IJDC | *Peer-Reviewed Paper*

Research Data Sharing and Reuse Practices of Academic Faculty Researchers: A Study of the Virginia Tech Data Landscape

Yi Shen
Virginia Polytechnic Institute and State University

Abstract

This paper presents the results of a research data assessment and landscape study in the institutional context of Virginia Tech to determine the data sharing and reuse practices of academic faculty researchers. Through mapping the level of user engagement in "openness of data," "openness of methodologies and workflows," and "reuse of existing data," this study contributes to the current knowledge in data sharing and open access, and supports the strategic development of institutional data stewardship. Asking faculty researchers to self-reflect sharing and reuse from both data producers' and data users' perspectives, the study reveals a significant gap between the rather limited sharing activities and the highly perceived reuse or repurpose values regarding data, indicating that potential values of data for future research are lost right after the original work is done. The localized and sporadic data management and documentation practices of researchers also contribute to the obstacles they themselves often encounter when reusing existing data.

Received 29 April 2015 | *Accepted* 25 November 2015

Correspondence should be addressed to Yi Shen, Virginia Polytechnic Institute and State University, 560 Drillfield Dr., Blacksburg, Virginia 24060, USA. Email: yishen18@vt.edu

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

 doi:10.2218/ijdc.v10i2.359

Introduction

Understanding research data sharing and reuse practices of academic faculty researchers is important to the development of data infrastructure, management, preservation, and curation systems at an academic institution. As faculty scholarship is experiencing a changing landscape that is more data-driven in methodologies and technologies, research data assessment and landscape study is necessary to identify core data practices and service requirements, determine obstacles and solutions for data use and discovery, and strategize data services, curation support, and training programs. Academic libraries must actively engage in this fast-developing landscape and lead data assessment, services, and sharing efforts (Walters and Skinner, 2011). With expertise in data, information, and archive fields, libraries have great values to offer in shaping the data culture and building a shared access research infrastructure.

These efforts must be contextualized within the specific research environment of an institution to identify the level of community engagement in data sharing and reuse in order to develop a user-centric, community-driven data repository and the required human resources. This project applies a newly engineered research data assessment tool in the institutional context of Virginia Tech (VT) to investigate how data are being stored, managed, shared, and reused by VT faculty and researchers. Two research papers are produced as the results of this data landscape study. The first paper presents the survey results regarding the faculty researchers' data holdings, current data management practices, as well as educational needs and service requirements related to data (Shen, 2016). This second paper focuses on the faculty researchers' data sharing and reuse practices in the institutional context.

The study helps determine the unique set of obstacles related to data production, use, and reuse. It reveals the potential future values of research data from both data producers' and data users' perspectives. It also identifies a major gap between the localized and limited data management and sharing activities, and the highly perceived reuse or repurpose values of data that often get lost in the transition of research practitioners and communities of practice. By mapping the level of user engagement in "openness of data," "openness of methodologies and workflows," and "reuse of existing data," this research contributes to the current knowledge in data sharing and open access and supports the strategic development of institutional data stewardship. A further investigation of college-level engagement in "openness of data" also reveals the different patterns and concentrations of activities in individual colleges and indicates the need to develop college-oriented strategies and approaches to a changing data culture.

Literature Review

Data Sharing and Open Access Movements

Against the backdrop of global efforts to build a scientific data infrastructure (e.g., Research Data Alliance) and a national initiative to develop the SHared Access Research Ecosystem (SHARE), data sharing and open access movements are quickly gaining significance in academic communities. Guedon (2015) has pointed out that "across the centuries, researchers have learned to share their papers, now they must

learn to share their data." In his view, data sharing is the "very essence of science if science is conceived as a gigantic system of distributed intelligence…Sharing data and sharing the interpretation of data in the form of published papers simply constitute the best way to optimize the whole research process."

With such understanding, "the leaders of the scientific community are recalibrating their requirements, pushing for the sharing of data and greater experimental transparency" (Achenbach, 2015). Open data and open access movements are quickly gaining momentum. For example, the global health community, including the World Health Organization, the National Institutes of Health, the Wellcome Trust, and the Research Councils UK, demonstrates a commitment to sharing research data and information. In November 2014, the Gates Foundation also announced its adoption of an Open Access policy to "enable the unrestricted access and reuse of all peer-reviewed published research funded by the foundation, including any underlying data sets." This movement puts a high priority not only on research, but also on the collection and sharing of data, so that other scientists and health experts can access the latest evidence, draw on it to advance their own research, and benefit from this knowledge (Mundel, 2014).

Among the many literature and press releases, reproducibility is one core argument for data sharing to avoid the so-called "data dredging" maneuver (Achenbach, 2015) or "overfitting" of data (Provost and Fawcett, 2013) in which researchers go on a deep dive for something publishable that may turn out to be a "statistical fluke" (Achenbach, 2015). Another core argument for data sharing is the tremendous value of reusing or repurposing data. Especially as new analytical techniques become available, academics may want to explore their data in ways that were not planned for in the original design of their data collections. Scientists and scholars may also be increasingly looking at how to integrate their structured data holdings with those of others and to explore links to both internal and external unstructured data sources (Hendler, 2014). As a result, discovery of and access to data outside their own control will become even more crucial. In order to support these efforts in an academic institution, we need to ask the important questions: what are faculty researchers' behaviors and attitudes towards data sharing and reuse? How accessible and discoverable are their data? And what are the major concerns of reusing and repurposing data among faculty researchers?

Data Sharing Practices

Several previous surveys have explored data sharing and withholding practices. Mostly framed from the perspective of data producers on the concerns or benefits of sharing, these studies explored where and how researchers are willing to share data, and what are the motivations and disincentives for sharing (e.g., Tenopir, et al., 2011). The results generally show minimal sharing practices and the historical lack of incentives, which involve time, funding, manual labor, policies and standards, competition and ownership, conventions and discourses, and technical capabilities, as well as other limiting factors. To facilitate broader participation in open access, Faniel and Zimmerman (2011) proposed a comprehensive research agenda to investigate data practices from the perspectives of scientists, non-scientists, and interdisciplinary researchers. With a focus on researchers and data, Borgman (2012) further examined the rationales for sharing data and discussed the associated complexities and difficulties in sharing, after taking into account of the different purposes for collecting data and the diverse approaches to handling data. Understanding scientists' views on "sharable forms of data," Cragin et al.

(2010) stressed that data curation services need to accommodate a wide range of sub-disciplinary data characteristics and sharing practices.

Building upon previous studies and mapping to the Community Capability Model Framework, this current study adopts newly designed measurements to identify and map the levels of individual, institutional, and community engagement in open data sharing activities — from nominal activity, pockets of activity, moderate activity, widespread activity, to complete engagement. Using a systematic approach, this study allows researchers to indicate their own data's special features, potential reuse and repurpose values, as well as their data documentation and reuse practices and concerns. Switching roles from data producers to data users enables researchers to have an insightful, personal, and meaningful reflection on the importance of data sharing, on the existing problems in managing data, and on the proper actions needed to preserve and "capitalize" the values of data.

A Changing Data and Research Landscape

Today, more and more government agencies are introducing and enforcing open access and data sharing policies. With the expanding worldwide recognition and global actions in building data infrastructure and sharing ecosystems, a changing data landscape is on the horizon. New studies should identify the emerging needs in this ever-changing data and academic research landscape and devise strategies for new development.

Hendler (2014) summarized, "the increasing amount of data available on the Web, the new technologies for linking data across datasets, and the increasing need to integrate structured and unstructured data" are driving the emerging trend in "broad data", in which a variety of heterogeneous data are being used. He also observed that "the ability to federate data across datasets, catalogs, domains, and cultures can provide data users with the ability to find, access, integrate, and analyze combinations of datasets based on their needs." In these regards, data management and curation play significant roles in collecting and preserving data, as well as making data accessible. Library and information communities, and the broader scholarly communities are developing controlled vocabularies, ontologies, and metadata standards to help with data integration efforts.

Among these new developments, understanding faculty scholars' research data practices, educational needs, and service requirements is essential to promoting appropriate data stewardship at the institutional level. A user-centered research data assessment also helps institutions strategize and manage the development of a data repository and required human infrastructure. The broader intent for such development is to give researchers, scholars, and other users readily accessible and easily analyzed data sets that contain key metadata elements, name entities, unique identifiers, and links to other derived data types.

Method

Targeted at a multifaceted and multilevel assessment, a data collection survey instrument was developed with the incorporation of multiple frameworks, models, and templates. These included the Data Asset Framework (Digital Curation Center, 2009) and its pilot studies (Jones et al., 2008), the Community Capability Model Framework (CCMF) (UKOLN, University of Bath and Microsoft Research Connections, 2013),

DataOne's scientists and research data survey (DataOne, 2013), as well as other institutional data management surveys (e.g., Emory University Research Data Management, 2012) and planning questionnaires (e.g., Johns Hopkins University Data Management Services, 2013).

By adopting multiple theoretical and practical frameworks, the survey instrument not only identifies individual data habits and needs, but also profiles institutional and community readiness and capability for lifecycle data management, discovery, and reuse. As detailed in the following sections, the adoption and modification of CCMF particularly contributes to the overall understanding of institution-wide and college-level engagement in open data activities. It helps identify major gaps and services needed to develop institutional capability for data stewardship. Before launching, the survey questionnaire was pretested by the cross-campus faculty representatives serving on the University Library Committee. The final survey has 32 questions and uses skip logic to streamline questions based on respondents' experiences or reactions.

The formal data collection using Qualtrics web survey took place in November 2014 and targeted at Teaching and Research faculty (T&R) and Research faculty at Virginia Tech (The Virginia Tech Office of the Senior Vice President and Provost, 2015). A total of 2,532 email invitations were distributed and 652 responses were received, among which are 423 completed entries. They are from eight colleges including the College of Agriculture and Life Sciences (CALS), College of Architecture and Urban Studies (CAUS), Pamplin College of Business, College of Engineering (COE), College of Liberal Arts and Human Sciences (CLAHS), College of Natural Resources and Environment (CNRE), College of Science (COS), and Virginia-Maryland College of Veterinary Medicine (VA-MD Vet Med).

Statistical analysis was performed to test hypothesis and recognize patterns. Content analysis was conducted on the qualitative responses to discern contexts and gather insights on the data-related practices, concerns, and needs of faculty researchers.

Findings

This section reports and discusses the findings about the faculty researchers' data special features, sharing and access activities, as well as their use and reuse practices. Note that the percentages reported are rounded numbers.

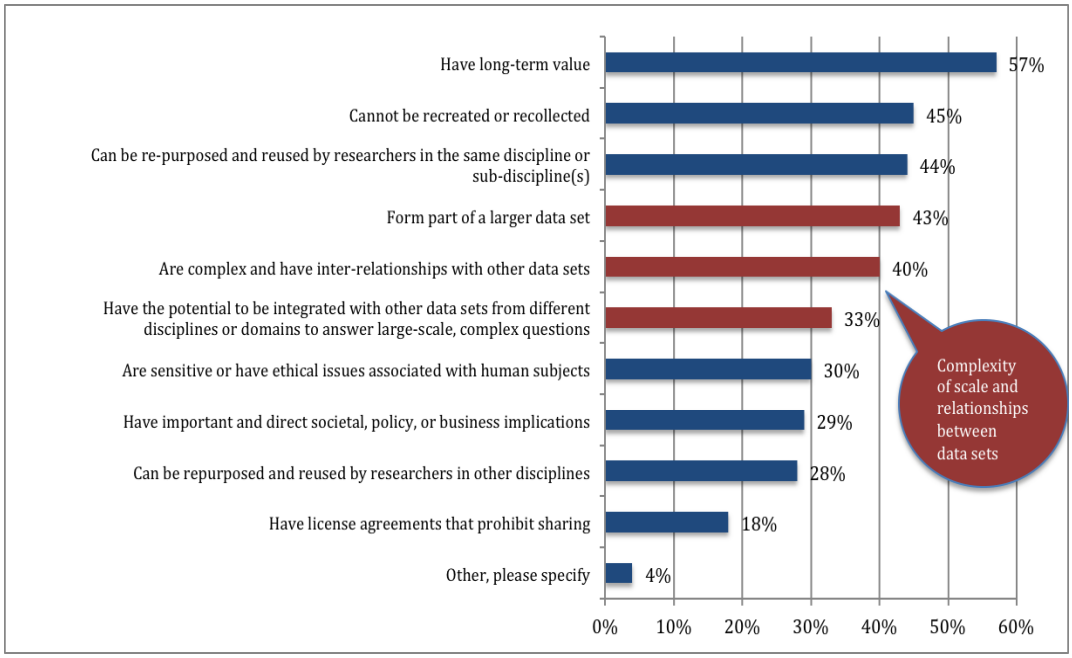
Data Special Features

The survey asked faculty researchers to indicate whether their research data have any of the special features described (see Figure 1). A total of 472 faculty responded and over half (57%) considered their data to have long term value. 45% indicated that their data cannot be recreated or recollected, and 44% reported that their data can be repurposed and reused by researchers in the same discipline or sub-discipline(s). Other special features indicated included: the data form part of a larger data set (43%), the data are complex and have inter-relationships with other data sets (40%), and the data have the potential to be integrated with other data sets from different disciplines or domains to answer large-scale, complex questions (33%). The results suggest that these data assets have huge potential value. They often contain valuable information for new investigations or longitudinal analysis, or support data integration using new analytical techniques. The complexity of scale and relationships between data sets requires

 doi:10.2218/ijdc.v10i2.359

metadata and indexing to support more granular levels of data description and access that can enable elastic discovery into larger or smaller or specific segments of the data sets.

In addition, 29% of the respondents believed their data to have important and direct societal, policy, or business implications. 28% indicated that their data can be repurposed and reused by researchers in other disciplines. As data translate to different values or transfer to different user communities, continued efforts are needed to curate the data for different purposes and communities of practice. These results all make for additional cases and heightened importance for data management, preservation, curation, and sharing efforts.



Total Responses (n) =472

Figure 1. Summary of data special features

Research Data Sharing and Access Activities

The survey asked a subset of questions regarding research data sharing and access activities. These concern "openness of data," "openness of methodologies and workflows," as well as "data discoverability and accessibility." The questions and their measurement statements were adopted, with modifications, from the corresponding sections of the CCMF Profile Tool Template. The results are shown in Figures 2, 4 and 5. Each wedge in these circular charts corresponds to a question statement or statements as listed. The answers are then grouped by the intensity of user engagement using the categories defined in the CCMF profile: Nominal Activity, Pockets of Activity, Moderate Activity, Widespread Activity, and Complete Engagement (UKOLN, University of Bath, and Microsoft Research Connections, 2013).

As to "openness of data," slight modifications were made to the "Widespread Activity" category (under CCMF section 3.3) with the addition of Item 4 as indicated in Figure 2. This new measurement statement was adopted from a faculty respondent's suggestion during the pretest. The results suggest the larger percentages of "Pockets of

Activity" (27%) and "Moderate Activity" (25%), totaling over 50%. In these cases, data are shared within limited scope or under limited conditions. There are smaller fractions of "Widespread Activity"(16%, combining Item 4 and 5) and "Complete Engagement" (8%) in sharing data. 12% indicated "no sharing and no details released" for data. Comments from the participants further described the limited scope of sharing activities either within collaborative teams, or by request, or under various conditions and restrictions. Interestingly, among those who chose the "other, please specify" option, a few respondents expressed their willingness or efforts to share data, which may indicate a transition from moderate, controlled activity to more widespread sharing activity. Special cases like these should be taken into account when further refining the survey measurements and categories in future research.

> "We currently share our data on request. We are moving most of our data to AQUARIUS so that we can share it more easily."

> "Data would be available once published."

> "[A] data sharing agreement is put in place."

> "[We] upload datasets onto International Tree-Ring Database run by NOAA."

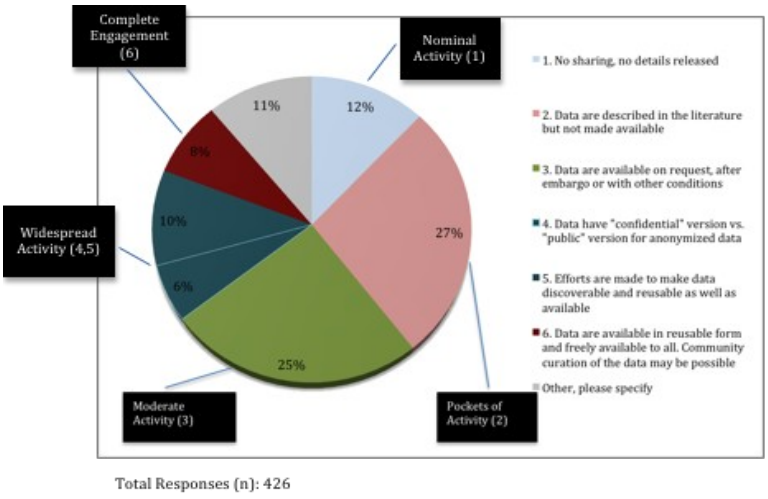


Figure 2. Openness of data

After looking at the overall participants' level of engagement in openness of data, it is meaningful to see if there are differences in openness within the college communities. The proposed null hypothesis is:

> **H0: There is no difference among the colleges in their community level of engagement in openness of data.**

 doi:10.2218/ijdc.v10i2.359

Myers-Lawson School of Construction is a joint school with CAUS and COE, so it was not included in the hypothesis testing. The statistical results below also exclude those who chose the "other, please specify" option. Table 1 indicates the college level engagement in openness of data.

Table 1. College level engagement in openness of data

	Nominal Activity (1)	Pockets of Activity (2)	Moderate Activity (3)	Widespread Activity (4)	Complete Engagement (5)	**Total**
CALS	6	18	20	16	5	65
CAUS	1	4	5	7	2	19
Pamplin	3	7	4	1	1	16
COE	5	20	22	10	8	65
CLAHS	15	17	7	6	3	48
CNRE	2	11	11	6	4	34
COS	5	10	23	7	8	53
VA-MD Vet Med	6	16	3	2	2	29
Total	43	103	95	55	33	329

Given the small sample sizes in CAUS and Pamplin, these two colleges were also eliminated from the hypothesis testing. Pearson's chi-squared test (χ^2) was performed based on the assumptions that each observation is independent of all the others; no more than 20% of the expected counts are less than five; and all individual expected counts are one or greater (Yates, Moore and McCabe, 1999). As a result, the p-value for the chi-square test statistic is $0.00101573 < 0.05 = \alpha$. Therefore, we reject the null hypothesis and conclude that there are significant differences among the colleges in their community level of engagement in openness of data.

Figure 3 demonstrates the distribution of user engagement in openness of data at each individual college. Notably, CALS has similar rates of engagement across "Pockets of Activity," "Moderate Activity," and "Widespread Activity," with each around 30%. Further investigation could determine the clustering effects and the mediating factors on the variances.

The significant differences among the colleges in openness of data suggest different cultures of data sharing activities and community practices. Based on the concentrations of activities identified here, further college-level investigation and different approaches are needed to promote the sharing and reuse of data. For example, for fields of research producing data that often involve human subjects, clinical trials, legal or ethical issues, the promotion activities may focus on how to facilitate the reuse of sensitive, safeguarded, and controlled data, as well as strategies and policies for ensuring the quality and security of data. In sciences (COS), questions may concern what causes the "bottleneck" from moderate sharing activity to complete engagement, as shown in Figure 3, and how to encourage more widespread sharing efforts among researchers in anticipation of broader community engagement.

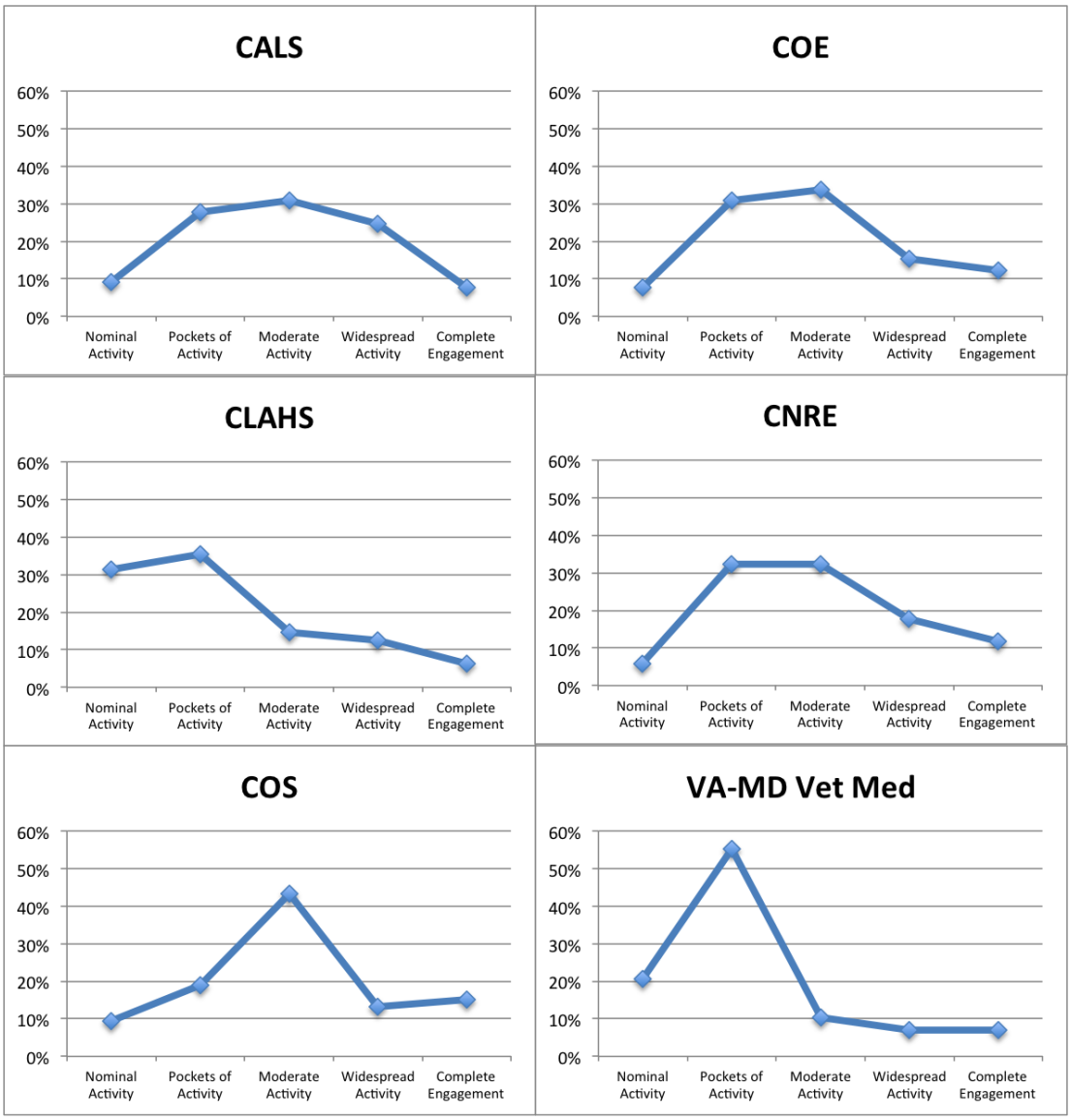


Figure 3. Distribution of user engagement in openness of data within individual colleges

As to "openness of methodologies and workflows," in the CCMF profiling tool (CCMF section 3.4), "Nominal Activity" was characterized as "no sharing, no details released." However, in actual practice, descriptions of methodologies and workflows are usually required components in scholarly publication. As such, this measurement statement would not fit and was eliminated from the current survey. "Pockets of Activity," originally characterized as "released within limited scope" in CCMF, was then replaced with the statement "released within the scope of research publications." The results are shown in Figure 4. Note that other word modifications were also made to the CCMF measurements to further specify and clarify the metrics. The responses suggest a dominant majority of "Pockets of Activity" (72%), indicating the limited release of methodologies and workflows information only within research publications, as is normally required. "Moderate" and "Widespread" sharing activities and "Complete Engagement" only range from 6% - 7% respectively, indicating very limited acts of openness.

 doi:10.2218/ijdc.v10i2.359

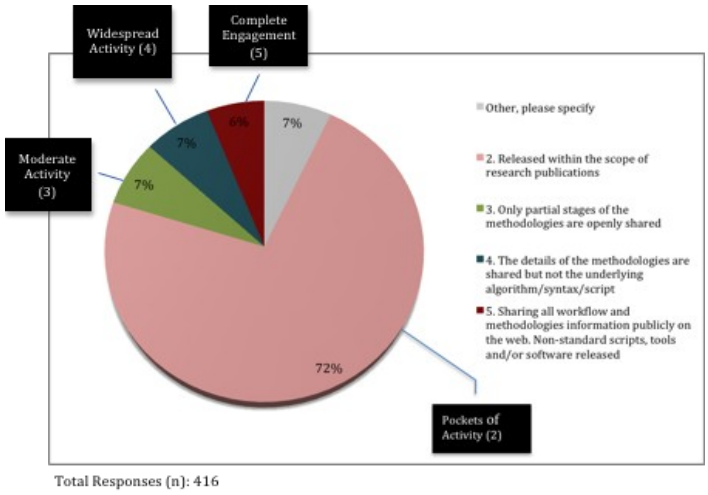


Figure 4. Openness of methodologies and workflows

A few comments demonstrate the researchers' willingness to or actual practices of sharing information broadly:

> "Not all workflow methodologies are available publicly (simply because they have not all been written in complete detail). Have a willingness to 'fill in gaps' to make information readily accessible when possible."

> "Upon publication, the experimental details and results are made available to all through submission to an International database."

Other comments suggest the limited scope of sharing either with collaborators or by request:

> "[Methodologies/workflows information are] shared selectively with collaborating research groups."

> "Details, scripts, software available on request. Some software available on the web."

> "Released with publications and more details given when contacted by other researchers."

Data discoverability and accessibility are another important aspect to ensure the actual value of data to be realized in sharing and reuse activities. The faculty respondents indicated the extent to which their data are discoverable and accessible to others after project completion, as shown in Figure 5. In CCMF, "Widespread Activity" (CCMF section 4.5) was originally characterized by "discovery opened to all but siloed - not interoperable or easy to customize." However, the pretest showed it is difficult to implement this from data owners' perspective, thus this category was eliminated from the current survey.

The results suggest once again that there are higher percentages of "Pockets" (43%) and "Nominal" (21%) activities. It is also notable that 17% of the respondents indicated that their data are discoverable and accessible to all with well-integrated services, as shown in the "Complete Engagement" category. It might be worthwhile to explore how this is realized in different domains or community contexts. Other comments of the participants suggest mixed activities, but overall reinforce the majority of "Pockets of Activity."

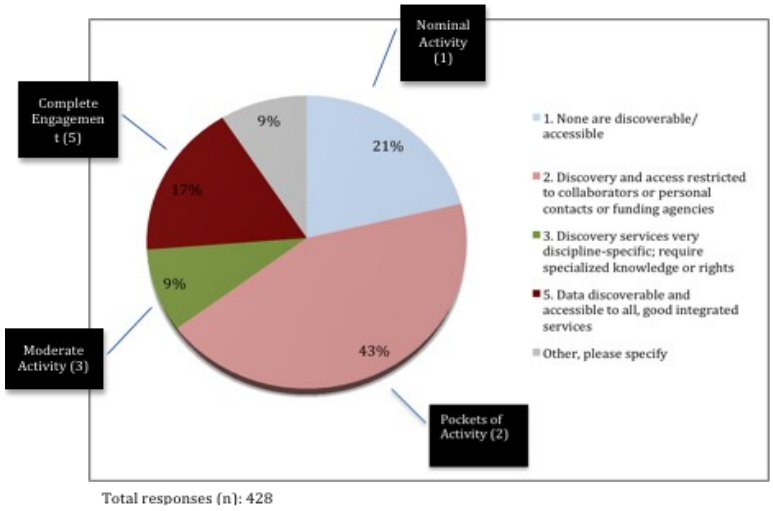


Figure 5. Discoverability and accessibility

The survey further explored the reasons as to why faculty researchers do not make their data openly available to others after project completion. The results suggest confidentiality or data protection issues (58%) are the major factors. After drilling it down by colleges, the results further indicate that agriculture and life sciences, engineering, liberal arts and human sciences are the top three showing concerns for confidentiality and data protection issues, which may be due to opportunities for commercial or industrial applications, technology transfers, protection of inventions, or sensitive issues associated with human subjects, and so on. In this space, data curation services could support putting data in a "dark archive" and prepare for possible future release according to policies (Guedon, 2015). "Dark archive" here refers an archive where access to the data is either limited to a few individuals or completely restricted to all. The next major reasons for not sharing are the time and effort required to prepare data for sharing (44%), and there is no incentive to do so because sharing is not required in some cases (36%). Other reasons and their rankings by response rates are illustrated in Figure 6.

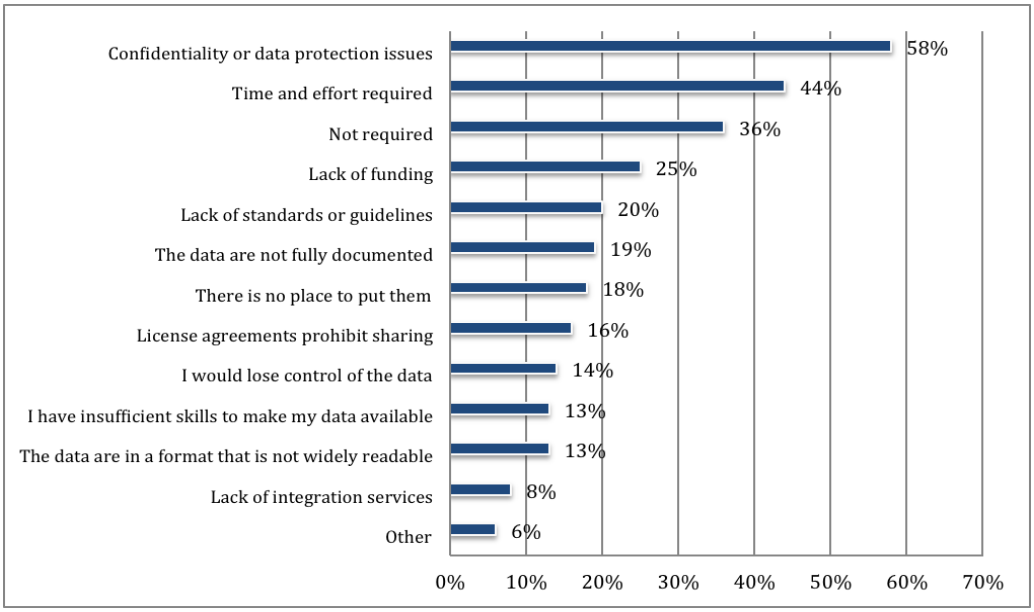
Aside from IRB restrictions, FERPA requirements, and intellectual property issues regarding patents, the participants specified other reasons for not sharing data:

> "It would be extremely difficult to anyone other than a handful of people to make sense of our data."

> "They currently lack sufficient impact to justify the efforts to share more widely."

 doi:10.2218/ijdc.v10i2.359

"Unreasonable to post to public, but accessible if someone contacts us."



Total responses (n): 353

Figure 6. Reasons for not sharing

Data Use and Reuse Practices

This section focuses on data use and reuse practices of the faculty researchers. First, a general question was asked on how faculty reuse existing data in research. Figure 7 shows the slightly modified CCFM categories on this topic (CCFM section 3.5).

The results show that the current reuse practices mostly stay at the "Nominal" (34%, combining Items 1 and 2) and "Pockets" (34%) levels. Reuse of data is very limited and data are only exchanged within limited scope, either with collaborators or personal contacts. There is a descending level of activities from "Moderate" (15%), to "Widespread" (7%), to "Complete Engagement" (6%), indicating that a decreasing number of researchers reuse data actively and systematically.

The survey further asked faculty researchers about the frequency of using existing data from other disciplines or using data collected by others beyond their immediate research teams. It then explored the major concerns keeping them from reusing or repurposing existing data. The results are shown in Figures 8-10. There are 55%-56% of the respondents who never or seldom reused existing data, either from other data producers or from other disciplines. The top three concerns about data reuse include the difficulty finding or accessing reusable data, difficulty integrating data, and possible misinterpretation of data. Over half of the respondents expressed all these concerns.

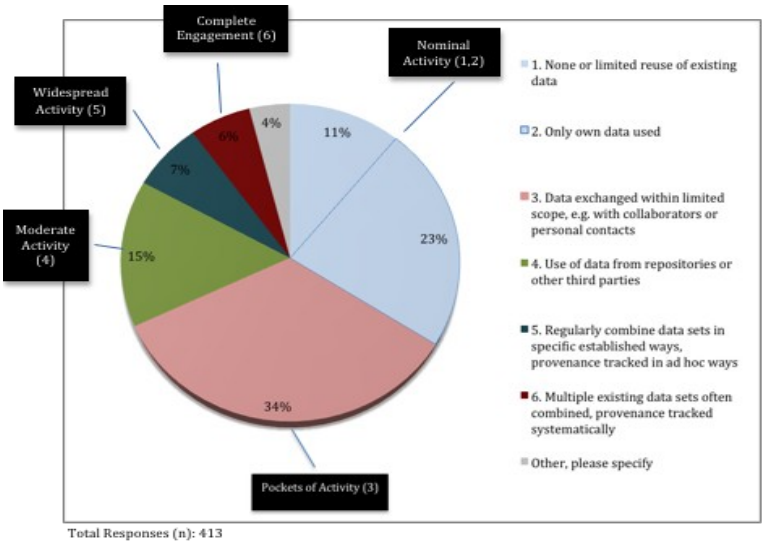


Figure 7. Reuse of existing data

Among those participants who chose the "other, please specify" option, some expressed concerns about "reliability and reproducibility" or data being "not easily accessible." Others described the long term academic culture of overlooking or disregarding the reusability of data:

> "Data of others is rarely applicable to new problems."

> "Collaborators share data but do not widely share unpublished data."

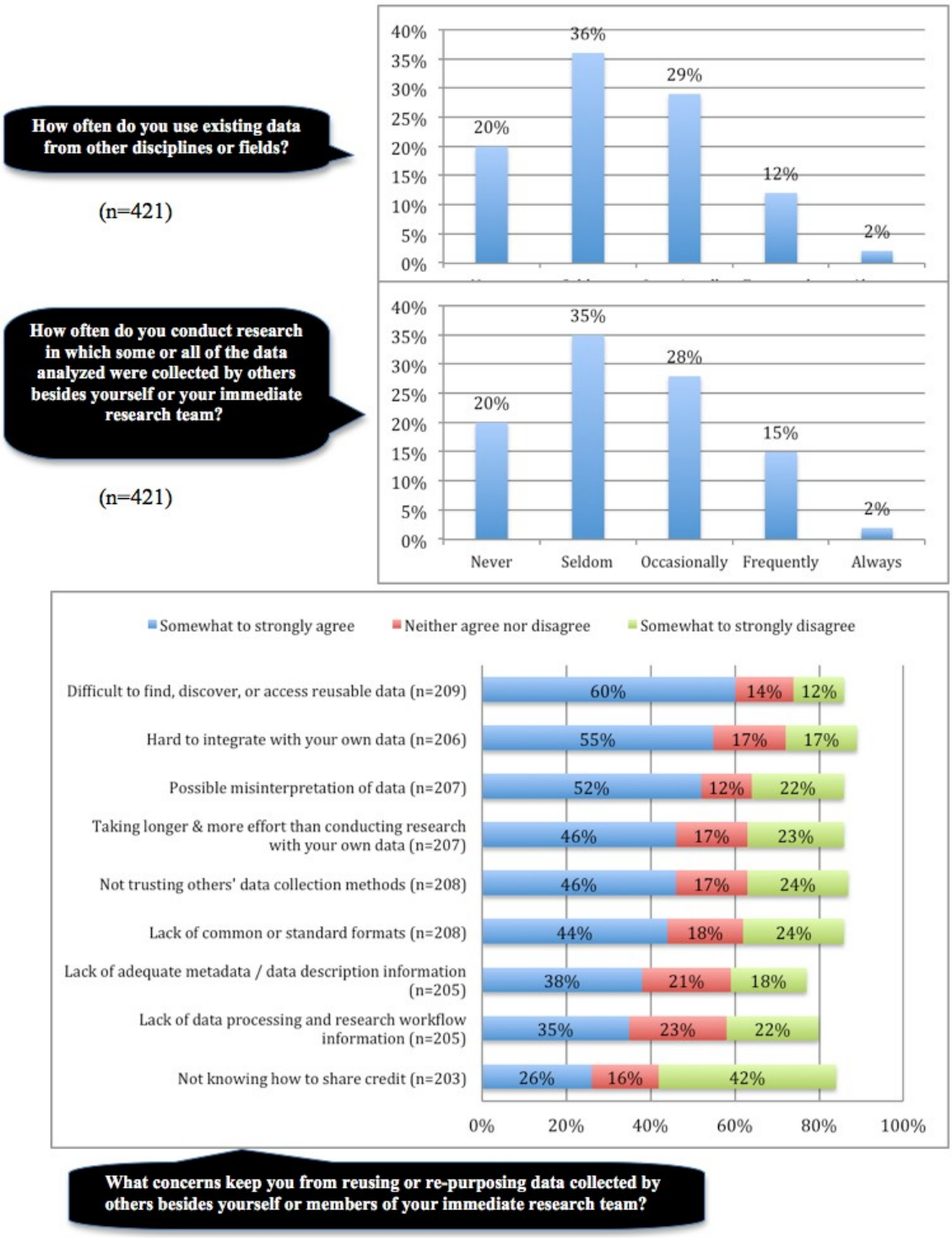
> "If someone collects data, presumably they analyze it. Therefore they don't need you to analyze it. I think this is most useful for global problem solving where you have united efforts to standardize data and assemble in a data set."

> "No access to lab data from other labs other than collaborators. And would take FORVER to go through all data generated by other labs. We stick to published results, and if any questions arise, we contact the authors."

Next, when asked about whether their own data could be reused or repurposed by others *if allowed*, 67% of the respondents identified that some of their data could be reused or repurposed, and 24% considered that all of their data could be reused or repurposed. Only 9% indicated that none of their data could be reused or repurposed. Figure 11 shows the results. It is notable that a total of 91% of the faculty respondents considered their data, more or less, to have potential reuse values. This is in sharp contrast to the small percentages of sharing and reusing activities reported in the previous sections of this paper.

Figure 2 showed that only 49% of the respondents indicated some levels of sharing activities, ranging from Moderate (25%), which is under various limited conditions, to Widespread (16%) and Complete Engagement (8%). This explains why a lower

 doi:10.2218/ijdc.v10i2.359



Figures 8-10. Data reuse practices and concerns

percentage of the faculty (44%, see Figure 1) identified that their data "can be reused or repurposed by researchers in the same discipline or sub-discipline(s)." While in contrast, a much higher rate of the respondents (91%, see Figure 11) considered their data to have potential reuse or repurpose values. However, only a small fraction (24%: 16% Widespread Activity + 8% Complete Engagement, see Figure 2) reported open data sharing activities. The gap between the high percentage of perceived data reuse

values and actual low level of sharing and reuse practices means substantial cost and lost opportunities in the research enterprise.

On a related note, the arrangement of these questions in the survey allows faculty scholars to self reflect on both their reuse practices and encountered obstacles involving others' data, as well as the reuse values and management practices regarding their own data. Hopefully, through switching roles between data users and producers in this context and considering reuse from both sides, the participants could better appreciate the importance of appropriate data stewardship to ensure the discoverability, accessibility, credibility, quality, and interoperability of data. Future research could further investigate the value propositions of these reusable data from the researchers' perspectives and decide what curation services could be devised to capture these values.

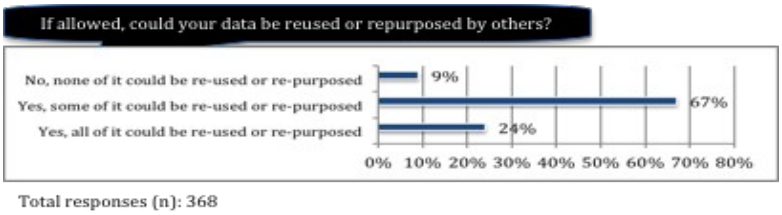


Figure 11. Reuse or repurpose values of faculty researchers' own data

Discussion and Conclusions

A high percentage of the faculty researchers considered their own data to have reuse or repurpose values. However, the openness of data, methodologies and workflows, as well as the discoverability and accessibility of data, continue to remain low. Widespread activity and community engagement in openly sharing data are not yet developed. Besides the highly pronounced confidentiality and data protection issues, the researchers' major reasons for not sharing rest with the required commitment of time and effort and a lack of funding and incentives. These agree with previous studies and are specified in the participants' comments:

> "The main problem is that faculty have to do their jobs (intellectual advancement, scholarship, etc.) and then also act as a full time administrative support staff for themselves. So time is the biggest factor. The physical things that are missing are: easily accessible server space, examples of how to meet the data sharing requirements of granting agencies, and collaborative networks where you can share your data without having to take on ANOTHER project to administer in perpetuity."

> "Providing data for widespread access requires effort. If there is no value to our research group in sharing this data, there is no incentive to do so. We readily share data with funding partners, but if data were shared freely beyond them, they would lose some of the competitive advantage that our data gives them, and they would be less likely to support us financially."

In practical applications, there has been a lack of regular or formalized data documentation practice. As reported in another paper (Shen, 2016), often there are no standard metadata or documentation schemes being applied by the researchers, or only simple, home-grown, self-developed metadata and documentation schemes are being used. It has been identified that guidance and assistance are needed to help researchers transit from localized micro-practices to more standardized, community-sensitive approaches. In particular, actions are needed to carefully select and apply metadata standards to enhance or supplement researchers' own documentation, to register and formalize data structures and formats for specific data types to support automated handling, and to develop logic-based methods to align or merge related disciplinary taxonomies and conceptual frameworks (Shen, 2016).

When in turn asked about their own data reuse practices and concerns, relatively few faculty researchers indicated the reuse of existing data collected by others or from other disciplines. Their major concerns about reusing others' data concentrate on the difficulty finding or accessing reusable data, difficulty integrating data, and possible misinterpretation of data. The poorly exercised data management and documentation practices of data producers are exactly the source of problems they themselves would encounter when thinking about reusing or repurposing others' data from the perspective of data users. Consequently, the highly perceived potential values of data for future research are often lost right after the original work is done.

As such, it is especially meaningful to have researchers self reflect on data reuse issues from both data producers' and data users' perspectives. Self-reflection and projection may heighten researchers' awareness of proper data management for their own sake and the community's benefit. It also offers a channel for us to cultivate data stewardship and bridge communication barriers between data producers and data users. Obviously, such opportunities exist given the high level of interest and demand of faculty researchers in a wide range of data-related educational opportunities and support services (Shen, 2016).

To the researchers, benefits of sharing information, exchanging knowledge, and extending digital platforms are apparent when it comes to research collaboration and scholarly publication, but not quite so when it comes to data. Without recognizing the importance of discoverability and accessibility, one respondent questioned the necessity for openly sharing data: "If someone wants to see the data from someone else's research, don't they just ask?" Similarly, others do not think of sharing data the same way as they have learned to share papers:

> "Unreasonable to post [data] to public, but accessible if someone contacts us."

> "I could make some data available, but no one has asked for any."

> "I'm not sure what you mean by "discoverable." People can get it by request. I don't spend a lot of time making raw data available on the web. It would be a lot of work and there's no point…"

However, if other researchers could not find or locate the data and if they do not know of the original data producers and associated works, they are not likely to know whom to ask and where to request such data to realize the additional reuse values.

Guedon (2015) pointed out that researchers should reflect on their role in the "greater scheme of scientific work" and "be socialized into the network vision of research activities," which is sharing and collaborating, now at the level of data, just as they've learned to share their papers. Libraries can play significant roles in making data retrievable and discoverable by curating the data, providing the metadata, and within existing policies, exposing the data.

When speaking of "scientific ethos," Guedon (2015) further pointed out the need to reward sharing by visibility and prestige, and to approach culture change through educating and socializing researchers into the importance of sharing. As we explore the challenges and opportunities in collecting, analyzing, and disseminating vast amounts of data through a technical lens, we should also act as a key facilitator to demonstrate the potential and impact of data stewardship, especially how such activity improves the way researchers think of, respond to, and understand research questions and scientific challenges. This resonates with the above "value" and "incentive" comment of a respondent.

To promote a university culture that values and rewards good data practice, we need to showcase research on whether data management and curation is indeed a credible contributor to new problem solving and more successful outcomes. One important strategy is to have researchers of related disciplines or with interconnected questions to engage in interactive data activities and focus group exercises to peer review each other's data and identify data reuse possibilities. Two approaches can be deployed. One is to engage stakeholders in making data fit for a given reuse scenario or for solving a grand research challenge. The other is to have stakeholders identify new research questions from data collections by visualizing the data and its connections, finding new patterns in the data, and seeking novel applications. The goal is to have researchers recognize data value and then add value by deciding necessary curatorial activities.

Researchers also need to know about current happenings and future trends in data management and curation. These could be exemplified by practical case studies from end-users and prominent experts in the community. A pragmatic approach is to draw on actual projects, such as scientific data curation efforts or digital humanities projects, to demonstrate how targeted, archive-based, and discovery-enabled projects can enhance research and pedagogy.

In conclusion, this study contributes to the understanding of how academic scholars practice data management and sharing and what problems, concerns, or challenges they encounter when reusing existing data. Future research should examine specific domains or academic disciplines to determine how to best capture and preserve the contextual information and reuse values of data.

References

Achenbach, J. (2015). The new scientific revolution: Reproducibility at last. Washington Post. Retrieved from http://www.washingtonpost.com/national/health-science/the-new-scientific-revolution-reproducibility-at-last/2015/01/27/ed5f2076-9546-11e4-927a-4fa2638cd1b0_story.html

Borgman, C.L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology, 63*(6), 1059-1078. doi:10.1002/asi.22634

Cragin, M.H., Palmer, C.L., Carlson, J.R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *368*(1926), 4023-4038. doi:10.1098/rsta.2010.0165

DataOne. (2013). Scientists and research data: Continuing to build an understanding of your data needs. Retrieved from http://www.dataone.org/news/help-us-understand-how-scientists-work-data

Digital Curation Center (DCC). (2009). Data asset framework implementation guide. Retrieved from http://www.data-audit.eu/docs/DAF_Implementation_Guide.pdf

Emory University Research Data Management. (2012). Faculty practices and perspectives on research data management. Retrieved from http://guides.main.library.emory.edu/content_mobile.php?pid=333927&sid=3327853#box_3327853

Faniel, I.M., & Zimmerman, A. (2011). Beyond the data deluge: A research agenda for large-scale data sharing and reuse. *International Journal of Digital Curation*, *6*(1), 58-69. Retrieved from http://www.ijdc.net/index.php/ijdc/article/viewFile/163/231

Guedon, J.C. (2015). Open data and science: Towards optimizing the research process. February 10th DataOne Webinar and Discussion Forum. Retrieved from https://www.dataone.org/webinars/open-data-and-science-towards-optimizing-research-process

Hendler, J. (2014). Data integration for heterogeneous datasets. *Big Data, 2*(4), 205-215. doi:10.1089/big.2014.0068

Johns Hopkins University Data Management Services. (2013). JHU DMS data management planning questionnaire. Retrieved from http://dmp.data.jhu.edu/assistance/nsf-data-management-plans/#Questionnaire

Jones, S., Ball, A., & Ekmekcioglu, Ç. (2008). The data audit framework: A first step in the data management challenge. *The International Journal of Digital Curation*, *3*(2), 112-120. doi:10.2218/ijdc.v3i2.62

Mundel. T. (2014). Knowledge is power: Sharing information can accelerate global health impact. Bill & Melinda Gates Foundation. Retrieved from http://www.impatientoptimists.org/Posts/2014/11/Knowledge-is-Power

Provost, F. & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, *1*(1), 51-59. doi:10.1089/big.2013.1508

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., et al. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, *6*(6), e21101. doi:10.1371/journal.pone.0021101

Shen, Y. (2016). Strategic planning for a data-driven, shared-access research enterprise: Virginia Tech research data assessment and landscape study. College & Research Libraries [Publication date: July 1, 2016].

UKOLN, The University of Bath, & Microsoft Research Connections. (2013). Community capability model for data-intensive research. Retrieved from https://communitymodel.sharepoint.com/Documents/CCMF-Profile.xlsx

The Virginia Tech Office of the Senior Vice President and Provost. (2015). Faculty handbook chapter 02: Policies and procedures for all faculty. Retrieved from http://www.provost.vt.edu/faculty_handbook/chapter02/chapter02.html

Walters, T., & Skinner, K. (2011). *New roles for new Times: Digital curation for preservation*. Washington, DC: Association of Research Libraries.

Yates, D., Moore, D., & McCabe, G. (1999). *The practice of statistics* (1st Ed.). New York: W.H. Freeman.