RESEARCH PAPER

Understanding Perspectives on Sharing Neutron Data at Oak Ridge National Laboratory

Devan Ray Donaldson[1], Shawn Martin[1] and Thomas Proffen[2]

[1] Department of Information and Library Science, School of Informatics and Computing, Indiana University, Bloomington, Indiana, US

[2] Neutron Data Analysis and Visualization, Neutron Sciences Directorate, Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, Tennessee, US

Corresponding author: Dr. Devan Ray Donaldson, Assistant Professor (drdonald@indiana.edu)

Even though the importance of sharing data is frequently discussed, data sharing appears to be limited to a few fields, and practices within those fields are not well understood. This study examines perspectives on sharing neutron data collected at Oak Ridge National Laboratory's neutron sources. Operation at user facilities has traditionally focused on making data accessible to those who create them. The recent emphasis on open data is shifting the focus to ensure that the data produced are reusable by others. This mixed methods research study included a series of surveys and focus group interviews in which 13 data consumers, data managers, and data producers answered questions about their perspectives on sharing neutron data. Data consumers reported interest in reusing neutron data for comparison/verification of results against their own measurements and testing new theories using existing data. They also stressed the importance of establishing context for data, including how data are produced, how samples are prepared, units of measurement, and how temperatures are determined. Data managers expressed reservations about reusing others' data because they were not always sure if they could trust whether the people responsible for interpreting data did so correctly. Data producers described concerns about their data being misused, competing with other users, and over-reliance on data producers to understand data. We present the Consumers Managers Producers (CMP) Model for understanding the interplay of each group regarding data sharing. We conclude with policy and system recommendations and discuss directions for future research.

Keywords: Scientific data reuse; data sharing; neutron data; data consumers; data managers; data producers

1 Introduction

Data sharing is a worldwide desideratum. Several community and policy as well as funder recommendations stress the importance of sharing and reusing data (Hodson & Molloy 2015). Yet, a mix of various socio-technical barriers and policies inhibit people from sharing data across technologies, disciplines, and countries. In response to this challenge, the Research Data Alliance (RDA), a group comprising over 4,000 members from 110 countries, was formed to build the social and technical bridges that enable open sharing of data. In 2014, the RDA Europe produced a report entitled, *The Data Harvest: How Sharing Research Data Can Yield Knowledge, Jobs, and Growth*. The report advocates an international effort centering on seven actions (RDA Europe 2014: 6–7): 1) creating and implementing data plans, 2) promoting data literacy across society, from researchers to citizens, 3) developing incentives and grants for data sharing, 4) developing tools and practices to build trust and data-sharing, 5) supporting international collaboration, 6) thoughtful regulation, and 7) sustaining successful projects. These actions are all critically important because, "it isn't just the existence of … data that counts: it's the ability to share and re-use the data across disciplines and institutions and countries" (RDA Europe 2014: 10).

This study advances the seven actions that RDA recommends with research on data sharing perspectives and practices. Specifically, we examine perspectives on data reuse for data that require a substantial amount of federal funds to produce. We argue that the cost associated with creating data could be further justified if researchers within and across disciplines would be willing to share and reuse those data. A particularly interesting case are data collected at national user facilities, in the case of this study, data collected at the Spallation Neutron Source (SNS) and the High Flux Isotope Reactor (HFIR) at Oak Ridge National Laboratory (ORNL). SNS and HFIR host over 3,000 users annually who collect neutron scattering data related to their own science projects. With the increasing amount of data generated by experimental user facilities and advanced materials modeling enabled by high performance computing, the value of neutron scattering data to scientists beyond those who initially produce the data is gaining attention. Experimental facilities such as SNS and HIFR are valuable and expensive science resources; the impact of the collected data can be significantly increased by reuse (Fienberg, Martin, & Straf 1985; Hodson & Molloy 2015; NAS 2009; RDA Europe 2014; RSSPC 2012; SI 2015). According to several US national laboratories' user agreements, the user (i.e., the experimenter) owns the data they produce and is ultimately responsible for making them publicly available to comply with a recent executive order by the Obama administration (Obama, 2013). As such, user facilities in the US are actively working with their user communities to help enable reuse of experimental data.

This mixed methods research study included a series of surveys and focus group interviews in which 13 data consumers, data managers, and data producers answered questions about their views on sharing neutron data. The research question guiding the study is: How do various classes of stakeholders (e.g., data consumers, data managers, and data producers) think about data sharing at an organization that has not typically focused on data reuse?

This paper is structured as follows. First, we provide background on data sharing research, including barriers to data reuse that multiple researchers have identified. Second, we describe our mixed methods research design and approach to the study. Third, we discuss findings resulting from analysis of the survey and focus group data we collected. We also present a new model that we developed as a result of our data analysis, the Consumers Managers Producers (CMP) Model. The CMP Model illustrates the interaction of different classes of stakeholders regarding data sharing. Fourth, we discuss how our findings relate to the existing literature on data sharing, provide recommendations for facilitating data reuse at ORNL and similar institutions, and conclude by exploring directions for future research.

2 Background

For decades, scholars from across a broad variety of academic disciplines have written about the topic of data reuse. In fact, the possibility of data reuse has been hailed by some as offering "a vast potential for scientific progress," because with open data, new avenues of research can be explored and new questions asked. Reuse also provides the potential for reproducing research results, allowing those data to apply to new situations (Fecher, Friesike, & Hebing 2015: 1). Yet, despite such enormous possibilities, there remain in some cases very significant barriers to reusing data (van Panhuis et al. 2014). Such problems are particularly acute in interdisciplinary research (Akers & Doty 2013), and are of interest to governments seeking to maximize national investments in research infrastructure (Dallmeier-Tiessen et al. 2014). In response to such tremendous interest in reuse of data, researchers have begun to study the reasons why the potential of data sharing has not yet been fully realized.

Literature within the field of data reuse is scattered among many different disciplines. For example, fields as diverse as information science, biology, medicine, public health, computer science, ecology, and geology have all investigated the barriers to sharing data both within particular disciplines and across disciplinary boundaries (Akers & Doty 2013; Coetzee 2015; Dallmeier-Tiessen et al. 2014; Duke & Porter 2013; Elwood 2008; Enke et al. 2012; Faniel, Kriesberg, & Yakel 2016; Federer et al. 2015; Hsu et al. 2015; McLure et al. 2014; Michener 2015; Nahar 2016; Pericas, Taura & Matsuoka 2014; Piwowar & Chapman 2010; Specht et al. 2015; van Panhuis et al. 2014; Wynholds et al. 2012). Additionally, there has been increased interest in the problems associated with data reuse because of increased emphasis on issues related to "big data" (Jin et al. 2015; Lee & Kang 2015). Overall, researchers who have studied obstacles to data reuse have identified three broad categories: social barriers, technical barriers, and legal barriers.

Social barriers, outlined in **Table 1**, include all of the difficulties that result from humans working together. Examples include institutional blocks when universities and governmental bodies have differing agendas. Additionally, a social barrier might involve an instance when two different disciplines have problems communicating about similar data. Within the academic literature on data reuse, social barriers seem to be the most mentioned, and the most important. Particular subcategories of social barriers include disciplinary

Table 1: Literature Citing Social Barriers to Data Reuse.

Issue	Literature Mentioned
Differences between disciplines in creating data	Akers & Doty 2013; Cieri 2014; Collins 1998; Elwood 2008; Fecher, Friesike, & Hebing 2015; Federer et al. 2015; Hsu et al. 2015; Jin et al. 2015; Lee & Kang 2015; Michener 2015; Nahar 2016; SI 2015; Specht et al. 2015; Tenopir et al. 2015; Treloar 2014; van Panhuis et al. 2014; Wynholds et al. 2012
Lack of incentives to share or reuse data	Cieri 2014; Coetzee 2015; Collins 1998; Dallmeier-Tiessen et al. 2014; Duke & Porter 2013; Enke et al. 2012; Fecher, Friesike, & Hebing 2015; Hsu et al. 2015; Michener 2015; Nahar 2016; NAS 2009; RSSPC 2012; Piwowar & Chapman 2010; Savage & Vickers, 2009; Tenopir et al. 2015; Treloar 2014; van Panhuis et al. 2014
Differences in quality of data for particular research purposes	Dallmeier-Tiessen et al. 2014; Elwood 2008; Faniel, Kriesberg, & Yakel 2016; Fecher, Friesike, & Hebing 2015; Michener 2015; Pepe et al. 2014; Sayogo & Pardo 2013; Wynholds et al. 2012
Lack of education about best practices	Enke et al. 2012; Hsu et al. 2015; Pepe et al. 2014; Sayogo & Pardo 2013; Specht et al. 2015
Journals incentivize different kinds of data sharing	Cieri 2014; Piwowar & Chapman 2010; Savage & Vickers, 2009; Treloar 2014; Wallis, Rolando & Borgman 2013
Lack of standards for peer review of data	Hsu et al. 2015; Michener 2015; Wynholds et al. 2012
Misuse of research data	Borgman 2012; Campbell et al. 2002; Hilgartner 1997; Hilgartner & Brandt-Rauf 1994; Savage & Vickers, 2009
Competing interests of researchers	Borgman 2012; Campbell et al. 2002; National Research Council, 1997; Hilgartner & Brandt-Rauf 1994; Ure et al. 2009
Too much time required to share data	Campbell et al. 2002; Fecher, Friesike, & Hebing 2015; Hilgartner & Brandt-Rauf 1994; NAS 2009; Savage & Vickers, 2009; van Panhuis et al. 2014

differences in creating data, a lack of incentive to share data, differences in quality of data for particular research purposes (e.g., one research question might need more or different data than for another research purpose), lack of education about best practices for creating data, different standards for data sharing by different academic journals, a lack of standards for peer review, misuse of research data, competing interests of researchers, and too much time required to share data. Some of these barriers such as lack of peer review and academic journal incentives are mentioned less frequently in our sample of the research literature (Cieri 2014; Hsu et al. 2015; Michener 2015; Piwowar & Chapman 2010; Treloar 2014; Wallis, Rolando & Borgman 2013; Wynholds et al. 2012). In contrast, issues such as the lack of incentives seem to be mentioned more frequently, even when researchers come from very different disciplines (Cieri 2014; Coetzee 2015; Collins 1998; Dallmeier-Tiessen et al. 2014; Duke & Porter 2013; Enke et al. 2012; Fecher, Friesike, & Hebing 2015; Hsu et al. 2015; Michener 2015; Nahar 2016; Piwowar & Chapman 2010; Savage & Vickeers, 2009; Tenopir et al. 2015; Treloar 2014; van Panhuis et al. 2014).

The second group of obstacles researchers have identified regarding data reuse are technical challenges. Technical difficulties include any kind of issue related to hardware, software, or other computer related problems that inhibit the free sharing of data. Subcategories of this larger set of barriers, outlined in **Table 2**, include the lack of data storage and support services, difficulties with interoperability, lack of discoverability on websites or in online catalogs, an inability for researchers to control access to data (e.g., limiting it only to fellow researchers on their team or others within the discipline for instance), inconsistent citation standards, and lack of metadata guidelines for including data in online repositories. Some of these technical barriers such as inconsistent citation standards and lack of metadata guidelines complement the social barriers identified in **Table 1**, for example, those regarding a lack of social standards developed for creating data. Yet, inconsistent citation standards and metadata guidelines are mentioned relatively rarely (e.g., Duke & Porter 2013; Hsu et al. 2015; Nahar 2016; Michener 2015; Specht et al. 2015; Wallis, Rolando & Borgman 2013; Wynholds et al. 2012) as compared to issues such as lack of data and support services (e.g., Akers & Doty 2013; Coetzee 2015; Fecher, Friesike, & Hebing 2015; Hsu et al. 2015; McLure et al. 2014; Michener 2015; Pepe et al. 2014; Pericas, Taura & Matsuoka 2014; Sayogo & Pardo 2013) which are mentioned quite frequently by researchers from many different disciplines.

Table 2: Literature Citing Technical Barriers to Data Reuse.

Issue	Literature Mentioned
Lack of data storage and support services	Akers & Doty 2013; Coetzee 2015; Fecher, Friesike, & Hebing 2015; Hsu et al. 2015; McLure et al. 2014; Michener 2015; Pepe et al. 2014; NAS 2009; Pericas, Taura & Matsuoka 2014; Sayogo & Pardo 2013
Difficulties with accessing data	Enke et al. 2012; Hsu et al. 2015; Jin et al. 2015; RSSPC 2012; SI 2015; Specht et al. 2015; van Panhuis et al. 2014
Lack of discoverability	Dallmeier-Tiessen et al. 2014; Elwood 2008; Enke et al. 2012; RSSPC 2012; SI 2015; Specht et al. 2015; Willibanks & Friend 2016
Inability of researchers to control access to data	Dallmeier-Tiessen et al. 2014; Enke et al. 2012; Federer et al. 2015; Pepe et al. 2014; van Panhuis et al. 2014
Inconsistent citation standards	Duke & Porter 2013; Nahar 2016; SI 2015; Wallis, Rolando & Borgman 2013; Wynholds et al. 2012
Lack of metadata guidelines	Hsu et al. 2015; Michener 2015; Specht et al. 2015

Table 3: Literature Citing Legal Barriers to Data Reuse.

Issue	Literature Mentioned
Differing government standards for archiving	Elwood 2008; Enke et al. 2012; Hodson & Molloy 2015; Michener 2015; NAS 2009; Piwowar & Chapman 2010; Specht et al. 2015; Treloar 2014; Wallis, Rolando & Borgman 2013; Wicherts, Bakker, & Molenaar 2011
Government requirements for privacy	Coetzee 2015; Dallmeier-Tiessen et al. 2014; Hodson & Molloy 2015; NAS 2009; RSSPC 2012; Sayogo & Pardo 2013; SI 2015
No ability to track down original source of data	Fecher, Friesike, & Hebing 2015; NAS 2009; RSSPC 2012; Sayogo & Pardo 2013

The group of barriers for data reuse mentioned least frequently relate to legal issues. These refer to any laws such as copyright, patent, or privacy that might inhibit the free sharing of data. By far, the most widely cited legal issue barrier to data sharing is the existence of differing government standards for archiving (Elwood 2008; Enke et al. 2012; Michener 2015; NAS 2009; Piwowar & Chapman 2010; Specht et al. 2015; Treloar 2014; Wallis, Rolando & Borgman 2013; Wicherts, Bakker, & Molenaar 2011). Different government agencies might have different rules for submitting data to an online government sponsored repository, or different universities might have other requirements for depositing data in a repository sponsored by their universities. Additionally, the research does identify some other legal obstacles such as government requirements for privacy, and an inability to track down the original source of the data. Yet these challenges seem to be less prevalent overall than the issue of differing requirements for archiving. **Table 3** outlines these legal obstacles to data reuse.

In all, social barriers to data reuse seem to be the most prevalent problem identified by academic researchers, followed by technical issues (some of which relate to the social barriers), and finally legal barriers. To understand what barriers to data reuse might exist in the neutron science community, we conducted a study to gain some insight into their perspectives on data sharing.

3 Methods

Our findings are drawn from data collected via surveys and focus groups conducted in July 2016 on site at Oak Ridge National Laboratory (ORNL) in Oak Ridge, TN, USA. We selected ORNL as our primary site of study for two main reasons. First, its Neutron Sciences Directorate (NScD) manages and operates the Spallation Neutron Source (SNS) and the High Flux Isotope Reactor (HFIR), two of the world's most advanced neutron scattering facilities. Second, staff at ORNL were interested in understanding potential barriers to data reuse and wanted to learn more about their stakeholders' views on the topic as an initial step forward. Staff at ONRL helped us to recruit participants for the study by forwarding a recruitment email that we prepared to individuals who lived in close proximity to ORNL. In total, 13 people participated: 3 data consumers, 5

data managers, and 5 data producers. We selected these data consumers, data managers, and data producers to gain an understanding of perspectives on sharing neutron data from a variety of different stakeholders.

All participants answered general questions about the types of information they need to know about a dataset before using it, their recent experiences with using datasets, whether they have ever reused data created by others, and what, in their opinion, makes datasets trustworthy. All participants also answered demographics questions related to their research interests, their field and professional titles/roles, how long they have been at their current institutions, their years of experience using neutron data, their years of experience using ORNL facilities, and their frequency of using data made accessible by any ORNL facilities.

Although all of the surveys and focus groups centered on the topic of data sharing at ORNL, we used three different surveys and focus group protocols to include additional questions that were relevant to each class of stakeholders. For example, data consumers examined a data set in real time and were asked about whether they could reuse that dataset, and if so, for what purpose(s). Data consumers also discussed any barriers to reuse as well as whether they thought it was worth the effort to reuse the dataset. Survey and focus group questions for data managers focused on understanding the types of data sets they manage as well as what activities they perform on data for their management and how they thought those activities help enable reuse. Survey and focus group questions for data producers focused on why they produce data and the extent to which they thought their data could be used by others.

Each group took the surveys first and then participated in a focus group immediately afterwards. The purpose for this mixed-method research design was two-fold: 1) to give participants an opportunity to provide more detail and context for their survey responses, and 2) to compare the data collected from both data collection methods to determine whether the data triangulate, thus underscoring the validity of the data (Creswell 2015). Finally, we conducted a joint focus group of data consumers, data managers, and data producers to discuss common issues related to data sharing. Surveys took approximately 15 minutes for respondents to complete; each focus group lasted approximately one to two hours. No incentives were provided for participation.

The data from the focus groups came in two forms: field notes and video recordings. A graduate student transcribed the video recordings. Afterwards, we compared the transcripts with our field notes to identify common themes and patterns. While there were too few survey participants to compute any inferential statistics, we analyzed the survey data by comparing participants' survey responses to their responses during the focus groups. The Indiana University Human Subjects Office approved this study (IRB Study #1605012591).

4 Findings

After briefly describing the study participants' demographic characteristics, this section provides details regarding data consumers', data managers', and data producers' perspectives on sharing neutron data.

4.1 Study Participants

Three data consumers, five data managers, and five data producers participated in this study. **Table 4** enumerates participants by their demographic characteristics including their occupation, field, experience with using neutron data and ORNL facilities.

Table 4: Study Participants.

Attributes	Data Consumers (n = 3)	Data Managers (n = 5)	Data Producers (n = 5)
Occupation	1 – Professor 2 – Research Scientists	4 – Research Scientists 1 – Software Engineer	3 – Research Scientists 2 – Post-Doctoral Fellows/Researchers
Field	3 – Research	3 – Government 1 – Research	5 – Government
Experience Using Neutron Data	1 – 5–10 years 2 – Over 10 years	1 – 1–4 years 1 – 5–10 years 3 – Over 10 years	1 – 1–4 years 2 – 5–10 years 2 – Over 10 years
Experience Using ORNL Facilities	1 – Less than 1 year 1 – 5–10 years 1 – Over 10 years	1 – Less than 1 year 3 – 5–10 years 1 – Over 10 years	2 – 1–4 years 2 – 5–10 years 1 – Over 10 years

The data consumers identified themselves as people who were interested in reusing data created by others. Of these three, one was a professor and two were research scientists. Their research interests ranged from condensed matter physics and the theory of magnetism of materials to soft matter physics. Two of the participants had over 10 years of experience with using neutron data; one participant had five to 10 years of experience with using neutron data. One had more than 10 years of experience using ORNL facilities, one had five to 10 years of experience, and one had less than one year of experience. Although everyone reported having an interest in data reuse, only one participant reported actually using someone else's data before.

The data managers identified themselves as people who add value to data created by others. Of these five, four were research scientists and one was a software engineer. Three described their field as government; one described his/her field as research. Two participants had worked at ORNL for more than 10 years, two for five to 10 years, and one for less than a year. Three participants had more than 10 years of experience with using neutron data, one participant had five to 10 years of experience using neutron data, and one participant had one to four years of experience. Three participants had five to 10 years of experience using ORNL facilities, one had more than 10 years of experience, and one had less than a year of experience. Four participants noted making use of ORNL data more than 50 times; one participant reported never having used data at ORNL.

The data producers' research interests ranged from molecular dynamics and structure in polymers and complex fluids, hard condensed matter physics, neutron total scattering studies of catalytic materials, and materials science, to mechanisms of heat transfer in thermoelectric materials and exotic coupling mechanisms in functional materials. Three were research scientists; two were post-doctoral fellows/researchers. All described themselves as government employees. Two had worked at their current institutions for over five years, three for one to four years, and one for less than a year. Two had more than 10 years of experience with using neutron data, two had five to 10 years of experience, and one had one to four years of experience. One had more than 10 years of experience with using facilities at ORNL, two had five to 10 years of experience, and two had one to four years of experience. All had used data made accessible via ORNL facilities more than 50 times.

4.2 Data Consumers Focus Group Findings

During focus groups centering on data consumers' perspectives on sharing neutron data, participants stressed the importance of creating and testing theoretical models. The main reason for reusing data is to test their models against existing data. To perform such tests, data consumers discussed information they needed to know about data that they were interested in reusing, such as how the materials were prepared and how the instruments were calibrated. Data consumers articulated the importance of journal articles, described barriers to reuse such as not knowing enough about a specific data set, and they expressed their desire for systems that could enable discoverability across data sets.

4.2.1 Reasons for Reusing Neutron Data

Participants reported interest in reusing neutron data for two main reasons:

1) To compare/verify a result against their own measurements.
2) To test a new theory using existing data.

4.2.2 Information Reusers Need to Know About Neutron Data

Participants mentioned the importance of establishing context for data. At ORNL, the data producers will help to prepare materials for testing, calibrate the instruments used to test the materials, and will record all of this information into individual log books. These logs are specific to the data producer, but may also be highly relevant if a researcher intends to go back at a later time to understand how samples were prepared and how the instrument was calibrated in order to measure specific properties. The data consumers who participated in this study reported wanting to know about these kinds of information. Specifically, they reported needing to know four things about any data set that they would consider reusing:

1) How the data were produced.
2) How the sample was prepared.
3) What the units of measurement are.
4) Details about experimental conditions and how they were determined (e.g., temperature).

4.2.3 The Importance of Journal Articles

Participants highlighted the importance of journal articles for reuse in two main respects. First, journal articles provided important context about the data that enabled participants to know if the data were of sufficient quality for reuse. Second, participants articulated interest in reproducing charts and graphs from previously published articles given publishers' approval.

4.2.4 Barriers to Reuse

Participants primarily reported technical barriers to reusing neutron data. For instance, data consumers reported not having the necessary expertise to utilize software needed to render the data that they otherwise would be interested in reusing.

4.2.5 Discoverability

Participants discussed the need for greater discoverability. They often would like to know what other measurements have been created related to particular scientific problems, or would be interested in pulling data with particular characteristics (e.g., temperature readings) to test against their models. Search capabilities across all data collected at ORNL neutron sources currently do not exist.

4.3 Data Managers Focus Group Findings

While none of the participants considered themselves as particularly involved with creation or consumption of data, they nevertheless had perspectives on data reuse that they were willing to share during the focus group. Generally speaking, they only trust data produced within their own lab, and often only coming from more "fastidious" researchers. Such researchers provide detailed notes on issues like temperature or how the instrument was calibrated.

The data managers expressed reservations about reusing others' data because one is not always sure whether the person reading the data at the point of their creation properly interpreted them. In particular, participants discussed how they manage different kinds of data both "raw" and "reduced." "Raw" data would include numbers coming off the machine without much context, and "reduced" data would help to contextualize those numbers in some meaningful way so that data could be interpreted by other researchers.

One participant said that getting data from a raw to a reduced state was "arty," meaning that a degree of intuition was needed to figure out what needed to be done. The initial data produced by neutron scattering instruments are a simple event list or histogram of neutron events detected after interacting with the sample of scientific interest. However, these representations of the raw data are not useful to the researchers who are carrying out the experiments or reusing the data. These raw data are reduced by averaging and combining signals, subtracting instrument background measurements and transforming data to scientific units common within a researcher's given science domain. These reduced data sets are used in further scientific analysis. The data managers who participated in this study argued that activities related to reducing data from their raw state, such as averaging and combining signals, subtracting instrument background measurements and transforming data to scientific units common within a researcher's given science domain, require intuition. Such intuition means that people at ORNL could easily trust data if they knew the source was from their own lab or from a known colleague, but they could not trust the intuition of people they did not know, because they felt they would be unable to assess whether those people had the expertise to determine what was necessary to move the data from a raw to a reduced form.

4.4 Data Producers Focus Group Findings

The data producers expressed similar concerns to both managers and consumers and expressed a need to understand the context of the data. They are primarily interested in comparing their data to other data that may have been gathered for similar experiments. For instance, if in an experiment they achieved a result that seems extremely strange, they want to see if there are others that have also seen similar results. If results are similar, then they have no reason to suspect that there are issues with the instrument or otherwise (e.g., insufficient calibration of the instrument, etc.). On the other hand, if they achieve results that are different from what others have achieved using similar instruments, they may need to go back and look at what they did to see if the instrument was calibrated properly or whether there may have been an issue with the sample that was used.

Data producers had particular concerns regarding data reuse, including the fear of misuse of their data by others, competition amongst other researchers with similar research interests, their role in facilitating data reuse, and the importance of creating metadata of sufficient quality to enable reuse.

4.4.1 The Fear of Their Data Being Misused

Participants discussed concerns about their data being misused, either inadvertently or intentionally. For example, some participants discussed past issues with other data producers who had been accused of incorrectly measuring phenomena. They also reported concerns about data consumers who are so obsessed with their own models that they might misuse others' data during reuse in order to better conform to their models.

4.4.2 Competition

Data producers discussed how research can be particularly competitive; they are often put into difficult positions when competing scientists use the same instruments to do slightly different, though complementary research. They worried about talking to potential data consumers about their data; for example, they worried that by discussing their data with potential data consumers they might give away information that could unduly advantage another researcher.

4.4.3 The Role of the Data Producer in Reuse

These data producers thought they would be needed in order to help other researchers understand their data. They said this based on prior experience where it was common for researchers within their own teams to come back to them and ask for help in understanding data that had been created at that lab years before. Given this, they believed it would be necessary for data consumers to interact with data producers in order to actually be able to reuse neutron data.

4.4.4 The Importance of Metadata

Data producers agreed that more metadata would be a benefit to data managers and consumers as well as to themselves. They also recognized that producing such metadata would be difficult. They discussed how much metadata are kept in log books, while acknowledging their inconsistency, even when they are managed by the same person. They felt that metadata collection should be as automated and easy as possible to reduce the burden on them when creating new datasets.

4.5 Joint Focus Group Findings

Discussion in the joint focus group with the data consumers, data managers, and data producers focused largely on metadata and greater discoverability. All parties agreed that there was a lack of metadata and that there need to be better ways of cataloging information about datasets as easily as possible (and if possible automatically). Participants also stressed that they needed better ways to access metadata which should not be tacit or in a person's brain but instead in a format that is readable by others. In cases where the format is not necessarily readable by all consumers, participants believed that it was important to create tools that could make the datasets more discoverable and accessible to everyone. Participants also discussed, though indirectly, how it was important to link publications to individual datasets because they felt it is easy to cite sources in a journal, but not elsewhere.

Finally, participants thought about ways in which data sharing would not only help others but could also influence their own research. Participants believed that data producers might be especially influenced by reusing others' data, including results that were "boring" in the sense that they did not discover anything new, but reconfirmed what others have seen. These data would be valuable to data producers because they would provide insight into how to (or not to) calibrate their instruments, or it could help them avoid conducting experiments that will not yield any new science.

We used the findings from focus groups with data consumers, data managers, and data producers to generate a workflow diagram for neutron data across each group, which we term the Consumers Managers Producers (CMP) Model (See **Figure 1**). In **Figure 1**, the blue boxes show the three types of users (data producers, data managers, and data consumers), and the blue arrows show the progress of the workflow between these groups. Similarly, the red boxes show the data types (raw data, reduced data, and modeled data), and the red arrows show the flow from one type of data to another. Each of these groups interacts with the data in particular ways, represented by the green arrows. This graph illustrates the interrelationships between data and their users at ORNL.

As shown in **Figure 1**, data producers carry out the experiment and generate raw data (e.g., unprocessed numbers and descriptions) from which reduced data are constructed (e.g., a data set that was transformed into a representation relevant to the corresponding science domain). Data managers facilitate the data reduction step as well as tools needed to produce model data derived from theoretical materials models. Data consumers utilize modeled data to create research and scholarship demonstrating how materials
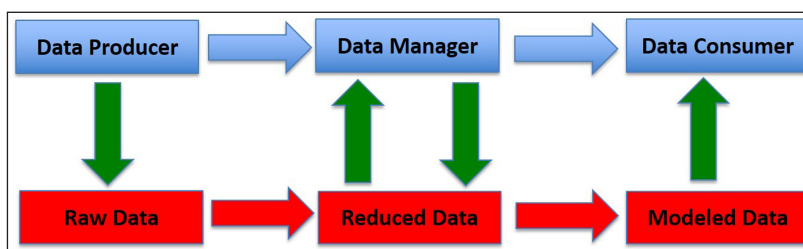
Figure 1: The Consumers Managers Producers (CMP) Model. The CMP Model illustrates the flow of neutron
data and the interactions of data consumers, data managers, and data producers with data at various
stages of production at Oak Ridge National Laboratory (ORNL). Users are represented in blue, data are
represented in red, and interactions between users and data are represented in green.

function on an atomic level. We found this model helpful for understanding each group's relationship to
neutron data as well as their perspectives on data and data sharing.

5 Discussion

5.1 Contributions to Research Literature on Data Reuse

This study makes three primary contributions to the research literature on data reuse. First, it underscores the
value and importance of data reuse. Although many technical reports, guidelines, and recommendations exist
which state that data reuse is important and valuable, not everyone is convinced that this is true, and even if
they are, this does not mean that they themselves actually share or reuse data (Borgman 2015). Studies with
findings demonstrating the perceived importance of sharing data by actual scientists are still necessary to
help convince different classes of stakeholders that creating and/or modifying existing policies and comput-
ing infrastructures to support data reuse is worth the investment and effort. Even though some neutron scien-
tists share and reuse data, not all are convinced that they should. This study provides empirical support for the
idea that the neutron scientists who participated in our study would reuse data if they had enough context to
understand them, trusted their accuracy, and were able to use the appropriate tools to access the data. This is
good news for the field of neutron science and for the field of scientific data reuse research in general.

Second, this study advances our understanding of barriers to data reuse by focusing on members of a
scientific community that has been understudied with regard to their data reuse practices—neutron
scientists. What we learn from this study and similar studies is not just identification of barriers to data
reuse, but more importantly, how different scientific communities of practice experience these barriers.
Gathering knowledge about how scientific researchers experience barriers to data reuse is an important first
step in being able to mitigate or remove those barriers for the sake of enabling reuse.

Third, we propose a new framework for understanding the interplay among three different classes of
stakeholders regarding data reuse, the Consumers Managers Producers (CMP) Model. The CMP Model illustrates
the workflow of neutron data amongst data consumers, data managers, and data producers at ORNL. Future
studies should test the extent to which the CMP Model applies to other neutron scientists besides those who
participated in this study, as well as other scientific disciplines with similar classes of stakeholders.

We argue that this study is of potential benefit for any organization that is similar to ORNL. There are
approximately 40 neutron science facilities worldwide (neutronsources.org 2012). Although a detailed
account of these facilities' data management policies and other measures to foster data reuse and sharing
is beyond the scope of this study, we argue that this paper offers a good starting point for some issues that
could be important to consider when creating or modifying policies and technical infrastructures at ORNL
and similar institutions to support data reuse. Overall, we recommend shifting focus from what we term
traditional data use (where researchers use the data they produce) toward facilitating data reuse (where
researchers utilize data produced by others). Based on the findings from this study, we propose the following
policy and system recommendations to ORNL and similar institutions to help facilitate data reuse.

5.1.1 Policy Recommendations

Current access to experimental neutron data at SNS and HFIR is limited to the team conducting the
experiment; however, a data management plan and development of a data portal facilitating data discovery
and access is currently being developed for enabling reuse of neutron scattering data. The findings of this
study provide critical input for developing, communicating and implementing this future data management
policy at ORNL. We recommend that similar organizations put data management plans in place and develop
data portals that can help facilitate data reuse.

5.1.2 System Recommendations

Systems that support sharing neutron data at ORNL and similar institutions should:

· Include comprehensive metadata about how the data were produced, how the sample was prepared, what the units of measurement are, and details about experimental conditions and how they were determined (e.g., temperature).
· Link data to any publications based on or otherwise related to those data.
· Link data to any software that was used to derive results, and
· Make data more discoverable, for example, by allowing characteristics of data (e.g., temperature readings) to be searchable across data sets.

5.2 Analysis of Barriers to Data Reuse Identified in this Study as They Relate to Barriers Identified in the Research Literature on Data Reuse

Our findings provide empirical support for the relevance of barriers to data reuse that have been identified in prior empirical studies to the neutron scientists who participated in this study. According to Fecher, Friesike, and Hebing (2015), research on data reuse has identified three broad categories of barriers encountered by researchers when they attempt to make available or utilize the data of other scientists. These include: social barriers (e.g., obstacles encountered because sharing of data might contravene community practices within certain academic fields), technical barriers (e.g., problems with computer systems, formatting of the data itself, or interoperability of data between different computer systems), and legal barriers (e.g., copyright or privacy laws that would prevent an investigator from openly sharing results). Furthermore, each of these three categories includes more specific subcategories. **Table 5** provides a summary of each data reuse barrier category and subcategory along with an assigned number (1 through 18).

Several, though not all, of the barriers mentioned in **Table 5** were also discussed by staff at the SNS at ORNL. During the focus groups with data managers, data consumers, data producers, and a final "wrap-up" focus group with representatives from all three groups, participants focused largely on technical issues, but

Table 5: Categorization of Barriers to Data Reuse.

Social Barriers
1. Differences between disciplines in creating data
2. Lack of incentives to share or reuse data
3. Differences in quality of data for particular research purposes
4. Lack of education about best practices
5. Journals incentivize different kinds of data sharing
6. Lack of standards for peer review of data
7. Misuse of research data
8. Competing interests of researchers
9. Too much time required to share data

Technical Barriers
10. Lack of data storage and support services
11. Difficulties with accessing data
12. Lack of discoverability
13. Inability of researchers to control access to data
14. Inconsistent citation standards
15. Lack of metadata guidelines

Legal Barriers
16. Differing government standards for archiving
17. Government requirements for privacy
18. No ability to track down original source of data

also addressed social issues present within their field. **Table 6** shows which barriers to data reuse (identified using the numbers from **Table 5**) that members of the three focus groups (data managers, data producers, and data consumers) and the final "wrap-up" session mentioned during their focus groups.

   Table 7 looks at the same data as **Table 6**, but organizes them in a different way. Rather than focusing on just which issues came up during the sessions, **Table 7** shows which issues were common either to multiple groups, or were unique to a single group. Generally, **Table 7** demonstrates that some of these barriers were discussed across groups while others were unique only to one group of stakeholders. Two barriers were discussed separately by all three groups (barriers 1 and 3). Two barriers were specifically mentioned only by data producers (barriers 5 and 10). Participants also discussed several barriers when they came together during the wrap-up session (including barrier 14, one that had not been discussed previously). In part, members of that focus group were reiterating what they had said during previous meetings. Nonetheless, what was discussed in the wrap-up session would also have been new information for some participants who were not present during the other focus group meetings.

   Although our findings reflected the relevance of some barriers identified in the literature, there was less empirical support for the relevance of other barriers. For example, social barriers were quite prevalent, much the same way as they are in other disciplines. Additionally, technical barriers also appear to be quite prevalent. Legal barriers, at least for our study participants, do not seem to be particularly prevalent. This most likely has to do with the type of data our study participants were considering. Although ORNL collects a broad range of data (e.g., proprietary data collected by companies for a fee, export controlled data, etc.), our study participants answered questions about open research data that fall under President Barack Obama's executive order on open data (Obama 2013). Since those data are required to be open, many legal restrictions that are typically associated with reuse do not apply here.

   Among the social barriers mentioned by participants, barriers 1 (Differences between disciplines in creating data) and 3 (Differences in quality of data for particular research purposes) were common to all three groups, indicating a particularly important commonality among data managers, producers, and consumers. In contrast, barrier 5 (Journals incentivize different kinds of data sharing) was mentioned only by the data producers, perhaps indicating that this issue is either not recognized by the other groups, or perhaps is more

Table 6: Barriers to data reuse mentioned in focus groups.

Session	Barriers
Data Managers	1, 3, 11, 13, 15
Data Producers	1, 3, 5, 7, 8, 9, 10, 11, 12
Data Consumers	1, 3, 15
Wrap-up	10, 11, 12, 14, 15

Table 7: Barriers common to multiple groups or unique to other groups.

Issue	Mentioned by
Differences between disciplines in creating data (1)	Data Managers, Data Producers, Data Consumers
Differences in quality of data for particular research purposes (3)	Data Managers, Data Producers, Data Consumers
Journals incentivize different kinds of data sharing (5)	Data Producers
Lack of data storage and support (10)	Data Producers
Misuse of research data (7)	Data Producers
Competing interests of researchers (8)	Data Producers
Too much time required to share data (9)	Data Producers
Difficulties with accessing data (11)	Data Managers, Data Producers, Wrap-up
Lack of discoverability (12)	Data Producers, Wrap-Up
Inconsistent citation standards (14)	Wrap-Up
Lack of metadata guidelines (15)	Data Managers, Data Consumers, Wrap-Up

completely understood by those generating neutron data rather than by the other groups of stakeholders. Barriers 7 (Misuse of research data), 8 (Competing interests of researchers), and 9 (Too much time required to share data) were mentioned only by data producers. Notably, participants did not mention many of these social issues in the final wrap-up session. Perhaps this is because it was more comfortable for participants to talk about technical issues. At any rate, despite the fact that there were some issues common to all three groups, these issues were not discussed when representatives of all three groups were present.

Perhaps not surprisingly, technical barriers were mentioned frequently. Barrier 11 (Difficulties with accessing data) seemed to be the most common and was mentioned by data managers, data producers, and in the wrap-up session. Barrier 15 (Lack of metadata guidelines) was also brought up relatively frequently by data managers, data consumers, and during the wrap-up meeting. Barrier 12 (Lack of discoverability) was the next most frequently mentioned technical issue by data producers and in the wrap-up session. Barriers 7 (Lack of data storage and support services) and 14 (Inconsistent citation standards) were mentioned only by data producers and in the wrap-up group. Interestingly, the wrap-up session provided a space for all of these stakeholders to mention common problems; for technical issues (more so than social issues), the wrap-up session provided a particularly fruitful way for all three groups to discuss common issues.

The primary limitation of this study is its sample size. Only 13 people participated. In addition, selection bias could be present in our sample because we recruited and selected participants based on our knowledge of individuals who fit the categories that we were interested in studying (i.e., data consumers, data managers, and data producers) as well as their proximity to ORNL; we selected individuals for our study who lived and worked nearby. While this recruitment strategy made it practically feasible to conduct the study, we acknowledge that the views of the participants may not reflect the perspectives of other data consumers, data managers, and data producers who use ORNL's neutron sources or manage data originating from ORNL's user facilities.

6 Conclusion

The results of this study underscore the value and importance of data reuse. They provide insight into how some members of the neutron science community perceive barriers to data reuse. We also present the Consumers Managers Producers (CMP) Model to explain the interplay among the data consumers, managers, and producers who participated in our study regarding data reuse. Future studies should consider the extent to which the CMP Model can characterize other fields of research. Will members of other scientific communities have similar attitudes about data sharing? In other words, does the structure and workflow that the CMP Model represents correspond to particular sets of barriers across different scientific disciplines? Candidates for this type of research include other scientific domains whose researchers depend on major research facilities to produce their data, such as astronomy and ocean science.

Future studies should also examine reuse of neutron data "in real time" to better understand what technical and social issues arise during the act of reuse. Results of such studies could be compared with the findings of this investigation and can also further inform the development of policies and systems designed to support sharing of neutron data.

Acknowledgements

Competing Interests
The authors have no competing interests to declare.

Author Information
Dr. Devan Ray Donaldson is an Assistant Professor in the School of Informatics and Computing at Indiana University, Bloomington. His research interests include digital repositories, data sharing practices, mass digitization, preservation management, preservation metadata, trust, and security. He holds a Ph.D. in Information from the University of Michigan, a M.S. in Library Science from the University of North Carolina at Chapel Hill and a B.A. in History from the College of William and Mary.

Shawn Martin is a doctoral student and an Integrated Doctoral Education with Application to Scholarly Communication (IDEASc) Fellow at Indiana University, Bloomington. His research focuses on scholarly communication and the history of academic publishing. Prior to coming to Indiana, he was a scholarly communication librarian at the University of Pennsylvania. He holds a B.A. in History from Ohio State University and a M.A. in History from the College of William and Mary.

Dr. Thomas Proffen is the Director for Neutron Data Analysis and Visualization in the Neutron Sciences Directorate at Oak Ridge National Laboratory (ORNL). He is responsible for data and scientific computing at the neutron user facilities at ORNL. He holds a Ph.D. in Physics from the Ludwig Maximilans University in Munich, Germany.

References

Akers, K G and **Doty, J** 2013 Disciplinary differences in faculty research data management practices and perspectives. *The International Journal of Digital Curation*, 8(2): 5–26. DOI: https://doi.org/10.2218/ijdc.v8i2.263

Borgman, C 2012 The conundrum of sharing research data. *Journal of the Association for Information Science and Technology*, 63(6): 1059–1078. DOI: https://doi.org/10.1002/asi.22634

Borgman, C 2015 *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press.

Campbell, E G, Clarridge, B R, Gokhale, M, Birenbaum, L, Hilgartner, S, Holtzman, N A and **Blumenthal, D** 2002 Data withholding in academic genetics: Evidence from a national survey. *Journal of the American Medical Association*, 287(4): 473–480. DOI: https://doi.org/10.1001/jama.287.4.473

Cieri, C 2014 Challenges and opportunities in sociolinguistic data and metadata sharing. *Languages and Linguistics*, 8(11): 472–485. DOI: https://doi.org/10.1111/lnc3.12112

Coetzee, T 2015 Sharing data from MS clinical trials: Opportunities, challenges, and future directions. *Multiple Sclerosis Journal*, 21(11): 1365–1368. DOI: https://doi.org/10.1177/1352458515608005

Collins, H M 1998 The meaning of data: Open and closed evidential cultures in the search for gravitational waves. *American Journal of Sociology*, 104(2): 293–338. DOI: https://doi.org/10.1086/210040

Creswell, J W 2015 *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. 4th ed. Los Angeles, CA: Sage.

Dallmeier-Tiessen, S, Darby, R, Gitmans, K, Lambert, S, Matthews, B, Mele, S, Suhonen, J and **Wilson, M** 2014 Enabling sharing and reuse of scientific data. *New Review of Information Networking*, 19: 16–43. DOI: https://doi.org/10.1080/13614576.2014.883936

Duke, C S and **Porter, J H** 2013 The ethics of data sharing and reuse in biology. *Bioscience*, 63: 483–489. DOI: https://doi.org/10.1525/bio.2013.63.6.10

Elwood, S 2008 Grassroots groups as stakeholders in spatial data infrastructures: Challenges and opportunities for local data development and sharing. *International Journal of Geographical Information Science*, 22(1): 71–90. DOI: https://doi.org/10.1080/13658810701348971

Enke, N, Thessen, A, Bach, A, Bendix, J, Seeger, B and **Gemeinholzer, B** 2012 The user's view on biodiversity data sharing – Investigating facts of acceptance and requirements to realize a sustainable use of research data. *Ecological Informatics*, 11: 25–33. DOI: https://doi.org/10.1016/j.ecoinf.2012.03.004

Faniel, I M, Kriesberg, A and **Yakel, E** 2016 Social scientists satisfaction with data reuse. *Journal of the Association for Information Science and Technology*, 67(6): 1404–1416. DOI: https://doi.org/10.1002/asi.23480

Fecher, B, Friesike, S and **Hebing, M** 2015 What drives academic data sharing? *PLoS ONE*, 10(2): e0118053. DOI: https://doi.org/10.1371/journal.pone.0118053

Federer, L M, Lu, Y, Joubert, D J, Welsh, J and **Brandys, B** 2015 Biomedical data sharing and reuse: Attitudes and practices of clinical research staff. *Plos ONE*, 10(6): e0129506. DOI: https://doi.org/10.1371/journal.pone.0129506

Fienberg, S E, Martin, M E and **Straf, M L** (eds.) 1985 *Sharing Research Data*. Washington, D.C.: National Academies Press.

Hilgartner, S 1997 Access to Data and Intellectual Property: Scientific Exchange in Genome Research. Intellectual Property Rights and the Dissemination of Research Tools in Molecular Biology. In: *Summary of a Workshop Held at the National Academy of Sciences, February, 15–16, 1996*, 28–39. Washington, DC: National Academies Press.

Hilgartner, S and **Brandt-Rauf, S I** 1994 Data access, ownership and control: Toward empirical studies of access practices. *Knowledge*, 15: 355–372. DOI: https://doi.org/10.1177/107554709401500401

Hodson, S and **Molloy, L** 2015 *Current Best Practice for Research Data Management Policies.* Paris, France: International Council for Science: Committee on Data for Science and Technology (CODATA).

Hsu, L, Martin, R L, McElroy, B, Litwin-Miller, K and **Wonsuck, K** 2015 Data management, sharing, and reuse in experimental geomorphology: Challenges, strategies, and scientific opportunities. *Geomorphology*, 244: 180–189. DOI: https://doi.org/10.1016/j.geomorph.2015.03.039

Jin, X, Wah, B W, Cheng, X and **Wang, Y** 2015 Significance and challenges of big data research. *Big Data Research*, 2: 59–64. DOI: https://doi.org/10.1016/j.bdr.2015.01.006

Lee, J and **Kang, M** 2015 Geospatial big data: Challenges and opportunities. *Big Data Research*, 2: 74–81. DOI: https://doi.org/10.1016/j.bdr.2015.01.003

McLure, M, Level, A V, Cranston, C L, Oehlerts, B and **Culbertson, M** 2014 Data curation: A study of researcher practices and needs. *portal: Libraries and the Academy*, 14(2): 139–164. DOI: https://doi.org/10.1353/pla.2014.0009

Michener, W K 2015 Ecological Data Sharing. *Ecological Informatics*, 29: 33–34. DOI: https://doi.org/10.1016/j.ecoinf.2015.06.010

Nahar, L H V 2016 Reuse of scientific data in academic publications. *Aslib Journal of Information Management*, 68(4): 478–494. DOI: https://doi.org/10.1108/AJIM-01-2016-0008

National Academies of Science, Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age (NAS) 2009 *Ensuring the Integrity, Accessibility, and Stewardship of Research Data.* Washington, D.C.: National Academies Press.

National Research Council (NRC) 1997 *Intellectual Property Rights and Research Tools in Molecular Biology.* Washington, D. C.: National Academies Press.

Neutronsources.org 2012 *Neutron Centres.* Available at: http://neutronsources.org/neutron-centres.html [Last accessed 16 April 2017].

Obama, B 2013 *Executive Order – Making Open and Machine Readable the New Default for Government Information.* Washington DC: The White House Office of the Press Secretary. Available at: https://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government- [Last Accessed 30 October 2016].

Pepe, A, Goodman, A, Muench, A, Crosas, M and **Erdmann, C** 2014 How do astronomers share data? Reliability and persistence of datasets linked in AAS publications and a qualitative study of data practices among US astronomers. *Plos ONE*, 9(8). DOI: https://doi.org/10.1371/journal.pone.0104798

Pericas, M, Taura, K and **Matsuoka, S** 2014 Scalable analysis of multicore data reuse and sharing. *Proceedings of the 28th ACM international conference on Supercomputing.* New York: Association for Computing Machinery, 353–362. DOI: https://doi.org/10.1145/2597652.2597674

Piwowar, H A and **Chapman, W** 2010 Public sharing of research datasets: A pilot study of associations. *Journal of Informetrics*, 4: 148–156. DOI: https://doi.org/10.1016/j.joi.2009.11.010

Research Data Alliance Europe (RDA) 2014 *The Data Harvest: How sharing research data can yield knowledge, jobs and growth.* Europe: RDA.

Royal Society Science Policy Centre (RSSPC) 2012 *Science as an Open Enterprise.* London: The Royal Society.

Savage, C J and **Vickers, A J** 2009 Empirical Study of Data Sharing by Authors Publishing in PLOS Journals. *PLoS ONE*, 4(9). DOI: https://doi.org/10.1371/journal.pone.0007078

Sayogo, D S and **Pardo, T A** 2013 Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. *Government Information Quarterly*, 30: S19–S31. DOI: https://doi.org/10.1016/j.giq.2012.06.011

Science International (SI) 2015 *Open Data in a Big Data World.* Paris: International Council for Science (ICSU), International Social Science Council (ISSC), the World Academy of Sciences (TWAS), InterAcademy Partnership (IAP).

Specht, A, Gurus, S, Houghton, L, Keniger, L, Driver, P, Ritche, E G, Lai, K and **Treloar, A** 2015 Data management challenges in analysis and synthesis in the eoysstem sciences. *Science of the Total Environment*, 534: 144–158. DOI: https://doi.org/10.1016/j.scitotenv.2015.03.092

Tenopir, C, Dalton, D, Allard, S, Frame, M, Pjesivac, I, Birch, B, Pollock, D and **Dorsett, K** 2015 Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS ONE*, 10(8). DOI: https://doi.org/10.1371/journal.pone.0134826

Treloar, A 2014 The research data alliance: Globally co-ordinated action against barriers to data publishing and sharing. *Learned Publishing*, 27: S9–S13. DOI: https://doi.org/10.1087/20140503

Ure, J, Procter, R, Lin, Y W, Hartswood, M, Anderson, S, Lloyd, S, Wardlaw, J, Gonzalez-Velez, H and **Ho, K** 2009. The development of data infrastructures for eHealth: A socio-technical perspective. *Journal of the Association for Information Systems, suppl. Special Issue on e-Infrastructure*, 10(5): 415–429.

van Panhuis, W G, Proma, P, Emerson, C, Grenfenstette, J, Wilder, R, Herbst, A J, Heymann, D and **Burke, D S** 2014 A systematic review of barriers to data sharing in public health. *BMC Public Health*, 14. Available at: http://www.biomedcentral.com/1471-2458/14/1144. DOI: https://doi.org/10.1186/1471-2458-14-1144

Wallis, J C, Rolando, E and **Borgman, C** 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. DOI: https://doi.org/10.1371/journal.pone.0067332

Wicherts, J M, Bakker, M and **Molenaar, D** 2011 Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE*, 6(11): e26828. DOI: https://doi.org/10.1371/journal.pone.0026828

Willibanks, J and **Friend, S H** 2016 First, design for data sharing. *Nature Biotechnology*. DOI: https://doi.org/10.1038/nbt.3516

Wynholds, L A, Wallis, J C, Borgman, C L, Sands, A and **Trawek, S** 2012 Data, data use, and scientific inquiry: Two case studies of data practices. *Proceedings of the 12th ACM/IEEE Joint Conference on Digital Libraries*. New York: Association for Computing Machinery, 19–22. DOI: https://doi.org/10.1145/2232817.2232822