11 pgs data.world Company Notes on Acquisition - 2 pgs data.world Notes on User Profiles - 1 pg data.world FAQ & Best Practices for D4D datasets on data.world - 5 pgs data.world Notes on Personas - 2 pgs data.world User Research Compilation - 2 pgs data.world Notes on Persona Dimensions - 1 pg data.world User Notes - 12 pgs, 10 pgs, 4 pgs data.world 6 User Introductory chats - 1 pg data.world Platform Test - 9 users - 9 pgs data.world Hackathon Best Practices - 7 pgs data.world Internal Hackathon Notes – 1 pg data.world Professional Use Takeaways - 7 pgs	These materials provide insight into the software development, user understanding, and customer support aspects of the platform.
Primary data source Statistics and Usage History	data.world – statistics and reports on usage, users, datasets, and social behaviors (likes, bookmarks, followers) data.world – Slack transcripts	This information shows the progress that the platform has made over the course of the study and provides insight about user behavior on the platform.
Secondary data sources	Published data.world case studies and white papers.	data.world published case studies on 2 customers and several white papers on the features & benefits of the platform. The case studies were helpful as I did not have direct access to customers.
Secondary data sources	Websites, reports, and blogs from governments, IGOs, NGOs, non-profit and for-profit organizations	These websites gave me a broad view of the many aspects of data work and the variety of players in this space. It also pointed out how far behind academic research is compared to the general public.
Secondary data sources Academic publications	344 articles on data, macro-economics, open data, e-government, data activism	Provided a foundation for understanding the various elements of data work and how they interact.

Table 8. Primary Interviews & Meetings

Name	Length	Speaker	Type	Name	Length	Speaker	Type
Employee1	17	Employee	Interview	Employee8	18	Employee	Interview
Employee10	16	Employee	Interview	Employee9	31	Employee	Interview
Employee11	33	Employee	Interview	Exec1	28	Exec	Interview
Employee12	20	Employee	Interview	Exec2	36	Exec	Interview
Employee13	31	Employee	Interview	Exec2	39	Exec	Interview
Employee2	35	Employee	Interview	Exec2	74	Exec	Interview
Employee2	24	Employee	Interview	Exec2	75	Exec	Interview
Employee3	12	Employee	Interview	Exec2	7	Exec	Interview
Employee4	42	Employee	Interview	Exec3	44	Exec	Interview
Employee4	53	Employee	Interview	Legal	37	Legal	Interview
Employee4	23	Employee	Interview	Team Meeting	53	Employees	Meeting
Employee4	34	Employee	Interview	Team Meeting	52	Employees	Meeting
Employee4	53	Employee	Interview	Team Meeting	59	Employees	Meeting
Employee4	34	Employee	Interview	Team Meeting	52	Employees	Meeting
Employee5	19	Employee	Interview	Team Meeting	63	Employees	Meeting
Employee5	9	Employee	Interview	Team Meeting	55	Employees	Meeting
Employee6	17	Employee	Interview	Team Meeting	53	Employees	Meeting
Employee7	24	Employee	Interview	User1	46	User	Interview
Employee7	28	Employee	Interview	User2	30	User	Interview
TOTAL MEETINGS & INTERVIEWS 38							
TOTAL MINUTES 1346							

Analysis Methods

To analyze the data, the recorded materials were transcribed into text by a transcription service, which along with Slack transcripts and other electronic communication were then coded in NVivo 11. After every five or so interviews in an iterative process, I used the software to identify repeating patterns and organize them into first order codes (in-vivo where possible), which were then gathered into broader axial codes or themes, and ultimately condensed into categories of processes. To build the case, I first wrote field notes while I was onsite at data.world. These were then developed into thick descriptions and ultimately integrated into a narrative (Balugun, et al., 2015; Schultze & Avital, 2011).

Setting: data.world

The company, comprised of thirty individuals during my study, originally billed itself as the "social network for data people" (data.world, 2016; Reader, 2016). This seemed odd for what at first glance appeared to be a data repository, but data.world was far more. It was one of the first data platforms. The product not only provided storage for data, along with APIs and integrations for uploading or downloading data, but it handled a range of query tools and even offered data science tutorials and practice data sets. There were several key aspects to the platform: sociality, projects, data movement, and semantic web links.

The social side of data.world allowed participants to create profiles, follow and like other participants and data projects, and collaborate through chats and messaging. It was described as "a social network geared toward helping data scientists connect and nerd out over collections of data" along with a "user experience that allows for the unanticipated glee of discovering of new data sets" (Reader, 2016). Yet the social aspect of data.world wasn't just for fun. The communal nature was designed from the beginning to foster collaboration and learning in order to make data more usable. The second key aspect was projects. Unlike data repositories and warehouses, data.world provided "data projects" where multiple data sets could be gathered along with supporting documentation and communications for the purpose of adding context and provenance. Projects contained not only multiple data sets, but also summaries, notes, supplementary files, data dictionaries, provenance, comments, and multimedia. data.world users could follow projects, staying abreast of changes and collaborating with others. If the project owner desired, any other member could be given rights to add or edit data, as well.

One of the key issues for data work has always been how to move the data. Getting it into and out of systems is still a challenge mainly because of size and/or format. This problem was dealt with early on with APIs to make data movement between systems easier and a developer toolkit. The APIs were described in this quote from the data.world website and the developer toolkit is illustrated in Figure 7:

> Creating an integration or application with data.world is the perfect way for you to empower your users to connect, explore, and share data....Even simple tasks like getting data out of data.world can be accomplished with our API, saving you the time and effort of manually recreating them... You'll have access to a suite of materials that will assist you in using our API and other features to do everything from completing small tasks to developing large-scale data apps....(data.world, 2018)

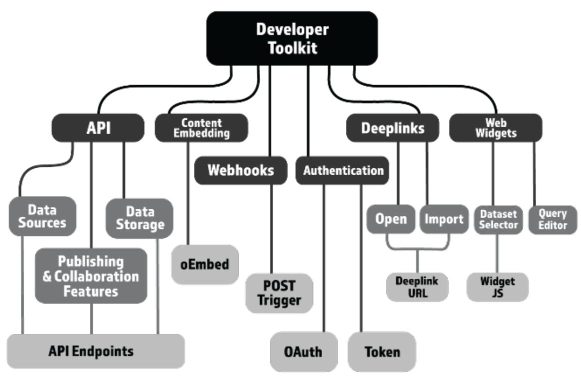

Figure 7. data.world Developer Toolkit

Beyond APIs, integrations with systems, such as Google, permitted data in online spreadsheets to be updated in their native app and automatically pushed to data.world. This was important for individual data users in particular because they often lacked the IT skills to use APIs. With an integration, the user only needed to link their Google spreadsheet to their data.world data project once. The other main component of data.world, and one of its strategic advantages, was its ability to provide linked data sets that could be used for semantic web publishing.

The business model of data.world had two streams: a free open data option for individuals and a paid subscription model for enterprises. Some of their clients include the Associated Press, Northwestern University, NGO Rare, eMarketer, Encast, and Square Panda, which demonstrates the wide range of industries they serve. The free option, originally conceived as a way to not only support open data but also seed the platform with data sets, became an important part of data.world's community as individuals published data sets, collaborated with other individuals, and began to want the same functionality that data.world offered in their workplace. The platform solved several of the challenges that still plagued data warehouses and repositories, such as data movement, trust, and lack of human resources. Data movement was solved through the API and over 40 integrations with systems ranging from Tableau and Salesforce to Facebook Ads and Canvas.

data.world Data Providers

Because data.world had both free and paid users, the data providers differed. The company began with a free, open data model. In order to build the community and achieve critical mass, all types of data sets and users were encouraged to use the platform for no charge. While there was a targeted effort to recruit academic and government institutions, all sorts of non-essential data was also encouraged to increase the entertainment and social value of the site and to increase stickiness. The individuals who contributed data provided perhaps the most entertaining of these data sets. The topic of beer is an example. When the keyword "beer" is entered on the platform, 151 data projects come up. A sampling of these include beer styles with ABV and bitterness units, beer distributor political contributions, beer reviews, best places to drink beer state-by-state, the names of US craft breweries, the cost of beer at all major league baseball stadiums, beer consumption by country, antioxidant levels by beer style, homebrew ingredient reviews, and many more. A number of the projects were crowdsourced, which increased the volume and quality of the data (data.world, 2018).

Individual Data Providers

While some data was simply republished from other open sources, other data was original. Data providers republished data on the platform for several reasons, including making the data easier to use and find, putting it into a project so that complementary documents and data could be added, and to find others interested in the topic, including potential collaborators. Original data from individuals showed a wide range of topics. Some examples included personal lists of books read, favorite restaurants, and recipes. One of the more significant types of individual data contributor was citizen science, with over 8000 data projects available for this keyword. Citizen science is crowd sourced data collection by private individuals for the purpose of gathering very large amounts of data for science that would be impossible for any single organization. Citizen scientists span the globe in collecting data on environments and habitats, species counts and distributions, and observations of natural phenomena (Bonney et al., 2009; Gura, 2013; Link et al., 2017). Examples of citizen science on the platform included bird counts, geological surveys, and a sea otter census.

A second type of individual data provider on the platform was data journalists. A data journalist develops a news story using data as the primary source. It is a combination of using “a nose for news” with the ability to use data to "tell a compelling story" (Gray, Chambers, & Bounegru, 2012). The data.world platform was a good tool for data journalists because it provided not only interesting existing data sets and a place to store new data sets, but collaboration and projects to organize complementary materials. The ability to provide provenance was also important to these data providers so they could document sources, a key aspect of responsible journalism.

Organizational Data Providers

The data.world organizational providers were more limited and were generally paid subscribers, not open data customers. These organizations used the platform to provide secure collaboration and data access to their members. Governance was a key feature so the organizations could control who had access to the data.

The Associated Press (AP) is an example of an organizational data provider. The AP was a cooperative non-profit news agency that offered subscriptions to its service, which provided breaking news, investigative reporting, and recently, data for news stories. The organization was quite large, and it was estimated that over half of the global population was exposed to AP content daily with 2,000 stories per day, 50,000 videos per year, and a million photos per year. Quality was paramount to AP, as well, and they prided themselves on earning 52 Pulitzer Prizes over their history. The AP subscriber base numbered over 3,000 agencies (Associated Press, 2019; data.world, 2018).

The AP was interested in growing its data journalism practice and the data.world platform seemed to provide a lot of what the organization needed. The problems the AP hoped to solve included members not leveraging the data well, data going to the wrong people and missing the right people who could use it, getting the data out to analytics tools, and simply finding the data the user wanted. In addition, the existing systems were expensive and did not seem to be paying off.

Government/Global Data Provider

From the beginning, data.world courted government agencies at the local, state, and national levels to use the platform for data publishing. There were currently over 100,000 government data sets available on the platform in early 2019. Considering that there are only 200,000-300,000 data sets on Data.gov (the US government open data site), the amount of government data on data.world is impressive considering their short history. Some of the government data on data.world was local or state and superseded older, limited data repositories. It also provided a solution for new data initiatives. Examples include the cities of New York, NY, Baltimore, MD, Tempe, AZ, and Bloomington, IN among others. Some data was published in multiple places, including Data.gov, government websites, and data.world. Organizations did this because data.world's platform offered features that were not available on data.gov, not the least of which is accessibility and the ability to add complementary materials. data.world was also deemed somewhat more reliable than data.gov. During the 2019 US government shutdown, for example, Data.gov went offline for weeks until the shutdown was resolved (Chappellet-Lanier, 2019).

data.world Data Users

Individual Data Users

Individuals used the data on the platform for both professional and personal reasons. The homebrewers might use the ingredient additive review dataset when deciding whether to add orange peels to their ale. Bird watchers might check recent bird counts to plan their travel, and home buyers might look up crime statistics in a new town. Another big group of individual users on data.world were data science students. Several universities such as Northwestern and the University of Texas at Austin used data.world for classes, and proactive people who were not in formal programs could find tutorials and practice data sets on the platform.

Organizational Data Users

Organizational data users included both members of paid corporate subscriptions and those who searched the open data for relevant content. Square Panda is an example of organizational data users. Square Panda was a tech-centric early reading startup developed with Stanford University neuroscience researchers. As an early-stage startup, the company had few employees who filled multiple roles. The company based their decisions on aggregated activity data such as the accuracy of spelling, words used, and other reading skill data with each user profile, and also used data on sales activity on shopping platforms and ads. However, getting and using that data was difficult particularly because things were run in a fast-moving agile environment and they had multiple sources of data coming from their own system and external sources such as Amazon, Shopify, and Facebook Ads.

Government/Global Data Users

Governments not only provided data, but also used data on the platform. The biannual employee survey from the City of Tempe, AZ is an example of a government data user. This survey covered six areas: professional development and career mobility; organizational support; supervisors and the work environment; compensation and benefits; employee engagement; and relationships between peers. The data was not only published publicly but was used internally by the city to measure employee satisfaction and determine where improvements could be made. The results of the survey provided recommendations to city managers to aid in improving working conditions.

In addition to data users and data providers, there were intermediaries on the platform. These were individuals, organizations, or government entities that provided additional services between data providers and users.

data.world Data Intermediaries

Individual Data Intermediaries

An interesting aspect of the data.world platform is the number of individuals who found data and republished it on data.world, often with value added complementary materials and much of the cleaning already done. I term these people individual data intermediaries. They differed from providers in that they worked on data provided from another source. They differed from users in that they only worked the data and/or made it more accessible. They did not employ the data for their own uses, such as economic evaluations or research. Some data sets published by individual data intermediaries were for entertainment value, such as the Halloween themed Salem Witches data set, while others demonstrated social activist agendas, such as political candidate donations by industry or President Trump's tweets harvested from Twitter. .

Organizational Data Intermediaries

The organizational data intermediaries on data.world were either data activist groups or paid subscriber organizations. The data activists, such as Data for Democracy, took raw open data from mostly government sources and worked it into clean data in usable formats, which they then published on openly the data.world platform along with complementary materials and information. These activities occurred during hackathons and also when members had to free time to devote to projects. Paid subscriber organizations provided similar tasks but the data was closed. These data workers manipulated the raw data in their organizations into useful data that could be used by other members of the organization.

Government/Global Data Intermediaries

Government and global intermediaries generally worked with government open data. There were a number of United Nations (UN) participants on data.world, for example, and they republished UN data for development programs, disaster research, health, immigration, and economics, among others. The data was also hosted on UN data repositories, but the data.world data permited multiple data sets within a project and the addition of complementary materials. Intermediaries put these multiple assets together and either used the data themselves or left it open for others.

CHAPTER FOUR

Analysis

I found three main aspects to the data.world case: the types of platform participants, how the platform was used, and the outcomes of platform utilization. I analyze these three aspects through nine examples in the data.world case and how they relate to the research question, how does intermediating through digital platforms impacts data? I define intermediating as dynamic digital activities that are performed in between the activities of data creation and data use (Ordanini & Pol, 2001; Zhao, 2012). Looking back to the data cycle described in the literature review section, intermediating is not one directional and it is not a mandatory route, as data may bypass this step completely. Intermediating is bidirectional because data may be created, go through intermediation, and then be employed by a user, however, the user may then provide feedback, which is then integrated by the intermediary.

Breaking down the case, I found nine types of platform participants, which included data providers, data users, and data intermediaries at the individual, organizational, and government levels. These participants demonstrated four main ways of using the platform: social learning, data wrangling, data complementing, and data liberalizing. Last, I found that the outcome of digital platform utilization was liquid data. These concepts are summarized in Table 9 and Figure 8. I define the constructs as follows: Social learning is the practice of learning new skills while engaging in digital social behaviors such as liking, following, commenting, and collaborating online. Data wrangling encompasses data tasks such as cleaning, reformatting, scraping, moving, substituting, redacting, and standardizing. Data complementing is the process of augmenting data sets with metadata and additional materials, which may include data dictionaries, summaries, provenance, previous versions, and visualizations. Data liberalizing is opening data to the largest possible group of relevant users. Liberalizing data is not the same as open data, although some data may result in that state, because open data is open to all. Rather than being open to the world, liberalized data may be data that is open within an organization. For example, in many companies, raw product data and sales data is available to only a few people, but liberalized data is made available to as many employees as possible in the hopes that it will stimulate innovation and efficiencies.

Table 9. Four Activities

Activity	Definition	Resources	Contribution
Social learning	The practice of learning new skills while engaging in digital social behaviors such as liking, following, commenting, and collaborating online	Barton & Tusting, 2005; Hildreth & Kimble, 2004; Lave & Wenger, 1991; Roberts, 2006; Wenger, 1999	Data.world demonstrates how a data skills are learned in social ecosystem using social media types of features
Data wrangling	Data tasks such as cleaning, reformatting, scraping, moving, substituting, redacting, and standardizing	Helland, 2011; Mayer-Schönberger and Cukier, 2014; Roose, 2018	Data.world demonstrates how a platform facilitates working with data more than other types of systems
Data complementing	Augmenting data sets with metadata and additional materials, which may include data dictionaries, summaries, provenance, previous versions, and visualizations.	Garsha, 2004; Sen, 2004; Sheth, 2003	The case demonstrates how platforms facilitate data complements that might be otherwise ignored
Data liberalizing	Opening data to as many relevant viewers as possible	Manyika, et al. 2013; van Schalkwyk, et al. 2016	The case illustrates how platforms increase data openness

Last, I bring in the construct of liquid data, which we discussed in the literature review. In this context, liquid data is 1) in a standardized machine-readable format; 2) fully cleaned and vetted; 3) augmented with complements that aid comprehension, and 4) open to the maximum number of relevant people possible. In the next sections I discuss each of the nine types of platform users, how they used the platform, and how the outcome of liquid data manifested itself in that context.

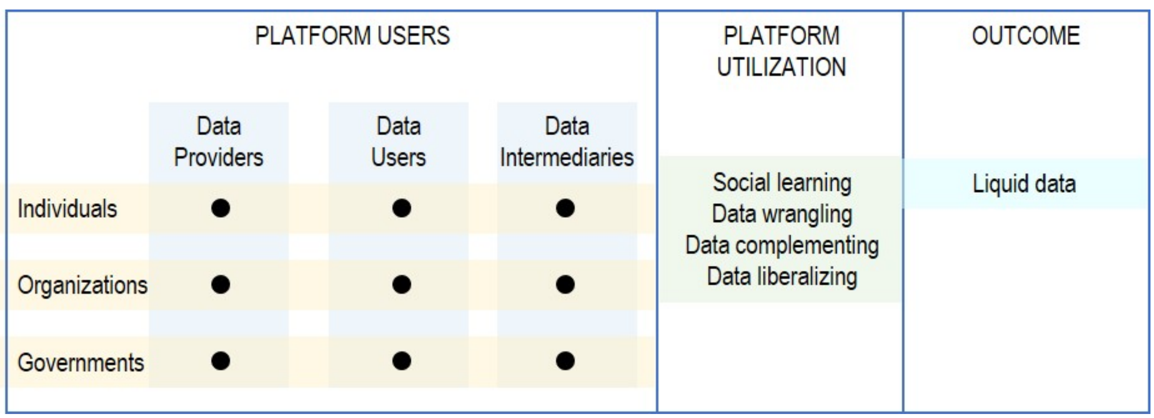

Figure 8. Platform Users, Utilization, Outcome

data.world Data Providers

Because data.world had both free and paid users, the data providers differed depending upon their own context. The company began with a free, open data model. In order to build the community and achieve critical mass, all types of data sets and users were encouraged to use the platform at no cost. While there was a targeted effort to recruit academic and government institutions, all sorts of non-essential data was also encouraged in order to increase the entertainment and social value of the site. This also increased stickiness, which is a means to encourage longer stays on the website. The individuals who contributed data provided perhaps the most entertaining of these data sets. The topic of beer is an example. When the keyword "beer" is entered on the platform, 151 data projects come up. A sampling of these include beer styles with ABV and bitterness units, beer distributor political contributions, beer reviews, best places to drink beer state-by-state, the names of US craft breweries, the cost of beer at all major league baseball stadiums, beer consumption by country, antioxidant levels by beer style, homebrew ingredient reviews, and many more. A number of the projects were crowdsourced, which increased the volume and quality of the data (data.world, 2018).

Individual Data Providers

While some data was simply republished from other open sources, other data was original. Data providers republished data on the platform for several reasons, including making the data easier to use and find, putting it into a project so that complementary documents and data can be added, and to find others interested in the topic, including potential collaborators. Original data from individuals had a wide range of topics. Some examples included personal lists of books read, favorite restaurants, and recipes. One of the more significant types of individual data contributor was citizen science, with over 8000 data projects available for this keyword. Citizen science is the crowd sourced data collection by private individuals for the purpose of gathering very large amounts of data for science that would be impossible for any single organization. Citizen scientists span the globe in collecting data on environments and habitats, species counts and distributions, and observations of natural phenomena (Bonney et al., 2009; Gura, 2013; Link et al., 2017). Examples of citizen science on the platform included bird counts, geological surveys, and a sea otter census. These individuals collaborated in gathering data and publishing it openly, sometimes cross-posting with other citizen science websites such as eBird.

Organizational Data Providers

The data.world organizational providers were more limited and were generally paid subscribers, not open data users. They also joined that platform later in the timeframe of the study when data.world hired on dedicated sales staff. Organizations used the platform to provide secure collaboration and data access to their members. Governance was a key feature, as well, as organizations could easily expand access to the data or limit confidential data where necessary. The data accessibility and collaboration features on the platform offered alternatives to data silos because members could not only see data from other departments, but also get complementary materials and ask questions of data set owners and their organization's data community. The ability to "document as you go" also helped with data reuse, which saved time and money for organizations. Users could document which data sets worked, which had problems, and how to address issues. This made data work more efficient because less time was spent preparing data, the data was more consistent, accurate, and higher quality, and more people could understand what the data represented, how it was used in the past, and how it might be used in the future.

Square Panda is an example of organizational data provider. Square Panda was a tech-centric early reading startup developed with Stanford University neuroscience researchers. As an early-stage startup, the company had few employees who filled multiple roles. The company based their decisions on aggregated activity data such as the accuracy of spelling, words used, and other reading skill data with each user profile, and also used data on sales activity on shopping platforms and ads. However, getting and using that raw data was difficult particularly because they operated in a fast-moving agile environment and had multiple sources of data, such as Amazon, Shopify, and Facebook Ads. The data.world platform allowed Square Panda to not only consolidate data, but automate integrations from external data sources which saved hundreds of hours of employee time. Data consolidation enabled queries from multiple data sets (called federated queries) to create new aggregated data sets, which made them even more efficient. The platform also liberalized data for all users in the organization, providing access to people who lacked access previously, and the user-friendliness and sociality of the platform encouraged greater activity on the platform and with other employees via the platform. The result of the switch to the new platform was a 2900% sales increase in one year (data.world, 2018).

Government/Global Data Providers

From the beginning, data.world courted government agencies at the local, state, and national levels to use the platform for open government data publishing. There were over 100,000 government data sets available on the platform by 2019. Considering that there are only 200,000-300,000 data sets on Data.gov (the US government open data site), the amount of government data on data.world is impressive considering their short history. Some of the government data on data.world was local or state and superseded older, limited data repositories or provided a solution for new data initiatives. Examples include the cities of Austin, TX, New York, NY, Baltimore, MD, Tempe, AZ, and Bloomington, IN among others. Some data was published in multiple places, including Data.gov, government websites, along with data.world. Organizations did this because data.world's platform offered features that were not available on data.gov, such as greater accessibility through APIs and integrations, the ability to aggregate data sets and complements in data projects, and linking data in the semantic web. Hosting data in multiple places is also safer. For example, during the 2019 US government shutdown, Data.gov went offline for weeks, but any Data.gov data sets co-published on data.world were still available (Chappellet-Lanier, 2019).

The City of Austin, Texas, is an example of a governmental data provider. The City of Austin had published over 800 open data sets on the platform as of 2019. The data was fairly diverse, including data from city animal shelters, traffic signals, dockless vehicles, disease outbreaks, traffic fatalities, arrests, and property subdivision applications. The City of Austin had over 200 followers on data.world, interested individuals who were notified of new data sets and comments concerning the city. Looking at one data set in particular, Animal Shelter Intakes, one can view the history of the data set including all changes, any comments made by the community, and who is following the City of Austin on the platform. The page also lists other data projects on data.world that used the animal shelter data set, along with any additional queries others have created. For the City of Austin, using the data.world platform meant that hundreds of data sets could be easily and quickly updated, opened to the public, and linked semantically to government websites. Updates were transparent and the public could easily see not just data, but all changes to the data and complements from the beginning. In the case of the animal shelter data, the audit trail was nearly two years long.

Publicly available data that is liquid has particular value for government data providers because of open information acts. Every year, public service entities spend thousands of hours providing data for open records requests. Recently, many entities are beginning to charge a fee for particularly large data requests to help recover costs and lost manhours (FOIA.gov, 2019; Texasattorneygeneral.gov, 2019). When liquid data is available to the public on data.world, the outcome is a significant cost savings for the government, as well as the philosophical benefit of greater government transparency.

data.world Data Users

Individual Data Users

Individuals use the data on the platform for both professional and personal reasons. Homebrewers might use a beer ingredient additive review dataset when deciding whether to add orange peels to their ale. Bird watchers might check recent counts to plan their visits, and home buyers might look up crime statistics in a new town. Another group of individual users on data.world are data science students. Several universities such as Northwestern and the University of Texas at Austin use data.world for classes, and proactive people who are not in formal programs can find tutorials and practice data sets on the platform. In this way, data.world works to bring data science to more people.

Data journalists were common data users on the platform, as well. A data journalist develops a news story using data as the primary source. It is a combination of using “a nose for news” with the ability to use data to "tell a compelling story" (Gray, Chambers, & Bounegru, 2012). The data.world platform was a good tool for data journalists because it provided not only interesting existing data sets and a place to store new data sets, but collaboration and projects to organize complementary materials. The ability to provide provenance was also important to these data providers so they could document sources, a key aspect of responsible journalism. The sociality of the platform also enabled lesser-skilled journalists to ask questions and learn from those who were more experienced. The outcome was the ability to create a wider range of data-based news stories with greater credibility along with the opportunity to improve data skills.

Organizational Data Users

Organizational data users included both members of paid corporate subscriptions and those who searched the open data for relevant content. However, the group I use in this example was comprised of data.world employees who used the platform for their own operations, which they termed "dog fooding." This comes from the phrase "eating your own dog food," which refers to a company using its own products. Dog fooding at data.world was popular and most employees spent considerable time on the platform. An employee described it, "A lot of our dashboards and our conversation focuses on stuff that is actually using our platform. And so, we have dashboards that are populated by our products that are based on actual user data that's collected from the vendors and stuff that we use on our platform. And so, we do a lot of dog fooding, as we like to call it, in that we have conversations about the data and about the work that's going into our platform." To summarize, as an organizational data user, data.world used the platform for data gathering and aggregating, to populate dashboards, to collaborate on product development, and the outcome ultimately sped up product delivery.

Government/Global Data Users

As an example of government data users, I use the biannual employee survey from the City of Tempe, AZ. Tempe is a data provider, but they also internally use the data they publish. The liquidity of the data on the platform made it easy for departments to make use of data they had never had access to before or found the data too difficult to deal with.

This particular employee survey covered six areas: professional development and career mobility; organizational support; supervisors and the work environment; compensation and benefits; employee engagement; and relationships between peers. The data was not only published publicly but was used internally to measure employee satisfaction and determine where improvements could be made. The results of the survey provided recommendations to city managers to aid in improving employee working conditions.

data.world Data Intermediaries

Individual Data Intermediaries

An interesting aspect of the data.world platform is the number of individuals who found data and republished it on data.world, often with value added complementary materials and much of the cleaning already done. Some data sets were for pure entertainment value, such as the Halloween themed Salem Witches data set, while others demonstrated social activism agendas, such as political candidate donations from certain industries or President Trump's tweets harvested from Twitter. One common type of individual data intermediary on the platform was the data activist. While some data activists work within organizations such as Data for Democracy or Open Data DC, others worked alone, rescuing and archiving open data from any future redaction. Some users took it upon themselves to make open records requests and then publish the data on the platform, as in the case of the data set USDA APHIS Inspection Reports. Others scraped data from non-machine readable sources such as PDFs or web pages and published it on the platform. These activists used the platform's tools to wrangle the data into a clean usable format, provided information on the source of the data, and opened it for others to use and view. The outcome was a greater number of open data sets being stored outside of government repositories, which resulted in higher availability of the data in case the original source was inaccessible along with greater use of the data because of the platform's usability and exporting features.

Organizational Data Intermediaries

The organizational data intermediaries on data.world were often paid subscribers. They worked the raw data in their organizations, aggregating and manipulating messy data into useful data that can be used by other members. The Associated Press (AP) was an example of an organizational data intermediary. The AP was a cooperative non-profit news agency that offered subscriptions to its service, providing breaking news, investigative reporting, and data for news stories. The organization was quite large, and it was estimated that over half of the global population was exposed to AP content daily with 2,000 stories per day, 50,000 videos per year, and a million photos per year. Quality was paramount to AP, as well, and they prided themselves on earning 52 Pulitzer Prizes over their history. The AP subscriber base numbered over 3,000 news agencies (Associated Press, 2019; data.world, 2018).

The AP was interested in growing its data journalism practice and the data.world platform seemed to provide a lot of what the organization needed. The problems the AP hoped to solve included members not leveraging the data well, data going to the wrong people and missing the right people who could use it, getting the data out into analytics tools, and simply finding the data the user wanted. In addition, the existing systems were expensive and did not seem to be paying off. When the AP went to the data.world platform they solved many of these problems, doubling both data production and data customers.

They credit these accomplishments to the data.world platform, which provided a single place for their customers to find and use data and a sophisticated means to get the data out, add complementary materials, and collaborate ("Data Solutions | AP," 2019). Figure 9 demonstrates the data.world AP data project for Opioid Prescriptions from 2010-2015. The website screenshot shows a preview of the data set, its metadata, documentation that describes what the data represents, a summary, queries that people have used, contributors, and comments from the community. Another key aspect on the platform was the ability for AP to work on data projects internally before releasing it to customers. This gave internal people an opportunity to improve the data projects by working together and putting complementary materials in place. All of this meant that customers could spend less time cleaning and managing the data. The data.world platform was also helpful for the data journalists who were often nascent data scientists with few technical skills. The data community was valuable to these people as they learned data analysis skills and it paid off with reduced time between raw data to news story publication. AP used the platform to house the data, wrangle and clean it, enrich it with context, and provide access to its subscribers. "I often say that most of the work in a data project is caught up in that unglamorous 80%— finding the data, vetting it, cleaning it, coming to understand its limitations. We were doing that for all of these stories anyway, so why not let our members benefit from that head start?" asked Troy Thibodeaux of AP (data.world, 2018).

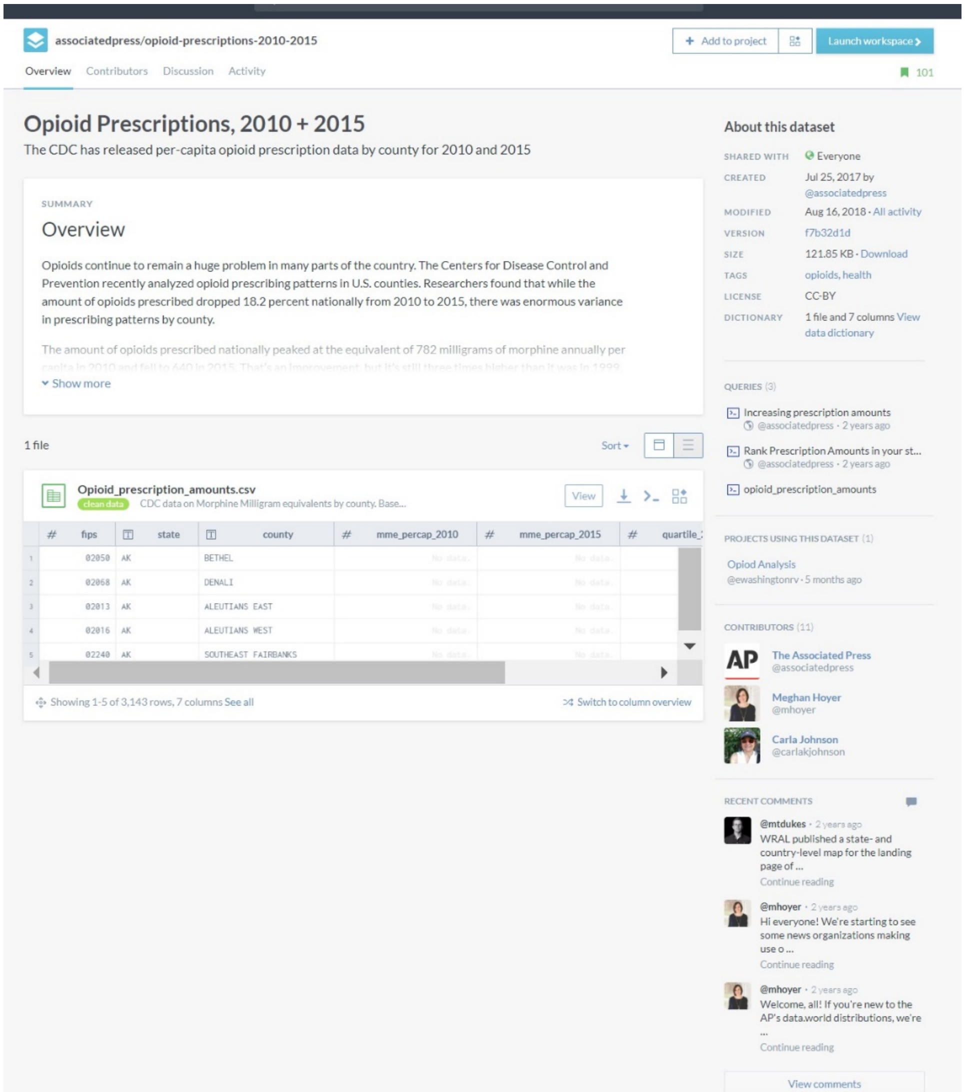

Figure 9. AP Data Project on Opioid Prescriptions

Government/Global Data Intermediaries

Government and global intermediaries generally work with government open data. There are a number of United Nations participants on data.world, for example, and they republish UN data for development programs, disaster research, health, immigration, and economics, among others. The data is also hosted on UN data repositories, but the data.world data permits multiple data sets within a project and the addition of complementary materials. Intermediaries put these multiple assets together and either use the data themselves or leave it open for others. Many government/global intermediaries are also data providers and data users.

The nine types of platform participants, data providers, data users, and data intermediaries at the individual, organizational, and government levels are summarized in Table 10, along with the four ways of using the platform: social learning, data wrangling, data complementing, and data liberalizing. Each instance also provides the manifestations of the liquid data outcome.

Table 10. Summary of Analysis

Example	Platform Participant Type	Utilization	Outcome
Citizen Scientists	Individual data provider	Social Learning: commenting and collaborating Data Wrangling: combining data sets, making data machine readable, semantic links Complementing: inclusion of maps and photos Liberalizing: making data fully open and copying to eBird platform	Liquid data: easy to search databases used by both scientists and birders for trends and queries, also popular with environmentalists and data journalists outside of the citizen science communities
Square Panda	Organizational data provider	Social Learning: asking questions/discussions Data Wrangling: aggregating data from multiple sources, standardizing Complementing: Notes about the various sources Liberalizing: more employees had access than before	Liquid data: able to use the data to pinpoint active users and potential users with greater accuracy, increased sales 2900%
City of Austin	Governmental data provider	Social Learning: n/a Data Wrangling: easier to clean data and publish Complementing: able to add supplemental information and comments Liberalizing: previously closed data was able to be published	Liquid data: able to publish data more easily, resulted in more data published, outcomes included reduced city costs for open records requests and greater transparency

Example	Platform Participant Type	Utilization	Outcome
Data Journalists	Individual data user	Social Learning: active community of people learning from each other Data Wrangling: lesser skilled users able to find greater success in working the data Complementing: able to append the data with different people's projects, able to add provenance Liberalizing: data available for any journalist to find a story	Liquid data: able to link the final article back to the original data, resulting in greater confidence in the truth of news articles
data.world Employees	Organizational data user	Social Learning: mostly used for suggestions on different pieces of data to look at Data Wrangling: less wrangling required because of data systems designed for the platform Complementing: able to add additional visualizations, comments Liberalizing: open to nearly all employees	Liquid data: with nearly all employees able to access liquid data, the pool of people looking at problems was greatly expanded, thereby exhibiting the phrase "with enough eyeballs all bugs are shallow." Also, company dashboards automatically updated with platform data made decision making fast and easy. No silos as nearly all employees had data access.
City of Phoenix	Governmental data user	Social Learning: n/a Data Wrangling: n/a Complementing: able to add the original survey questions to the results Liberalizing: open to all employees and to the public	Liquid data: able to use the data regardless of what department employees worked in, able to make decisions from the data and show how/why those decisions were made
Data Activists	Individual intermediary	Social Learning: lots of asking and guiding within the community Data Wrangling: able to manipulate messy data into usable formats Complementing: able to provide provenance, add metadata, raw data files (precleaning). Liberalizing: Made open records requests to increase open data availability, archived open data on multiple platforms.	Liquid data: data was easier to work with on the platform and easier to move and collaborate on, resulting in greater amounts of open data published and easier search for those looking for it. Data stored on the platform in addition to other places increased the availability went other repositories went down.

Example	Platform Participant Type	Utilization	Outcome
Associated Press	Organizational intermediary	Social Learning: AP data workers collaborated together prior to publishing to clients, clients collaborated with each other once data was published. Data Wrangling: facilitated cleaning and manipulating data, saving many man-hours. Complementing: able to add provenance, raw data files, and other articles using the data. Liberalizing: able to make the data available to more people within client organizations. Data was also easier to find.	Liquid data: the platform allowed AP to publish data in better condition in less time. This resulted in a greater number of clients, who were also better able to find what they needed and share it more easily with others in their organization. Also resulted in greater general confidence in news reliability.
United Nations	Governmental intermediary	Social Learning: n/a Data Wrangling: able to publish more data because of ease of use. Complementing: able to provide provenance, summaries, and related materials. Liberalizing: able to publish more quickly on the platform than on UN servers. Publishing on multiple platforms protected the data. Publishing on the platform made it easier to get the data out.	Liquid data: the UN data was protected by being published on both UN sites and on the platform because of redundancy. Platform data was easier to publish on, able to handle complementary materials (which was not the case on the UN sites) easier to search, and much easier to export from. The semantic links were also unique to the platform and important for UN data.

CHAPTER SEVEN

Discussion

After many conversations with interviewees, it became apparent to me that the goal of data.world wasn't just linking data or storing open data or providing a safe place for data nerds to congregate and chat online. It was about taking viscous data and transforming it into liquid data. A data.world employee explained it: "I think that's what makes us really unique and compelling in the marketplace. Just because I feel like we're the only company that's focusing on both sides of the spectrum. Data in and of itself is very well understood, but everything that happened around the data is not. And so when you bring those two things together in the context of the people who put the data in and the analysis and empowering the people to have bigger conversations around it and much faster work flows, I think it makes a big difference in the effectiveness of that report that you're getting each month."

In terms of a process, I identified platform users participating in platform utilization which resulted in liquid data. This is illustrated in Figure 10.

Based on my analysis of the case and the process model, I identify several propositions that reflect my research question of how intermediating through digital platforms impacts data. First, throughout the case I found that although platforms users varied considerably, most were able to find value in using the platform, albeit with differences in which features provided the greatest value.

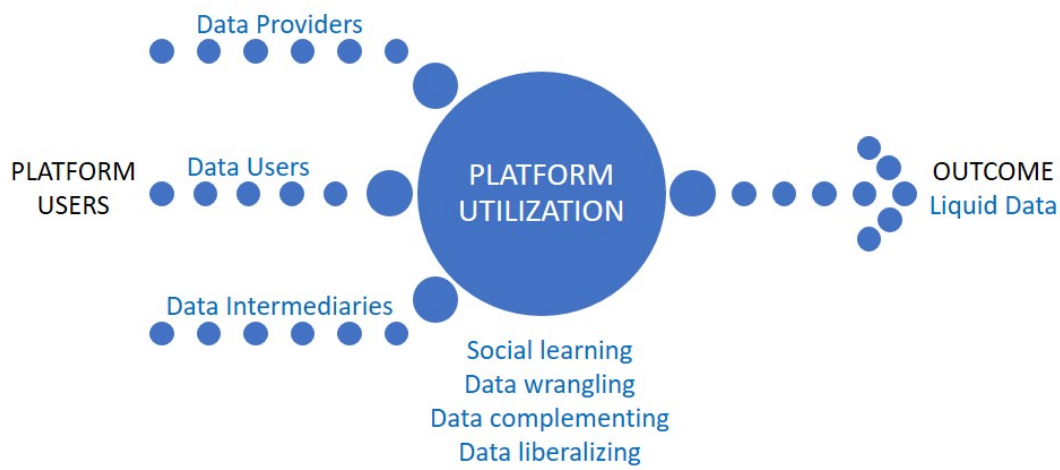

Figure 10. Liquid Data Process

Most platform users employed most of the features to some extent, especially data wrangling, but there were differences in which groups flocked to which features the most:

> P1a. Individuals, organizations, and governments receive benefits from utilizing a digital data platform in different ways.
>
> P1b. Individuals are more likely to receive value from social learning than other groups.
>
> P1c. Organizations are more likely to receive value from data complementing than other groups.
>
> P1d. Governments are more likely to receive value from data liberalizing than other groups.

Another way to look at platform users is by function: data providers, data users, and data intermediaries. With this perspective, I also found differences in how each group used the platform the most. All groups relied upon data wrangling features, but also had other favorite features:

> P2a. Data providers, data users, and data intermediaries receive the most benefit from data wrangling on the platform.
>
> P2b. Data providers are more likely to use data complementing than other groups.

P2c. Data users are more likely to use social learning than other groups.

P2d. Data intermediaries are more likely to use data liberalization than other groups.

The specific tasks undertaken on the platform impacted data differently. Social learning, for example, was useful for asking questions related to the data set or how to resolve a problem experienced when working with the data set. Viewing social learning through the communities of practice lens, I found substantial evidence that the social features found on the platform were beneficial in helping people learn from each other through sociality. This resulted in new topics learned more quickly and aided in the transfer of tacit knowledge. Data wrangling was likely the most practical activity, as users employed the platform's tools for a wide range of tasks. Some users did all their work in other systems and simply used the data.world platform to import, transform, and export back to the original system. Others used data.world for nearly all their data work. A unique feature to data.world was data complementing, but I found less evidence of it being used. When it was used, it was valuable and appreciated by the data community, but fewer participants took the time to provide complements. Last, data liberalization was commonly employed, perhaps because it was easy to achieve on the platform. This suggests the following propositions:

3a. Social learning on a digital data platform improves data skills through asking questions and receiving answers.

3b. Social learning on a digital data platform increases efficiency through collaboration.
3c. Data wrangling on a digital data platform increases efficiency through automated transformations.

3d. Data wrangling on a digital data platform increases efficiency through moving data via APIs and integrations.

3e. Data wrangling on a digital data platform enables lesser skilled users to perform data tasks more easily than through other means because of its tools and user interface.

3f. Data complementing provides greater veracity and confidence in data through provenance.

3g. Data complementing provides greater understanding of data through supplements.

3h. Data liberalizing on a digital data platform enables data access to a wide audience through its cloud platform.

3i. Data liberalizing on a digital data platform enables data access to a wide audience through free pricing for open data.

Last, in looking at platform utilization, I found that platform users who participated in all the activities (social learning, data wrangling, data complementing, and data liberalizing) finished projects faster, provided greater quality, and had their work used by more people than those who only used one or two activities. This suggests that the ultimate outcome of liquid data is best served through a combination of all the activities:

P4a: Liquid data is best achieved through a combination of all platform activities including social learning, data wrangling, data complementing, and data liberalizing.

P4b: Using only one or two activities on the platform results in relatively little change in data viscosity.

CHAPTER EIGHT

Implications

Data, one of the original topics researched in the early days of information systems, has enjoyed a fair amount of research over the years, including data structures, open data, and data systems such as repositories and data warehouses (Chaudhuri, et al., 2011; Chen, et al., 2012; Inmon, 1993; Janssen & Zuiderwijk, 2014). Yet challenges in using data still persist, as decades-old problems remain while new era issues add to the complexity of using data. The original research question for this study asked how does intermediating through digital platforms impact data. Through a grounded theory approach, I was able to contextualize abstract ideas about data into real world examples and show how data challenges can be overcome through the use of a platform. I found that while data was generally improved on platforms, it was impacted differently depending on which platform activities were utilized, and which type of platform user was involved (individual, organizational, governmental, data provider, data user, or data intermediary). The study also brought to light issues of how people use data, how people learn about data, and where platforms are more likely to provide the best outcome. This leads to several implications for research and practice, further research questions about data usage and management, and questions about the people who use data. Table 11 summarizes the main findings of this study and how they relate to the literature.

Table 11. Summary of Findings

Key Finding from Analysis	Relation to Literature
Data providers, data users, and data intermediaries make up a data cycle. These groups are further broken into three tiers: individual, organizational, governmental.	No studies that I am aware of discuss variances in those who work with data. While some studies focus on organizations and other on governments, none explicate differences.
Individuals, organizations, and governments receive benefits from utilizing a digital data platform in different ways.	No studies that I am aware of discuss variances in those who work with data.
The main activities of a data platform include social learning, data wrangling, data complementing, and data liberalizing.	Extension of Constantinides, Henfridsson, & Parker, 2018; Greenberg, 2018 descriptions of platform activities.
All features of the platform used in combination resulted in the greatest transformation from viscous to liquid data.	Extension of Eisenmann, et al., 2009; Parker, Van Alstyne, & Jiang, 2016 in platform usage and combination of activities. Congruent with viscous and liquid data research (e.g. Mallah, 2018; van Schalkwyk, Willmers, & McNaughton, 2016; Chui, Manyika, & Kuiken, 2014.)
Social learning on the platform was particularly effective for data work.	Congruent with theories of communities of practice (e.g. Lave & Wenger, 1991; Wenger, 1999)

Implications for Understanding Data Liquidity

Beginning with the primary research topic, I note that viscous and liquid data is a subjective description. If it were possible to develop a scale to measure data liquidity it may provide guidance for improving data usage. Such a scale could be used in both research settings and in organizations to estimate data workloads and costs. The idea of a data quality scorecard is not new (Piprani & Ernst, 2008), however, data quality is only one facet of liquid data. Data liquidity includes standard machine-readable formats, access/openness, and context, as well as timeliness, accuracy, completeness, and other quality measures. This suggest the following research questions.

1. What would an objective data liquidity scale look like?
2. How could a data liquidity scale benefit data users?
3. What steps would be needed to improve scores in each category?

Implications for Working with Data

One of the most common issues with data, especially with big data, is volume (Chen, et al., 2012; Marr, 2016). Volume creates three main problems: getting the data in, working the data, and getting the data out. The data.world case demonstrated how strong API and integration features help manage the movement of data and how automated tools and formatting eased data wrangling. Many data systems have import and export tools and a number have APIs. However, using these features is often confusing for those with limited technical skills. These users often prefer general business spreadsheets such as Excel or Google Sheets. This brings up the following concerns.

4. While data scientists may use Hadoop, R, SAS, or other specialty systems, what is the impact of Google Sheets and Excel for lesser skilled data platform users? How do system integration opportunities and limitations impact data usage? Logic suggests that platform integrations to Google Sheets and Excel should increase data usage because it lowers the bar for technical knowledge, but it would be interesting to understand if this is indeed so and under which circumstances. There may also be risks to using lesser specialized data tools. Which integrations/APIs benefit the greatest number of users would also provide direction.

5. The case study illustrated how the platform was able to combine multiple types of data with less work than through other methods. This implies that data aggregations are positive and desirable, but are they? Are there risks to aggregating data, such as misinterpretation or aggregating incompatible fields? If aggregation is beneficial, what types of data are best worked aggregated on a platform? Which types of data should not be aggregated on a platform or receive little benefit from it?

The data.world research suggests that one of the most important pieces of the data liquidity puzzle is the addition of context, primarily through meta data and provenance, but also through the addition of summaries, alternate versions of the data, and other related media. Yet we know little about which types of complementary data are the most helpful. The literature on data complements suggests providing meta data such as file characteristics and author, but we know little about how additional context changes data usage (Garsha, 2004; Sen, 2004). The following questions stem from this line of thought.

6. What are the various types of data complements and how do they impact data usage?
7. How does data complement interaction affect data usage?

Another side of the data issue is the semantic web. Once data is in place via the platform, new situations will arise that we have little experience with. Data platforms can link data through the semantic web, which is not new but is not currently a mainstream practice, either. If linking data becomes easier, it is likely to become more common. The internet is already full of dead web pages and expired links, hence the following questions.

7. What are the risks of massive amounts of open linked data provided through platforms?
8. How would extending the semantic web impact the number stale links?
9. Who will be responsible for maintaining accuracy and keeping links current on the semantic web?

Implications for Understanding People Who Work with Data

One of the more interesting facets of the data.world case is the data community where people learn data skills from each other through the social features on the platform. The communities of practice literature states that learning is a social activity and many benefit from interactions with others doing the same activities (Barton & Tusting, 2005; Hildreth & Kimble, 2004; Lave & Wenger, 1991). If we accept that sociality combined with data work improves data skills and data usage, then it behooves us to learn more about the relationships between participants, data, and social features on the platform through the following questions.

10. Which social features drive the greatest interaction among participants?
11. How can platforms best be used to build data skills?
12. Should data work be "dumbed down" on platforms in order to increase the number of data workers?
13. What are the pros and cons of this approach?
14. Which types of data and data work tend to be favored by data communities and why?
15. How do different platform features impact how people use data?
16. How can reticent platform users be encouraged to actively participate in a community?
17. Are there social issues in data communities, such as a lack of diversity or sexism, that need to be addressed?
18. What are the antecedents of successful data communities?

Implications for Open Government Data

Open government data (OGD), although not new, has still a long way to go to reach its potential (Bertot, et al., 2010; Jetzek, et al., 2013). The problems of OGD include difficulty getting data in and out, delays in updating new data, inconsistent data, incomplete data, non-standard formats, untimely data, confusing column-headings, and lack of context. The data.world case demonstrated how a platform could ease a number of OGD issues, which ultimately made it easier and less expensive for governments to provide open data.

The standardized formats, automatic linked data, and easy import/export features alone improved OGD efficiency for the governments mentioned in the case. Looking at this sector, several potential research questions are suggested.

19. What types of OGD are best presented on a platform?
20. How can data platforms increase citizen engagement?
21. What are the costs and benefits for governments to use data platforms instead of self-hosting OGD?

Implications for Personal Data Use

One of the surprising findings in this research was the number of individuals who use data as part of a hobby. The case showed a range of personal data sets from book collections to wine lists to sports and users engaged in many of the same activities as their professional counterparts, including data analysis and visualizations. It is possible that there is a market for simplified data products designed for hobbyists, as opposed the current products designed for professionals, or a potential for partnering with other hobby platforms such as Pinterest. This suggests the following questions.

22. In which personal activities do people tend to track data?
23. What are the antecedents of tracking hobby data?
24. Which hobby/entertainment platforms might lend themselves to data extensions?
25. How does data impact how people participate in a hobby?
26. What data products do individuals use and how do they use them?

Practical Implications

Managers working with data should consider digital data platforms for several reasons. First, firms commonly experience a shortage of employees with data skills, which results in delays in getting data out to information systems. Second, many managers do not possess data skills beyond some Excel experience and are not able to understand the data or how to make use of it for decisions. Third, existing data systems often have silos and data is treated with a need-to-know attitude rather than an open data perspective. Data platforms can solve these problems. The cloud architecture of data platforms provides easy access through a browser while still providing security. This means that IT departments do not have to load special software on employee computers. Data is also easily imported and exported via integrations and APIs. Moving data around has long been a problem for organizations and these tools simplify the task. Platforms also understand that many users will not be data scientists, therefore extra help screens, tutorials, and examples are available to guide users. The collaboration and social features are particularly helpful for lesser skilled employees to ask questions directly on the data project and receive answers that are then available for other users. Another problem in corporate data use is not understanding cryptic column headings, what the data is related to, or who else might have worked on it and what their results were. Complementing data with summaries, visualizations, history, and past work all help to make data more understandable. When data is more understandable, managers are better able to use it for making decisions. Last, data platforms come from an open data concept, where data should be open unless it is necessary to lock it up. This is in contrast to most existing data systems where data is considered closed unless an individual has a need to know. This is a paradigm change for many organizations but making data available to more people in the organization can increase innovation and efficiency. However, making viscous data available is likely to have little impact. Liquid data that can be easily employed by the greatest number of people will provide the greatest value.

Future Research

There are many opportunities for future research on data platforms. The take-aways from this work include the following: 1) While new data problems are coming to light we still haven't solved all the issues from 20 years ago; 2) Data skills are still rare and increasing them should be a high priority; 3) Social learning appears to be particularly helpful for learning data skills; 4) Digital platforms are well suited to collaborating on data; 5) Digital platforms are well suited to moving data and transforming data; and 6) Digital platforms are particularly well suited to democratizing data through liberalization. These take-aways may be applied to future research to expand what we know about the phenomenon. Several suggestions for future work are discussed below.

First, delving more deeply into the propositions and looking at a replication with another data platform would produce more insights. Second, understanding the relationships between the five four mechanisms activities could be a fruitful direction. It would provide a better understanding of how they work together, different manifestations of the activities, and what happens when the balance of the activities changes. Third, case work with a firm that adopts a digital data platform could illuminate the process from the organizational side, including barriers not visible in this study.

The data gathered during this project included a great deal of material that was out of scope for this paper but would be relevant for future research. For example, a quantitative study drawn from internal Data for Democracy user surveys might provide insights on the antecedents and success factors for data activists. Using community social interactions, we might understand if data communities vary from other online communities, and if so, how and under what circumstances. There is also a large amount of case data on the start up phase of a data company that focuses on business operations and entrepreneurship. This part of the research could inform questions about the viability and success factors for data economy focused new enterprises, technology employee acquisition and retention factors, and the practice of dog-fooding.

CHAPTER NINE

Conclusion

As with any research, there are limitations to this study. First, there are potential weaknesses with the data. In the interviews, poorly worded questions, participants telling the interviewer what they want to hear, and poor recall can impact responses (Yin, 2003). Archival data, artifacts, and documents may have selective bias as I was only able to use the content I was given access to. Archival data may also reflect the biases of the authors (Yin, 2003). Even direct observation has its limitations as participants may behave differently because I was there, although over time I believe this lessened considerably as employees grew used to me (Yin, 2003). By using a wide variety of data sources, it was my intention to mitigate these limitations through source diversity and triangulation (Benbasat, Goldstein, & Mead, 1987; Myers, 1997, Yin, 2003).

This work contributes to the IS literature in several ways. First, it provides a detailed method for conducting grounded theory research in information systems which should be of interest to those who are learning this method. Second, the literature review provides a broader view of data than that found in many studies, incorporating open and semantic data, data warehousing, and data characteristics. The literature review also describes a data cycle of data providers, data users, and data intermediaries, each with additional individual, organizational, or governmental tiers. These descriptions and categorizations should be of help to data scholars and those who research data systems and management. The findings of this research offer a contribution through new examples that extend or support existing theory on platforms and communities of practice. Last, this research provides new theory on how digital platform activities aid the transformation of viscous data into liquid data and how groups use data platforms in different ways.

Data work is still in its infancy as individuals struggle to learn data skills and organizations learn how to use the data it collects. Many have jumped on the data bandwagon but have yet to realize value from it. Data collection and storage seems to have progressed at a faster rate than data usage, but hopefully this research is step towards democratizing data.

APPENDIX

APPENDIX

Coding Details

Table A.1. Coding Details

Theme	Platform Utilization	Social Learning	Data Wrangling
Selective Codes	Building the platform	Collaboration	Data
	Company & operations	Community	Data challenges
	Competitors	Data people	Data privacy & security
	Data users	Data training	Data tools
	data.world Platform	Data users	Data types
	External parties		Working with data
	How people use the system		Metadata & provenance
Theme	Data Liberalization	Data Liquidation	Data Democratization
Selective Codes	Data providers	Consolidating data	Moving data
	Integrations	Data Intermediaries	Opening data
	Moving data	Data Intermediaries	Data into information
		Data into information	Value of the platform

Table A.2. Platform Utilization Selective & Open Codes

Open Codes	Selective codes
Partners	Building the platform
Paying Customers	Building the platform
Platform engine	Company & operations
Product	Company & operations
Competitors	Competitors
Associated Press	Data users
eMarketer	Data users
Encast	Data users
Hobbyists on the platform	Data users
NGO Rare	Data users
Northwestern University	Data users
Square Panda	Data users
Product development	data.world Platform
Transparency	data.world Platform
Github	External parties
Google	External parties
Microsoft	External parties
Tableau	External parties
Data as entertainment	How people use the system
How users use the system	How people use the system

Table A.3. Social Learning Selective & Open Codes

Open Codes	Selective codes
Finding platform collaborators	Collaboration
Platform collaboration	Collaboration
Platform crowdsourcing	Collaboration
Data community	Community
Data people	Community
Following	Community
Learning through the community	Community
Liking	Community
Members of the platform community	Community
New to data on the platform	Community
New users on the platform	Community
Platform commenting	Community
Platform community	Community
Profiles	Community
Types of users on the platform	Community
Data scientists	Data people
Geeks	Data people
Academic data science programs	Data training
Practice data sets on the platform	Data training
Training	Data training
Training on the platform	Data training
Tutorials on the platform	Data training
Users lack data skills on the platform	Data training
Users on the platform	Data users

Table A.4. Data Wrangling Selective & Open Codes

Open Codes	Selective codes
Data	Data
Data challenges	Data challenges
Data formats	Data challenges
Silos	Data challenges
Which data to load	Data challenges
Data Privacy	Data privacy & security
Data Security	Data privacy & security
Health data	Data privacy & security
Privacy on the platform	Data privacy & security
Private data being posted on the platform	Data privacy & security
Data tools	Data tools
Excel on the platform	Data tools
JSON on the platform	Data tools
SPARQL on the platform	Data tools
SQL on the platform	Data tools
Tools	Data tools
Data types	Data types
Multiple data formats on the platform	Data types
Data cleaning	Working with data
Data consolidation	Working with data
Data wrangling	Working with data
Queries	Working with data
Working with data	Working with data

Table A.5. Data Complementing Selective & Open Codes

Open Codes	Selective codes
Adding complements	Metadata & provenance
Data projects with multiple media on the platform	Metadata & provenance
Data warehouses	Metadata & provenance
Lack of complements	Metadata & provenance
Metadata	Metadata & provenance
Multiple data sources on the platform	Metadata & provenance
Provenance	Metadata & provenance
Reidentification on the platform	Metadata & provenance
Trust	Metadata & provenance

Table A.6. Data Liberalization Selective & Open Codes

Open Codes	Selective codes
Citizen scientists on the platform	Data providers
Governments on the platform	Data providers
Platform data providers	Data providers
Facebook Ads on the platform	Integrations
Github and the platform	Integrations
Google on the platform	Integrations
Integrations on the platform	Integrations
Oracle on the platform	Integrations
Software on the platform	Integrations
Tableau on the platform	Integrations
APIs on the platform	Moving data

Table A.7. Data Liquidation Selective & Open Codes

Open Codes	Selective codes
Federated queries on the platform	Consolidating data
Data contests	Data Intermediaries
Platform data intermediaries	Data Intermediaries
Making data liquid	Data into information

Table A.8. Data Democratization Selective & Open Codes

Open Codes	Selective codes
Consolidation on the platform	Moving data
Getting data into the platform	Moving data
Getting data out of the platform	Moving data
Moving data on the platform	Moving data
Access on the platform	Opening data
Data for Democracy	Opening data
Data hackathons	Opening data
Data Rescue	Opening data
Democratization of data on the platform	Opening data
No silos on the platform	Opening data
Semantic web	Opening data

REFERENCES

5-Star Open Data. (2010). 5-star Open Data. Retrieved March 17, 2019, from http://5stardata.info/en/

Aalst, W. v d, & Damiani, E. (2015). Processes Meet Big Data: Connecting Data Science with Process Science. *IEEE Transactions on Services Computing*, *8*(6), 810–819. https://doi.org/10.1109/TSC.2015.2493732

Abadi, D. (2009). Data Management in the Cloud: Limitations and Opportunities. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, *32*(1), 3–12.

Abbasi, A., Sarker, S., Aalto University, & Chiang, R. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems*, *17*(2), I–XXXII. https://doi.org/10.17705/1jais.00423

Abrahamson, E. (1991). Managerial Fads and Fashions: The Diffusion and Refection of Innovations. *Academy of Management Review*, *16*(3), 586–612. https://doi.org/10.5465/AMR.1991.4279484

Access to information - Kenya Law Reform Commission (KLRC). (2018). Retrieved November 28, 2017, from http://www.klrc.go.ke/index.php/constitution-of-kenya/112-chapter-four-the-bill-of-rights/part-2-rights-and-fundamental-freedoms/201-35-access-to-information

Agarwal, R., & Dhar, V. (2014). Editorial—Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research. *Information Systems Research*, *25*(3), 443–448. https://doi.org/10.1287/isre.2014.0546

Agrawal, M., Hariharan, G., Kishore, R., & Rao, H. R. (2005). Matching intermediaries for information goods in the presence of direct search: an examination of switching costs and obsolescence of information. *Decision Support Systems*, *41*(1), 20–36. https://doi.org/10.1016/j.dss.2004.05.002

Alavi, M., Kayworth, T. R., & Leidner, D. E. (2005). An Empirical Examination of the Influence of Organizational Culture on Knowledge Management Practices. *Journal of Management Information Systems*, *22*(3), 191–224. https://doi.org/10.2753/MIS0742-1222220307

Alavi, M., & Leidner, D. E. (1999). Knowledge Management Systems: Issues, Challenges, and Benefits. *Commun. AIS*, *1*(2es), 37.

Alavi, M., & Leidner, D. E. (2001). Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues. *MIS Quarterly*, *25*(1), 107–136. https://doi.org/10.2307/3250961

Alexander, F. (2016, August 16). The dark side of the source: Bringing unused customer data into the light. Retrieved March 19, 2018, from IBM Business Analytics Blog website: https://www.ibm.com/blogs/business-analytics/the-dark-side-of-the-source-bringing-unused-customer-data-into-the-light/

Allen, A. L. (2016). Protecting One's Own Privacy in a Big Data Economy. *Harvard Law Review*, *130*(2), 71–78.

Allen, F., & Santomero, A. M. (1997). The theory of financial intermediation. *Journal of Banking & Finance*, *21*(11–12), 1461–1485.

Almklov, P., Østerlie, T., & Haavik, T. (2014). Situated with Infrastructures: Interactivity and Entanglement in Sensor Data Interpretation. *Journal of the Association for Information Systems*, *15*(5), 263–286. https://doi.org/10.17705/1jais.00361

American Chemical Society. (2017). Crystallography. Retrieved January 31, 2018, from Crystallography website: https://www.acs.org/content/acs/en/careers/college-to-career/chemistry-careers/cystallography.html

Amiri, A. (2007). Dare to share: Protecting sensitive knowledge with data sanitization. *Decision Support Systems*, *43*(1), 181–191. https://doi.org/10.1016/j.dss.2006.08.007

Amit, R., Brander, J., & Zott, C. (1998). Why do venture capital firms exist? theory and canadian evidence. *Journal of Business Venturing*, *13*(6), 441–466. https://doi.org/10.1016/S0883-9026(97)00061-X

Amsterdam Museum. (2015, August 25). Online Collection. Retrieved March 17, 2019, from Amsterdam Museum website: https://www.amsterdammuseum.nl/en/collection/online-collection

An example of how to perform open coding, axial coding and selective coding. (2013, July 22). Retrieved November 14, 2018, from The PR Post website: https://prpost.wordpress.com/2013/07/22/an-example-of-how-to-perform-open-coding-axial-coding-and-selective-coding/

Anderson, F. M. (1908). *The Constitutions and Other Select Documents Illustrative of the History of France, 1789-1907*. Retrieved from https://books.google.com/books?id=FJwMAQAAMAAJ&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Anderson, J. Q., & Rainie, L. (2012). Big Data: Experts say new forms of information analysis will help people be more nimble and adaptive, but worry over humans' capacity to understand and use these new tools well. Washington, DC: Pew Research Center.

Arthur, C. (2010, June 29). Businesses unwilling to share data, but keen on government doing it. *The Guardian*. Retrieved from http://www.theguardian.com/technology/2010/jun/29/business-data-sharing-unwilling

Artz, D., & Gil, Y. (2007). A survey of trust in computer science and the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web*, *5*(2), 58–71. https://doi.org/10.1016/j.websem.2007.03.002

Aslam, S. (2017, January 23). Pinterest by the Numbers (2017): Stats, Demographics & Fun Facts. Retrieved December 5, 2017, from https://www.omnicoreagency.com/pinterest-statistics/

Assaf, A., & Senart, A. (2012). Data Quality Principles in the Semantic Web. *2012 IEEE Sixth International Conference on Semantic Computing*, 226–229. https://doi.org/10.1109/ICSC.2012.39

Association of Business Schools Journal List 2015. (2016). Retrieved from http://gsom.spbu.ru/files/abs-list-2015.pdf

Attard, J., Orlandi, F., Scerri, S., & Auer, S. (2015). A systematic review of open government data initiatives. *Government Information Quarterly*, *32*(4), 399–418. https://doi.org/10.1016/j.giq.2015.07.006

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). DBpedia: A Nucleus for a Web of Open Data. In K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, … P. Cudré-Mauroux (Eds.), *The Semantic Web* (pp. 722–735). https://doi.org/10.1007/978-3-540-76298-0_52

Austin's Best Places to Work for 2017. (2017). Retrieved January 31, 2018, from Austin Business Journal website: https://www.bizjournals.com/austin/news/2017/05/12/austin-s-best-places-to-work-for-2017.html

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data & Society*, *2*(2), 2053951715594634. https://doi.org/10.1177/2053951715594634

Bacchetta, P., & van Wincoop, E. (2016). The Great Recession: A Self-Fulfilling Global Panic. *American Economic Journal: Macroeconomics*, *8*(4), 177–198. https://doi.org/10.1257/mac.20140092

Baker, T., & Nelson, R. E. (2005). Creating Something from Nothing: Resource Construction through Entrepreneurial Bricolage. *Administrative Science Quarterly*, *50*(3), 329–366. https://doi.org/10.2189/asqu.2005.50.3.329

Bakhashi, H., Mateos-Garcia, J., & Whitby, A. (2014). *Model workers: how leading companies are recruiting and managing their data talent* (p. 45). Retrieved from National Centre for Vocational Education Research website: http://hdl.voced.edu.au/10707/325216

Bakos, J. Y. (1991). A Strategic Analysis of Electronic Marketplaces. *MIS Quarterly*, *15*(3), 295. https://doi.org/10.2307/249641

Bakos, Y. (1998). The emerging role of electronic marketplaces on the Internet. *Communications of the ACM*, *41*(8), 35–42.

Bakos, Y., Henry C. Lucas, J., Oh, W., Simon, G., Viswanathan, S., & Weber, B. W. (2005). The Impact of E-Commerce on Competition in the Retail Brokerage Industry. *Information Systems Research*, *16*(4), 352–371. https://doi.org/10.1287/isre.1050.0064

Bakos, Y., & Katsamakas, E. (2008). Design and Ownership of Two-Sided Networks: Implications for Internet Platforms. *Journal of Management Information Systems*, *25*(2), 171–202. https://doi.org/10.2753/MIS0742-1222250208

Balugun, J., Bartunek, J. M., & Do, B. (2015). Senior Managers' Sensemaking and Responses to Strategic Change. *Organization Science*, *26*(4), 960–979. https://doi.org/10.1287/orsc.2015.0985

Banisar, D. (2002). *Freedom of Information and Access To Government Records Around The World* (p. 45). Retrieved from World Bank website: www1.worldbank.org/publicsector/learningprogram/.../AccessInfoLaw%20Survey.rtf

Banta, T. W., Pike, G. R., & Hansen, M. J. (2009). The use of engagement data in accreditation, planning, and assessment. *New Directions for Institutional Research*, *2009*(141), 21–34. https://doi.org/10.1002/ir.284

Barrett, M., Oborn, E., & Orlikowski, W. (2016). Creating Value in Online Communities: The Sociomaterial Configuring of Strategy, Platform, and Stakeholder Engagement. *Information Systems Research*, *27*(4), 704–723. https://doi.org/10.1287/isre.2016.0648

Barrett, M., Oborn, E., Orlikowski, W. J., & Yates, J. (2012). Reconfiguring Boundary Relations: Robotic Innovations in Pharmacy Work. *Organization Science*, *23*(5), 1448–1466. https://doi.org/10.1287/orsc.1100.0639

Barrett, R., & Maglio, P. P. (1999). Intermediaries: An approach to manipulating information streams. *IBM Systems Journal; Armonk*, *38*(4), 629–641.

Barton, D., & Tusting, K. (2005). Beyond Communities of Practice: Language Power and Social Context. Cambridge University Press.

Batagan, L. P., Constantin, D.-L., & Moga, L. M. (2017). Facts and Prospects of Open Government Data Use. A Case Study in Romania. In C. Certoma, M. Dyer, L. Pocatilu, & F. Rizzi (Eds.), *Citizen Empowerment and Innovation in the Data-Rich City* (pp. 195–208). Cham: Springer International Publishing Ag.

Batini, C., & Scannapieco, M. (2016). *Data Quality Issues in Linked Open Data*. Berlin: Springer-Verlag Berlin.

Batra, D. (2017). Adapting Agile Practices for Data Warehousing, Business Intelligence, and Analytics. *Journal of Database Management (JDM)*, *28*(4), 1–23. https://doi.org/10.4018/JDM.2017100101

Baumer, B. S. (2017). *A Grammar for Reproducible and Painless Extract-Transform-Load Operations on Medium Data*. Retrieved from https://arxiv.org/abs/1708.07073

Bechmann, A. (2013). Internet profiling: The economy of data intraoperability on Facebook and Google. *MedieKultur: Journal of Media and Communication Research*, *29*(55), 19.

Bélanger, F., & Carter, L. (2008). Trust and risk in e-government adoption. *The Journal of Strategic Information Systems*, *17*(2), 165–176. https://doi.org/10.1016/j.jsis.2007.12.002

Belissent, J. (2017, February 3). The Data Economy Is Going To Be Huge. Believe Me. Retrieved December 5, 2017, from Forrester website: https://go.forrester.com/blogs/17-02-03-the_data_economy_is_going_to_be_huge_believe_me/

Belissent, J., Cullen, E., Leganza, G., & Lee, J. (2017). *The "Data For Good" Movement Delivers Social Impact*. Retrieved from Forrester website: https://www.forrester.com/report/The+Data+For+Good+Movement+Delivers+Social+Impact/-/E-RES138592

Belissent, J., Green, C., Leganza, G., Kisker, H., Iannopollo, E., & Datta, D. (2014). *It's Time To Take Your Data To Market*. Retrieved from Forrester website: https://www.forrester.com/report/Its+Time+To+Take+Your+Data+To+Market/-/E-RES115339

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly*, *11*(3), 369–386. https://doi.org/10.2307/248684

Benjamin, R., & Wigand, R. (1995). Electronic Markets and Virtual Value Chains on the Information Superhighway. *Sloan Management Review; Cambridge, Mass.*, *36*(2), 62–72.

Benkler, Y. (2006). The Wealth of Networks: How Social Production Transforms Markets and Freedom. Yale University Press.

Benner, K. (2017, June 6). Pinterest Raises Valuation to $12.3 Billion With New Funding. *The New York Times*. Retrieved from https://www.nytimes.com/2017/06/06/technology/pinterest-raises-valuation-to-12-3-billion-with-new-funding.html

Bera, P., Burton-Jones, A., & Wand, Y. (2011). Guidelines for Designing Visual Ontologies to Support Knowledge Identification. *MIS Quarterly, 35*(4), 883-A11.

Berente, N., & Yoo, Y. (2011). Institutional Contradictions and Loose Coupling: Postimplementation of NASA's Enterprise Information System. *Information Systems Research, 23*(2), 376–396. https://doi.org/10.1287/isre.1110.0373

Berners-Lee, T. (2009). Linked Data - Design Issues. Retrieved March 17, 2019, from https://www.w3.org/DesignIssues/LinkedData

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American, 284*(5), 34–43.

Bernstein, A., Hendler, J., & Noy, N. (2016). A new look at the semantic web. *Communications of the ACM, 59*(9), 35–37. https://doi.org/10.1145/2890489

Berrone, P., Ricart, J. E., & Carrasco, C. (2016). The open kimono: toward a general framework for open data initiatives in cities. *California Management Review, 59*(1), 39–70. https://doi.org/10.1177/0008125616683703

Berry, M. J., & Linoff, G. (1997). *Data Mining Techniques: For Marketing, Sales, and Customer Support.* New York, NY, USA: John Wiley & Sons, Inc.

Bertot, John C., Jaeger, P. T., & Grimes, J. M. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly, 27*(3), 264–271. https://doi.org/10.1016/j.giq.2010.03.001

Bertot, John Carlo, Gorham, U., Jaeger, P. T., Sarin, L. C., & Choi, H. (2014). Big data, open government and e-government: Issues, policies and recommendations. *Information Polity, 19*(1, 2), 5–16.

Bertot, John Carlo, McDermott, P., & Smith, T. (2012). Measurement of Open Government: Metrics and Process. *Proceedings of the 45th Hawaii International Conference on System Sciences*, 2491–2499. https://doi.org/10.1109/HICSS.2012.658

Bezuidenhout, L. M., Leonelli, S., Kelly, A. H., & Rappert, B. (2017). Beyond the digital divide: Towards a situated approach to open data. *Science and Public Policy, 44*(4), 464–475. https://doi.org/10.1093/scipol/scw036

Bharadwaj, S., Bharadwaj, A., & Bendoly, E. (2007). The Performance Effects of Complementarities Between Information Systems, Marketing, Manufacturing, and Supply Chain Processes. *Information Systems Research, 18*(4), 437–453. https://doi.org/10.1287/isre.1070.0148

Bhargava, H. K., & Choudhary, V. (2004). Economics of an Information Intermediary with Aggregation Benefits. *Information Systems Research, 15*(1), 22–36. https://doi.org/10.1287/isre.1040.0014

Biglaiser, G., & Friedman, J. W. (1994). Middlemen as guarantors of quality. *International Journal of Industrial Organization, 12*(4), 509–531. https://doi.org/10.1016/0167-7187(94)90005-1

Birchall, C. (2015). "Data.gov-in-a-box": Delimiting transparency. *European Journal of Social Theory, 18*(2), 185–202. https://doi.org/10.1177/1368431014555259

Birks, D. F., Fernandez, W., Levina, N., & Nasirin, S. (2012). Grounded theory method in information systems research: its nature, diversity and opportunities. *European Journal of Information Systems, 22*(1), 1–8. https://doi.org/10.1057/ejis.2012.48

Bolin, G., & Schwarz, J. A. (2015). Heuristics of the algorithm: Big Data, user interpretation and institutional translation. *Big Data & Society, 2*(2). https://doi.org/10.1177/2053951715608406

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. *BioScience, 59*(11), 977–984. https://doi.org/10.1525/bio.2009.59.11.9

Booton, J. (2016, June 15). IBM finally reveals why it bought The Weather Company. Retrieved January 15, 2018, from MarketWatch website: http://www.marketwatch.com/story/ibm-finally-reveals-why-it-bought-the-weather-company-2016-06-15

Borgatti, S. P., & Halgin, D. S. (2011). On Network Theory. *Organization Science, 22*(5), 1168–1181. https://doi.org/10.1287/orsc.1100.0641

Boudreau, K. J. (2017). Platform Boundary Choices & Governance: Opening-Up While Still Coordinating and Orchestrating. In *Advances in Strategic Management: Vol. 37. Entrepreneurship, Innovation, and Platforms* (Vol. 37, pp. 227–297). https://doi.org/10.1108/S0742-332220170000037009

Boudreau, M.-C., Serrano, C., & Larson, K. (2014). IT-driven identity work: Creating a group identity in a digital environment. *Information and Organization, 24*(1), 1–24. https://doi.org/10.1016/j.infoandorg.2013.11.001

Boulton, G., Rawlins, M., Vallance, P., & Walport, M. (2011). Science as a public enterprise: the case for open data. *The Lancet, 377*(9778), 1633–1635. https://doi.org/10.1016/S0140-6736(11)60647-8

Boychuk, M., Cousins, M., Lloyd, A., & MacKeigan, C. (2016). Do We Need Data Literacy? Public Perceptions Regarding Canada's Open Data Initiative. *Dalhousie Journal of Interdisciplinary Management, 12*. https://doi.org/10.5931/djim.v12.i1.6449

Brett Hurt. (2016). Retrieved from https://soundcloud.com/tex-talks/brett-hurt

Broderick, F. (2018, January 23). 5 ways cities can use data to drive outcomes for citizens. Retrieved January 30, 2018, from WEFLive website: https://www.weflive.com/?utm_campaign=carto%205&utm_source=hs_email&utm_medium=email&utm_content=60287771&_hsenc=p2ANqtz-9q095IYw0FQKJhu4FeWQUBP54Zf8FFNuGrTqfL8EAA01O96DAIm1sKqoyuw-6PgICbQM1G-FmgUKeqjMrSjdD17V9vCQ&_hsmi=60288330#!/story/e0b6caf0003611e8befbd721f80de8f5

Brown, J. S., & Duguid, P. (1991). Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation. *Organization Science*, *2*(1), 40–57. Retrieved from JSTOR.

Brown, J. S., & Duguid, P. (2000). *The Social Life of Information*. Harvard Business Press.

Bryman, A. (2001). *Social research methods*. Oxford ; New York: Oxford University Press.

Brynjolfsson, E., Geva, T., & Reichman, S. (2016). Crowd-squared: amplifying the predictive power of search trend data. *MIS Quarterly*, *40*(4), 941–961.

Bughin, J., Chui, M., & Manyika, J. (2010). Clouds, big data, and smart assets: Ten tech-enabled business trends to watch. *McKinsey Quarterly*, *56*(4), 26–43.

Buntain, C., & Golbeck, J. (2014). Identifying Social Roles in Reddit Using Network Structure. *Proceedings of the 23rd International Conference on World Wide Web*, 615–620. https://doi.org/10.1145/2567948.2579231

Bur, J. (2018, January 26). White House nominates federal CIO. Retrieved January 29, 2018, from Federal Times website: https://www.federaltimes.com/management/leadership/2018/01/26/white-house-nominates-federal-cio/

Byron, K., & Laurence, G. A. (2015). Diplomas, Photos, and Tchotchkes as Symbolic Self-Representations: Understanding Employees' Individual Use of Symbols. *Academy of Management Journal*, *58*(1), 298–323. https://doi.org/10.5465/amj.2012.0932

Caillaud, B., & Jullien, B. (2003). Chicken & Egg: Competition among Intermediation Service Providers. *The RAND Journal of Economics*, *34*(2), 309–328. https://doi.org/10.2307/1593720

Calzolari, N., Del Gratta, R., Francopoulo, G., Mariani, J., Rubino, F., Russo, I., & Soria, C. (2012). *The LRE Map. Harmonising Community Descriptions of Resources* (N. Calzolari, K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, … S. Piperidis, Eds.). Paris: European Language Resources Assoc-Elra.

Cannon-Albright, L. A., Dintelman, S., Maness, T., Backus, S., Thomas, A., & Meyer, L. J. (2013). Creation of a national resource with linked genealogy and phenotypic data: the Veterans Genealogy Project. *Genetics in Medicine*, *15*(7), 541–547. https://doi.org/10.1038/gim.2012.171

Cao, L. (2017). Data Science: A Comprehensive Overview. *ACM Computing Surveys*, *50*(3), 43. https://doi.org/10.1145/3076253

Carter, L., & Bélanger, F. (2005). The utilization of e-government services: citizen trust, innovation and acceptance factors*. *Information Systems Journal*, *15*(1), 5–25. https://doi.org/10.1111/j.1365-2575.2005.00183.x

Cave, A. (2017, April 13). What Will We Do When The World's Data Hits 163 Zettabytes In 2025? Retrieved January 31, 2018, from Forbes website: https://www.forbes.com/sites/andrewcave/2017/04/13/what-will-we-do-when-the-worlds-data-hits-163-zettabytes-in-2025/

Ceccagnoli, M., Forman, C., Huang, P., & Wu, D. J. (2012). Cocreation of Value in a Platform Ecosystem: The Case of Enterprise Software. *MIS Quarterly*, *36*(1), 263–290. https://doi.org/10.2307/41410417

Centre for Science and Technology Studies (CWTS) | Leiden University | Elsevier. (2017). *Open Data The Researcher Perspective*. Retrieved from https://www.elsevier.com/__data/assets/pdf_file/0004/281920/Open-data-report.pdf

Chahal, M. (2014). Taking back control: the personal data economy. *Marketing Week*, 9–9.

Chamberlin, B. (2016). *The Data Economy*. Retrieved from IBM website: http://www.slideshare.net/horizonwatching

Chan, C. (2013). From Open Data to Open Innovation Strategies: Creating E-Services Using Open Government Data. *Proceedings of the 2013 46th Hawaii International Conference on System Sciences*, 10. https://doi.org/10.1109/HICSS.2013.236

Chan, C. M. L., Lau, Y., & Pan, S. L. (2008). E-government implementation: A macro analysis of Singapore's e-government initiatives. *Government Information Quarterly*, *25*(2), 239–255. https://doi.org/10.1016/j.giq.2006.04.011

Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, *63*, 67–80. https://doi.org/10.1016/j.dss.2013.08.008

Chappellet-Lanier, T. (2019, January 11). Data.gov is offline because of the shutdown. Retrieved March 19, 2019, from FedScoop website: https://www.fedscoop.com/data-gov-open-data-offline-shutdown/

Chatfield, A. T., & Reddick, C. G. (2017). A longitudinal cross-sector analysis of open data portal service capability: The case of Australian local governments. *Government Information Quarterly*, *34*(2), 231–243. https://doi.org/10.1016/j.giq.2017.02.004

Chattapadhyay, S. (2013). Towards an expanded and integrated open government data agenda for India. *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance*, 202–205. https://doi.org/10.1145/2591888.2591923

Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, *54*(8), 88–98.

Chen, A. (2018, January 26). The Google Arts & Culture App and the Rise of the "Coded Gaze." *The New Yorker*. Retrieved from https://www.newyorker.com/tech/elements/the-google-arts-and-culture-app-and-the-rise-of-the-coded-gaze-doppelganger

Chen, H., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, *36*(4), 1165–1188.

Chen, L., Ma, X., Nguyen, T.-M.-T., Pan, G., & Jakubowicz, J. (2017). Understanding bike trip patterns leveraging bike sharing system open data. *Frontiers of Computer Science*, *11*(1), 38–48. https://doi.org/10.1007/s11704-016-6006-4

Chen, Y., Narasimhan, C., & Zhang, Z. J. (2001). Individual Marketing with Imperfect Targetability. *Marketing Science*, *20*(1), 23–41. https://doi.org/10.1287/mksc.20.1.23.10201

Chiang, R. H. L., Grover, V., Liang, Z., & Zhang, D. (2018). Special Issue: Strategic Value of Big Data and Business Analytics. *Journal of Management Information Systems*, *35*(2), 383–387. https://doi.org/10.1080/07421222.2018.1451950

Chui, M., Manyika, J., & Kuiken, S. V. (2014). What executives should know about open data. *McKinsey Quarterly, January, 2014*.

Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2016). GenBank. *Nucleic Acids Research*, *44*(D1), D67–D72. https://doi.org/10.1093/nar/gkv1276

CNBC.com, C. M., Special to. (2013a, July 23). Debbie does data: Porn gets statistical. Retrieved October 3, 2017, from CNBC website: https://www.cnbc.com/id/100904399

CNBC.com, C. M., Special to. (2013b, October 9). 10 strangest data findings you need to know. Retrieved October 3, 2017, from CNBC website: https://www.cnbc.com/2013/10/09/0-strangest-findings.html

Coddington, M. (2015). Clarifying Journalism's Quantitative Turn. *Digital Journalism*, *3*(3), 331–348. https://doi.org/10.1080/21670811.2014.976400

Conboy, K., Feller, J., Leimeister, J. M., Morgan, L., & Schlagwein, D. (2015). Openness and Information Technology. *Journal of Information Technology*. Retrieved from https://pdfs.semanticscholar.org/9a77/63c3ce2bffd06f151b639d7469992a595523.pdf

Connelly, C. E., Zweig, D., Webster, J., & Trougakos, J. P. (2012). Knowledge hiding in organizations. *Journal of Organizational Behavior*, *33*(1), 64–88. https://doi.org/10.1002/job.737

Conrad, C. C., & Hilchey, K. G. (2011). A review of citizen science and community-based environmental monitoring: issues and opportunities. *Environmental Monitoring and Assessment; Dordrecht*, *176*(1–4), 273–291. http://dx.doi.org/10.1007/s10661-010-1582-5

Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly*, *31*(Supplement 1), S10–S17. https://doi.org/10.1016/j.giq.2014.01.003

Constantinides, P., Henfridsson, O., & Parker, G. G. (2018). Introduction—Platforms and Infrastructures in the Digital Age. *Information Systems Research*. https://doi.org/10.1287/isre.2018.0794

Constantiou, I. D., & Kallinikos, J. (2015). New games, new rules: big data and the changing context of strategy. *Journal of Information Technology*, *30*(1), 44–57. https://doi.org/10.1057/jit.2014.17

Cooper, B. L., Watson, H. J., Wixom, B. H., & Goodhue, D. L. (2000). Data Warehousing Supports Corporate Strategy at First American Corporation. *MIS Quarterly*, *24*(4), 547–567. https://doi.org/10.2307/3250947

Corley, K. G., & Gioia, D. A. (2004). Identity Ambiguity and Change in the Wake of a Corporate Spin-off. *Administrative Science Quarterly*, *49*(2), 173–208. https://doi.org/10.2307/4131471

Correndo, G., & Alani, H. (2007). Survey of Tools for Collaborative Knowledge Construction and Sharing. *Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops*, 7–10. Retrieved from http://dl.acm.org/citation.cfm?id=1339264.1339647

Corsar, D., & Edwards, P. (2017). Challenges of Open Data Quality: More Than Just License, Format, and Customer Support. *ACM Journal of Data and Information Quality*, *9*(1), UNSP 3. https://doi.org/10.1145/3110291

Cortada, J. (1998). *Rise of the Knowledge Worker*. https://doi.org/10.1016/B978-0-7506-7058-6.50012-0

Council, N. D. (2015, January 5). National Development Council [Website]. Retrieved November 28, 2017, from 國發會全球資訊網(英文網) website: https://www.ndc.gov.tw/en/News_Content.aspx?n=8C362E80B990A55C&sms=1DB6C6A8871CA043&s=09819AC7E099BCD5

Crabtree, A., Lodge, T., Colley, J., Greenhalgh, C., Mortier, R., & Haddadi, H. (2016). Enabling the new economic actor: data protection, the digital economy, and the Databox. *Personal and Ubiquitous Computing*, *20*(6), 947–957. https://doi.org/10.1007/s00779-016-0939-3

Crowdbabble. (2017). Professional Social Media Analytics. Retrieved November 21, 2017, from Crowdbabble website: https://www.crowdbabble.com/

Cyranoski, D. (2017). China cracks down on fake data in drug trials. *Nature News*, *545*(7654), 275. https://doi.org/10.1038/nature.2017.21977

Dahlander, L., & Gann, D. M. (2010). How open is innovation? *Research Policy*, *39*(6), 699–709. https://doi.org/10.1016/j.respol.2010.01.013

Dai, J., & Li, Q. (2016). Designing Audit Apps for Armchair Auditors to Analyze Government Procurement Contracts. *Journal of Emerging Technologies in Accounting*, *13*(2), 71–88. https://doi.org/10.2308/jeta-51598

Daizadeh, I. (2006). An example of information management in biology: Qualitative data economizing theory applied to the human genome project databases. *Journal of the American Society for Information Science and Technology*, *57*(2), 244–250. https://doi.org/10.1002/asi.20270

Danneels, L., Viaene, S., & Van den Bergh, J. (2017). Open data platforms: Discussing alternative knowledge epistemologies. *Government Information Quarterly*, *34*(3), 365–378. https://doi.org/10.1016/j.giq.2017.08.007

Darzentas, D. P., Brown, M. A., Flintham, M., & Benford, S. (2015). The Data Driven Lives of Wargaming Miniatures. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 2427–2436. https://doi.org/10.1145/2702123.2702377

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. *Reuters*. Retrieved from https://www.reuters.com/article/amazoncom-jobs-automation/rpt-insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSL2N1WP1RO

DATA Act. (2016, July 19). Retrieved November 19, 2016, from Data Coalition website: http://www.datacoalition.org/issues/data-act/

databox.com. (2017, September 20). The Must-Read Guide To Data Journalism | Databox Blog. Retrieved February 7, 2018, from Databox website: https://databox.com/must-read-guide-data-journalism

DataCoalition.org. (2018). Data Coalition. Retrieved January 29, 2018, from Data Coalition website: https://www.datacoalition.org/

Data-Economy.com. (2017). Data Economy. Retrieved October 3, 2017, from Data Economy website: https://data-economy.com/

datafordemocracy.or. (2018). Data for Democracy. Retrieved February 16, 2018, from http://datafordemocracy.org/

Data.gov. (2018). Data.gov. Retrieved February 16, 2018, from Data.gov website: https://www.data.gov/

Data.gov. (2019). Retrieved March 19, 2019, from Data.gov website: https://www.data.gov/

Data.gov.uk. (2018). data.gov.uk. Retrieved January 17, 2018, from Data.gov.uk website: https://data.gov.uk/

datajournos.com. (2018). Data Journalism DC. Retrieved February 7, 2018, from Data Journalists website: http://www.datajournos.com/about1.html

Data Solutions | AP. (2019). Retrieved April 27, 2019, from Associated Press website: https://www.ap.org/en-us/formats/data

data.world. (2018a). Home | data.world. Retrieved February 16, 2018, from https://data.world/

data.world. (2018b). How the Associated Press puts data journalism in the heart of local newsrooms. *Data.World Case Study*, 18.

data.world. (2018c, March 6). data.world Launches Productive, Secure Platform for Modern Data Teamwork. Retrieved March 18, 2019, from GlobeNewswire News Room website: http://www.globenewswire.com/news-release/2018/03/06/1415905/0/en/data-world-Launches-Productive-Secure-Platform-for-Modern-Data-Teamwork.html

Davenport, T., Barth, P., & Bean, R. (2012). How 'Big Data' Is Different. *MIT Sloan Management Review*, (Fall 2012), 5.

Davenport, T. H., & Patil, D. J. (2012). The Sexiest Job of the 21st Century. *Harvard Business Review*, 9.

Dawes, S. S., Vidiasova, L., & Parkhimovich, O. (2016). Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly*, *33*(1), 15–27. https://doi.org/10.1016/j.giq.2016.01.003

Day, J., Junglas, I., University of Houston - Clear Lake, Silva, L., & University of Houston. (2009). Information Flow Impediments in Disaster Relief Supply Chains. *Journal of the Association for Information Systems*, *10*(8), 637–660. https://doi.org/10.17705/1jais.00205

de Boer, V., Wielemaker, J., van Gent, J., Oosterbroek, M., Hildebrand, M., Isaac, A., ... Schreiber, G. (2013). Amsterdam Museum Linked Open Data. *Semantic Web*, *4*(3), 237–243. https://doi.org/10.3233/SW-2012-0074

de Kool, D., & Bekkers, V. (2016). The perceived value-relevance of open data in the parents' choice of Dutch primary schools. *International Journal of Public Sector Management*, *29*(3), 271–287. https://doi.org/10.1108/IJPSM-02-2016-0022

Degli Esposti, S. (2014). When big data meets dataveillance: The hidden side of analytics. *Surveillance & Society*, *12*(2), 209.

Delina, R. (2016). *Digital Single Market Conceptual Model for Intelligent Supply Chains* (Vol. 45; D. Petr, C. Gerhard, & O. Vaclav, Eds.). Linz: Trauner Verlag.

Delina, R., & Sukker, A. A. M. (2015). *The Significance of Transparency in Electronic Procurement* (Vol. 44; D. Petr, C. Gerhard, & O. Vaclav, Eds.). Linz: Universitatsverlag Rudolf Trauner.

Deloitte Canada. (2015). *Driving economic and social growth: Designing the right open data strategy for public sector organizations* (p. 14). Retrieved from Deloitte website: https://www2.deloitte.com/content/dam/Deloitte/ca/Documents/public-sector/ca-en-ps-driving-economic-social-growth.PDF

Deloitte UK. (2012). *Open data Driving growth, ingenuity, and innovation* (p. 36). Retrieved from Deloitte LLP website: https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-data-driving-growth-ingenuity-and-innovation.pdf

DeLone, W. H., & McLean, E. R. (1992). Information Systems Success: The Quest for the Dependent Variable. *Information Systems Research*, *3*(1), 60–95. https://doi.org/10.1287/isre.3.1.60

Delone, W. H., & McLean, E. R. (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. *Journal of Management Information Systems*, *19*(4), 9–30. https://doi.org/10.1080/07421222.2003.11045748

Denis, J., & Goëta, S. (2017). Rawification and the careful generation of open government data. *Social Studies of Science (Sage Publications, Ltd.)*, *47*(5), 604–629. https://doi.org/10.1177/0306312717712473

Destremau, B., Giacaman, R., Harb, M., Herrera, L., Mexi, M., Mitwalli, S., ... Zambelli, E. (2017). The dilemmas of the European Union's open access to data policy. *The Lancet*, *390*(10090), 122–123. https://doi.org/10.1016/S0140-6736(17)31458-7 digital-activism.org. (2017). Digital Activism Research Project | investigating the global impact of digital media on political contention. Retrieved September 20, 2017, from digital-activism.org website: http://digital-activism.org/

DiMarco, M. (2016). *Data Marketplaces: A Panel with data.world and Infochimps*. Retrieved from https://vimeo.com/184276626

Diphoko, W. (2017, September). Opinion: It's time for African 'data economy' to rise | IOL Business Report. Retrieved December 5, 2017, from Independent Online Business Report website: https://www.iol.co.za/business-report/opinion-its-time-for-african-data-economy-to-rise-11317371

D'Onfro, J. (2014). Here's Exactly Why Pinterest Is Worth Its $5 Billion Valuation. Retrieved December 5, 2017, from Business Insider website: http://www.businessinsider.com/why-pinterest-is-worth-5-billion-2014-5

Dong, S., Xu, S. X., & Zhu, K. X. (2009). Research Note—Information Technology in Supply Chains: The Value of IT-Enabled Resources Under Competition. *Information Systems Research, 20*(1), 18–32. https://doi.org/10.1287/isre.1080.0195

Donovan, E., Guzman, I. R., Adya, M., & Wang, W. (2018). A CLOUD UPDATE OF THE DELONE AND MCLEAN MODEL OF INFORMATION SYSTEMS SUCCESS. (3), 13.

Drenik, G. (2017, February). The Emerging Challenge Of "Fake Data" [Forbes.com]. Retrieved September 20, 2017, from Forbes website: https://www.forbes.com/sites/forbesinsights/2017/02/22/the-emerging-challenge-of-fake-data/

Drucker, P. F. (1988). The Coming of the New Organization. *Harvard Business Review*. Retrieved from https://hbr.org/1988/01/the-coming-of-the-new-organization

Du, A. Y., Geng, X., Gopal, R., Ramesh, R., & Whinston, A. B. (2008). Capacity Provision Networks: Foundations of Markets for Sharable Resources in Distributed Computational Economies. *Information Systems Research*, *19*(2), 144–160. https://doi.org/10.1287/isre.1070.0145

Dubé, L., & Paré, G. (2003). Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations. *MIS Quarterly*, *27*(4), 597–636. https://doi.org/10.2307/30036550

eBird.org. (2018). eBird - Discover a new world of birding... Retrieved February 16, 2018, from https://ebird.org/home

Edwards, P. K., O'Mahoney, J., & Vincent, S. (2014). *Studying Organizations Using Critical Realism: A Practical Guide*. Oxford, New York: Oxford University Press.

Eisenmann, T., Parker, G., & Van Alstyne, M. (2011). Platform envelopment. *Strategic Management Journal, 32*(12), 1270–1285. https://doi.org/10.1002/smj.935

Eisenmann, T. R., Parker, G., & Alstyne, M. V. (2009). Opening Platforms: How, When and Why? In *Chapters* (p. 29). Retrieved from https://ideas.repec.org/h/elg/eechap/13257_6.html

Elbaz, G. (2012, September 30). Data Markets: The Emerging Data Economy. Retrieved October 3, 2017, from TechCrunch website: http://social.techcrunch.com/2012/09/30/data-markets-the-emerging-data-economy/

Elsbach, K. D., & Kramer, R. M. (2015). Handbook of Qualitative Organizational Research: Innovative Pathways and Methods. Routledge.

Elvy, S.-A. (2017). Paying for and the Personal Data Economy. *Columbia Law Review, 117*(6), 1369–1459.

Emaldi, M., Pena, O., Lazaro, J., & Lopez-de-Ipina, D. (2015). Linked Open Data as the Fuel for Smarter Cities. In F. Xhafa, L. Barolli, A. Barolli, & P. Papajorgji (Eds.), *Modeling and Processing for Next-Generation Big-Data Technologies: With Applications and Case Studies* (Vol. 4, pp. 443–472). Cham: Springer Int Publishing Ag.

Energy.com. (2013). Who Uses Open Data? Retrieved February 3, 2018, from Open Energy Data website: https://energy.gov/data/articles/who-uses-open-data

European Commission. (2014). *Towards a thriving data-driven economy* (p. 12). Brussels: European Commission.

European Commission. (2017a). *Building the European Data Economy* (p. 2). Retrieved from https://ec.europa.eu/digital-single-market/en/news/building-european-data-economy

European Commission. (2017b). Enter the Data Economy: EU Policies for a Thriving Data Ecosystem. *EPSC Strategic Notes*, (21), 16. https://doi.org/doi: 10.2872/33746

European Commission. (2017c). *Final results of the European Data Market study measuring the size and trends of the EU data economy*. Retrieved from European Commission website: https://ec.europa.eu/digital-single-market/en/news/final-results-european-data-market-study-measuring-size-and-trends-eu-data-economy

European Commission. (2017d). Statistics [Text]. Retrieved October 3, 2017, from European Commission - European Commission website: https://ec.europa.eu/info/statistics_en

European Commission. (2017e, September). Elements of the European data economy strategy. Retrieved December 4, 2017, from Digital Single Market website: https://ec.europa.eu/digital-single-market/en/towards-thriving-data-driven-economy

European Data Portal. (2017). Open Data Maturity in Europe - European Data Portal. Retrieved February 10, 2018, from European Data Portal website: /en/highlights/open-data-maturity-europe

Evans, D. S. (2009). How Catalysts Ignite: The Economics of Platform-Based Start-Ups. In A. Gawer, *Platforms, Markets and Innovation*. https://doi.org/10.4337/9781849803311.00011

Fan, B., & Zhao, Y. (2017). The moderating effect of external pressure on the relationship between internal organizational factors and the quality of open government data. *Government Information Quarterly*, *34*(3), 396–405. https://doi.org/10.1016/j.giq.2017.08.006

Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. (2011). Knowledge Collaboration in Online Communities. *Organization Science*, *22*(5), 1224–1239. https://doi.org/10.1287/orsc.1100.0614

Faraj, S., Kudaravalli, S., & Wasko, M. (2015). Leading Collaboration in Online Communities. *MIS Quarterly*, *39*(2), 393–412. https://doi.org/10.25300/MISQ/2015/39.2.06

Farhadloo, M., Patterson, R. A., & Rolland, E. (2016). Modeling customer satisfaction from unstructured data using a Bayesian approach. *Decision Support Systems*, *90*, 1–11. https://doi.org/10.1016/j.dss.2016.06.010

Fayard, A.-L., Gkeredakis, E., & Levina, N. (2016). Framing Innovation Opportunities While Staying Committed to an Organizational Epistemic Stance. *Information Systems Research*, *27*(2), 302–323. https://doi.org/10.1287/isre.2016.0623

Federal Depository Libraries. (2017). Retrieved November 28, 2017, from https://www.fdlp.gov/about-the-fdlp/federal-depository-libraries

Federal Depository Library Program. (2012). Federal Depository Library Program. Retrieved November 28, 2017, from Federal Depository Library Program website: https://www.fdlp.gov/

Federal Depository Library Program. (2017). Federal Depository Library Program (FDLP) | Association of Research Libraries® | ARL®. Retrieved November 28, 2017, from http://www.arl.org/focus-areas/public-access-policies/federal-depository-library-program

Federal Trade Commission. (2014a). *Data Brokers: A Call for Transparency and Accountability* (p. 110). Retrieved from Federal Trade Commission website: https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf

Federal Trade Commission. (2014b, May 27). FTC Recommends Congress Require the Data Broker Industry to be More Transparent and Give Consumers Greater Control Over Their Personal Information. Retrieved January 15, 2018, from FTC Recommends Congress Require the Data Broker Industry to be More Transparent and Give Consumers Greater Control Over Their Personal Information website: https://www.ftc.gov/news-events/press-releases/2014/05/ftc-recommends-congress-require-data-broker-industry-be-more

Federal Trade Commission. (2014c, December 24). FTC Charges Data Broker with Facilitating the Theft of Millions of Dollars from Consumers' Accounts. Retrieved January 15, 2018, from Federal Trade Commission website: https://www.ftc.gov/news-events/press-releases/2014/12/ftc-charges-data-broker-facilitating-theft-millions-dollars

Federal Trade Commission. (2016, February 18). Data Broker Defendants Settle FTC Charges They Sold Sensitive Personal Information to Scammers. Retrieved January 15, 2018, from Federal Trade Commission website: https://www.ftc.gov/news-events/press-releases/2016/02/data-broker-defendants-settle-ftc-charges-they-sold-sensitive

Feldman, J., & Tung, R. (2001). Whole School Reform: How Schools Use the Data-Based Inquiry and Decision Making Process. *Proceedings of the Annual Meeting of the American Educational Research Association*, 26. Retrieved from https://eric.ed.gov/?id=ED454242

Feller, J., Finnegan, P., & Nilsson, O. (2011). Open innovation and public administration: transformational typologies and business model impacts. *European Journal of Information Systems*, *20*(3), 358–374. https://doi.org/10.1057/ejis.2010.65

Fenichel, E. P., & Skelly, D. K. (2015). Why Should Data Be Free; Don't You Get What You Pay For? *BioScience*, *65*(6), 541. https://doi.org/10.1093/biosci/biv052

Fensel, A., Tomic, D. K., & Koller, A. (2017). Contributing to appliances' energy efficiency with Internet of Things, smart data and user engagement. *Future Generation Computer Systems-the International Journal of Escience*, *76*, 329–338. https://doi.org/10.1016/j.future.2016.11.026

Ferguson, N. (2008). The Ascent of Money: A Financial History of the World. Penguin.

Fernando, P. (2017, January). Why NGO's, Charities and Third sector organisations need to wake up and embrace open data. Retrieved January 17, 2018, from Liquid Light website: https://www.liquidlight.co.uk/blog/article/why-ngos-charities-and-third-sector-organisations-need-to-wake-up-and-embrace-open-data/

Ferro, E., & Osella, M. (2013). Eight Business Model Archetypes for PSI Re-Use. *ResearchGate*. Presented at the Open Data on the Web, London. Retrieved from https://www.researchgate.net/publication/265684074_Eight_Business_Model_Archetypes_for_PSI_Re-Use

Fisher, C. W., Chengalur-Smith, I., & Ballou, D. P. (2003). The Impact of Experience and Time on the Use of Data Quality Information in Decision Making. *Information Systems Research*, *14*(2), 170–188. https://doi.org/10.1287/isre.14.2.170.16017

Floridi, L. (2009). Web 2.0 vs. the Semantic Web: A Philosophical Assessment. *Episteme*, *6*(1), 25–37. https://doi.org/10.3366/E174236000800052X

FOIA.gov. (2017a). Freedom of Information Act. Retrieved November 28, 2017, from FOIA.gov - Freedom of Information Act website: https://www.foia.gov/faq.html

FOIA.gov. (2017b). Freedom of Information Act. Retrieved February 16, 2018, from https://www.foia.gov/

FOIA.gov. (2018). Freedom of Information Act. Retrieved January 16, 2018, from United States Department of Justic website: https://www.foia.gov/

Folger, R., & Konovsky, M. A. (1989). Effects of Procedural and Distributive Justice on Reactions to Pay Raise Decisions. *Academy of Management Journal*, *32*(1), 115–130. https://doi.org/10.2307/256422

Food and Agriculture Organization of the United Nations. (2017). Carbon footprint of the banana supply chain. Retrieved January 30, 2018, from World Banana Forum website: http://www.fao.org/world-banana-forum/projects/good-practices/carbon-footprint/en/

Food and Agriculture Organization of the United Nations. (2018). Carbon footprint of the banana supply chain | World Banana Forum |. Retrieved January 30, 2018, from Food and Agriculture Organization of the United Nations website: http://www.fao.org/world-banana-forum/projects/good-practices/carbon-footprint/en/

Fox, L. J., & Ockene, K. D. (2017, February 6). *2017-02-06-Notice-of-Violation_HSUS-v-USDA-No.-15-0197.pdf*. Retrieved from http://blog.humanesociety.org/wp-content/uploads/2017/02/2017-02-06-Notice-of-Violation_HSUS-v-USDA-No.-15-0197.pdf

Freeze, R. D., & Kulkarni, U. (2007). Knowledge management capability: defining knowledge assets. *Journal of Knowledge Management*, *11*(6), 94–109. https://doi.org/10.1108/13673270710832190

French, K. (2018). The DATA Act Section 5 Pilot Program—A 6th Test Model? Retrieved January 29, 2018, from https://www.streamlinksoftware.com/amplifund/blog/the-data-act-section-5-pilot-program-a-6th-test-model

Friedman, M. (1994). Money Mischief: Episodes in Monetary History. HMH.

Fung, A. (2013). Infotopia: Unleashing the Democratic Power of Transparency. *Politics & Society*, *41*(2), 183–212. https://doi.org/10.1177/0032329213483107

Fürber, C. (2016). Data Quality in the Semantic Web. In C. Fürber (Ed.), *Data Quality Management with Semantic Technologies* (pp. 69–77). https://doi.org/10.1007/978-3-658-12225-6_5

Furman, J., Gawer, A., Silverman, B. S., & Stern, S. (2017). *Entrepreneurship, Innovation, and Platforms*. Emerald Group Publishing.

Gangadharan, S. P., Eubanks, V., & Barocas, S. (2014). *Data and discrimination: Collected essays*. Retrieved from https://rws511.pbworks.com/w/file/fetch/88176947/OTI-Data-an-Discrimination-FINAL-small.pdf

Garg, R., & Telang, R. (2013). Inferring App Demand from Publicly Available Data. *MIS Quarterly*, *37*(4), 1253–1264.

Garshol, L. M. (2004). Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all. *Journal of Information Science*, *30*(4), 378–391. https://doi.org/10.1177/0165551504045856

Gartner. (2016). Dark Data - Gartner IT Glossary. Retrieved March 19, 2018, from Gartner IT Glossary website: https://www.gartner.com/it-glossary/dark-data

Garud, R., & Karnøe, P. (2003). Bricolage versus breakthrough: distributed and embedded agency in technology entrepreneurship. *Research Policy*, *32*(2), 277–300. https://doi.org/10.1016/S0048-7333(02)00100-2

Gawer, A. (2011). *Platforms, Markets and Innovation*. Edward Elgar Publishing.

Gellman, R. (1996). Disintermediation and the Internet. *Government Information Quarterly*, *13*(1), 1–8.

George, J., & Leidner, D. (2018). Digital Activism: a Hierarchy of Political Commitment. *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS)*. Presented at the 51st Hawaii International Conference on System Sciences (HICSS), Waikoloa, Hawaii. Retrieved from http://scholarspace.manoa.hawaii.edu/handle/10125/50176

Gerunov, A. (2017). Understanding Open Data Policy: Evidence from Bulgaria. *International Journal of Public Administration*, *40*(8), 649–657. https://doi.org/10.1080/01900692.2016.1186178

Geva, T., Oestreicher-Singer, G., Efron, N., & Shimshoni, Y. (2017). Using Forum and Search Data for Sales Prediction of High-Involvement Products. *MIS Quarterly*, *41*(1), 65–82.

Ghasemaghaei, M., Hassanein, K., & Turel, O. (2017). Increasing firm agility through the use of data analytics: The role of fit. *Decision Support Systems*, *101*, 95–105. https://doi.org/10.1016/j.dss.2017.06.004

Ghazawneh, A., & Henfridsson, O. (2013). Balancing platform control and external contribution in third-party development: the boundary resources model: Control and contribution in third-party development. *Information Systems Journal*, *23*(2), 173–192. https://doi.org/10.1111/j.1365-2575.2012.00406.x

Giddens, L., Gonzalez, E., & Leidner, D. (2016). I Track, Therefore I Am: Exploring the Impact of Wearable Fitness Devices on Employee Identity and Well-being. *AMCIS 2016 Proceedings*. Retrieved from https://aisel.aisnet.org/amcis2016/ITProj/Presentations/13

Gilbert, J. C. (2017, August 23). Is Data The New Oil? How One Startup Is Rescuing The World's Most Valuable Asset. Retrieved January 31, 2018, from Forbes Social Entrepreneurs website: https://www.forbes.com/sites/jaycoengilbert/2017/08/23/rescuing-the-worlds-most-valuable-stranded-asset-the-company-democratizing-data-the-new-oil/

Gkoulalas-Divanis, A., & Mac Aonghusa, P. (2014). Privacy protection in open information management platforms. *IBM Journal of Research and Development*, *58*(1), 2. https://doi.org/10.1147/JRD.2013.2285853

Glaser, B. G., & Holton, J. (2004). Remodeling Grounded Theory. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *5*(2). https://doi.org/10.17169/fqs-5.2.607

Glaser, B. G., & Strauss, A. L. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine.

Global Open Data Index. (2017). Retrieved February 9, 2018, from https://index.okfn.org/

Global Open Data Index. (2018). *Global Open Data Index*. Retrieved from Global Open Data Index website: https://index.okfn.org/place/

Globe Newswire. (2017, August 2). data.world Survey Reveals Consumer Attitudes Around Data and Trust in Online News. *GlobeNewswire News Room*. Retrieved from http://globenewswire.com/news-release/2017/08/02/1071114/0/en/data-world-Survey-Reveals-Consumer-Attitudes-Around-Data-and-Trust-in-Online-News.html

Globe Newswire. (2019, February 21). Mirum partners with data.world to deliver data-driven design thinking to client engagements. Retrieved March 20, 2019, from GlobeNewswire News Room website: http://www.globenewswire.com/news-release/2019/02/21/1739954/0/en/Mirum-partners-with-data-world-to-deliver-data-driven-design-thinking-to-client-engagements.html

Gómez-Pérez, A., & González-Cabero, R. (2004). A Framework for Design and Composition of Semantic Web Services. *American Association for Artificial Intelligence*, 8.

Gomez-Perez, A., Gonzalez-Cabero, R., & Lama, M. (2004). ODE SWS: a framework for designing and composing semantic Web services. *IEEE Intelligent Systems*, *19*(4), 24–31. https://doi.org/10.1109/MIS.2004.32

Gonzalez-Zapata, F., & Heeks, R. (2015). The multiple meanings of open government data: Understanding different stakeholders and their perspectives. *Government Information Quarterly*, *32*(4), 441–452. https://doi.org/10.1016/j.giq.2015.09.001

Goode, S., Hoehle, H., Venkatesh, V., & Brown, S. A. (2017). User Compensation as a Data Breach Recovery Action: An Investigation of The Sony Playstation Network Breach. *MIS Quarterly*, *41*(3). Retrieved from http://www.vvenkatesh.com/wp-content/uploads/dlm_uploads/2016/05/Goode-et-al.-MISQ-2017_Upload-13.pdf

Google Trends. (2018). Retrieved January 17, 2018, from Google Trends website: https://g.co/trends/FDVVW

Government, O. (2013, May 17). Open Data 101. Retrieved November 19, 2016, from http://open.canada.ca/en/open-data-principles

Graham, E. A., Henderson, S., & Schloss, A. (2011). Using mobile phones to engage citizen scientists in research. *EOS*, *92*(38), 313–315. https://doi.org/10.1029/2011EO380002

Granados, N., Gupta, A., & Kauffman, R. J. (2010). Research Commentary—Information Transparency in Business-to-Consumer Markets: Concepts, Framework, and Research Agenda. *Information Systems Research*, *21*(2), 207–226. https://doi.org/10.1287/isre.1090.0249

Grant, R. M. (1996). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, *17*(S2), 109–122.

Gray, J., Chambers, L., & Bounegru, L. (2012). The Data Journalism Handbook: How Journalists Can Use Data to Improve the News. O'Reilly Media, Inc.

Greenberg, P. (2018, September 13). What to do with the data? The evolution of data platforms in a post big data world. Retrieved September 20, 2018, from Social CRM: The Conversation website: https://www.zdnet.com/article/evolution-of-data-platforms-post-big-data/

Gregg, M. (2015). Inside the data spectacle. *Television & New Media*, *16*(1), 37–51. https://doi.org/10.1177/1527476414547774

Gressin, S. (2017, September 8). The Equifax Data Breach: What to Do. Retrieved September 20, 2017, from Federal Trade Commission Consumer Information website: https://www.consumer.ftc.gov/blog/2017/09/equifax-data-breach-what-do

Griffith, E. (2019, April 19). Pinterest Prices I.P.O. at $19 a Share, for a $12.7 Billion Valuation. The New York Times. Retrieved from https://www.nytimes.com/2019/04/17/technology/pinterest-ipo-stock.html

Gross, A. (2011). The economy of social data: exploring research ethics as device. *The Sociological Review*, *59*, 113–129. https://doi.org/10.1111/j.1467-954X.2012.02055.x

Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. (2017, March 21). Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

Gura, T. (2013). Citizen science: Amateur experts. *Nature*, *496*(7444), 259–261. https://doi.org/10.1038/nj7444-259a

Haddadi, H., Mortier, R., McAuley, D., & Crowcroft, J. (2013). *Human-data interaction* (No. UCAM-CL-TR-837; p. 9). Retrieved from University of Cambridge, Computer Laboratory website: http://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-837.html

Hafen, E. (2015). MIDATA cooperatives - citizen-controlled use of health data is a pre-requiste for big data analysis, economic success and a democratization of the personal data economy. *Tropical Medicine & International Health*, *20*, 129–129.

Haghighat, M., Rastegari, H., & Nourafza, N. (2013). A review of data mining techniques for result prediction in sports. *Advances in Computer Science: An International Journal*, *2*(5), 7–12.

Hand, M., & Hillyard, S. (2014). *Big Data?: Qualitative Approaches to Digital Research*. Emerald Group Publishing.

Hardy, Q. (2015, October 28). IBM to Acquire the Weather Company. *The New York Times*. Retrieved from https://www.nytimes.com/2015/10/29/technology/ibm-to-acquire-the-weather-company.html

Harmon, A. (2017, March 6). Activists Rush to Save Government Science Data — If They Can Find It. *The New York Times*. Retrieved from https://www.nytimes.com/2017/03/06/science/donald-trump-data-rescue-science.html

Harold, J. (2013, March 20). Nonprofits: Master "Medium Data" Before Tackling Big Data. Retrieved April 23, 2018, from Harvard Business Review website: https://hbr.org/2013/03/nonprofits-master-medium-data-1

Hartmann, P. M., Zaki, M. E., Feldmann, N., & Neely, A. D. (2016). *Capturing value from big data – a taxonomy of data-driven business models used by start-up firms*. https://doi.org/10.17863/CAM.5976

Harvard Business Review Analytic Services. (2017). *Digital Dilemma - Turning Data Security and Privacy Concerns Into Opportunities*. Retrieved from Harvard Business Review website: https://hbr.org/resources/pdfs/comm/cognizant/CognizantPulseReport.pdf

Hauptman, O. (2003). Platform Leadership: How Intel, Microsoft and Cisco Drive Industry Innovation. *Innovation*, *5*(1), 91–94. https://doi.org/10.5172/impp.2003.5.1.91

Heimowitz, D. (2018). Square Panda had the biggest campaign to date and was able to increase sales by 2919.6% year over year on Amazon Prime Day with the help of data.world's talented team and technolo . 2.

Heimstaedt, M. (2017). Openwashing: A decoupling perspective on organizational transparency. *Technological Forecasting and Social Change*, *125*, 77–86. https://doi.org/10.1016/j.techfore2017.03.037

Heinrich, B., Klier, M., Schiller, A., & Wagner, G. (2018). Assessing data quality – A probability-based metric for semantic consistency. *Decision Support Systems*. https://doi.org/10.1016/j.dss.2018.03.011

Hekkala, R., & Urquhart, C. (2013). Everyday power struggles: living in an IOIS project. *European Journal of Information Systems*, *22*(1), 76–94. https://doi.org/10.1057/ejis.2012.43

Helland, P. (2011). If You Have Too Much Data, then "Good Enough" is Good Enough. *Communications of the ACM*, *54*(6), 40–47. https://doi.org/10.1145/1953122.1953140

Henriksen-Bulmer, J., & Jeary, S. (2016). Re-identification attacks—A systematic literature review. *International Journal of Information Management*, *36*(6, Part B), 1184–1192. https://doi.org/10.1016/j.ijinfomgt.2016.08.002

Hernández, M. A., & Stolfo, S. J. (1998). Real-world Data is Dirty: Data Cleansing and The Merge/Purge Problem. *Data Mining and Knowledge Discovery*, *2*(1), 9–37. https://doi.org/10.1023/A:1009761603038

Hey, T., Tansley, S., & Tolle, K. (2009). The Fourth Paradigm: Data-Intensive Scientific Discovery. In The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft.

Hildreth, P. M., & Kimble, C. (2004). Knowledge Networks: Innovation Through Communities of Practice. Idea Group Inc (IGI).

Hirschheim, R., & Klein, H. (2012). A Glorious and Not-So-Short History of the Information Systems Field. *Journal of the Association for Information Systems*, *13*(4), 188–235.

Holm, S., & Ploug, T. (2017). Big Data and Health Research-The Governance Challenges in a Mixed Data Economy. *Journal of Bioethical Inquiry*, *14*(4), 515–525. https://doi.org/10.1007/s11673-017-9810-0

Holovaty, A. (2009, May 21). The definitive, two-part answer to "is data journalism?" | Holovaty.com. Retrieved February 7, 2018, from The definitive, two-part answer to "is data journalism?" website: http://www.holovaty.com/writing/data-is-journalism/

Howard, A. (2013, January 28). Open data economy: Eight business models for open data and insight from Deloitte UK. Retrieved February 4, 2018, from O'Reilly Media website: https://www.oreilly.com/ideas/open-data-business-models-deloitte-insight

Howard, A. B. (2014). *The Art and Science of Data-Driven Journalism* (p. 145). Retrieved from The Tow Center for Digital Journalism website: http://towcenter.org/wp-content/uploads/2014/05/Tow-Center-Data-Driven-Journalism.pdf

http://bananalink.org.uk/. (2018). Home | Banana Link. Retrieved February 16, 2018, from http://bananalink.org.uk/ website: http://bananalink.org.uk/

Huang, P., Tafti, A., & Mithas, S. (2018). Platform Sponsor Investments and User Contributions in Knowledge Communities: The Role of Knowledge Seeding. *MIS Quarterly*, *42*(1), 213–240. https://doi.org/10.25300/MISQ/2018/13490

Huang, Q., Davison, R. M., & Gu, J. (2011). The impact of trust, guanxi orientation and face on the intention of Chinese employees and managers to engage in peer-to-peer tacit and explicit knowledge sharing. *Information Systems Journal*, *21*(6), 557–577. https://doi.org/10.1111/j.1365-2575.2010.00361.x

Hubauer, T., Lamparter, S., Roshchin, M., Solomakhina, N., & Watson, S. (2013). Analysis of data quality issues in real-world industrial data. *Proceedings of The Annual Conference of the Prognostics and Health Management Society 2013*, 9.

Huberman, M., & Miles, M. B. (2002). *The Qualitative Researcher's Companion*. SAGE.

Hurt, B. (2016, October 30). Data donation: What's at stake? Retrieved November 6, 2016, from distinct values: data.world website: https://blog.data.world/data-donation-whats-at-stake-9e631c0dde4c#.q0jj5u2pi

IBM. (2017). *10 Key Marketing Trends for 2017*. Retrieved from IBM Marketing Cloud website: https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WRL12345USEN

If Your Company Has Data, Consider "Donating It." (2017, February 3). Retrieved February 6, 2017, from Techonomy website: http://techonomy.com/2017/02/data-donation-whats-at-stake/

Immonen, A., Palviainen, M., & Ovaska, E. (2014). Requirements of an Open Data Based Business Ecosystem. *IEEE Access*, *2*, 88–103. https://doi.org/10.1109/ACCESS.2014.2302872

Inc, S. P. (2018). Multisensory Phonics Learning System | Square Panda. Retrieved March 19, 2019, from Square Panda Inc. website: https://squarepanda.com/

Information wants to be free. (2017). In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Information_wants_to_be_free&oldid=810197636

Information wants to be free ... and expensive. (2009). Retrieved December 4, 2017, from Fortune website: http://fortune.com/2009/07/20/information-wants-to-be-free-and-expensive/

Ingram Bogusz, C., Teigland, R., & Vaast, E. (2018). Designed entrepreneurial legitimacy: the case of a Swedish crowdfunding platform. *European Journal of Information Systems*, 1–18. https://doi.org/10.1080/0960085X.2018.1534039

Ingrams, A. (2017a). Managing governance complexity and knowledge networks in transparency initiatives: the case of police open data. *Local Government Studies*, *43*(3), 364–387. https://doi.org/10.1080/03003930.2017.1294070

Ingrams, A. (2017b). The legal-normative conditions of police transparency: A configurational approach to open data adoption using qualitative comparative analysis. *Public Administration*, *95*(2), 527–545. https://doi.org/10.1111/padm.12319

Inmon, W. H. (1993). *Building the Data Warehouse*. New York: Wiley-QED Publication.

Inmon, W. H. (2005). *Building the Data Warehouse*. Wiley.

Institute of Medicine. (2013). Sharing Clinical Research Data: Workshop Summary. *Sharing Clinical Research Data: Workshop Summary*. Presented at the Barriers to Data Sharing, Washington. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK137824/

International Monetary Fund. (2018). IMF Data. Retrieved February 3, 2018, from IMF website: http://www.imf.org/en/Data

Introduction to Grounded Theory. (2018). Retrieved November 14, 2018, from http://www.analytictech.com/mb870/introtogt.htm

Investopedia. (2007). Rival Good. In *Investopedia*. Retrieved from https://www.investopedia.com/terms/r/rival_good.asp

Islam, N., Abu Zafar Abbasi, & Shaikh, Z. A. (2010). Semantic Web: Choosing the right methodologies, tools and standards. *2010 International Conference on Information and Emerging Technologies*, 1–5. https://doi.org/10.1109/ICIET.2010.5625736

Ismayilov, A., Kontokostas, D., Auer, S., Lehmann, J., & Hellmann, S. (2017). Wikidata through the eyes of DBpedia. *Semantic Web*, *Preprint*(Preprint), 1–11. https://doi.org/10.3233/SW-170277

Itkowitz, C. (2016, November 18). Fake news on Facebook is a real problem. These college students came up with a fix in 36 hours. *The Washington Post*. Retrieved from https://www.washingtonpost.com/news/inspired-life/wp/2016/11/18/fake-news-on-facebook-is-a-real-problem-these-college-students-came-up-with-a-fix/

Jacobs, J. A., & Humphrey, C. (2004). Preserving research data. *Communications of the ACM*, *47*(9). Retrieved from http://escholarship.org/uc/item/09g2f43j

Jaeger, P. T., & Bertot, J. C. (2010). Transparency and technological change: Ensuring equal and sustained public access to government information. *Government Information Quarterly*, *27*(4), 371–376. https://doi.org/10.1016/j.giq.2010.05.003

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. https://doi.org/10.1080/10580530.2012.716740

Janssen, M., & Zuiderwijk, A. (2014). Infomediary Business Models for Connecting Open Data Providers and Users. *Social Science Computer Review*, *32*(5), 694–711. https://doi.org/10.1177/0894439314525902

Jarman, H., Luna-Reyes, L. F., & Zhang, J. (2016). *Public Value and Private Organizations* (Vol. 26; H. Jarman & L. F. LunaReyes, Eds.). Cham: Springer Int Publishing Ag.

Jarvenpaa, S. L., & Lang, K. R. (2011). Boundary Management in Online Communities: Case Studies of the Nine Inch Nails and ccMixter Music Remix Sites. *Long Range Planning*, *44*(5), 440–457. https://doi.org/10.1016/j.lrp.2011.09.002

Jetzek, T. (2016). Managing complexity across multiple dimensions of liquid open data: The case of the Danish Basic Data Program. *Government Information Quarterly*, *33*(1), 89–104. https://doi.org/10.1016/j.giq.2015.11.003

Jetzek, T. (2017). Innovation in the Open Data Ecosystem: Exploring the Role of Real Options Thinking and Multi-sided Platforms for Sustainable Value Generation through Open Data (E. G. Carayannis & S. Sindakis, Eds.). Dordrecht: Springer.

Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2013). Generating Value from Open Government Data. *ICIS 2013 Proceedings*. Retrieved from http://aisel.aisnet.org/icis2013/proceedings/GeneralISTopics/5

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, *2*(2), 59–64. https://doi.org/10.1016/j.bdr.2015.01.006

Johnson, C. M. (2001). A survey of current research on online communities of practice. *The Internet and Higher Education*, *4*(1), 45–60. https://doi.org/10.1016/S1096-7516(01)00047-1

Johnson, P. A., Sieber, R., Scassa, T., Stephens, M., & Robinson, P. (2017). The Cost(s) of Geospatial Open Data. *Transactions in GIS*, *21*(3), 434–445. https://doi.org/10.1111/tgis.12283

Johnson, P., & Robinson, P. (2014). Civic Hackathons: Innovation, Procurement, or Civic Engagement? *Review of Policy Research*, *31*(4), 349–357. https://doi.org/10.1111/ropr.12074

Joo, J. (2011). Adoption of Semantic Web from the perspective of technology innovation: A grounded theory approach. *International Journal of Human-Computer Studies*, *69*(3), 139–154. https://doi.org/10.1016/j.ijhcs.2010.11.002

Jr, E. G. A., Parker, G. G., & Tan, B. (2014). Platform Performance Investment in the Presence of Network Externalities. *Information Systems Research*, 22.

Juell-Skielse, G., Hjalmarsson, A., Juell-Skielse, E., Johannesson, P., & Rudmark, D. (2014). Contests as innovation intermediaries in open data markets. *Information Polity: The International Journal of Government & Democracy in the Information Age*, *19*(3/4), 247–262. https://doi.org/10.3233/IP-140346

Jung, K., & Park, H. W. (2015). A semantic (TRIZ) network analysis of South Korea's "Open Public Data" policy. *Government Information Quarterly*, *32*(3), 353–358. https://doi.org/10.1016/j.giq.2015.03.006

Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January). *Big Data: Issues and Challenges Moving Forward*. 995–1004. https://doi.org/10.1109/HICSS.2013.645

Kalimo, H., & Majcher, K. (2017). The Concept of Fairness: Linking EU Competition and Data Protection Law in the Digital Marketplace. *European Law Review*, *42*(2), 210–233.

Kambies, T., Roma, P., Mittal, N., & Sharma, S. K. (2017, February 7). Dark analytics: Illuminating opportunities hidden within unstructured data. Retrieved March 19, 2018, from Deloitte Insights website: https://www2.deloitte.com/insights/us/en/focus/tech-trends/2017/dark-data-analyzing-unstructured-data

Kane, A. A., & Levina, N. (2017). 'Am I Still One of Them?': Bicultural Immigrant Managers Navigating Social Identity Threats When Spanning Global Boundaries. *Journal of Management Studies*, *54*(4), 540–577. https://doi.org/10.1111/joms.12259

Kaplan, S., & Orlikowski, W. J. (2012). Temporal Work in Strategy Making. *Organization Science*, *24*(4), 965–995. https://doi.org/10.1287/orsc.1120.0792

Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*, *30*(4), 508–513. https://doi.org/10.1016/j.giq.2013.05.012

Kasymova, J., Marques Ferreira, M. A., & Piotrowski, S. J. (2016). Do Open Data Initiatives Promote and Sustain Transparency? A Comparative Analysis of Open Budget Portals in Developing Countries (Vol. 20; J. Zhang, L. F. LunaReyes, T. A. Pardo, & D. S. Sayogo, Eds.). Cham: Springer Int Publishing Ag.

Kauffman, G. (2017, February 5). Why did the USDA shut down an online animal abuse database? *Christian Science Monitor*. Retrieved from http://www.csmonitor.com/USA/2017/0205/Why-did-the-USDA-shut-down-an-online-animal-abuse-database

Kavis, M. (2017). Forget Big Data -- Small Data Is Driving The Internet Of Things [Forbes.com]. Retrieved April 23, 2018, from Forbes website: https://www.forbes.com/sites/mikekavis/2015/02/25/forget-big-data-small-data-is-driving-the-internet-of-things/

Kelly, J. (2014, February 4). The Data Economy Manifesto. Retrieved October 3, 2017, from Wikibon website: http://wikibon.org/wiki/v/The_Data_Economy_Manifesto

Kennedy, J. (2014, October 2). The Data-Driven Economy [Foundation]. Retrieved October 3, 2017, from U.S. Chamber of Commerce Foundation website: https://www.uschamberfoundation.org/data-driven-economy

Kerr, R. A. (2011). Time to Adapt to a Warming World, But Where's the Science? *Science*, *334*(6059), 1052–1053. https://doi.org/10.1126/science.334.6059.1052

Khan, N. A., Sohrabzadeh, S., & Jhamb, G. (2016). Open Data Repositories in Knowledge Society. In *Encyclopedia of Information Science and Technology*. Retrieved from https://www.researchgate.net/publication/308344146_Open_Data_Repositories_in_Knowledge_Society

Kindleberger, C. P. (2015). *A Financial History of Western Europe*. Routledge.

Kirkpatrick, R. (2013, March 21). A New Type of Philanthropy: Donating Data. Retrieved February 28, 2018, from Harvard Business Review website: https://hbr.org/2013/03/a-new-type-of-philanthropy-don

Kirzner, I. M. (1997). Entrepreneurial Discovery and the Competitive Market Process: An Austrian Approach. *Journal of Economic Literature*, *35*(1), 60–85.

Kitchin, R. (2014a). The data revolution: big data, open data, data infrastructures & their consequences. Los Angeles, California: SAGE Publications.

Kitchin, R. (2014b). The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. SAGE.

Kitsios, F., & Kamariotou, M. (2018). Open Data and High-Tech Startups Towards Nascent Entrepreneurship Strategies. Hersey: Igi Global.

Klich, T. B. (2015, February 19). VC 100: The Top Investors in Early-Stage Startups. Retrieved November 19, 2016, from Entrepreneur website: https://www.entrepreneur.com/article/242702

Knowledge@Wharton. (2018). How data.world Wants to Unify the Data World. Retrieved March 20, 2019, from Knowledge@Wharton website: http://knowledge.wharton.upenn.edu/article/data-world-wants-unify-data-world/

KodakOne. (2017). KodakOne. Retrieved January 31, 2018, from https://kodakcoin.com/

Kraemer, K., & King, J. L. (2006). Information Technology and Administrative Reform: Will E-Government Be Different? *International Journal of Electronic Government Research*, *2*(1), 1–20. https://doi.org/10.4018/jegr.2006010101

Krotofil, M., Larsen, J., & Gollmann, D. (2015). The Process Matters: Ensuring Data Veracity in Cyber-Physical Systems. *Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security*, 133–144. https://doi.org/10.1145/2714576.2714599

Kumar, +Vishal. (2017, March 26). Big Data Facts. Retrieved January 31, 2018, from AnalyticsWeek website: https://analyticsweek.com/content/big-data-facts/

Kushmaro, P. (2017, January 4). How small data became bigger than big data. Retrieved April 23, 2018, from CIO website: https://www.cio.com/article/3153389/it-industry/how-small-data-became-bigger-than-big-data.html

Kwark, Y., Chen, J., & Raghunathan, S. (2017). Platform or Wholesale? A Strategic Tool for Online Retailers to Benefit from Third-Party Information. *MIS Quarterly*, *41*(3), 763–785.

Lai, J. (2009). Information wants to be free ... and expensive. Retrieved February 16, 2018, from Fortune website: http://fortune.com/2009/07/20/information-wants-to-be-free-and-expensive/

Lakomaa, E., & Kallberg, J. (2013). Open Data as a Foundation for Innovation: The Enabling Effect of Free Public Sector Information for Entrepreneurs. *IEEE Access*, *1*, 558–563. https://doi.org/10.1109/ACCESS.2013.2279164

Lämmerhirt, D., Rubinstein, M., & Montiel, O. (2017). *The State of Open Government Data in 2017* (p. 15). Retrieved from Open Knowledge International website: https://blog.okfn.org/files/2017/06/FinalreportTheStateofOpenGovernmentDatain2017.pdf

Laney, D. (2001). *3D data management: controlling data volume, velocity, and variety* (No. 949; p. 4). Retrieved from Meta Group Inc. website: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Laney, D. B. (2017). Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage. Routledge.

Langley, A. (1999). Strategies for Theorizing from Process Data. *The Academy of Management Review*, *24*(4), 691–710. https://doi.org/10.2307/259349

Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, *36*(5), 700–710. https://doi.org/10.1016/j.ijinfomgt.2016.04.013

Lassinantti, J., Bergvall-Kareborn, B., & Stahlbrost, A. (2014). Shaping Local Open Data Initiatives: Politics and Implications. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(2), 17–33. https://doi.org/10.4067/S0718-18762014000200003

Latif, A., Saeed, A. U., Hoefler, P., Stocker, E., & Wagner, C. (2009). The Linked Data Value Chain: A Lightweight Model for Business Engineers. *Proceedings of I-KNOW '09 and I-SEMANTICS '09*. Presented at the I-KNOW '09 and I-SEMANTICS '09, Graz, Austria. Retrieved from https://www.researchgate.net/profile/Alexander_Stocker/publication/210332168_The_Linked_Data_Value_Chain_A_Lightweight_Model_for_Business_Engineers/links/02e7e52174a0b6566c000000/The-Linked-Data-Value-Chain-A-Lightweight-Model-for-Business-Engineers.pdf

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. In *Situated Learning: Legitimate Peripheral Participation*. https://doi.org/10.1017/CBO9780511815355

Layne, K., & Lee, J. (2001). Developing fully functional E-government: A four stage model. *Government Information Quarterly*, *18*(2), 122–136. https://doi.org/10.1016/S0740-624X(01)00066-1

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The Parable of Google Flu: Traps in Big Data Analysis. *Science*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506

Lee, J., Upadhyaya, S. J., Rao, H. R., & Sharman, R. (2005). Secure knowledge management and the semantic web. *Communications of the ACM*, *48*(12), 48. https://doi.org/10.1145/1101779.1101808

Leidner, D. E. (2018). Review and Theory Symbiosis: An Introspective Retrospective. *Journal for the Association for Information Systems*.

Leidner, D. E., Gonzalez, E., & Koch, H. (2018). An affordance perspective of enterprise social media and organizational socialization. *The Journal of Strategic Information Systems*. https://doi.org/10.1016/j.jsis.2018.03.003

Leidner, D. E., Pan, G., & Pan, S. L. (2009). The role of IT in crisis response: Lessons from the SARS and Asian Tsunami disasters. *The Journal of Strategic Information Systems*, *18*(2), 80–99. https://doi.org/10.1016/j.jsis.2009.05.001

Levina, N., & Vaast, E. (2008). Innovating or Doing as Told? Status Differences and Overlapping Boundaries in Offshore Collaboration. *Management Information Systems Quarterly*, *32*(2), 307–332.

Levina, & Ross. (2003). From the Vendor's Perspective: Exploring the Value Proposition in Information Technology Outsourcing. *MIS Quarterly*, *27*(3), 331. https://doi.org/10.2307/30036537

Li, D.-C., Chang, C.-C., & Liu, C.-W. (2012). Using structure-based data transformation method to improve prediction accuracies for small data sets. *Decision Support Systems*, *52*(3), 748–756. https://doi.org/10.1016/j.dss.2011.11.021

Li, I., Dey, A. K., & Forlizzi, J. (2011). Understanding My Data, Myself: Supporting Self-reflection with Ubicomp Technologies. *Proceedings of the 13th International Conference on Ubiquitous Computing*, 405–414. https://doi.org/10.1145/2030112.2030166

Lillrank, P. (2003). The quality of information. *International Journal of Quality & Reliability Management*, *20*(6), 691–703. https://doi.org/10.1108/02656710310482131

Linder, J. C., Jarvenpaa, S., & Davenport, T. H. (2003). Toward an innovation sourcing strategy. *MIT Sloan Management Review; Cambridge*, *44*(4), 43–49.

Link, G., Lumbard, K., Conboy, K., Feldman, M., Feller, J., George, J., … Willis, M. (2017). Contemporary Issues of Open Data in Information Systems Research: Considerations and Recommendations. *Communications of the Association for Information Systems*, *41*(1). Retrieved from http://aisel.aisnet.org/cais/vol41/iss1/25

Liow, M. L.-F., & Lee, H.-Y. (2016). Instilling Trust and Confidence in a Data Driven Economy - Preliminary Design of a Data Certification Scheme. New York: Ieee.

Llorca-Abad, G., & Cano-Oron, L. (2016). How Social Networks and Data Brokers trade with Private Data. *Redes Com-Revista De Estudios Para El Desarrollo Social De La Comunicacion*, (14), 84–103.

Lomotey, R. K., & Deters, R. (2014). *Data Mining from NoSQL Document-Append Style Storages* (D. DeRoure, B. Thuraisingham, & J. Zhang, Eds.). New York: Ieee.

Lugmayr, A., & Grueblbauer, J. (2017). Review of information systems research for media industry-recent advances, challenges, and introduction of information systems research in the media industry. *Electronic Markets; Heidelberg*, *27*(1), 33–47. http://dx.doi.org/10.1007/s12525-016-0239-9

Luo, W. (2013). Enterprise Data Economy: A Hadoop-Driven Model and Strategy. In X. Hu, T. Y. Lin, V. Raghavan, B. Wah, R. BaezaYates, G. Fox, … R. Nambiar (Eds.), *2013 IEEE International Conference on Big Data* (pp. 65–70). https://doi.org/10.1109/BigData.2013.6691690

Lupton, D. (2014). The commodification of patient opinion: the digital patient experience economy in the age of big data. *Sociology of Health & Illness*, *36*(6), 856–869. https://doi.org/10.1111/1467-9566.12109

Lupton, D. (2016). The diverse domains of quantified selves: self-tracking modes and dataveillance. *Economy and Society*, *45*(1), 101–122. https://doi.org/10.1080/03085147.2016.1143726

Luxton, E. (2016, May 30). This is how cross-border bandwidth has grown since 2005 | World Economic Forum. Retrieved February 3, 2018, from https://www.weforum.org/agenda/2016/05/this-is-how-cross-border-bandwidth-has-grown-since-2005/

Luxton, E. (2018). Telecommunications market research that's data-driven. Retrieved February 3, 2018, from https://www.telegeography.com website: https://www.telegeography.com/index.html

Lynch, P., & Quealy, S. (2017). Finding the Value in an Animal Health Data Economy: A Participatory Market Model Approach. *Frontiers in Veterinary Science*, *4*. https://doi.org/10.3389/fvets.2017.00145

Lyytinen, K., Grover, V., & Clemson University. (2017). Management Misinformation Systems: A Time to Revisit? *Journal of the Association for Information Systems*, *18*(3), 206–230. https://doi.org/10.17705/1jais.00453

Mabillard, V., & Pasquier, M. (2016). Transparency and Trust in Government (2007-2014): A Comparative Study. *Nispacee Journal of Public Administration and Policy*, *9*(2), 69–92. https://doi.org/10.1515/nispa-2016-0015

Machova, R., & Lnenicka, M. (2017). Evaluating the Quality of Open Data Portals on the National Level. *Journal of Theoretical and Applied Electronic Commerce Research*, *12*(1), 21–41. https://doi.org/10.4067/S0718-18762017000100003

Magalhaes, G., Roseira, C., & Manley, L. (2014). Business Models for Open Government Data. *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance*, 365–370. https://doi.org/10.1145/2691195.2691273

Majchrzak, A., & Markus, M. L. (2013). Methods for Policy Research: Taking Socially Responsible Action. SAGE Publications.

Makela, E., Hypen, K., & Hyvonen, E. (2013). Fiction literature as linked open data-The BookSampo dataset. *Semantic Web*, *4*(3), 299–306. https://doi.org/10.3233/SW-120093

Mallah, R. (2018). The Other Five V's of Big Data: An Updated Paradigm. Retrieved February 9, 2018, from CIO Review website: https://bigdata.cioreview.com/cxoinsight/the-other-five-v-s-of-big-data-an-updated-paradigm-nid-10287-cid-15.html

Mandel, M. (2013, July 25). The Data Economy Is Much, Much Bigger Than You (and the Government) Think - The Atlantic. *The Atlantic*. Retrieved from https://www.theatlantic.com/business/archive/2013/07/the-data-economy-is-much-much-bigger-than-you-and-the-government-think/278113/

Mandel, M. (2014). Data, Trade and Growth. *Proceedings of the 2013 TPRC Research Conference on Communications, Information and Internet Policy*, 16. Retrieved from http://www.progressivepolicy.org/wp-content/uploads/2014/04/2014.04-Mandel_Data-Trade-and-Growth.pdf

Mannes, J. (2017). Data.world raises $18.7 million to remedy our post-fact society. Retrieved January 31, 2018, from TechCrunch website: http://social.techcrunch.com/2017/02/21/data-world-raises-18-7-million-to-remedy-our-post-fact-society/

Manninen, J. (2006). Anders Chydenius and the origins of world's first freedom of information act. The World's First Freedom of Information Act: Anders Chydenius' Legacy Today. Kokkola, Finland: Anders Chydenius Foundation.

Manyika, J., Chui, M., Farrell, D., Kuiken, S. V., Groves, P., & Doshi, E. A. (2013). *Open data: Unlocking innovation and performance with liquid information* (p. 116). Retrieved from McKinsey Global Institute website: http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information

March, S. T., & Hevner, A. R. (2007). Integrated decision support systems: A data warehousing perspective. *Decision Support Systems*, *43*(3), 1031–1043. https://doi.org/10.1016/j.dss.2005.05.029

Marcial, L. H., & Hemminger, B. M. (2010). Scientific data repositories on the Web: An initial survey. *Journal of the American Society for Information Science and Technology*, *61*(10), 2029–2048. https://doi.org/10.1002/asi.21339

Marczak, P., & Sieber, R. (2017). Linking legislative openness to open data in Canada. *The Canadian Geographer / Le Géographe Canadien*, *0*(0). https://doi.org/10.1111/cag.12408

Marjanovic, O., & Cecez-Kecmanovic, D. (2017). Exploring the tension between transparency and datification effects of open government IS through the lens of Complex Adaptive Systems. *Journal of Strategic Information Systems*, *26*(3), 210–232. https://doi.org/10.1016/j.jsis.2017.07.001

Markovich, R. (2017, December 20). Interrogating Your Ride: Data Analytics for the Car Enthusiast. Retrieved February 9, 2018, from Wavefront by VMware website: https://www.wavefront.com/interrogating-ride-data-analytics-car-enthusiast/

Markus, M. L. (2001). Toward a Theory of Knowledge Reuse: Types of Knowledge Reuse Situations and Factors in Reuse Success. *Journal of Management Information Systems*, *18*(1), 57–93.

Marr, B. (2015, May). Big Data At Caesars Entertainment - A One Billion Dollar Asset? Retrieved December 5, 2017, from Forbes website: https://www.forbes.com/sites/bernardmarr/2015/05/18/when-big-data-becomes-your-most-valuable-asset/

Marr, B. (2016, April 28). Big Data Overload: Why Most Companies Can't Deal With The Data Explosion [Forbes.com]. Retrieved September 27, 2017, from Forbes website: https://www.forbes.com/sites/bernardmarr/2016/04/28/big-data-overload-most-companies-cant-deal-with-the-data-explosion/

Marr, B. (2017). What Is Data Democratization? A Super Simple Explanation And The Key Pros And Cons. Retrieved March 20, 2019, from Forbes website: https://www.forbes.com/sites/bernardmarr/2017/07/24/what-is-data-democratization-a-super-simple-explanation-and-the-key-pros-and-cons/

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making Sense of Data-Driven Decision Making in Education* [Product Page]. Retrieved from RAND website: https://www.rand.org/pubs/occasional_papers/OP170.html

Martin, K. E. (2015). Ethical issues in the Big Data industry. *MIS Quarterly Executive*, *14*, 2.

Mawer, C. (2015, November 10). Valuing Data is Hard. Retrieved December 5, 2017, from Silicon Valley Data Science website: https://svds.com/valuing-data-is-hard/

Mayer-Schönberger, V., & Cukier, K. (2014). *Learning with Big Data: The Future of Education*. Houghton Mifflin Harcourt.

Mazmanian, A., & Jan 26, 2018. (2018). Trump picks federal CIO -. Retrieved January 29, 2018, from FCW website: https://fcw.com/articles/2018/01/26/kent-new-federal-cio.aspx

Mazmanian, M., Orlikowski, W. J., & Yates, J. (2013). The Autonomy Paradox: The Implications of Mobile Email Devices for Knowledge Professionals. *Organization Science*, *24*(5), 1337–1357. https://doi.org/10.1287/orsc.1120.0806

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review, 90*(10), 60–66, 68, 128.

McCallum, Q. E., & Gleason, K. (2013). *Business Models for the Data Economy*. O'Reilly Media, Inc.

McCubbrey, D. J. (1999). Disintermediation and reintermediation in the US air travel distribution industry: a Delphi study. *Communications of the AIS, 1*(5es), 3.

McDonald, J., & Léveillé, V. (2014). Whither the retention schedule in the era of big data and open data? *Records Management Journal*, *24*(2), 99–121. https://doi.org/10.1108/RMJ-01-2014-0010

McGarry, T. (2009). Applied and theoretical perspectives of performance analysis in sport: Scientific issues and challenges. *International Journal of Performance Analysis in Sport*, *9*(1), 128–140. https://doi.org/10.1080/24748668.2009.11868469

McIntosh, W., & Schmeichel, B. (2004). Collectors and Collecting: A Social Psychological Perspective. *Leisure Sciences*, *26*(1), 85–97. https://doi.org/10.1080/01490400490272639

McKinsey Global Institute. (2016). *Digital Globalization: The New Era of Global Flows* (p. 156). London: McKinsey Global Institute.

Meijer, R., Conradie, P., & Choenni, S. (2014). Reconciling Contradictions of Open Data Regarding Transparency, Privacy, Security and Trust. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(3), 32–44. https://doi.org/10.4067/S0718-18762014000300004

Melissa Lee, Almirall, E., & Wareham, J. (2016). Open Data and Civic Apps: First-Generation Failures, Second-Generation Improvements. *Communications of the ACM, 59*(1), 82–89. https://doi.org/10.1145/2756542

Menon, S., & Sarkar, S. (2016). Privacy and Big Data: Scalable Approaches to Sanitize Large Transactional Databases for Sharing. *MIS Quarterly*, *40*(4). Retrieved from http://aisel.aisnet.org/cgi/viewcontent.cgi?article=3319&context=misq

Michener, G., & Ritter, O. (2017). Comparing Resistance to Open Data Performance Measurement: Public Education in Brazil and the Uk. *Public Administration*, *95*(1), 4–21. https://doi.org/10.1111/padm.12293

Miles, M. B., & Huberman, M. A. (1994). Qualitative Data Analysis: An Expanded Sourcebook. SAGE.

Miller, R. (2017). The battle for control of data could be just starting. Retrieved November 21, 2017, from TechCrunch website: http://social.techcrunch.com/2017/11/19/the-battle-for-control-of-data-could-be-just-starting/

Mirani, L., & Mirani, L. (2014). The nine companies that know more about you than Google or Facebook. Retrieved January 15, 2018, from Quartz website: https://qz.com/213900/the-nine-companies-that-know-more-about-you-than-google-or-facebook/

Moges, H.-T., Vlasselaer, V. V., Lemahieu, W., & Baesens, B. (2016). Determining the use of data quality metadata (DQM) for decision making purposes and its impact on decision outcomes — An exploratory study. *Decision Support Systems*, *83*, 32–46. https://doi.org/10.1016/j.dss.2015.12.006

Monino, J.-L., & Sedkaoui, S. (2016a). *Big Data, Open Data and Data Development*. John Wiley & Sons.

Monino, J.-L., & Sedkaoui, S. (2016b). *Big Data, Open Data and Data Development*. John Wiley & Sons.

Monino, J.-L., & Sedkaoui, S. (2016c). Open Data: A New Challenge. In *Big Data, Open Data and Data Development* (pp. 23–42). https://doi.org/10.1002/9781119285199.ch2

Montiel-Ponsoda, E., Gracia del Río, J., Aguado de Cea, G., & Gómez-Pérez, A. (2011). Representing Translations on the Semantic Web. *Proceedings of the 2nd International Workshop on the Multilingual Semantic Web (MSW 2011), 775*. Retrieved from http://iswc2011.semanticweb.org/fileadmin/iswc/Papers/Workshops/MSW/Elena.pdf

Montiel-Ponsoda, E., Gracia, J., Aguado-de-Cea, G., & Gomez-Perez, A. (2011). Representing Translations on the Semantic Web. *Lecture Notes in Computer Science*, 13.

Moon, M. J. (2002). The Evolution of E-Government among Municipalities: Rhetoric or Reality? *Public Administration Review*, *62*(4), 424–433.

Morago, G. (2017, September 5). H-E-B's lessons learned from Hurricane Harvey - HoustonChronicle.com. Houston Chronicle. Retrieved from https://www.chron.com/life/food/article/H-E-B-s-lessons-learned-from-Hurricane-Harvey-12175190.php

Moreau, L. (2010). The Foundations for Provenance on the Web. *Foundations and Trends in Web Science*, *2*(2–3), 99–241. https://doi.org/10.1561/1800000010

Moreno, F. (2016, September 20). Dark data discovery: Improve marketing insights to increase ROI. Retrieved March 19, 2018, from IBM Business Analytics Blog website: https://www.ibm.com/blogs/business-analytics/dark-data-discovery-improve-marketing-insights-to-increase-roi/

Morris, C. (2013, October 11). Rock music's new heroes: Lady Gaga and...big data? Retrieved November 21, 2017, from https://www.cnbc.com/2013/10/10/-new-heroes-lady-gaga-andbig-data.html

Munne, R. (2013). *BIG Work in Progress: Big Data Public Private Forum and Public Sector* (W. Castelnovo & E. Ferrari, Eds.). Nr Reading: Acad Conferences Ltd.

Munro, R. (2011). Subword and Spatiotemporal Models for Identifying Actionable Information in Haitian Kreyol. *Proceedings of the Fifteenth Conference on Computational Natural Language Learning*, 68–77. Retrieved from http://dl.acm.org/citation.cfm?id=2018936.2018945

Murtagh, M. J., Demir, I., Jenkings, K. N., Wallace, S. E., Murtagh, B., Boniol, M., ... Burton, P. R. (2012). Securing the Data Economy: Translating Privacy and Enacting Security in the Development of DataSHIELD. *Public Health Genomics; Basel*, *15*(5), 243–253. http://dx.doi.org/10.1159/000336673

Musgrave, S. (2003). The metadata dynamic - Ensuring data has a long and healthy life. In R. Banks, J. Currall, J. Francis, L. Gerrard, R. Khan, T. Macer, ... A. Westlake (Eds.), *Proceedings of the Fourth ASC International Conference* (pp. 119–129). Chesham: Assoc Survey Computing.

Myers, M. D. (1997). Qualitative Research in Information Systems. *MIS Quarterly*, *21*(2), 241–242.

Myers, M. D. (2013). *Qualitative research in business & management* (2nd ed). London: SAGE.

National Freedom of Information Coalition. (2017). International FOI Laws. Retrieved January 19, 2018, from https://www.nfoic.org/international-foi-laws

National Neighborhood Indicators Partnership. (2017). National Neighborhood Indicators Partnership. Retrieved February 10, 2018, from National Neighborhood Indicators Partnership website: https://www.neighborhoodindicators.org/

Nature. (2017). Recommended Data Repositories | Scientific Data. Retrieved September 20, 2017, from Nature website: https://www.nature.com/sdata/policies/repositories#general

Nebot, V., & Berlanga, R. (2012). Building data warehouses with semantic web data. *Decision Support Systems*, *52*(4), 853–868. https://doi.org/10.1016/j.dss.2011.11.009

neighborhoodindicators.org. (2017). National Neighborhood Indicators Partnership. Retrieved February 16, 2018, from https://www.neighborhoodindicators.org/

Nelson, R. R., Todd, P. A., & Wixom, B. H. (2005). Antecedents of Information and System Quality: An Empirical Examination Within the Context of Data Warehousing. *Journal of Management Information Systems*, *21*(4), 199–235. https://doi.org/10.1080/07421222.2005.11045823

Nemati, H. R., Steiger, D. M., Iyer, L. S., & Herschel, R. T. (2002). Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. *Decision Support Systems*, *33*(2), 143–161. https://doi.org/10.1016/S0167-9236(01)00141-5

Newman, D. (2011). *How to Plan, Participate and Prosper in the Data Economy*. Retrieved from Gartner website: https://www.gartner.com/doc/1610514/plan-participate-prosper-data-economy

Nicholson, S. W., & Bennett, T. B. (2009). Transparent Practices: Primary and Secondary Data in Business Ethics Dissertations. *Journal of Business Ethics*, *84*(3), 417–425. https://doi.org/10.1007/s10551-008-9717-0

Nonaka, I., Toyama, R., & Nagata, A. (2000). A firm as a knowledge-creating entity: a new perspective on the theory of the firm. *Industrial and Corporate Change*, *9*(1), 1–20. https://doi.org/10.1093/icc/9.1.1

Nørretranders, T. (1999). The User Illusion: Cutting Consciousness Down to Size. Penguin.

Nwabueze, T. U. (2010). Basic steps in adapting response surface methodology as mathematical modelling for bioprocess optimisation in the food systems. *International Journal of Food Science and Technology*, *45*(9), 1768–1776. https://doi.org/10.1111/j.1365-2621.2010.02256.x

O'Leary, D. E. (2015). Armchair Auditors: Crowdsourcing Analysis of Government Expenditures. *Journal of Emerging Technologies in Accounting*, *12*(1), 71–91. https://doi.org/10.2308/jeta-51225

Olshansky, S. J., Carnes, B. A., Yang, Y. C., Miller, N., Anderson, J., Beltran-Sanchez, H., & Ricanek, K. (2016). The Future of Smart Health. *Computer*, *49*(11), 14–21. https://doi.org/10.1109/MC.2016.336

O'Neal, S. (2016). The Personal-Data Tsunami And the Future of Marketing A Moments-Based Marketing Approach For the New People-Data Economy. *Journal of Advertising Research*, *56*(2), 136–141. https://doi.org/10.2501/JAR-2016-027

O'Neil, C. (2017). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown.

Open Data 500. (2017). Open Data 500. Retrieved September 12, 2017, from http://www.opendata500.com/

Open Data at NSF | NSF - National Science Foundation. (2016). Retrieved November 17, 2016, from https://www.nsf.gov/data/

Open Data Institute. (2018). Open Data Institute. Retrieved April 22, 2018, from https://theodi.org/topic/data-infrastructure/

Open Government. (2016). Retrieved November 19, 2016, from Data.gov website: https://www.data.gov/open-gov/

Opengovdata.ru. (2017). Open data in Russia | Open data - public, commercial, public. Retrieved November 28, 2017, from Open Government Data website: https://opengovdata.ru/

Opher, A., Chou, A., & Onda, A. (2016). The Rise of the Data Economy: Driving Value through Internet of Things Data Monetization. 16.

Opher, A., Chou, A., Onda, A., & Sounderrajan, K. (2016). *The Rise of the Data Economy: Driving Value through Internet of Things Data Monetization* (p. 15). Retrieved from IBM Global Services website: https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WWW12367USEN

Ordanini, A., & Pol, A. (2001). Infomediation and competitive advantage in b2b digital marketplaces. *European Management Journal*, *19*(3), 276–285. https://doi.org/10.1016/S0263-2373(01)00024-X

O'Riain, S., Curry, E., & Harth, A. (2012). XBRL and open data for global financial ecosystems: A linked data approach. *International Journal of Accounting Information Systems*, *13*(2), 141–162. https://doi.org/10.1016/j.accinf.2012.02.002

Oriko, S. (2017, March 1). Presenting the value of open data to NGOs: A gentle introduction to open data for Oxfam in Kenya. Retrieved January 17, 2018, from Open Knowledge International Blog website: https://blog.okfn.org/2017/03/01/presenting-the-value-of-open-data-to-ngos-a-gentle-introduction-to-open-data-for-oxfam-in-kenya/

Otto, B., & Aier, S. (2013). Business models in the data economy: A case study from the business partner data domain. *Proceedings of the 11th International Conference on Wirtschaftsinformatik*, 475–489. Leipzig, Germany.

Our Story | AP. (2019). Retrieved March 19, 2019, from Associated Press website: https://www.ap.org/about/our-story/

Oxford Dictionaries. (N.D.). economy | Definition of economy in English by Oxford Dictionaries. In *Oxford Dictionaries | English*. Retrieved from https://en.oxforddictionaries.com/definition/economy

Pancras, J., & Sudhir, K. (2007). Optimal marketing strategies for a customer data intermediary. *Journal of Marketing Research*, *44*(4), 560–578. https://doi.org/10.1509/jmkr.44.4.560

Pantano, E., Priporas, C.-V., & Stylos, N. (2017). "You will like it!" using open data to predict tourists' response to a tourist attraction. *Tourism Management*, *60*, 430–438. https://doi.org/10.1016/j.tourman.2016.12.020

Parker, G., Van Alstyne, M. W., & Jiang, X. (2016). Platform Ecosystems: How Developers Invert the Firm. *MIS Quarterly*, *41*(1), 255–266.

Paré, G., Trudel, M.-C., Jaana, M., & Kitsiou, S. (2015). Synthesizing information systems knowledge: A typology of literature reviews. Information & Management, 52(2), 183–199. https://doi.org/10.1016/j.im.2014.08.008

Patterson, E. S., Roth, E. M., & Woods, D. D. (2001). Predicting Vulnerabilities in Computer-Supported Inferential Analysis under Data Overload. *Cognition, Technology & Work*, *3*(4), 224–237. https://doi.org/10.1007/s10111-001-8004-y

Pattuelli, M. C. (2012). Personal name vocabularies as linked open data: A case study of jazz artist names. *Journal of Information Science*, *38*(6), 558–565. https://doi.org/10.1177/0165551512455989

Pavlou, P. A., Liang, H., & Xue, Y. (2007). Understanding and Mitigating Uncertainty in Online Exchange Relationships: A Principal-Agent Perspective. *MIS Quarterly*, *31*(1), 105–136. https://doi.org/10.2307/25148783

Pekala, S. (2017). Privacy and User Experience in 21st Century Library Discovery. *Information Technology and Libraries*, *36*(2), 48–58. https://doi.org/10.6017/ital.v36i2.9817

Peled, A. (2011). When Transparency and Collaboration Collide: The USA Open Data Program. *Journal of the American Society for Information Science and Technology*, *62*(11), 2085–2094. https://doi.org/10.1002/asi.21622

Pentland, B. T. (1999). Building Process Theory with Narrative: From Description to Explanation. *The Academy of Management Review*, *24*(4), 711–724. https://doi.org/10.2307/259350

Pereira, G. V., Macadar, M. A., Luciano, E. M., & Testa, M. G. (2017). Delivering public value through open government data initiatives in a Smart City context. *Information Systems Frontiers*, *19*(2), 213–229. https://doi.org/10.1007/s10796-016-9673-7

Petter, S., & McLean, E. R. (2009). A meta-analytic assessment of the DeLone and McLean IS success model: An examination of IS success at the individual level. *Information & Management*, *46*(3), 159–166. https://doi.org/10.1016/j.im.2008.12.006

Phillips, S. D. (2013). Shining Light on Charities or Looking in the Wrong Place? Regulation-by-Transparency in Canada. *Voluntas*, *24*(3), 881–905. https://doi.org/10.1007/s11266-013-9374-5

Piprani, B., & Ernst, D. (2008). A Model for Data Quality Assessment. In R. Meersman, Z. Tari, & P. Herrero (Eds.), On the Move to Meaningful Internet Systems: OTM 2008 Workshops (pp. 750–759). Springer Berlin Heidelberg.

Pointsoflight.com. (2012, April 13). President George H. W. Bush. Retrieved November 15, 2016, from Points of Light website: http://www.pointsoflight.org/people/board-members/president-george-h-w-bush

Popp, R., Armour, T., Senator, T., & Numrych, K. (2004). Countering Terrorism Through Information Technology. *Commun. ACM*, *47*(3), 36–43. https://doi.org/10.1145/971617.971642

Popvox. (2017a). Retrieved February 6, 2018, from Popvox website: https://www.popvox.com

Popvox. (2017b, September 7). POPVOX. Retrieved February 6, 2018, from Wikipedia website: https://en.wikipedia.org/w/index.php?title=POPVOX&oldid=799434966

Porat, M. U. (1977). *The Information Economy: Definition and Measurement* (No. 77-12(1); p. 319). Retrieved from US Office of Telecommunications website: https://files.eric.ed.gov/fulltext/ED142205.pdf

Privacy Rights Clearinghouse | Data Breaches. (2017). Retrieved September 20, 2017, from https://www.privacyrights.org/data-breaches

Provenance Incubator Group. (2009). Retrieved March 17, 2019, from https://www.w3.org/2005/Incubator/prov/charter

Quintanilla, G., & Gil-Garcia, J. R. (2016). Open Government and Linked Data: Concepts, Experiences and Lessons Based on the Mexican Case. *Revista Del Clad Reforma Y Democracia*, (65), 69–102.

Raconteur. (2016). *The Data Economy* (No. 0324; p. 16). Retrieved from Raconteur website: https://www.raconteur.net/the-data-economy-2015

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Cardamone, C., Murray, P., ... Vandenberg, J. (2013). Galaxy Zoo: Motivations of Citizen Scientists. *Astronomy Education Review, 9*(1). Retrieved from http://arxiv.org/abs/1303.6886

Rafaeli, S., & Ravid, G. (2003). Information sharing as enabler for the virtual team: an experimental approach to assessing the role of electronic mail in disintermediation. *Information Systems Journal*, *13*(2), 191–206.

Ram, S., Hwang, Y., & Khatri, V. (2003). Infomediation for E-business enabled Supply Chains: A Semantics Based Approach. In *IFIP - The International Federation for Information Processing. Semantic Issues in E-Commerce Systems* (pp. 205–219). https://doi.org/10.1007/978-0-387-35658-7_13

Ramamurthy, K. (Ram), Sen, A., & Sinha, A. P. (2008). An empirical investigation of the key determinants of data warehouse adoption. *Decision Support Systems*, *44*(4), 817–841. https://doi.org/10.1016/j.dss.2007.10.006

Ramoglou, S., & Tsang, E. W. K. (2016). A Realist Perspective of Entrepreneurship: Opportunities As Propensities. *Academy of Management Review*, *41*(3), 410–434. https://doi.org/10.5465/amr.2014.0281

Ranade, P., Scannell, D., & Stafford, B. (2014, January). APIs: Three steps to unlock the data economy's most promising new go-to-market channel | McKinsey & Company. Retrieved December 5, 2017, from McKinsey & Co website: https://www.mckinsey.com/business-functions/marketing-and-sales/our-insights/apis-three-steps-to-unlock-the-data-economy039s-most-promising-new-go-to-market-channel

Ravel, B., & Newville, M. (2005). Athena, Artemis, Hephaestus: data analysis for X-ray absorption spectroscopy using IFEFFIT. *Journal of Synchrotron Radiation*, *12*(4), 537–541. https://doi.org/10.1107/S0909049505012719

Raymond, E. S. (1999). The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. O'Reilly.

Reader, R. (2016, July 11). Data.World Aims To Be The Social Network For Data Nerds. Retrieved September 9, 2016, from Fast Company website: https://www.fastcompany.com/3061712/dataworld-aims-to-be-the-social-network-for-data-nerds

Reese, B. (2016, November 7). data.world CEO Brett Hurt Interview. Retrieved November 19, 2016, from https://gigaom.com/2016/11/07/data-world-ceo-brett-hurt-interview/

Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and Opportunities of Open Data in Ecology. *Science*, *331*(6018), 703–705. https://doi.org/10.1126/science.1197962

Reitz, C. (2018, January 26). There's A Legit Reason Why Google Arts & Culture Isn't Available In Texas. Retrieved January 27, 2018, from Elite Daily website: https://www.elitedaily.com/p/why-is-the-google-arts-culture-app-not-available-in-texas-the-reason-is-legit-8026321

Report: Hate crimes increase after election - CNN Video. (2016). Retrieved from http://www.cnn.com/videos/politics/2016/11/19/hate-crimes-after-election-increase-sandoval-newday-pkg.cnn

Reuters. (2017, August 2). Exclusive: Venezuelan vote data casts doubt on turnout at Sunday poll. *Reuters*. Retrieved from https://www.reuters.com/article/us-venezuela-politics-vote-exclusive/exclusive-venezuelan-vote-data-casts-doubt-on-turnout-at-sunday-poll-idUSKBN1AI0AL

Roberts, J. (2006). Limits to Communities of Practice. 17.

Rogers, J. (2017, March 31). Privacy activists want to sell Trump's browsing data [Text.Article]. Retrieved September 20, 2017, from Fox News website: http://www.foxnews.com/tech/2017/03/31/privacy-activists-want-to-sell-trumps-browsing-data.html

Rogoff, B., & Lave, J. (Series Ed.). (1984). Everyday cognition: Its development in social context. In Everyday Cognition: Its Development in Social Context. Cambridge, MA, US: Harvard University Press.

Rohunen, A., Markkula, J., Heikkila, M., & Heikkila, J. (2014). Open Traffic Data for Future Service Innovation - Addressing the Privacy Challenges of Driving Data. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(3), 71–89. https://doi.org/10.4067/S0718-18762014000300007

Roof, K., & Lynley, M. (2015). Leaked Pinterest Documents Show Revenue, Growth Forecasts. Retrieved December 5, 2017, from TechCrunch website: http://social.techcrunch.com/2015/10/16/leaked-pinterest-documents-show-revenue-growth-forecasts/

Roose, K. (2018, January 30). Kodak's Dubious Cryptocurrency Gamble. *The New York Times*. Retrieved from https://www.nytimes.com/2018/01/30/technology/kodak-blockchain-bitcoin.html

Rosenbaum, E. (2013a, October 12). Twitter: Forget IPO value, what's the data worth? Retrieved October 3, 2017, from CNBC website: https://www.cnbc.com/2013/10/11/twitter-forget-ipo-value-whats-the-data-worth.html

Rosenbaum, E. (2013b, October 26). Bra designed with algorithms to hit stores. Retrieved October 3, 2017, from CNBC website: https://www.cnbc.com/2013/10/26/bra-designed-with-algorithms-to-hit-stores.html

Rosenbaum, E. (2013c, November). Your data is worth close to nothing. Retrieved December 5, 2017, from Data Economy website: https://www.cnbc.com/data-economy/

Rosenbaum, E. (2013d, November 30). The end of the social media bubble in a $2,500 check. Retrieved October 3, 2017, from CNBC website: https://www.cnbc.com/2013/11/29/the-end-of-the-social-media-bubble-in-a-2500-check.html

Rosenberg, D. (2013). Data before the Fact. In *"Raw Data" Is an Oxymoron* (pp. 15–40). Retrieved from https://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6462162

Ross, M. L. (2001). Does Oil Hinder Democracy? *World Politics*, *53*(03), 325–361. https://doi.org/10.1353/wp.2001.0011

Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., … Jacobs, D. (2012). Dynamic Changes in Motivation in Collaborative Citizen-science Projects. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, 217–226. https://doi.org/10.1145/2145204.2145238

Rubinstein, A., & Wolinsky, A. (1987). Middlemen. *The Quarterly Journal of Economics*, *102*(3), 581–594. https://doi.org/10.2307/1884218

Ruijer, E., Grimmelikhuijsen, S., Hogan, M., Enzerink, S., Ojo, A., & Meijer, A. (2017). Connecting societal issues, users and data. Scenario-based design of open data platforms. *Government Information Quarterly*, *34*(3), 470–480. https://doi.org/10.1016/j.giq.2017.06.003

Ruijer, E., Grimmelikhuijsen, S., & Meijer, A. (2017). Open data for democracy: Developing a theoretical framework for open data use. *Government Information Quarterly*, *34*(1), 45–52. https://doi.org/10.1016/j.giq.2017.01.001

Russell, M. A. (2013). Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More. O'Reilly Media, Inc.

Russom, P. (2016). *Best Practices Report | Data Warehouse Modernization in the Age of Big Data Analytics*. Retrieved from TDWI website: https://tdwi.org/research/2016/03/best-practices-report-data-warehouse-modernization.aspx

Sa, C., & Grieco, J. (2016). Open Data for Science, Policy, and the Public Good. *Review of Policy Research*, *33*(5), 526–543. https://doi.org/10.1111/ropr.12188

Safarov, I., Meijer, A., & Grimmelikhuijsen, S. (2017). Utilization of open government data: A systematic literature review of types, conditions, effects and users. *Information Polity: The International Journal of Government & Democracy in the Information Age*, *22*(1), 1–24. https://doi.org/10.3233/IP-160012

Sagiroglu, S., & Sinanc, D. (2013). Big data: A review. *2013 International Conference on Collaboration Technologies and Systems (CTS)*, 42–47. https://doi.org/10.1109/CTS.2013.6567202

Sahoo, N., Singh, P. V., & Mukhopadhyay, T. (2012). A Hidden Markov Model for Collaborative Filtering. *MIS Quarterly*, *36*(4), 1329–1356.

SameWorks. (2018). First Three Companies Achieve SameWorks Equal Pay Certification. Retrieved January 31, 2018, from https://www.prnewswire.com/news-releases/first-three-companies-achieve-sameworks-equal-pay-certification-300589806.html

Sankaranarayanan, R., & Sundararajan, A. (2010). Electronic Markets, Search Costs, and Firm Boundaries. *Information Systems Research*, *21*(1), 154–169. https://doi.org/10.1287/isre.1090.0235

Sarasvathy, S. D. (2001). Causation and effectuation: Toward a theoretical shift from economic inevitability to entrepreneurial contingency. *Academy of Management Review*, *26*(2), 243–263. https://doi.org/10.5465/AMR.2001.4378020

Sarker, S., & Lee, A. S. (2002). Using a Positivist Case Research Methodology to Test Three Competing Theories-in-Use of Business Process Redesign. *Journal for the Association for Information Systems*, *2*(1), 7–79.

Sarker, S., Xiao Xiao, & Beaulieu, T. (2013). Qualitative Studies in Information Systems: A Critical Review and Some Guiding Principles. *MIS Quarterly*, iii–xviii.

Sawy, O. A. E., Malhotra, A., Gosain, S., & Young, K. M. (1999). IT-Intensive Value Innovation in the Electronic Economy: Insights from Marshall Industries. *MIS Quarterly*, *23*(3), 305. https://doi.org/10.2307/249466

Sayogo, D. S., Zhang, J., Pardo, T. A., Tayi, G. K., Hrdinova, J., Andersen, D. F., & Luna-Reyes, L. F. (2014). Going Beyond Open Data: Challenges and Motivations for Smart Disclosure in Ethical Consumption. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(2), 1–16. https://doi.org/10.4067/S0718-18762014000200002

Schlagwein, D., Conboy, K., Feller, J., Leimeister, J. M., & Morgan, L. (2017). "Openness" with and without Information Technology: a framework and a brief history. *Journal of Information Technology*, *32*(4), 297–305. https://doi.org/10.1057/s41265-017-0049-3

Schoenherr, T., & Speier-Pero, C. (2015). Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential. *Journal of Business Logistics*, *36*(1), 120–132. https://doi.org/10.1111/jbl.12082

Scholes, M., Benston, G. J., & Smith, C. W. (1976). A transactions cost approach to the theory of financial intermediation. *The Journal of Finance*, *31*(2), 215–231. https://doi.org/10.1111/j.1540-6261.1976.tb01882.x

Scholtens, B., & Van Wensveen, D. (2000). A critique on the theory of financial intermediation. *Journal of Banking & Finance*, *24*(8), 1243–1251.

Schultze, U., & Avital, M. (2011). Designing interviews to generate rich data for information systems research. *Information and Organization*, *21*(1), 1–16. https://doi.org/10.1016/j.infoandorg.2010.11.001

Schumpeter, J. A. (1934). The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle. Transaction Publishers.

Schweidel, D. A. (2014). Profiting from the Data Economy: Understanding the Roles of Consumers, Innovators and Regulators in a Data-Driven World (1 edition). Upper Saddle River, New Jersey: Pearson FT Press.

Seddon, P. B. (1997). A Respecification and Extension of the DeLone and McLean Model of IS Success. *Information Systems Research*, *8*(3), 240–253. https://doi.org/10.1287/isre.8.3.240

Seidel, S., Recker, J., & vom Brocke, J. (2013). Sensemaking and Sustainable Practicing: Functional Affordances of Information Systems in Green Transformations. *MIS Quarterly*, *37*(4), 1275-A10.

Seidel, S., & Urquhart, C. (2013). On emergence and forcing in information systems grounded theory studies: the case of Strauss and Corbin. *Journal of Information Technology*, *28*(3), 237–260. https://doi.org/10.1057/jit.2013.17

Seife, C. (2015, May 28). Who's to blame when fake science gets published? *Los Angeles Times*. Retrieved from http://www.latimes.com/nation/la-oe-0528-seife-scientific-research-process-20150528-story.html

Sen, A., Ramamurthy, K., & Sinha, A. P. (2012). A Model of Data Warehousing Process Maturity. *IEEE Transactions on Software Engineering*, *38*(2), 336–353. https://doi.org/10.1109/TSE.2011.2

Sen, Arun. (2004). Metadata management: past, present and future. *Decision Support Systems*, *37*(1), 151–173.

Sen, Arun, & Sinha, A. P. (2005). A comparison of data warehousing methodologies. *Communications of the ACM*, *48*(3), 79–84. https://doi.org/10.1145/1047671.1047673

Sensemaking Under Pressure: The Influence of Professional Roles and Social Accountability on the Creation of Sense. (2011). *Organization Science*, *23*(1), 118–137. https://doi.org/10.1287/orsc.1100.0640

Sernovitz, A. (2015). *Books I've Read Dataset*. Retrieved from https://data.world/sernovitz/books

Shah, S., Horne, A., & Capellá, J. (2012, April 1). Good Data Won't Guarantee Good Decisions. *Harvard Business Review*, (April 2012). Retrieved from https://hbr.org/2012/04/good-data-wont-guarantee-good-decisions

Shaikh, M., & Vaast, E. (2016). Folding and Unfolding: Balancing Openness and Transparency in Open Source Communities. *Information Systems Research*, *27*(4), 813–833. https://doi.org/10.1287/isre.2016.0646

Shapiro, C., & Varian, H. R. (1998). *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business Press.

Sharif, B., Lundin, R. M., Morgan, P., Hall, J. E., Dhadda, A., Mann, C., ... Szakmany, T. (2016). Developing a digital data collection platform to measure the prevalence of sepsis in Wales. *Journal of the American Medical Informatics Association*, *23*(6), 1185–1189. https://doi.org/10.1093/jamia/ocv208

Shattuck, J. (1988). The right to know: Public access to federal information in the 1980s. *Government Information Quarterly*, *5*(4), 369–375. https://doi.org/10.1016/0740-624X(88)90025-1

Sheth, A. (2003). Semantic Meta Data for Enterprise Information Integration. *DM Review*, 52–54.

Shim, D. C., & Eom, T. H. (2008). E-Government and Anti-Corruption: Empirical Analysis of International Data. *International Journal of Public Administration*, *31*(3), 298–316. https://doi.org/10.1080/01900690701590553

Shin, B. (2003). An Exploratory Investigation of System Success Factors in Data Warehousing. *Journal of the Association for Information Systems*, *4*(1), 141–170. https://doi.org/10.17705/1jais.00033

Shu, C. (2018, January). Why inclusion in the Google Arts & Culture selfie feature matters. Retrieved January 31, 2018, from TechCrunch website: http://social.techcrunch.com/2018/01/21/why-inclusion-in-the-google-arts-culture-selfie-feature-matters/

Shum, S. B., Clark, T., de Waard, A., Groza, T., Handschuh, S., & Sandor, A. (2010). Scientific discourse on the semantic web: A survey of models and enabling technologies. *Semantic Web Journal: Interoperability, Usability, Applicability*. Retrieved from http://semantic-web-journal.net/sites/default/files/swj127.pdf

Sieber, R. E., & Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly*, *32*(3), 308–315. https://doi.org/10.1016/j.giq.2015.05.003

Siegel, P. A., & Hambrick, D. C. (2005). Pay Disparities Within Top Management Groups: Evidence of Harmful Effects on Performance of High-Technology Firms. *Organization Science*, *16*(3), 259–274. https://doi.org/10.1287/orsc.1050.0128

Silva, T., Wuwongse, V., & Sharma, H. N. (2013). Disaster mitigation and preparedness using linked open data. *Journal of Ambient Intelligence and Humanized Computing*, *4*(5), 591–602. https://doi.org/10.1007/s12652-012-0128-9

Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology & Evolution*, *24*(9), 467–471. https://doi.org/10.1016/j.tree.2009.03.017

Simoudis, E. (1996). Reality check for data mining. *IEEE Expert*, *11*(5), 26–33. https://doi.org/10.1109/64.539014

Singleton, Alex David, Spielman, S., & Brunsdon, C. (2016). Establishing a framework for Open Geographic Information science. *International Journal of Geographical Information Science*, *30*(8), 1507–1521. https://doi.org/10.1080/13658816.2015.1137579

Singleton, Alexander D., & Spielman, S. E. (2014). The Past, Present, and Future of Geodemographic Research in the United States and United Kingdom. *Professional Geographer*, *66*(4), 558–567. https://doi.org/10.1080/00330124.2013.848764

Sivarajah, U., Weerakkody, V., Waller, P., Lee, H., Irani, Z., Choi, Y., … Glikman, Y. (2016). The role of e-participation and open data in evidence-based policy decision making in local government. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 64–79. https://doi.org/10.1080/10919392.2015.1125171

Small, B. (2014, May 27). FTC report examines data brokers. Retrieved January 15, 2018, from Federal Trade Commission Consumer Information website: https://www.consumer.ftc.gov/blog/2014/05/ftc-report-examines-data-brokers

Snyder, W. W. (1978). Horse Racing: Testing the Efficient Markets Model. *The Journal of Finance*, *33*(4), 1109–1118. https://doi.org/10.2307/2326943

Soldner, M. (2016). *TED@UPS: What If Data was Donated?* Retrieved from https://longitudes.ups.com/tedups-what-if-data-was-donated/

Spiegler, I. (2003). Technology and knowledge: bridging a “generating” gap. *Information & Management*, *40*(6), 533–539. https://doi.org/10.1016/S0378-7206(02)00069-1

Spiekermann, S., & Korunovska, J. (2017). Towards a value theory for personal data. *Journal of Information Technology*, *32*(1), 62–84. https://doi.org/10.1057/jit.2016.4

Spielman, S. E. (2017). The Potential for Big Data to Improve Neighborhood-Level Census Data. In P. Thakuriah, N. Tilahun, & M. Zellner (Eds.), *Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics* (pp. 99–111). Cham: Springer International Publishing Ag.

Spulber, D. F. (1996). Market Microstructure and Intermediation. *The Journal of Economic Perspectives*, *10*(3), 135–152.

Spulber, D. F. (1999). Market Microstructure: Intermediaries and the Theory of the Firm. Cambridge University Press.

Sreenivasan, R. R. (2017). *Characteristics of Big Data – A Delphi study* (Masters, Memorial University of Newfoundland). Retrieved from https://research.library.mun.ca/13080/1/thesis.pdf

Srinivasan, S., Barchas, I., Gorenberg, M., & Simoudis, E. (2014). Venture Capital: Fueling the Innovation Economy. *Computer*, *47*(8), 40–47. https://doi.org/10.1109/MC.2014.230

Staff, I. (2003, November 18). Economy. Retrieved January 30, 2018, from Investopedia website: https://www.investopedia.com/terms/e/economy.asp

Standing, S., Standing, C., & Love, P. E. D. (2010). A review of research on e-marketplaces 1997–2008. *Decision Support Systems*, *49*(1), 41–51. https://doi.org/10.1016/j.dss.2009.12.008

Storey, V. C., Burton-Jones, A., Sugumaran, V., & Purao, S. (2008). CONQUER: A Methodology for Context-Aware Query Processing on the World Wide Web. *Information Systems Research*, *19*(1), 3–25.

Storey, V. C., Dewan, R. M., & Freimer, M. (2012). Data quality: Setting organizational policies. *Decision Support Systems*, *54*(1), 434–442. https://doi.org/10.1016/j.dss.2012.06.004

Strauss, A., & Corbin, J. (1998). Basics of Qualitative Research: Second Edition: Techniques and Procedures for Developing Grounded Theory (2nd ed.). SAGE Publications, Inc.

Strauss, C. L. (1962). *Savage mind*. Retrieved from http://thenewschoolhistory.org/wp-content/uploads/2013/10/levistrauss_savagemind.pdf

Styhre, A. (2010). Organizing technologies of vision: Making the invisible visible in media-laden observations. *Information and Organization*, *20*(1), 64–78. https://doi.org/10.1016/j.infoandorg.2010.01.002

Suhaka, K., & Tauberer, J. (2012, July). *Business Models for Reuse of Open Legislative Data*. Presented at the IOGDC 2012 Virtual Conference: International Open Government Data Conference, Washington, DC.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D., & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, *142*(10), 2282–2292. https://doi.org/10.1016/j.biocon.2009.05.006

Sullivan, M. (2012). Data Snatchers! *PCWorld*, *30*(8), 77–85.

Sung, A. H., Ribeiro, B., & Liu, Q. (2016). *Sampling and Evaluating the Big Data for Knowledge Discovery* (M. Ramachandran, G. Wills, R. Walters, V. M. Munoz, & V. Chang, Eds.). Setubal: Scitepress.

Sunlight Foundation. (2017, February 6). How Data Refuge works, and how YOU can help save federal open data. Retrieved February 10, 2018, from Sunlight Foundation website: https://sunlightfoundation.com/2017/02/06/how-data-refuge-works-and-how-you-can-help-save-federal-open-data/

Survey: How access to data affects trust in news - dataset by ddjdemos. (2017). Retrieved February 7, 2018, from data.world website: https://data.world/ddjdemos/survey-how-access-to-data-affects-trust-in-news

Susha, I., Zuiderwijk, A., Janssen, M., & Grönlund, Å. (2015). Benchmarks for Evaluating the Progress of Open Data Adoption: Usage, Limitations, and Lessons Learned. *Social Science Computer Review*, *33*(5), 613–630. https://doi.org/10.1177/0894439314560852

Syngenta publishes data to unlock environmental, social and economic value. (2015). Retrieved November 6, 2016, from http://www4.syngenta.com/media/media-releases/yr-2015/23-04-2015

Szulanski, G., Ringov, D., & Jensen, R. J. (2016). Overcoming Stickiness: How the Timing of Knowledge Transfer Methods Affects Transfer Difficulty. Organization Science, 27(2), 304–322. https://doi.org/10.1287/orsc.2016.1049

Tableau Software. (n.d.). Medium Data is the New Sweet Spot. Retrieved April 23, 2018, from Tableau Software website: https://www.tableau.com/medium-data

Tan, B., Pan, S., Lu, X., & Huang, L. (2015). The Role of IS Capabilities in the Development of Multi-Sided Platforms: The Digital Ecosystem Strategy of Alibaba.com. *Journal of the Association for Information Systems*, *16*(4), 248–280. https://doi.org/10.17705/1jais.00393

Tempini, N. (2017). Till data do us part: Understanding data-based value creation in data-intensive infrastructures. *Information and Organization*, *27*(4), 191–210. https://doi.org/10.1016/j.infoandorg.2017.08.001

Templier, M., & Paré, G. (2015). A Framework for Guiding and Evaluating Literature Reviews. Communications of the Association for Information Systems, 37(1). Retrieved from http://aisel.aisnet.org/cais/vol37/iss1/6

Tene, O., & Polonetsky, J. (2012). Privacy in the Age of Big Data. *Stanford Law Review*. Retrieved from https://www.stanfordlawreview.org/online/privacy-paradox-privacy-and-big-data/

Tene, O., & Polonetsky, J. (2013). Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property*, *11*, xxvii.

Thakur, R. (2014). What keeps mobile banking customers loyal? *International Journal of Bank Marketing*, *32*(7), 628–646. https://doi.org/10.1108/IJBM-07-2013-0062

Thakuriah, Piyushimita, Dirks, L., & Keita, Y. M. (2017). Digital Infomediaries and Civic Hacking in Emerging Urban Data Initiatives. In P. Thakuriah, N. Tilahun, & M. Zellner (Eds.), *Seeing Cities Through Big Data: Research, Methods and Applications in Urban Informatics* (pp. 189–207). Cham: Springer International Publishing Ag.

The 8 Principles of Open Government Data (OpenGovData.org). (2016). Retrieved November 19, 2016, from https://opengovdata.org/

The Association of Religion Data Archives. (2017). The Association of Religion Data Archives. Retrieved February 10, 2018, from Quality Data on Religion website: http://www.thearda.com/

The Economist. (2017a, May 6). Fuel of the future; The data economy. *The Economist, 423*(9039), 22.

The Economist. (2017b, May 6). The world's most valuable resource; Regulating the data economy. *The Economist, 423*(9039), 9.

The Guardian. (2018). The Guardian Data Blog. Retrieved February 7, 2018, from The Guardian website: http://www.theguardian.com/data

The Open Community Data Exchange: Advancing Data Sharing and Discovery in Open Online Community Science. (2017). Retrieved January 24, 2017, from https://www.researchgate.net/publication/310424606_The_Open_Community_Data_Exchange_Advancing_Data_Sharing_and_Discovery_in_Open_Online_Community_Science

The Top 20 Venture Capital Investors Worldwide. (2016, March 13). *The New York Times*. Retrieved from http://www.nytimes.com/interactive/2016/03/13/technology/venture-capital-investor-top-20.html

Thibodeaux, T. (2018). AP data is more valuable on data.world. 18.

Thirani, V. (2017). The value of data. Retrieved December 5, 2017, from World Economic Forum website: https://www.weforum.org/agenda/2017/09/the-value-of-data/

Thomas, S., & Rosenman, D. (2006, March). Information Overload. Retrieved September 27, 2017, from The Hospitalist website: http://www.the-hospitalist.org/hospitalist/article/123083/information-overload

Thomé, H. (2017). *Businesses in Europe and the data economy* (No. 423; p. 5). Robert Schuman Foundation.

Thompson, C. (2013, October 30). What companies are doing with your intimate social data. Retrieved October 3, 2017, from CNBC website: https://www.cnbc.com/2013/10/30/what-companies-are-doing-with-your-intimate-social-data.html

Thorsby, J., Stowers, G. N. L., Wolslegel, K., & Tumbuan, E. (2017). Understanding the content and features of open data portals in American cities. *Government Information Quarterly, 34*(1), 53–61. https://doi.org/10.1016/j.giq.2016.07.001

Tiwana, A. (2015). Platform Desertion by App Developers. *Journal of Management Information Systems, 32*(4), 40–77. https://doi.org/10.1080/07421222.2015.1138365

Tolbert, C. J., & Mossberger, K. (2006). The Effects of E-Government on Trust and Confidence in Government. *Public Administration Review, 66*(3), 354.

Tomic, S. D. K., & Fensel, A. (2013). OpenFridge: A Platform for Data Economy for Energy Efficiency Data. In X. Hu, T. Y. Lin, V. Raghavan, B. Wah, R. BaezaYates, G. Fox, ... R. Nambiar (Eds.), *2013 IEEE International Conference on Big Data* (pp. 43–47). https://doi.org/10.1109/BigData.2013.6691686

Tow Center. (2013a, March 1). Data Journalism Resources. Retrieved March 21, 2019, from Medium website: https://medium.com/tow-center/data-journalism-resources-e075f53387b

Tow Center. (2013b, August 6). Algorithmic Defamation: The Case of the Shameless Autocomplete. Retrieved March 21, 2019, from Medium website: https://medium.com/tow-center/algorithmic-defamation-the-case-of-the-shameless-autocomplete-11e686ef7555

Tow Center. (2013c, September 17). Storytelling with Data Visualization: Context is King. Retrieved March 21, 2019, from Medium website: https://medium.com/tow-center/storytelling-with-data-visualization-context-is-king-ff25e6165d1e

Tow Center. (2014a, April 27). OpenVis is for Journalists! Retrieved March 21, 2019, from Medium website: https://medium.com/tow-center/openvis-is-for-journalists-e87a84354de3

Tow Center. (2014b, June 12). The Anatomy of a Robot Journalist. Retrieved March 21, 2019, from Medium website: https://medium.com/tow-center/the-anatomy-of-a-robot-journalist-1aa879bc6769

Tow Center. (2015a, October 20). Tow Tea: Computational Journalism in Practice. Retrieved March 21, 2019, from Medium website: https://medium.com/tow-center/tow-tea-computational-journalism-in-practice-91fc16ef5f7a

Tow Center. (2015b, November 10). Tow Tea: Parsing Tech Talk. Retrieved March 21, 2019, from Medium website: https://medium.com/tow-center/tow-tea-parsing-tech-talk-e4bc92ea4c5e

Tow Center. (2015c, December 28). Reducing barriers between programmers and non-programmers in the newsroom. Retrieved March 21, 2019, from Medium website: https://medium.com/tow-center/reducing-barriers-between-programmers-and-non-programmers-in-the-newsroom-eb696ae3341e

Tow Center. (2016a, July 12). Data Cleaning With Muck. Retrieved March 21, 2019, from Medium website: https://medium.com/tow-center/data-cleaning-with-muck-56821320ed6b

Tow Center. (2016b, July 14). What will the Internet of Things do to journalism? Retrieved March 21, 2019, from Medium website: https://medium.com/tow-center/what-will-the-internet-of-things-do-to-journalism-2ec2b453eaf2

Trauth, E. (2000). Qualitative Research in IS: Issues and Trends: Issues and Trends. Idea Group Inc (IGI).

trends.google.com. (2018). Google Trends. Retrieved February 16, 2018, from Google Trends website: https://g.co/trends/FDVVW

Tsang, E. W. K. (1999). Replication and Theory Development in Organizational Science: A Critical Realist Perspective. *Academy of Management Review*, *24*(4), 759–780. https://doi.org/10.5465/AMR.1999.2553252

Tsang, Eric W.K. (2014). Case studies and generalization in information systems research: A critical realist perspective. *The Journal of Strategic Information Systems*, *23*(2), 174–186. https://doi.org/10.1016/j.jsis.2013.09.002

Tsianos, V. S., & Kuster, B. (2016). Eurodac in Times of Bigness: The Power of Big Data within the Emerging European IT Agency. *Journal of Borderlands Studies*, *31*(2), 235–249. https://doi.org/10.1080/08865655.2016.1174606

Turner, F. (2006). *From Counterculture to Cyberculture*. Retrieved from http://www.press.uchicago.edu/ucp/books/book/chicago/F/bo3773600.html

Ullmann, A. A. (1985). Data in Search of a Theory: A Critical Examination of the Relationships Among Social Performance, Social Disclosure, and Economic Performance of U.S. Firms. *Academy of Management Review*, *10*(3), 540–557. https://doi.org/10.5465/AMR.1985.4278989

Update: More Than 400 Incidents of Hateful Harassment and Intimidation Since the Election. (2016). Retrieved November 19, 2016, from Southern Poverty Law Center website: https://www.splcenter.org/hatewatch/2016/11/15/update-more-400-incidents-hateful-harassment-and-intimidation-election

Ure, J., Procter, R., Lin, Y., Hartswood, M., Anderson, S., Lloyd, S., ... Ho, K. (2009). The Development of Data Infrastructures for eHealth: A Socio-Technical Perspective. *Journal of the Association for Information Systems*, *10*(5), 415–429. https://doi.org/10.17705/1jais.00197

Urquhart, C. (2013). Grounded theory for qualitative research: a practical guide. Los Angeles, Calif. ; London: SAGE.

Urquhart, C., & Fernández, W. (2013). Using grounded theory method in information systems: the researcher as blank slate and other myths. *Journal of Information Technology*, *28*(3), 224–236. http://dx.doi.org/10.1057/jit.2012.34

Urquhart, C., Lehmann, H., & Myers, M. D. (2010). Putting the 'theory' back into grounded theory: guidelines for grounded theory studies in information systems. *Information Systems Journal*, *20*(4), 357–381. https://doi.org/10.1111/j.1365-2575.2009.00328.x

Urquhart, C., & Vaast, E. (2012). Building social media theory from case studies: A new frontier for IS research. *ICIS 2012 Proceedings*. Retrieved from http://aisel.aisnet.org/icis2012/proceedings/ResearchMethods/8

US Department of Education. (2017). No Child Left Behind - ED.gov. Retrieved February 16, 2018, from No Child Left Behind Act website: https://www2.ed.gov/nclb/landing.jhtml

Vaghely, I. P., & Julien, P.-A. (2010). Are opportunities recognized or constructed? *Journal of Business Venturing*, *25*(1), 73–86. https://doi.org/10.1016/j.jbusvent.2008.06.004

van Schalkwyk, F., Willmers, M., & McNaughton, M. (2016). Viscous Open Data: The Roles of Intermediaries in an Open Data Ecosystem. *Information Technology for Development*, *22*, 68–83. https://doi.org/10.1080/02681102.2015.1081868

Vassilakopoulou, P., Skorve, E., & Aanestad, M. (2018). Enabling openness of valuable information resources: Curbing data subtractability and exclusion. *Information Systems Journal*, *0*(0), 1–19. https://doi.org/10.1111/isj.12191

Veit, D., & Huntgeburth, J. (2014). Open Government. In *Foundations of Digital Government: Leading and Managing in the Digital Era* (pp. 87–100). Berlin: Springer-Verlag Berlin.

Veljković, N., Bogdanović-Dinić, S., & Stoimenov, L. (2014). Benchmarking open government: An open data perspective. *Government Information Quarterly*, *31*(2), 278–290. https://doi.org/10.1016/j.giq.2013.10.011

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly*, *37*(1), 21–54.

Vertesi, J., & Dourish, P. (2011). The Value of Data: Considering the Context of Production in Data Economies. *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, 533–542. https://doi.org/10.1145/1958824.1958906

Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, *33*(2), 325–337. https://doi.org/10.1016/j.giq.2016.02.001

Vigliarolo, B. (2019). Big data adoption exploding, but companies struggle to extract meaningful information. Retrieved March 20, 2019, from TechRepublic website: https://www.techrepublic.com/article/big-data-adoption-exploding-but-companies-struggle-to-extract-meaningful-information/

Volkoff, O., & Strong, D. M. (2013). Critical Realism and Affordances: Theorizing It-Associated Organizational Change Processes. *MIS Quarterly*, *37*(3), 819–834.

Vorhies, W. (2014, October 31). How Many "V's" in Big Data? The Characteristics that Define Big Data. Retrieved March 15, 2019, from Data Science Central website: https://www.datasciencecentral.com/profiles/blogs/how-many-v-s-in-big-data-the-characteristics-that-define-big-data

Vostrovsky, V., Tyrychtr, J., & Halbich, C. (2015). Correctness of open data in the agricultural sector. In L. Smutka & H. Rezbova (Eds.), *Agrarian Perspectives Xxiv: Global Agribusiness and the Rural Economy* (pp. 528–535). Prague 6: Czech University Life Sciences Prague.

W3C. (2019). LinkedData - W3C Wiki. Retrieved March 17, 2019, from https://www.w3.org/wiki/LinkedData

Wadman, M. (2017, February 3). USDA blacks out animal welfare information. Retrieved February 6, 2017, from Science website: http://www.sciencemag.org/news/2017/02/trump-administration-blacks-out-animal-welfare-information

Waitelonis, J., & Sack, H. (2012). Towards exploratory video search using linked data. *Multimedia Tools and Applications*, *59*(2), 645–672. https://doi.org/10.1007/s11042-011-0733-1

Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLOS ONE*, *8*(7), e67332. https://doi.org/10.1371/journal.pone.0067332

Walsham, G. (1995). Interpretive case studies in IS research: nature and method. *European Journal of Information Systems*, *4*(2), 74–81. https://doi.org/10.1057/ejis.1995.9

Walsham, Geoff. (2006). Doing interpretive research. *European Journal of Information Systems*, *15*(3), 320–330. https://doi.org/10.1057/palgrave.ejis.3000589

Wang, R. "Ray." (2012, February 27). Monday's Musings: Beyond The Three V's of Big Data - Viscosity and Virality. Retrieved March 15, 2019, from Forbes website: https://www.forbes.com/sites/raywang/2012/02/27/mondays-musings-beyond-the-three-vs-of-big-data-viscosity-and-virality/

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099

Wang, Y.-S., & Liao, Y.-W. (2008). Assessing eGovernment systems success: A validation of the DeLone and McLean model of information systems success. *Government Information Quarterly*, *25*(4), 717–733. https://doi.org/10.1016/j.giq.2007.06.002

Ward, M. (2017, March 21). How fake data could lead to failed crops and other woes. *BBC News*. Retrieved from http://www.bbc.com/news/business-38254362

Washington, 1440 G. Street NW, & Skype, D. 20005 202-742-1520 C. with. (2018). Sunlight Foundation. Retrieved February 6, 2018, from Sunlight Foundation website: http://sunlightfoundation.com

Wasko, M., Faraj, S., & Teigland, R. (2004). Collective Action and Knowledge Contribution in Electronic Networks of Practice. *Journal of the Association for Information Systems*, *5*(11), 493–513. https://doi.org/10.17705/1jais.00058

Wasko, M McLure, & Faraj, S. (2000). It is what one does: why people participate and help others in electronic communities of practice. *Journal of Strategic Information Systems*, 19.

Wasko, Molly McLure, & Faraj, S. (2005). Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice1. *MIS Quarterly*, *29*(1), 35–57.

Waters, R. (2016, July 11). US start-up aims to steer through flood of data. *Financial Times*. Retrieved from http://www.ft.com/cms/s/0/e11c098c-46a7-11e6-8d68-72e9211e86ab.html?ftcamp=published_links%2Frss%2Fcompanies_technology%2Ffeed%2F%2Fproduct#axzz4JnXG8Y26

Watson, H., Ariyachandra, T., & Matyska, R. J. (2001). Data Warehousing Stages of Growth. *Information Systems Management*, *18*(3), 42–50. https://doi.org/10.1201/1078/43196.18.3.20010601/31289.6

Watson, H. J. (2002). Recent Developments in Data Warehousing. *Communications of the Association for Information Systems*, *8*. https://doi.org/10.17705/1CAIS.00801

Watson, H. J., Annino, D. A., Wixom, B. H., Avery, K. L., & Rutherford, M. (2001). Current Practices in Data Warehousing. *Information Systems Management*, *18*(1), 47–55. https://doi.org/10.1201/1078/43194.18.1.20010101/31264.6

Watson, H. J., & Gray, P. (1997). *Decision Support in the Data Warehouse* (1 edition). Upper Saddle River, NJ: Prentice Hall.

Webber, R. (2015). MRS Census and Geodemographic Group (CGG) Conference: Harnessing Open Data for Business Advantage. *International Journal of Market Research*, *57*, 145–149. https://doi.org/10.2501/IJMR-2015-008

Weber, N. M., Baker, K. S., Thomer, A. K., Chao, T. C., & Palmer, C. L. (2012). Value and context in data use: Domain analysis revisited. *Proceedings of the American Society for Information Science and Technology*, *49*(1), 1–10. https://doi.org/10.1002/meet.14504901168

Weber, T. A. (2014). Intermediation in a Sharing Economy: Insurance, Moral Hazard, and Rent Extraction. *Journal of Management Information Systems*, *31*(3), 35–71. https://doi.org/10.1080/07421222.2014.995520

Weber, T. A., & Zheng, Z. (Eric). (2007). A Model of Search Intermediaries and Paid Referrals. *Information Systems Research*, *18*(4), 414–436. https://doi.org/10.1287/isre.1070.0139

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly, 26(2), XIII–XXIII.

Weerakkody, V., Irani, Z., Kapoor, K., Sivarajah, U., & Dwivedi, Y. K. (2017). Open data and its usability: an empirical view from the Citizen's perspective. *Information Systems Frontiers*, *19*(2), 285–300. https://doi.org/10.1007/s10796-016-9679-1

Welcome | B Corporation. (2017). Retrieved November 4, 2016, from https://www.bcorporation.net/

Wenger, E. (1999). Communities of Practice: Learning, Meaning, and Identity. Cambridge University Press.

Wenger, E. (2010). Communities of Practice and Social Learning Systems: the Career of a Concept. In C. Blackmore (Ed.), *Social Learning Systems and Communities of Practice* (pp. 179–198). https://doi.org/10.1007/978-1-84996-133-2_11

Wenger, E. C., & Snyder, W. M. (2000). Communities of Practice: The Organizational Frontier. *Harvard Business Review*, *78*(1), 139–145.

Wenger, E., McDermott, R. A., & Snyder, W. (2002). *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Harvard Business Press.

West, J., & Gallagher, S. (2006). Challenges of open innovation: the paradox of firm investment in open-source software. *R&D Management*, *36*(3), 319–331. https://doi.org/10.1111/j.1467-9310.2006.00436.x

Westlake, A. (2006). Provenance and Reliability: Managing Metadata for Statistical Models. *18th International Conference on Scientific and Statistical Database Management (SSDBM'06)*, 204–216. https://doi.org/10.1109/SSDBM.2006.42

What is the Freedom of Information Act? (2017, July 28). Retrieved November 28, 2017, from https://ico.org.uk/for-organisations/guide-to-freedom-of-information/what-is-the-foi-act/

Whitmore, A. (2014). Using open government data to predict war: A case study of data and systems challenges. *Government Information Quarterly*, *31*(4), 622–630. https://doi.org/10.1016/j.giq.2014.04.003

Wiesche, M., Jurisch, M. C., Yetton, P. W., & Krcmar, H. (2017). Grounded Theory Methodology in Information Systems Research. *MIS Quarterly*, *41*(3), 685-A9.

Wikimedia. (2018). Retrieved January 17, 2018, from https://www.wikimedia.org/

Wikimedia Commons. (2017). Retrieved January 17, 2018, from Wikimedia website: https://commons.wikimedia.org/wiki/Main_Page

Wikimedia.org. (2017). Wikimedia. Retrieved February 16, 2018, from https://www.wikimedia.org/

Wikipedia. (2017). Data collection. In *Wikipedia*. Retrieved from https://en.wikipedia.org/w/index.php?title=Data_collection&oldid=813346585

Wladawsky-Berger, I. (2017, May 26). The Rise of the Data Economy Is Triggering More Powerful Network Effects. Retrieved October 3, 2017, from Wall Street Journal website: https://blogs.wsj.com/cio/2017/05/26/the-rise-of-the-data-economy-is-triggering-more-powerful-network-effects/

Wood, M. (2017, June 7). Can data.world become the Github of enterprise data sharing? Retrieved January 31, 2018, from Diginomica website: https://diginomica.com/2017/06/07/can-data-world-become-github-data-sharing/

World Wide Web Foundation. (2015). *The Open Data Barometer Global Report* (No. 3; pp. 1–48). Retrieved from Open Data for Development and the World Wide Web Foundation website: http://opendatabarometer.org/doc/3rdEdition/ODB-3rdEdition-GlobalReport.pdf

Worthy, B. (2015). The Impact of Open Data in the Uk: Complex, Unpredictable, and Political. *Public Administration*, *93*(3), 788–805. https://doi.org/10.1111/padm.12166

Wray, L. R. (1998). *Understanding modern money* (Vol. 11). Cheltenham: Edward Elgar.

Wu, Y. (2014). Protecting personal data in E-government: A cross-country study. *Government Information Quarterly*, *31*(1), 150–159. https://doi.org/10.1016/j.giq.2013.07.003

Wunderground.com. (2018). API | Weather Underground. Retrieved February 3, 2018, from Weather Underground website: https://www.wunderground.com/weather/api/d/pricing.html

Xia, S. (2017). E-Governance and Political Modernization: An Empirical Study Based on Asia from 2003 to 2014. *Administrative Sciences*, *7*(3), 25. http://dx.doi.org/10.3390/admsci7030025

Yang, M., Adomavicius, G., Burtch, G., & Ren, Y. (2018). Mind the Gap: Accounting for Measurement Error and Misclassification in Variables Generated via Data Mining. *Information Systems Research*, *29*(1), 4–24. https://doi.org/10.1287/isre.2017.0727

Yang, T.-M., Lo, J., & Shiang, J. (2015). To open or not to open? Determinants of open government data. *Journal of Information Science*, *41*(5), 596–612. https://doi.org/10.1177/0165551515586715

Yi, J. (2009). A measure of knowledge sharing behavior: scale development and validation. *Knowledge Management Research & Practice*, *7*(1), 65–81. https://doi.org/10.1057/kmrp.2008.36

Yin, R. K. (1984). *Case study research: design and methods*. Sage Publications.

Yin, R. K. (2003). Case Study Research: Design and Methods. SAGE.

Yin, R. K. (2004). *Case Study Research and Methods*. Retrieved from http://cemusstudent.se/wp-content/uploads/2012/02/YIN_K_ROBERT-2.pdf

Yoo, B., Choudhary, V., & Mukhopadhyay, T. (2002). A model of neutral B2B intermediaries. *Journal of Management Information Systems*, *19*(3), 43–68.

Yoo, Y., Henfridsson, O., & Lyytinen, K. (2010). Research Commentary: The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research. *Information Systems Research*, *21*(4), 724–735. Retrieved from JSTOR.

Zech, H. (2016). A legal framework for a data economy in the European Digital Single Market: rights to use data. *Journal of Intellectual Property Law & Practice*, *11*(6), 460–470. https://doi.org/10.1093/jiplp/jpw049

Zech, H. (2017). Building a European Data Economy. *International Review of Intellectual Property and Competition Law*, *48*(5), 501–503. https://doi.org/10.1007/s40319-017-0604-z

Zeleti, F. A., Ojo, A., & Curry, E. (2014). Emerging Business Models for the Open Data Industry: Characterization and Analysis. *Proceedings of the 15th Annual International Conference on Digital Government Research*, 215–226. https://doi.org/10.1145/2612733.2612745

Zeleti, F. A., Ojo, A., & Curry, E. (2016). Exploring the economic value of open government data. *Government Information Quarterly*, *33*(3), 535–551. https://doi.org/10.1016/j.giq.2016.01.008

Zhang, P., Zhu, X., Shi, Y., Guo, L., & Wu, X. (2011). Robust ensemble learning for mining noisy data streams. *Decision Support Systems*, *50*(2), 469–479. https://doi.org/10.1016/j.dss.2010.11.004

Zhao, X. (2012). Service design of consumer data intermediary for competitive individual targeting. *Decision Support Systems*, *54*(1), 699–718. https://doi.org/10.1016/j.dss.2012.08.015

Zhu, H., & Wu, H. (2014). Assessing the quality of large-scale data standards: A case of XBRL GAAP Taxonomy. *Decision Support Systems*, *59*, 351–360. https://doi.org/10.1016/j.dss.2014.01.006

Zhu, X. (2017). The failure of an early episode in the open government data movement: A historical case study. *Government Information Quarterly*, *34*(2), 256–269. https://doi.org/10.1016/j.giq.2017.03.004

Zikopoulos, P., & Eaton, C. (2011). Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data (1st ed.). McGraw-Hill Osborne Media.

Zillow Research. (2017). Zillow Research. Retrieved February 16, 2018, from Zillow Research website: https://www.zillow.com/research/

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, *30*(1), 75–89. https://doi.org/10.1057/jit.2015.5

Zuiderwijk, A., Helbig, N., Gil-García, J. R., & Janssen, M. (2014). Special Issue on Innovation through Open Data - A Review of the State-of-the-Art and an Emerging Research Agenda: Guest Editors' Introduction. *Journal of Theoretical and Applied Electronic Commerce Research; Curicó*, *9*(2), 1–8.

Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, *31*(1), 17–29. https://doi.org/10.1016/j.giq.2013.04.003

Zuiderwijk, A., Janssen, M., & Dwivedi, Y. K. (2015). Acceptance and use predictors of open data technologies: Drawing upon the unified theory of acceptance and use of technology. *Government Information Quarterly*, *32*(4), 429–440. https://doi.org/10.1016/j.giq.2015.09.005

Zuiderwijk, A., Janssen, M., van de Kaa, G., & Poulis, K. (2016). The wicked problem of commercial value creation in open data ecosystems: Policy guidelines for governments. *Information Polity: The International Journal of Government & Democracy in the Information Age*, *21*(3), 223–236. https://doi.org/10.3233/IP-160391