12-29-2017

Connecting Researchers to Repositories IMLS Project Report

D. Scott Brandt
Purdue University, techman@purdue.edu

Jake R. Carlson
University of Michigan - Ann Arbor, jakecar@umich.edu

Sherry Lake
University of Virginia, sah@virginia.edu

Connecting Researchers to Repositories

IMLS Project Report[1]
29 December 2017

D. Scott Brandt, Senior Research Data Specialist, Purdue University Libraries
Jake Carlson, Research Data Services Manager, University of Michigan Library
Sherry Lake, Scholarly Repository Librarian, University of Virginia Library

Introduction

Despite a general consensus that making research data available is beneficial to many stakeholders, data sharing/curation is still not performed as an integrated step in most research lifecycles or common practice in the academic setting. (Fecher, Friesike & Hebing, 2015) This is true for a range of qualitative, quantitative and mixed methods researchers, from science, (Tenopir, Dalton, et al., 2015), social science, (Miguel et al., 2014) medicine, (Margolis et al., 2014) and humanities. (Kaplan, 2015) It should be noted that the cultures and practices surrounding the treatment of research data vary by field of study. Funders of research in health and medicine regularly require depositing data, but it has been noted recently by authors in the US that geology, ecology, and archaeology "lag behind some laboratory sciences in making data and samples available." (McNutt et al., 2016) Recent studies indicate that researchers in agriculture did not deposit regularly or occasionally in repositories. (Chang & Milligan, 2016; Andrews, Young, Ochs, Shea, & Morris-Knower, 2016) And a study in Data Science Journal comparing five repositories showed modest increases and totals of data sets published over the past five years. (Assante, Candela, Castelli & Tani, 2016)

This situation is not due to a lack of effort. Many stakeholders have championed the cause of making research data more accessible and taken steps to encourage researchers to share more of their data. The National Science Foundation's (NSF) requirement that researchers draft a two page data management plan (DMP) outlining how they intended to make their data available beyond the duration of award was an important milestone in the push to make research data more accessible. Many academic libraries responded by developing services and support for researchers faced with understanding and navigating through what the DMP requirement meant for them. Helping researchers with their DMPs was their first research data service and was seen as an opportunity for librarians to play a larger role in supporting the research mission of their institutions. (Fearon, Gunia, Lake, Pralle & Sallans, 2013) However, reviewing/revising DMPs does not seem to have been as big a need as first thought, perhaps because of the conventional wisdom that DMPs are rarely scrutinized in most grant proposals. There are examples of libraries helping develop data workflows on projects, best practices in labs, and courses/workshops for graduate students, but this does not seem to be a high level of activity across all academic libraries. (Hudson-Vitale et al., 2017)

[1] IMLS NLG Planning Grant: Enhancing the Data Curation Profiles to help Bridge the Gap between Researcher and Repository (IMLS LG-55-14-0098-14)

Although the response to the NSF's data management plan requirement and other efforts to prompt data sharing has been limited, the pressure to make data more available is quite real. Studies have found that when journal publishers, in addition to funding agencies, put pressure on scientists to share it influences data sharing behaviors.(Kim & Zhang, 2015) Collaborations and integration between repositories and journal publishers, such as with the Dryad data repository, is making it easier for journal publishers to facilitate better data deposit, rather than simply treat data as a supplement file.(http://datadryad.org/pages/submissionIntegration)

Given this need, why aren't repositories used more by researchers? We sought to explore this question in a series of workshops as a means to consider the next steps in developing the Data Curation Profiles (DCP) Toolkit. As an instrument for understanding researchers' data, the Data Curation Toolkit was useful for understanding the data researchers had and what they wanted to do with it, and we were interested in expanding it towards helping librarians take action to increase data deposits. Though we recognized that this would be a complicated question to try and address, we believed that the DCP Toolkit provided a solid starting point and that it could be leveraged to create a means for librarians to take action to increase deposits. However, by the end of the two workshops we came to realize that maybe we need to approach this problem in a different way.

Background

The Data Curation Profiles (DCP) Toolkit was designed to assist libraries in developing and offering data services through enabling librarians to engage with researchers to better understand a particular data set and its components, to learn about a researcher's current practices in managing, sharing and curating the data set, and to identify areas of unmet needs in managing, sharing or curating the data set to inform possible services. With funding from the Institute of Museum and Library Services (IMLS), the DCP Toolkit was developed by librarians and library faculty at Purdue University and the University of Illinois with the intent of better understanding "who will share what (kinds of data) with whom, and when." (Brandt, Witt, Carlson, Palmer, & Cragin, 2007). Answering this question went beyond exploring motivations or barriers of sharing by looking at what it would take "to support deposit of data into shared repositories." (Cragin, Palmer, Carlson & Witt, 2010) The resulting DCP Toolkit was designed with collection building and management in mind, "to help gather information to make local data development policies and selection and deselection decisions." (Witt, Carlson, Cragin & Brandt, 2009)

Since its release in 2010, the DCP Toolkit has been repurposed in a variety ways, from helping do short data interviews to scoping campus wide needs, and resulted in the Data Curation Profiles Directory, a series of Profiles which provide insight into how research data is managed in different disciplines (http://docs.lib.purdue.edu/dcp/).The Data Curation Profiles project has 32 Profiles published in the Directory, and is a resource used in library schools, such as at the University of North Carolina, University of Illinois Urbana-Champaign and the University of Michigan.

We held two workshops at Purdue University to explore the challenges of increasing data deposits into repositories and to better understand the changes in cultural practices needed to make data deposit a natural component of the research workflow. The impetus behind holding these two workshops came from an IMLS planning grant to redesign the Data Curation Profiles Toolkit (DCPT) and to improve its capabilities in strengthening connections between the needs of researchers and the services offered by data repositories. Outcomes from these workshops were intended to inform a redesign of the Data Curation Profiles Toolkit (DCPT). Specifically, we were seeking to improve upon the DCPT as a means to better facilitate the deposit process through bridging the gap between the "active" stages of the data lifecycle management where the data are under the purview of the researcher to the "curation" stages of discovery, access and preservation where stewardship of the data is transferred to a third party operating a data repository. More broadly, the outcomes of these workshops were intended to further define and clarify the issues and barriers faced by the data curation community in attracting and facilitating deposits. It was our hope that a thorough articulation and in-depth examination of the issues surrounding the deposit of data from multiple vantage points would serve as a foundation for developing the next iteration of the DCPT and to further community efforts to bridge the gap between researchers and repositories.

In the first workshop in 2015 we strove to identify the issues surrounding the transition from active use to 3rd party curation. We sought to do this "from the perspective of data" to identify possible responses to address these issues. In the process of discussing getting data into a repository, we looked at facilitators and inhibitors to depositing data, and looked at activities of consumers and producers of data. We sought to put this together in a Business Model Canvas (see for example https://strategyzer.com/canvas/business-model-canvas), but found it difficult to reconcile partners, activities and resources to develop value propositions. We were able to develop a long list of possibilities, but found it rather difficult to build solid a business case out of our work.

In a second workshop in 2016 we took a different approach. What if we looked at repositories from an entrepreneurial perspective and treated them as start up ventures? What would librarian interactions with researchers about their data look like if librarians took on the role of entrepreneurs seeking to identify and respond to the needs of researchers, as a market segment, with their data? Could applying the strategies and approaches of start ups enable libraries to develop services that would solve the real world problems faced by researchers so much so that they would be eager to use them? Recent work in an area called Lean Launchpad put an interesting spin on customer discovery, identifying a viable solution for problems, and creating a market fit to address needs. We walked through a startup process with an investigator of an NSF grant who teaches Lean Launchpad to faculty and graduate students. We believe what we learned can inform future explorations on connecting researchers to repositories. We know that making research data widely available can benefit the public, the research community and the individual researcher him or herself. The challenge is in finding ways in which data sharing and data deposits will become a normative part of the research process in all fields rather than an exception.

Literature Review - Benefits of ready access to open data

Before we get into describing the workshops, it's worth reviewing the work that has been done on the benefits of open access to research data. Benefits of access to readily available open data are numerous. Christine Borgman, in her 2012 article, "The Conundrum of Sharing Research Data", describes four reasons for sharing data: 1) To reproduce and to verify the results of past research; 2) To make products and the results of publicly funded research available to the public; 3) To enable others to ask new questions of the existing data; 4) To advance the state of research and innovation. (Borgman, 2012)

Perhaps the strongest argument for sharing research data is the ability to verify and reproduce research results. Being able to reproduce a study validates analysis and confirms the science and thus increases the value of the investment made by the funders. Sharing data encourages others to use it and investigate new uses and helps to identify errors and discourages fraud and also increases the value of funding dollars by avoiding duplication of data collection. Reusing shared data has the potential to increase research efficiency and quality.

However, although it is clear that while "most researchers appreciate the benefits of sharing research data, on an individual basis they may be reluctant to share their own data". (Van den Eynden & Bishop, 2014) Data sharing is often difficult to do based on the complexity of data, current research practices, a lack of meaningful and direct incentives, costs, intellectual property, and public policy. (Borgman, 2012) As a result, making data open and freely available is not yet a routine part of researchers' workflow or process. In a 2007 editorial in *Nature Neuroscience*, stated that unless researchers are given "credit for good citizenship in promotion decisions and give preference in awarding grants", data sharing will not happen. The editorial concludes that the "scientific community needs to develop better incentives to encourage compliance and reward those who share". (Nature.com, 2007)

Efforts at requiring researchers to share data: U.S. Funding Agencies

In the U.S., the Office of Management and Budget (OMB) Circular A-110 provides the federal administrative requirements for grants and agreements with institutions of higher education. In 1999, OMB Circular A-110 was revised to provide public access under some circumstances to research data through the Freedom of Information Act (FOIA). U.S. funding agencies implemented the OMB requirement in various ways by encouraging or asking that data from federally funded awards be "shared".

In 2002 the National Science Foundation (NSF) implemented its sharing requirement based on the OMB Circular A-110 statement by updating its policy requiring data sharing: "Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing." This requirement did not include guidelines on how data sharing should be done.

In 2003 the National Institutes for Health (NIH) implemented their data sharing requirement "Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data". (NIH 2007) All NIH proposals after October 2003, seeking $500,000 or more in direct costs, were to include a plan for sharing final research data, or state why data sharing was not possible.

These two data sharing requirements did not result in much of an increase in the amount of research data being shared. (Piowar, 2011) So in 2005 the National Science Board called for greater access to data from federally funded research of the National Science Foundation recommending that the NSF develop a strategy to provide an "effective framework for planning and managing NSF investments". The report also recommended that the NSF require research proposals contain a data management plan for review. (NSB, 2005)

Although it took awhile to respond, in 2010 the NSF announced its Data Management Plan (DMP) requirement. The guideline now states, "Proposals must include a supplementary document of no more than two pages labeled 'Data Management Plan'. This supplement should describe how the proposal would conform to NSF policy on the dissemination and sharing of research results." This policy went into effect in January 2011.

In 2012 the NSF made a change in their instructions for preparing the researchers' "Biographical Sketch". One section was renamed from "Publications" to "Products" and included instructions that "products" could include, but not limited to: publications, datasets, software, and patents. In 2014 NIH followed suit, instructing researchers to "emphasize accomplishments" instead of just listing publications.

On February 22, 2013, the White House's Office of Science and Technology Policy (OSTP) took action to strengthen the data-sharing requirement further by issuing an executive directive. The directive stated purpose was "to increase access the results of federally funded scientific research" by requiring the results of taxpayer-funded research – both articles and data – be made freely available to the general public. This requirement extended the NSF DMP requirement to other federal agencies (those making over $100 Million in annual external contracts). The goal of the directive was for the plans "to have clear and coordinated policies for increasing [public] access". (Holdren, 2013) Since then, 28 funders have established policies for data sharing and management requirements. (https://purr.purdue.edu/start/funder-requirements)

Efforts at requiring researchers to share data: Journals

Funding agencies are not the only groups interested in making data open and available. Journals have a responsibility to ensure that other researchers can replicate and build on the studies that they have published. Journal publishers have argued that making data available fosters scientific progress and allows others to benefit from it, and believe, researchers want to see their work used and cited by others. (Klump, 2017)

The Nature Publishing Group is one example of how academic publishers have been adopting data sharing policies. Publishing in any *Nature* journal requires authors to make the materials, data and associated protocols underlying the paper available. It's early efforts to "police" sharing of data (when the requirement was only to share when asked) resulted in Editors resolving complaints. In the Fall of 2016, the Nature Publishing Group initiated their latest data availability policies which included the following statements: "First, the sharing of research data is a condition of publication in Nature journals and second, each article must have a data availability statement". Data availability statements are meant to provide more transparent and consistent information about where and how data supporting published articles are available. This supports the reuse, where possible, of data for further research and validation or reanalysis of findings by other researchers. Data availability statements also support researchers' compliance with the requirements of funding agencies. (Vasilevsky, Minnier, Haendel, & Champieux, 2017)

The Public Library of Science (PLOS) is another example of a publisher taking steps to promote public data sharing. In 2014, PLOS journals clarified their data availability policy to "make all data underlying the findings described in their manuscript fully available, at the time of submission" and encouraged depositing and sharing data in PLOS suggested repositories. (PLoS ONE, 2015) The requirement goes further to say that refusal to share data, related metadata and methods will be grounds for rejection of future submissions. When data requests or questions about the data go unanswered by the authors of a published article, PLOS has issued expressions of concern alerting their readership that their data policy is not being followed. (PLoS ONE Editors, 2017)

Efforts at requiring researchers to share data: Societies

Scholarly societies also have an important role in leading and facilitating discussions about the future development of open access to data. These types of discussions require input from multiple stakeholders including researchers, funders, policy makers, data repositories, and publishers. (Norman, 2014)

The British Ecological Society (BES) introduced a mandatory data archiving policy for its journals at the beginning of 2014 to increase accessibility and improve preservation. BES thought it was important to mandate making data published in its journals publicly accessible; hoping to encourage a behavioral change in the ecological community. BES reports that since the introduction of the mandate the journals have seen an average 6.7% increase in submissions. (Norman, 2014) To increase the legitimacy, credibility, and openness of intellectually diverse research communities, the American Political Science Association, in 2014, integrated "Data-Access & Research Transparency" "DA-RT" principles into their Ethics Guidelines. (APSA, 2016) And likewise the American Geophysical Union, in its "Scientific Integrity & Professional Ethics" guidelines, states that members have a responsibility to share data & findings openly and promptly, which is detailed in the "Publication Data Policy": all data necessary to understand, evaluate, replicate, and build upon the reported research must be made available and accessible whenever possible. (AGU, 2017)

Why Attempts at Requiring Sharing Data have Not Succeeded

Unfortunately, these efforts by federal agencies, journal publishers and scholarly societies to get researchers to share data have not yet resulted in a substantial increase in data deposits. Even in fields with mature policies, repositories and standards, research data sharing levels are low and increasing only slowly, and data is least available in areas where it could make the biggest impact. (Piowar, 2011) It is evident that it is not just policies and stated requirements that impact researchers' decisions to share data; other factors are also likely under consideration. As Fecher explained, data sharing in academia as a "multidimensional effort that includes a diverse set of stakeholders, entities and individual interests." Barriers to sharing data are best understood as a convergence of multiple factors including: social/cultural "norms", technology barriers, and economic barriers (including time). (Fecher et al., 2015)

Fecher is one of several to explore the data sharing process from the researcher's point of view. He concludes that clear research policies with incentives for data sharing do have an effect on improving the quality of research that is shared. Researchers not only need to have a clear understanding of "why" they should share data, but also need to know "how" to do so. (Fecher, et al., 2015) Roche and his colleagues propose ways to increase the use and re-usability of data published in repositories by allowing for flexible data embargoes, encourage communication between data collectors and data re-users, make re-use policies clear, and encourage recognition by funders and institutions for publicly sharing data. (Roche et al., 2014)

One particularly notable barrier for researchers is the lack of rewards for managing data and making it usable by others outside of its creators, and in sharing it. For most researchers career rewards come from publications, not data sharing. There is no universally accepted mechanism for data creators to obtain academic credit, especially in the sights of Promotion and Tenure committees, for their creating and then sharing data. Without such incentives, researchers tend to only invest minimal time and effort to manage and share their data openly with others outside of their research team, if any such effort is made at all, which leads to poor documentation and datasets that are hard to find or reuse. (Friesike, 2015) Furthermore, most research communities have been slow to recognize data as a “first class output” of research, deserving the same level of attention as a journal article, book or other formal publications. No one is checking on the quality of the data, because there is no requirement to make data “useful”. (Roche et al., 2014)

The lack of clear and strong expectations from publishers has been identified as another barrier to data sharing. A recent study showed that a large number of journals provide no policy for data sharing. (Sturges et al., 2015) One study went a step further in stating, “journal publishers do not currently provide adequate direction through policy documentation and guidance” and need to work more closely “with data repositories to provide specific procedures concerning data deposit.” (Charbonneau & Beaudoin, 2015) As far as societies go, some, like the Ecological Society of America take on the responsibility of making authors submit data sets with paper, but most, “have neither the mechanism for authors to submit supplementary data nor a way to share such data.” (Herold, 2015) And while some, such as the American Society of Naturalists work with the Dryad data repository to facilitate deposit, have admitted that there mandates are “loose by design.”

Ultimately there are many explanations as to why researchers are not yet sharing their data. York, Gutmann and Berman (2017) conducted an extensive literature review on the subject and found six overarching factors behind what they deemed to be a “stewardship gap”, the amount of valuable data created versus the amount that is protected through active stewardship. The six factors they identified are: culture (attitude and norms on the value of data stewardship), a lack of knowledge about stewardship, low commitment, confusion on responsibility, lack of resources and lack of stewardship action (York et. al., 2017). The NIH has also been interested in identifying the issues behind the low rates of data sharing. The NIH have expressed alarm in the frequency of published reports that claim a significant result, but then cannot reproduce it, and so they are exploring ways to provide greater transparency of the data that are the basis of published manuscripts. They have found a complex array of factors that seem to contribute to the lack of research reproducibility, including the need for additional training for researchers in managing their data. (Collins & Tabak, 2014)

Other studies on data sharing and reuse explore the process from the consumer’s point of view. Separate studies by Curty and Faniel describe the factors researchers consider when deciding to use data produced by others. Curty uncovered that the “more practical and social benefits [researchers] perceive from reusing data, the more likely they would reuse data”. Her study concluded that actual data reuse is poorly accomplished due to the lack of incentives by funding agencies and policy makers to leverage the reuse of data.(Curty, 2015) Faniel concluded that in order to reuse data, researchers must understand the context the data was collected; assess that it is relevant to them. Researches need to be able to make judgments on the data, trust the data before they would reuse it (Faniel & Jacobsen, 2010).

Efforts to understand researchers and the data they produce along with needs

Though data sharing is not yet a widespread practice in most fields, there are researchers who have made their data publicly available. Understanding how and why these researchers have shared their data and the data reuse practices of researchers are critically important to the development of data infrastructure, management, preservation and curation systems at an academic institution.

One of the earliest studies on data sharing and reuse, was a longitudinal (10 year) study by Wallis, et al. that explored data sharing practices focused on the willingness of researcher to share data and their motivations to share. (Wallis, Rolando & Borgman, 2013) Research from 2011 by Carol Tenopir, et al out of the DataONE project further explored data sharing and withholding practices, from the perspective of the data producers. This seminal survey of 1329 earth scientists identified that barriers to sharing data are deeply rooted in research cultures, and that data sharing would more liked be served by creating new and easy to use infrastructure and tools than changing culture. Specifically it explored where and how researchers are willing to share data and what the motivations for sharing. (Tenopir et al., 2011) They found that ''While the majority of researchers believe that colleagues should share their data, only a minority of respondents actually share their own data with individuals who did not help in gathering the data''. On the other hand, a majority of those respondents are amenable to sharing at least some of their data; they also favor reusing others' data given certain stipulations.

Tenopir's 2015 follow-up survey discussed the changes in data sharing and reuse practices as well as perceptions and examined how these practices and perceptions changed, or not, over the four years since the baseline study. The follow-up survey was taken well after the 2011 NSF requirements of a data management plan had been implemented. These new results showed an increased acceptance of and willingness to engage in data sharing, as well as a modest increase in actual data sharing. Tenoir's study concluded that for researchers, the tendency to share data is context-dependent. Differences in researchers' attitudes in willingness to share depend upon research domain, age and country of origin. (Tenopir et al., 2011; Tenopir, Dalton, et al., 2015) Variations in institutional support, and the available technological infrastructure were also factors that affected researchers' desire and ability to share their data.

The impact of research domain on data sharing practices has formed the basis of several studies. For example, Kim has studied researchers in two different fields: STEM and Social Science and found that both groups' sharing behaviors were influenced by perceived career benefits and risks. This result was similar to that of Willis' longitudinal study where "researchers are willing to share data if they receive credit to publish their results". (Kim & Adler, 2015; Kim & Zhang, 2015) But the two groups differed in factors that would encourage them to share. (Kim & Zhang, 2015) STEM researchers said they would share more if risks were eliminated and if the benefits of sharing were emphasized more. Social Scientists would need better career benefits and more obvious benefits for their reputation before they would increase their sharing.

A European study commissioned by Knowledge Exchange, gathered evidence, examples and opinions through conducting interviews and focus groups, on current and future incentives for research data sharing from the researcher's point of view. (Van den Eynden et al. 2014) The results of the study produced recommendations for policy and practice on how various stakeholders (research funders, societies, research institutions, data repositories and publishers) could best incentivize data access and reuse. It was recommended, in part, that funders invest in infrastructure and promote reuse of existing data resources and that research institutions incorporate data impact on PhD assessments and set expectations for data sharing within the institution.

Understanding research data sharing and reuse practices of researchers is important to the development of data infrastructure, management, preservation and curation systems at an academic institution. Crowston's "Personas" project combined results from earlier studies and results including sources such as usage scenarios from DataONE and the Data Conservancy project and the Data Curation Profiles. Personas were found to be useful to understand users and their needs and useful for others trying to develop systems or services for data sharing. (Crowston, 2015) Currently under development is research by Shen, which incorporates multiple frameworks, models and templates to create a complex survey instrument to identify data sharing habits and needs of researchers, but also identifies gaps and services needed at an institution level. (Shen, 2016)

Another set of institutional studies focused on surveying local researchers and stakeholders in order to inform local library services, using tools such as the Data Curation Profiles toolkit, the Data Asset Framework, DataONE's research data survey and other institutional data management surveys (McLure, Level, Cranston, Oehlerts & Culbertson, 2014; Parham, Bodnar, & Fuchs, 2012; Peters & Dryden, 2011; Westra, 2014; Whitmire, Boock & Sutton, 2015) Goals of these different efforts were similar: to understand data management and sharing practices, to inform services at a local level regarding the behaviors, needs, interests and concerns of data and to make recommendations including policy recommendations.

Conclusions from studies like these on research on data sharing and reuse behavior demonstrate a wide range of data sharing and reuse practices that suggest variance in practices, but also show a commonality in needing of for better tools, more support services, training to develop and skills to manage data, and incentives for sharing and reuse. Studies like these provide insights for informational professionals to enable them to better support and facilitate data sharing. The results from these studies can provide guidelines for policymakers, open data advocates, and data repository stakeholders to better attune policies and repositories to researchers needs.

Library Services and Support for Data Sharing

Involvement of academic libraries in e-science and e-research has been seen as a natural extension of their electronic resource management and digital stewardship responsibilities. Libraries have been able to connect research data management with historical and contemporary areas of professional practice, including materials selection, metadata creation and collection management; reference services, information literacy, and research consultation; and scholarly communication, open access, and institutional repositories. Libraries have recognized that they should start supporting researchers in managing and sharing data and some institutions have done so through advice and support with data management plans others have contributed to the use and reuse of research data by teaching techniques for sharing research data and promoting open data access. Libraries create value by extending their stewardship and service activities to the management and sharing of research datasets. (Corrall, Kennan & Afzal, 2013) Librarians can't force change, but can help facilitate it through identifying needs and aligning services to stakeholder these needs accordingly.

Multiple librarians have made the case for libraries providing services to support the data management, sharing and curation needs of researchers. In her introduction to her edited book

Research data management: practical strategies for information professionals, Ray states that "library and archival communities have been deeply involved with developing best practices for managing digital data for long-term use… These protocols are now being used as the basis for library services for research data." (Ray, 2014, intro) In his conclusion in his paper *The Emerging Role of Libraries in Data Curation and E-science*, Heidorn states "libraries have the skill sets, longevity, and most of the infrastructure needed to accomplish this task for many types of data. If libraries do not actively engage in the task, then society may choose to create a new type of institution to curate digital data". (Heidorn, 2011)

To keep up with changes in the data landscape, librarians have been racing to reinvent themselves. Librarians have reacted quickly to the funder requirements developing research data management support services and repurposing institutional repositories to take in datasets. Funder mandates have been a major driver for the establishment of data management support services. It has been recognized that libraries cannot provide data management and sharing solutions totally on their own, but need to collaborate with other institutional departments such as research support and IT services. (Pinfield, Cox, & Smith, 2014)

Case studies centered on library engagement in e-science began to emerge in 2008. Services during these early years were built on existing practices across the libraries, in areas from the reference interview and information literacy to digital preservation and repository development, as well as developing new models of practice, especially in relation to assessing data curation needs. (Corrall et al., 2013) In 2009, the ARL eScience Task Force surveyed ARL institutions on their e-Services and data support services to understand the changing requirements for professional skills and infrastructure to address the "new data stewardship". This survey found twenty-one libraries were already providing infrastructure or support services for e-science, and another 23 intended to do so. (Soehner, Steeves & Ward, 2010)

The NSF DMP mandate went into effect in January 2011, thus prompting a move from supporting eScience to a more direct focus for research data management. Using the 2009 survey as a baseline data about institutional about planning structures project, program and services, Fearon, et al conducted a follow up survey in the 2013 ARL SPEC Kit #334: *Research data management services*. The SPEC kit helped librarians compare services across institutions and by peers, and to inform creation of new services. It provided a snapshot of what research data management activities ARL libraries are currently involved in, what human resources are being used to provide these services, and projected service provision. (Fearon et al., 2013)

As was seen in the Fearon, et al's ARL SPEC Kit, the 2011 implementation of the NSF data management plan requirement was the impetus for a significant number of university libraries to create data services. Briney et al's 2015 survey showed, that within only a few years of the requirement going into effect, half of the major research universities offered data services. Briney noted that this was a large increase from the approximately 20% of ACRL libraries offering data-related services previously observed by Tenopir, Birch & Allard (2012). By the time of Briney's study, the results suggested, "Data services at libraries have passed the point of novelty and are becoming mainstream". (Briney, Goben & Zilinski, 2015) In 2017, Hudson-Vitale, et. al. produced an ARL SPEC Kit (#354 *Data Curation)* that focused on the state of data curation services offered by ARL libraries. Using the Center for Informatics Research in Science and Scholarship's definition for data curation as "the active and on-going management of data through its lifecycle of interest and usefulness to scholarly and educational activities", (CIRSS, 2006) this SPEC Kit sought to understand the level of investment made by libraries not just to help researchers manage their data but to prepare it for a life beyond its point of origin. They found that of the 80 libraries that responded 51 were already offering data curation services and another 13 were in the process of developing services (Hudson-Vitale et al., 2017).

Librarians have developed techniques and tools to identify and develop support services for managing and sharing research data. Tools such as the Data Curation Profile Toolkit and have helped librarians uncover the data and types that researchers generate. By identifying and naming transformations of data stages, information professionals can target services that address real world scenarios. Mapping these stages together to create life cycle models has helped librarians to identify potential areas of need, develop services around these needs and then to communicate the data services the library has to offer. (Carlson, 2014)

Assessment of Data Services in Libraries

With the advent of data management and similar services in the library comes the need to understand how and to what extent these services are successful in meeting the needs of researchers. It was noted in 2013 that while cultural changes toward data sharing were indeed needed, research data management services were an "unfolding patchwork of challenges" and there was a large "gap between service provision and customer needs." (Pryor, 2013) In the ARL SPEC Kit later that year it was noted that many libraries were still experiencing "'growing pains' of new service development" around research data, and that uptake is slow. (Fearon et al. 2013) In 2014, librarians at the University of Michigan Library did interviews with librarians at eight institutions focusing on their research data management support services and how they were developed. The results of these interviews were then plotted on a timeline to determine the key steps in developing data services, which were defined as: garnering administrative backing, conducting needs assessments, developing campus partnerships, crafting services and defining staffing and job responsibilities. (Akers, Sferdean, Nicholls, & Green, 2014). Carol Tenopir and her coauthors report that in ARL libraries many librarians have professional interest in, and feel equipped for, future engagement in research data services. Tenopir's study assessed the extent of libraries involved in research data management from technical infrastructure development to support and advisory services. (Tenopir, Hughes et al., 2015) Pinfield used a qualitative approach from interviews of UK staff to examine the roles and relationships involved in research data management. Through this study, he created a model to identify the layers of activity, multiple stakeholders and drivers and the factors of implementing research data management. (Pinfield et al., 2014) The model helped clarify different issues in research data management and identified layers of activity, multiple stakeholders and drivers & a large number of factors. At the time of Pinfield's study library services were still emerging. But the findings provided a starting point for prioritization by suggesting themes and a model to be used to benchmark current Library activities against the model.

Service assessments are an inherently a local process intended to reveal discontinuities between resources and stakeholder needs. One of the unexpected results from a study done by Stephan Kutay at California State University at Northridge was that most scientists did not believe that their organization was doing a sufficient job in helping them with data preservation. Some didn't know if their library was offering help at all. This local assessment revealed a need to further promote the librarians as information experts, partners, collaborators and consultants in the areas of content management and access of faculty-owned research assets. (Kutay, 2014) Another survey of data management practices at the University of Houston revealed that there was more than one unit on campus providing data management support. (Peters & Dryden, 2011) This led to the creation of a campus-wide working group, lead by the library, to promote more efficient coordination of data management initiatives and to increase communication among campus offices and library departments. At the University of Minnesota, Lisa Johnston used the results of a needs assessment and a workflow assessment to create their institutional data repository. The curation workflow model for repositories had two goals: to figure out curation/repository services and help to figure out library services. Including a workflow model in addition to assessment allowed Minnesota to test and expand technical capacities and support data management. The outcome gave a more realistic sense of the overall capacities and expertise needed to develop a sustainable data curation service model. (Johnston, 2014)

To help institutions “boost institutional support of e-research and the management and preservation of our scientific and scholarly record,” the Association of Research Libraries and the Digital Library Federation developed the E-Science Institute in 2011. To strengthen its emerging research data services, Oregon State University (OSU) Libraries participated in the E-Science Institute in 2012. A goal of the Institute was to create a strategy that would guide the development of identified services at OSU. Part of the strategy was a campus survey to help the library move from basic research data services to providing more focused services that meet specific local needs. (Sutton, Barbara & Whitmire, 2013) The OSU Library is using the results of their faculty survey results to further discern campus needs and direct an expansion of library and technology support services (Whitmire et. al., 2015).

Changing our Approach to Changing the Culture Around Data

Despite the heavy investments made by researchers, academic libraries and others in providing resources and services to support data management, sharing and curation, we have not yet seen researchers routinely depositing their data into repositories. This was noted in a panel at the 2015 International Digital Curation Conference titled, “Why Is It Taking So Long?” (IDCC10, 2015) Torsten Reimer described a point of view that the diffusion of research data management requires a massive culture change, and given how difficult that is, things aren’t really going that smoothly. He postulated that perhaps there is a perception among researchers that the cost-benefit ratio for curating data isn’t “right”—that it takes a lot of effort, but there is little benefit for doing so. Given that data sharing, as it has been defined and promoted to researchers, has not yet caught on, perhaps we need to explore employing a different approach. One possible approach is to recast our view of data management activities as a series of tasks to be completed and towards a model of innovation to be adopted.

Data Management as Innovation

In his book, *Diffusion of Innovations*, Everett Rogers defined innovation as an “idea, practice or object that is perceived as new by an individual or other unit of adoption” which “need not just involve new knowledge.” (Rogers, 2010) He acknowledges that because he mostly analyzed technology, he used the terms “innovation” and “technology” synonymously. Thus, sometimes his theory is called “diffusion of technology.” However, Rogers acknowledged that technology could include a philosophy, event or process. Diffusion of innovation looks beyond technology immersion, to the adoption of ideas. Take for instance online shopping; the activity may be performed using various technologies (websites, encryption, e-checkout/payment, etc.) but online shopping is an innovative idea in and of itself.

Vaughan Jason, in attempting to define what innovation means for libraries, shows that libraries use the word “innovation” a lot—in positions descriptions, awards, strategic plans and conference planning. He notes that innovation relates not only to emerging technologies, but user-focused projects and their resulting impact. (Jason, 2013) Almeida further notes methodological approaches, such as the use of MOOCs for library instruction, can be seen as classic “disruptive technology” in that they disrupt established pedagogical practices. (Almeida, 2013) In this context we believe that the current push for research data management and curation can be seen as innovation. It fits the definition Rogers put forward, and it has context with emerging practices and tools in the library world, as well as adoption within the academic research community. Librarians have not only been early adopters of research data management as an innovation of service, advocating for its application, but given the immersion of data services in academic libraries, it’s reasonable to think that academic libraries have been early majority adopters (see figure 1).

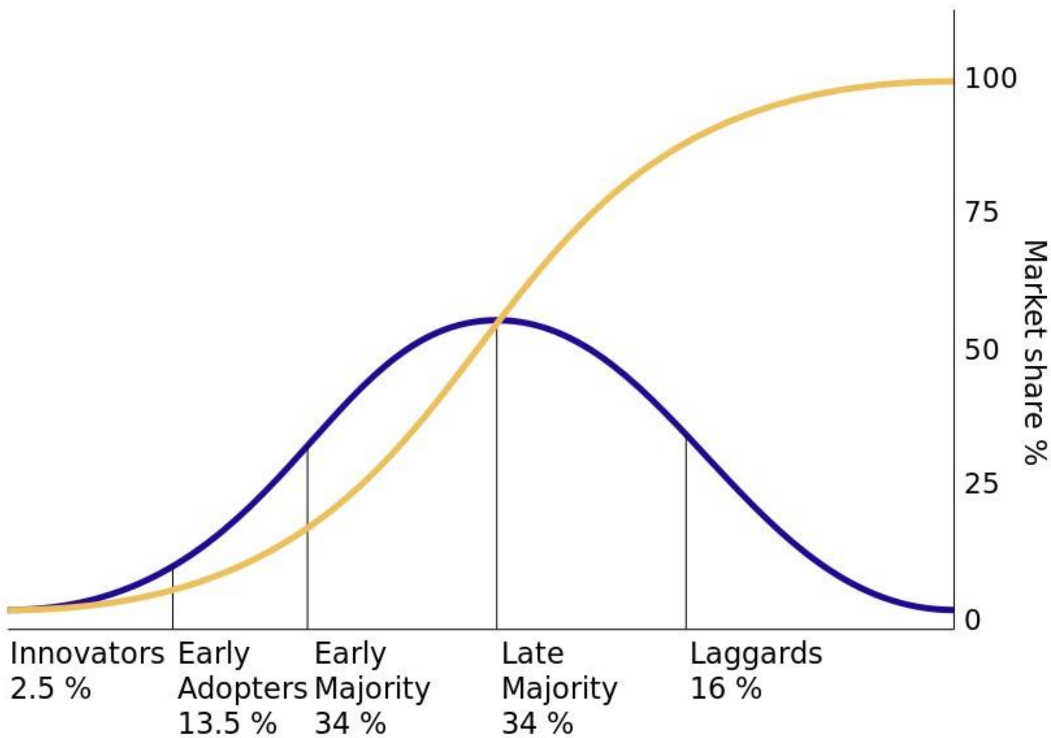


Figure 1 - The diffusion of innovations according to Rogers
Permission by copyright holder https://commons.wikimedia.org/wiki/File:Diffusion_of_ideas.svg

Although it is apparent that most researchers have not adopted data management as an innovation, looking at data management from this perspective allows us to reconsider at why they haven’t easily adopted more practices and technologies into their research. Rogers identified five stages to adoption: being exposed to innovation (knowledge), becoming interested in it (persuasion), accepting the concept (decision), beginning to use it (implementation), and making it part of one’s work (confirmation). Thus we set out to explore the idea of diffusion of research data management and curation by talking with early adopters in the library field for their understanding of whether/how researchers have been exposed to innovation and are becoming interested in it.

Rodgers also developed several tools to help understand innovation. We wanted to explore whether one of these tools, Business Canvas Model, could help us identify where/why researchers might not be accepting the concept of data management and curation as beneficial to their research. This tool can be used by organizations to collect information to work through essentials aspect of a business, or similar enterprise, to identify areas to focus on to create value (and profit). The model lays out nine building blocks in which information and data is collected, and then linkages are made to determine where to focus. These nine areas are: key partners, key activities, key resources, value propositions, consumer relationships, consumer segments, channels, cost structures and revenue streams. The ultimate goal of the Business Canvas Model is to identify how to deliver value while optimizing or reducing risk. According to Osterwalder, a business model can help describe the rationale behind how organizational structures, processes and systems can be organized into a blueprint. (Osterwalder & Pigneur 2010) The Nine building blocks help see the bigger picture, and the Business Canvas Model helps to show how pieces fit together.

Workshops on Connecting Researchers to Repositories

In 2014 we were awarded a planning grant from the Institute of Museum and Library Services to explore how librarians could further bridge the gap between data management in the "active" stages of the data lifecycle to the data curation stages of discovery, access and preservation. The results of our investigation would be used to inform the next iteration of the Data Curation Profile Toolkit. The work would be accomplished through an in depth environmental scan and literature review, and through holding two workshops to bring experts in the field together to clarify further and respond to the challenges of bridging the data deposit gap.

Prior to holding the first workshop, the coordinators met with information professionals at the University of Tennessee, Dryad and University of North Carolina, the University of Virginia, and the Digital Curation Center and University of Edinburgh to engage experts in the library community in their own settings on issues and needs in bridging the gap between the active stages of the data management lifecycle and those of the curation lifecycle. Topics included looking at the value of data deposit at both an institutional and individual level, trying to realize the potential of data as a product and the possibility of identifying intervention points to maximize data deposit and dissemination. Outcomes led to a plan for a workshop to address the library community on issues and needs in bridging the gap between the active stages of the data management lifecycle to those of the curation lifecycle.

For our first workshop, we brought in together experts from disciplinary repositories, iSchools and libraries to come together and discuss the current state of connections between research practices and data repositories. In our invitation to the workshop we asked attendees to consider the following questions: What advances been made in identifying researcher needs for their data and where does work still need to be done? Are data service providers properly in tune with the needs of researchers and if not what resources, education and support could be provided to better communicate between data producers and curators? Are there models or approaches that should be considered to increase the flow of data from active use into stewardship in ways that would reduce the high level of investment that is often required of both data producers and curators? We also compiled our literature review into a bibliography of articles and other materials that addressed problems, issues or barriers relating to connecting researchers to repositories for attendees to review and consider prior to the workshop. This bibliography included as an appendix to this white paper, Connecting Researchers to Repositories.

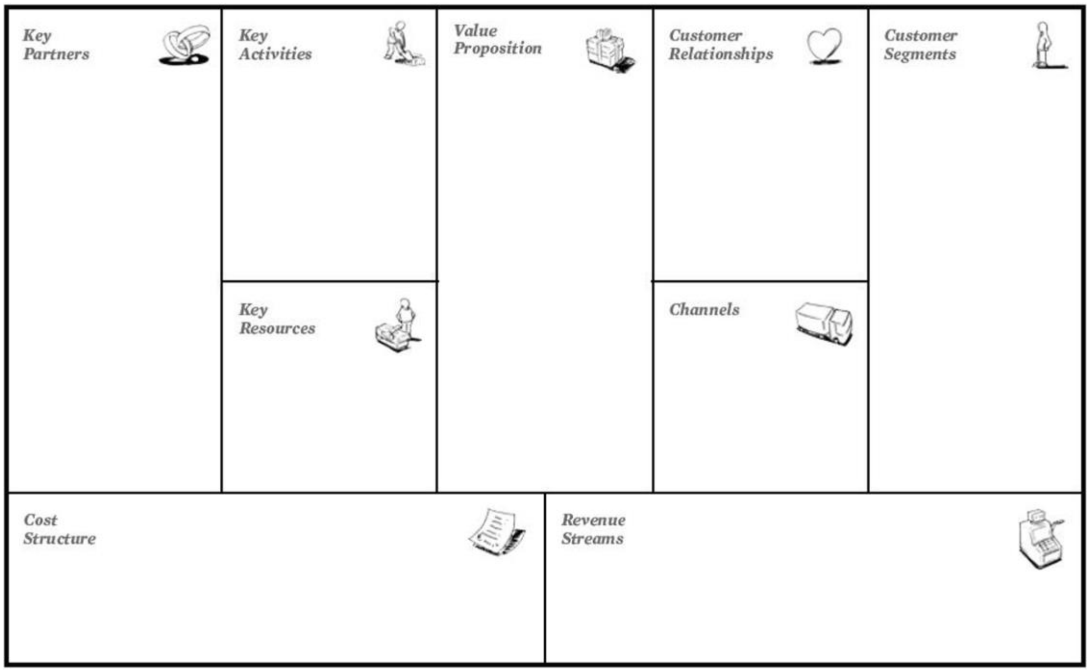
Workshop #1 - The Business Canvas Model

The first workshop was held at Purdue University on June 15-16th 2015. [See Appendix 1] Our goal for this workshop was to articulate the existing barriers surrounding the transfer of data from its state of active development and use by the researchers who produced or acquired it to a curated state where it would be disseminated, stewarded and/or preserved by a 3rd party (librarians or others providing data curation services and resources). As a part of being in this curated space, the data would be made discoverable and accessible to people outside of the environment in which the data were originally generated or acquired to view or make use of in some fashion. In articulating this goal we identified three basic actors in this process: data producers, data curators and data consumers.

First the group examined where librarians fit into the data landscape by examining well known research (JISC) and data (DataONE) lifecycles. A “sticky notes” exercise was used to identify the top items librarians felt (1) inhibited data in the lifecycle and (2) facilitated data in the lifecycle. After sharing ideas from this activity, participants worked to synthesize thematic areas that emerged as inhibitors or facilitators (i.e., presence of or lack of) needs for research data deposit, use and preservation. These included how researchers trust letting go of control of data, the time investment involved for all parties involved in deposit, the work and knowledge it takes to create metadata and documentation, the ability to apply standards and best practices, and how to identify and implement education and training related to these needs.

Participants then broke into groups to discuss stakeholders in the data lifecycle process and examined issues from producer, consumer and repository/data lifecycle perspectives. Stakeholders we identified included researchers (faculty, students, others), publishers, institutions, funders, librarians, and the public. A further step was taken to identify possible actions that could be taken to develop recommendations or solutions to convince researchers to answer a “call to arms” to deposit data by the stakeholders. [See Appendix 2]

Participants then worked through the Business Canvas Model tool to articulate value propositions for specific approaches.



The Business Canvas Model
Source: https://steveblank.com/2014/10/24/17577/

While libraries are often agile and tenacious in responding to needs, they also need planning to anticipate and meet future needs. We used the Business Model Canvas as a tool to walk through strategic planning. Who are our "customers?" What are our value propositions? Or our key partners and resources that help achieve those propositional goals? Key Partners/Activities/Resources, Value Propositions, Customer Relationships, Customer Segments, Channels, etc. were identified from the perspectives noted above (producer, consumer, repository) and filled in on the canvas.

At the end of the second day of the workshop participants were asked to consider what outcome or product of this discussion would be of most interest or use to their respective communities of data librarians and data repository people. The workshop resulted in identifying possible avenues that would take further study and assessment to determine how or whether they would work. For instance, could we track best practices through access to research data (i.e., what would usage data of datasets tell us)? Or, could we profile use cases or case studies that demonstrate coordination of services that solve problems about which researchers have dataset organization questions?

The workshop also helped us to identify many elements, aspects and issues from different perspectives. Many of these suggestions are useful in and of themselves in articulating areas of need for the data curation community:

- Can we promote trust in the data, not just the repository?
- Could libraries sponsor/subsidize training, adoption and use of e-lab notebooks that link to institutional data repositories

- And promote policies where research data belongs to the university, has full ownership and requires deposit locally in addition to disciplinary or national repositories

Overall, we found the Business Model Canvas to be useful a useful tool in the particular, but not in the aggregate. In other words, it seemed possible to brainstorm what the issues were from each perspective, and even suggest actions to take to get to the solution of depositing data, but it was difficult to map across the canvas when trying to integrate the three perspectives. We were able to understand and articulate activities, resources, relationships and value propositions, for each of the stakeholder groups. But we had trouble better understanding the bigger picture and seeing patterns and connections. Instructions https://canvanizer.com/how-to-use/business-model-canvas-tutorial for using the tool make it look somewhat easy and straightforward to fill in the elements of the canvas. But it did not provide insight into mapping the elements together to make sense of them, or in actually helping to build a business case. If a business case simply provides the reasons to form and carry out a plan, it seemed like this approach was putting the cart before the horse. As newcomers to the business model tool, that we neither had experience in using the model for all it’s intricacies, nor the insight to understand that it could not be used to develop immediate solutions by plugging answers into the template. We realized we could use more helping in understanding and using this tool.

How this led to looking more closely at “lean startup”

Clearly, there are some cautions in using the Business Canvas Model. One perhaps obvious one is that the model seeks to create a better business model and improve profit, which, for better or worse, is not a priority for academic libraries. It is designed for use for a specific product or service, not necessarily improving a complex service model. (Fielt, 2011) Its visual nature, blocks on a canvas are often filled in with Post-It ® notes that have been brainstormed by a group, can mislead users into thinking the Business Canvas is a simple, straight-forward product to use. In fact, identifying partners, resources and customers may seem to be easy, but understanding the underlying value propositions may require extensive market analysis prior to completing the model.

In 2005, Steven Blank published, “The Four Steps to the Epiphany,” which argued against a product-based approach, which is how the Business Canvas model is sometimes used (i.e., start with a product or service and try to justify it using the model). (Blank, 2005) Blank argued for looking more closely at customer development, specifically customer discovery, validation, creation, and building before developing prototypes. The detailed nature of Blank’s Customer Discovery model is evident in its 18 steps as shown in Figure 2 below.

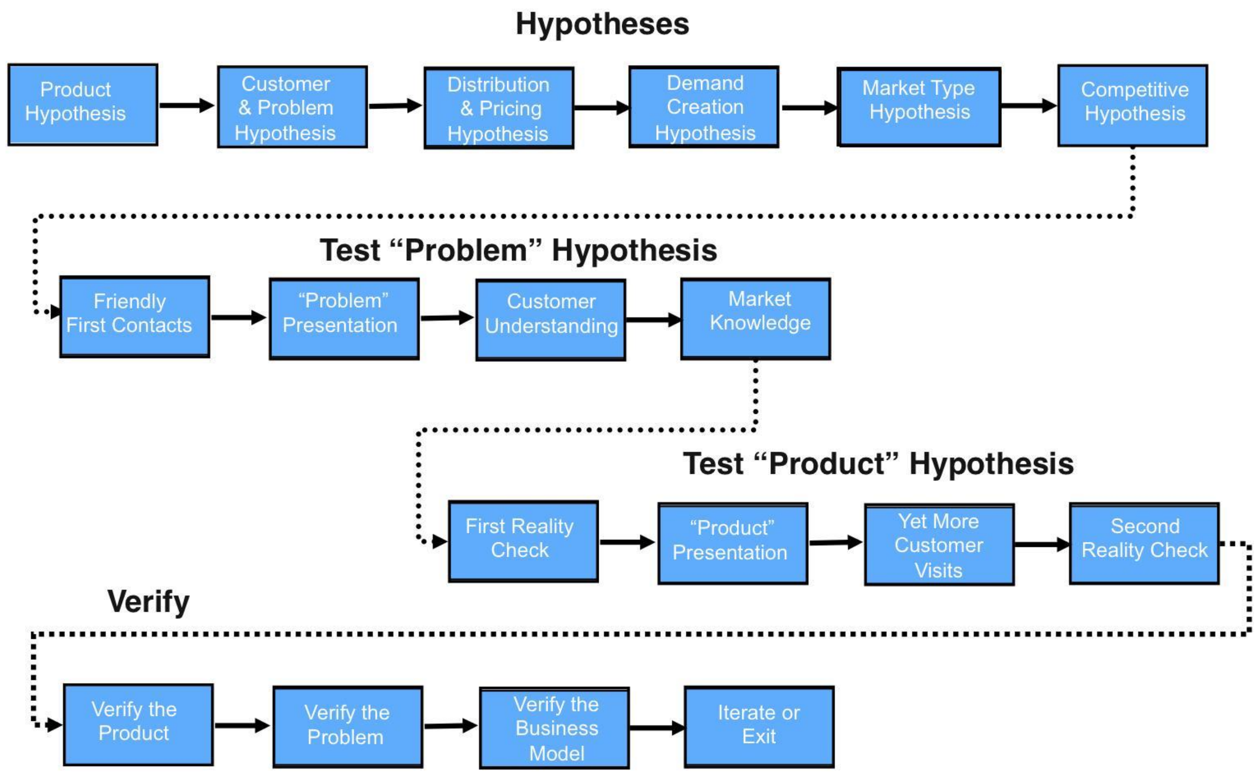


Figure 2 - Customer Discovery

Source: https://steveblank.files.wordpress.com/2010/02/customer-discovery-for-the-enterprise.jpg

Blank called this approach "a Lean Launchpad" to assessing startup ideas. He later contributed to an NSF program, Innovation Corps Program (I-Corps), a program to better facilitate scientific discovery in technology development. (NSF, 2011) Blank's contribution has included helping develop teaching objectives for the I-Corps program. (Blank, 2012a) Blank later turned his curriculum into a series of courses at the Udacity learning site, called Lean Launchpad. (Blank, 2012b)

Purdue University participates in the I-Corps program in which teaching faculty and graduate students learn the Lean Launchpad approach "to identify valuable product opportunities that can emerge from academic research." (Purdue, 2015) Dr. Matthew Lynall teaches the curriculum, and has extensive knowledge of research start-ups, the Business Canvas model, and Lean Launchpad. In discussions with him, he revealed that many times people jump right into the Business Canvas model. Instead, he recommended starting with a more preliminary step of working with the Value Proposition Canvas.

In reviewing our experience with the Business Canvas Model, we realized that we had tried to accomplish too much too quickly. We needed to pull back and to focus in on the value propositions for the various stakeholders in sharing data. Namely, articulating what were the motivations for each stakeholder type in sharing data or in supporting this practice. Matching value propositions with customer segments is the key of the Business Model Canvas. Analysis requires digging deep to understand customer “pains and gains,” rather than simply guessing or assuming as to what they might be. Likewise, looking at a product or service requires understanding its potential use and benefits from the user’s perspective. Finding a match between the two that fills a need not currently available helps identify a Value Proposition (defined as a key service that customers want) through a value mapping exercise to identify a market fit. (Osterwalder, Pigneur, Bernarda & Smith, 2014)

The Value Proposition Canvas was developed as a tool to help entrepreneurs identify products and services that their customers would want. It is a component of the Business Model Canvas tool that focuses on the “Customer Segments” and “Value Propositions” elements as depicted in Figure 3.

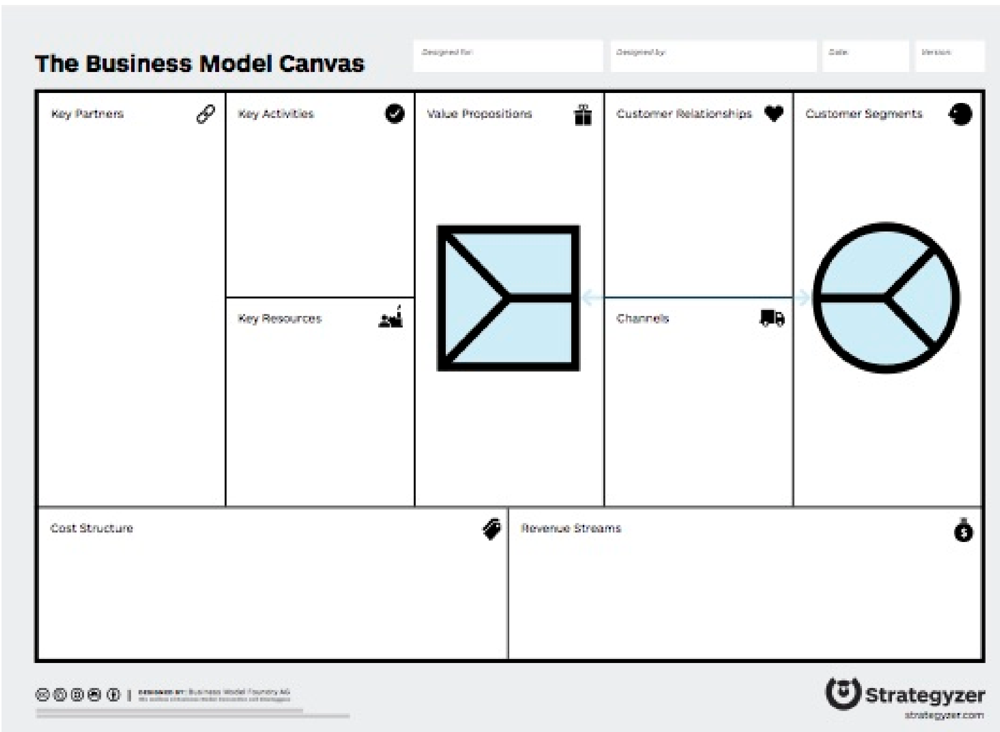


Figure 3 - The Business Canvas Model

Source: https://steveblank.com/2014/10/24/17577/

The customer segments are the people whom you intend to create value for and the value propositions are the elements of your product that you believe will attract these people. The Value Proposition Canvas is a means to explore each in more depth to create a better understanding of what your potential customers want and how what you have to offer fits customer needs.

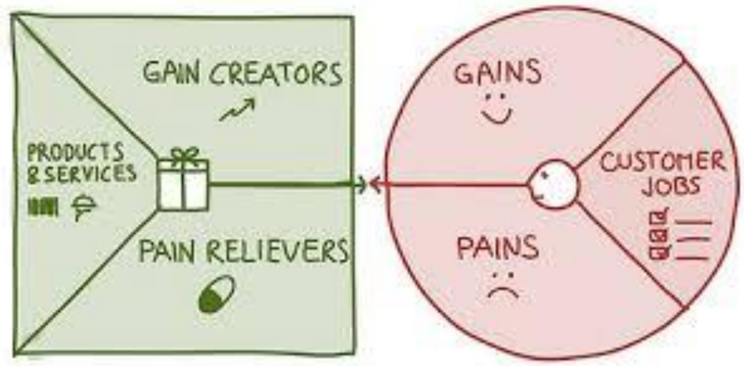


Figure 4 - The Value Proposition Canvas
Source: http://businessmodelalchemist.com/blog/2012/08/achieve-product-market-fit-with-our-brand-new-value-proposition-designer.html

In Figure 4 above the circle represents the customer segment, and the square represents the value proposition.

The customer segment is comprised of three components: customer jobs, gains and pains.

- Customer jobs are the things that your customers are the things they are trying to accomplish over the course of doing their jobs or living their lives. For researchers, these could be things like securing the funding need to carry out their research, or publishing their findings to disseminate their work.
- Pains are the challenges or barriers encountered by your customers as they carry out their jobs. These are things that they would rather avoid or not have to do themselves, before, during or after carrying out a job. For researchers, these could be things like filling out paperwork as a part of applying for grants or paying author fees to have their article published.
- Gains are the positive impacts or benefits that make customers successful, make their jobs easier to do, reduce expenses or other barriers, or otherwise produce a positive result or emotion. For researchers this could include things like having ready-made templates to plug into grant applications or having access to an author's' fund to cover publication fees.

Learning more about customer jobs is an essential part of the value proposition canvas. Customer jobs are broadly defined in this model and could include tasks, issues, or needs they are trying to satisfy. In addition, customer jobs could be comprised of functional, social or emotional elements.

In learning more about the jobs performed by customers it is important that care be taken to identify which tasks are critical to the work and which are more trivial in nature. The same holds true for understanding their pains and gains. Some the pains and gains of the customer are more important, relevant or impactful on their jobs than others.

The value proposition is also comprised of three components: products & services, gain creators, pain relievers. All three of the components of the customer segment are things that can be observed.

- Products and Services are all of the things that your value proposition is built around. These are the things that you would offer the customer to get their jobs done in ways that address their pain and/or maximize their gain.
- Pain relievers are descriptions of how the products and services would provide some measure of relief for the customer's pains. How exactly the products and services would address and alleviate pain before, during and after the customer completes his/her jobs. Pain relievers should explicitly reference which of the customer's pains they are addressing, mitigating or removing and how.
- Gain creators are descriptions of how the products and services would create customer gains through a positive outcome or result. Here too, gain creators should explicitly list which gains are being addressed.

If you look closely at the graphic, you will notice an arrow from the circle and an arrow from the square coming together. This is meant to represent the "fit" of the product or service to the situation and needs of the customer. A problem-solution fit is achieved when the pain relievers and gain creators of your product or service align exactly with the pain and gains identified by the customer. When this match is validated by the market (i.e. people buy and use your product or service) it is called a "product market fit".

This is certainly not all that an entrepreneur needs to do in order to be successful. A lot depends on having a great business model, access to resources and on extraneous factors to say the least. However, using the value proposition model can help focus attention on the customers and in designing products and services that meet their needs.

Workshop #2 - The Value Proposition Canvas

Professor Lynall agreed to provide an overview of the Value Proposition Canvas and instruction in its use to librarians at our second workshop, held at Purdue University June 6-7th, 2016. The goal of the second workshop was to understand how the Value Proposition Canvas might be applied to better understand and respond to researcher needs in sharing and curating their data. In the first workshop we looked at the challenge of moving data from its active state into a curated environment for stewardship, dissemination and preservation from the perspective of the data—what must be done and how does it get done. In the second workshop we used a "deep dive" approach to look more closely at researchers as customers, and their needs for very specific services, as opposed to larger all encompassing data services. (Note: We did not interview researchers as part of this process to gather more on-the-ground data.)

In a review of previous work we discussed the idea of research data management as an innovation and its progression along an innovation diffusion curve. There are two contexts to understand when doing so. For librarians, especially in many ARL libraries which have had some kind of research data services for several years, we are at least in the middle of the curve where a majority of early (and some late) adopters accept the need, and provide resources, for data services. However, for the most part, researchers are still in the very early adopters stage of employing data management practices. And in looking at the related five stages of adoption, you can't implement or adopt something before you've made the decision to do so. One must be persuaded to make a decision, and to be persuaded requires a sufficient level of understanding and knowledge about the decision. Librarians can't "jump the stack" and expect researchers to implement data solutions if they haven't gone through the other steps first.

Professor Lynall argued that when looking at improvement for data services, one must first go back and look closely at the pains of researchers in dealing with data management and curation. In particular, rather than looking at it from the data's point of view (i.e., what should happen to the data), to focus on the researcher's perspective. What are their specific pains as regards to their research, and what would alleviate them? What are they striving for in their research, professional career or life, and what would help them achieve their goals? In asking the question about research and not just data, the pains revealed may be related to workflow, processes, or other factors that may seem somewhat removed from the data itself. So we looked broadly at researchers' work first before ever looking at data collection/generation, management, dissemination or curation.

In introducing us to the Value Proposition Canvas, Lynall demonstrated how it was much simpler to look at the customer to identify pains, gains, and specific jobs to alleviate them. Then we could start looking at specific products or services that relieve pain or become a gain creator for researchers. Eventually this would lead us back to the Business Model Canvas where we could match customer segment to value proposition to create a market fit. This process would be completed over a series of small, incremental steps rather than jumping in headfirst and trying to complete everything at once.

The first exercise was to look at "what they are really trying to do," and was meant to understand and identify potential customer segments out of the larger generic group of researchers. The results may seem obvious, but serve as reminder that we first need to focus on customers, not on pushing services onto them. Based on previous work, experience and literature reviews, we started with generating several familiar "researchers want to…" goal statements:

- Produce results that impact my field
- Increase funding to further research
- Attract collaborators
- Bring the best graduate students into labs
- Get credit that counts for promotion and tenure through publication and citation
- Get awards and other recognition from peers or others in their field
- Raise the reputation or profile of lab and institution
- Secure legacy and reputation
- Contribute to society and the "greater good"

This discussion helped us steer away from putting data and services in the forefront without context, and led us to a step to analyze perspectives of customers. The second exercise was then to determine archetypes, or personas, of researchers as our customers to help us reveal likely paths or connections to better understand where our focus should be.

Our brainstorming on possible archetypes resulted in four broad categories:

- Disciplines: researchers have different methods and deal with research much differently across broad disciplines (STEM vs Humanities), and for instance, some collect data while others generate it
- Roles: researchers roles may involve those directly involved in research, such as faculty, postdoc, graduate student, lab manager, etc., as well as those somewhat peripheral to the research, such as administrator, vendor/supplier, or librarian
- Type of data: which can range from: experimental vs observation vs simulation; in small to large quantities; in sensitive or restricted access areas; with static, dynamic, or streaming data; with images, videos, or physical samples
- Responsibility: additionally, researchers may have a variety of responsibilities in the discipline, role, and type of research: PI or co-PIs; the ones who provide (collect or generate) data, clean or process, or analyze it; someone who determines ownership or authority, or ensures compliance or privacy/security

The next step for us was to determine jobs to be done (JTBD). As noted above, specific pains must be identified and an analysis must be done to understand what would relieve them or what gains could be identified that would help researchers achieve their goals. To accomplish this we broke into pairs to discuss the types of pains of researchers that we had identified. As with all of our work from this workshop, Lynall noted this exercise could only lead to hypothesis building, not solving the problems. Problem solving could not be done in the abstract, on a whiteboard or with sticky notes. He was adamant that the only way to test a hypothesis would be to interview many, many researchers to hear directly from them if our solutions (services, tools or resources we created) actually address their problems.

For customers pains we had to understand undesired costs and situations that caused problems or negative emotions for the customer. For instance, we might ask about things taking too long or costing too much money, and things that annoy, frustrate or give a headache. For customer gains we had to understand benefits that would be expected, such as use satisfaction, cost savings, relief, and social gains. To better understand potential customer gains from another lengthy list we might ask questions about what they are looking for, what would make their life easier, what might be the result of an ideal solution? Typically these are not the kinds of questions we ask directly in DMP consultations or Data Curation Profile interviews and so we are likely overlooking critical pieces of information in our drive to provide services and resources.

Only after you have conducted many interviews, up to a hundred, are you able to start defining what could be pain relievers and gain creators for customers with some degree of confidence. And then you can begin to identify a product or service that helps them achieve something functionally or socially or emotionally that makes life better. At this point we were confronted with what Lynall called “eating the elephant in the room.” This expression is a mixed metaphor meaning one has to deal with the big thing that we gathered to understand, research data services, but how could anyone possibly eat/solve it all at once? Lynall explained that looking at small discrete steps was really the only way to move forward in creating a new product or service with any degree of confidence, even if meant limiting yourself to small discrete successes. To illustrate his point, he explained how in the early days of dot com start-ups, someone would come up with a big idea, and try to get a lot of venture capital to build it or do it. Many of these start ups failed from trying to do too much too soon without a real sense of the market or need they were trying to serve. Lynall used the company pets.com as an example. It started with a wildly successful marketing mascot (a sock puppet dog) but failed because it "was weak on fundamentals and lost money on most of its sales." (It was eventually bought by PetSmart.) https://en.wikipedia.org/wiki/Pets.com

In a final post-it note exercise we drilled down to identify a small niche in which we could propose a hypothesis. We created a hypothetical archetype of mechanical engineers with grants who have trouble meeting funder requirements in managing and sharing data as a possible type of customer. We then set about defining what our interactions and processes would be to test out an hypothesis. The interaction would be to engage in a conversation to see if leads to a discussion about problems with sharing or using data. The process would be to approach many customers, iterate and fine-tune the questions, but avoid leading the conversation to data intentionally.

As a next step we generated a series of possible questions to ask and developed a very loose script to use in talking to researchers:

- Intro: Hi, I'm interested in learning about your pain points in your research… (you want to find out what they want to talk about)
- We're looking at ways institution can help with research [I'm here to help…]
- What is like doing research here? What are the big challenges you face? What are the requirements of the job that you have to fulfill?
- What does a successful day look like for you?
- Can/how can the institution help you?
- Anything else I can ask? Is there someone else to talk to? Can I come back?
- If topics of external funding, students, publication, etc. come up that can't be addressed easily or right away, you might ask whether you can do a follow up…

Lynall reiterated that talking to people was not only key, but also talking to as many of them as possible and quickly was important as well. Using Lean Launchpad techniques, typically one person talked to a hundred people or so in a week to ten days. Throughout the process it would be likely that some questions might change or get deleted based upon what was learned as these interactions and as the potential customer base progressed (e.g., by gaining insight one might learn how to get to the heart of the matter). The goal is to reveal two or three big things you've learned or insights into problems they have, and then to report them out to the team, preferably using an online tool or space where people can review and add comments. The information learned from the interviews and discussions about them would then be used to develop the product or service. Once a prototype was created the interviewees could be revisited to react to it. Questions about the extent to which the product or service met expectations and addressed the pains and gains identified would be asked. The answers would inform another iteration of the product or service and the process would continue until the team was confident that their result matched customer needs.

Finally Lynall wanted to impress upon us a little pessimism—that not only might the process be slow going, it might result in only a small thing that turned out to be a pain reliever or gain creator (or conversely, no pain may lead to stopping a service). For instance, what if researchers only wanted examples of other DMPs and that was it? Or to have their students simply learn better file naming and/or directory structure? Or what if they did not see data sharing as their problem at all? He pointed out that the Value Proposition Canvas and other Lean Launchpad techniques were about making products and services that were sure to work based on evidence rather than assumptions, no matter how small. But that in doing so, small successes might lead to additional gains, such as examples of DMPs leading to guidelines for description or standardization of metadata needed to publish data, perhaps leading to deposit of data in a repository.

Discussion

The Data Curation Profile Toolkit was designed to understand the story of the data in a project and to provide information professionals with enough information to respond to the specific needs expressed by the researcher(s) being interviewed. When we launched the DCPT we had visions of librarians generating series of Profiles that could be compared, contrasted and ultimately used to develop a better understanding of researcher needs on a larger scale. Having multiple Profiles on Mechanical Engineers, for example, would allow libraries and other agencies who provide support to researchers to identify common practices and specific needs related to data management, organization, description, sharing, and preservation with the intent of developing larger scale responses. However, in talking with librarians it became obvious that the amount of time and effort required to complete a Profile was prohibitive for many and so we could not expect a sufficient number of Profiles to do the large-scale analysis that we had initially envisioned. Instead, our study on the usability of the DCPT revealed librarians wanted “a lighter and more adjustable version with less time requirements.” (Zhang, Zilinski, Brandt & Carlson, 2015) This study used a survey to determine what influenced the use of the DCPT, and identified factors of perceived usability, specifically: the amount of time required using the tool and its format and structure were seen as deterrents to use.

Given our findings from the DCPT usability study, the idea of asking broad questions (“what does a successful day in the lab look like?”) that go further than the scope of the questions asked in the DCPT (“could you tell me about the data you create or use in your project?”) may seem counter-intuitive. However, the practice of doing a lot of information gathering and analysis before coming up with a hypothesis to test, as the Lean Launch Pad does, make good sense. The literature has many use cases and case studies in which librarians developed approaches and tried to market them and implement them as services for their constituents, without defining the level of success that was desired or expected. The challenge of course is in finding the time and the capacity to be able to gather the information that is required to truly understand the needs of our users and the environments in which they work and live, and to analyze and derive meaning from this information in ways that can be applied through our products and services. Although many librarians conduct research it is not the focus of most positions, which makes it hard for librarians to actively engage in the kind of activities needed to make use of tools like the DCPT or the Value Proposition Canvas. Herein lies one of the fundamental challenges for libraries. We desire an easy way to develop an in-depth understanding of researcher's environment and needs for their data that enable us to provide services of value to them, but the complexities of data and research necessitate a significant investment of time and resources to gain a sufficient enough understanding to respond.

Through this endeavor we've come to see that maybe we are not asking the right questions. For instance, perhaps the question to ask is not, "why won't researchers deposit?" For one thing, such questions have been asked in numerous surveys over the past decade. And the answers are generally fairly standardized around: time, knowledge and skills, credit and resources. As part of this grant work we researched the literature and examined expert experiences to look at possible solutions and they too seem to turn up common themes: provide training, tools and help.

It seems that we might not even ask, "what would make deposit easier?" Lynall helped us see in our second workshop that even asking that more innocuous question probably isn't the right approach. First we must show researchers that we are interested in understanding the "pain points" in carrying out their research, and ask what those might be (i.e., not pre-suppose they are data management related). This is the first job of a librarian or liaison: to understand the information needs of researchers. Years ago this might have been done through interactions at a reference desk, but now requires outreach, or rather, reaching out, to faculty, and literally meeting them in their spaces, where they work, teach, and drink coffee. (Delaney & Bates, 2015)

Asking a more general question—"how's your research going?"—is similar to the opening of the DCPT process, although as its name implies, even the first question pushes the conversation toward discussing data—"*Could you please provide me with a brief overview of the research project associated with the data that we will be discussing in this interview?*" (Carlson, 2010) But broader variations of this might be, "We're looking at ways institution/library can help with research…" or "The library would like to know what are the big challenges you face in your research?"

It is possible that this would lead down a rabbit hole of responses totally unrelated to data or to frustrations beyond our or their control ("If we only had localized IT support this would be so much easier!"). But still, it is information about the research done in the institution that provides information on needed services. It can be argued that such a "bottom-up" approach would not scale—that there aren't enough librarians or liaisons around to engage in such discussions. We've learned that information or data management problems can vary not only vary by discipline and sub-discipline, but also by lab and project as well. (Brandt et al., 2007) Therefore, the question shouldn't be "how do we reach all researchers?" but rather "How can we reach some researchers and help them?" And hopefully responses or approaches that solve similar problems can be turned into guides or resources or tools.

Perhaps the bigger takeaway from these workshops is the idea of scaling back and looking only for little things that are sure to be successes, at least at first. Not asking “How can we do everything for researchers?” but “What would be one thing that would make life easier for them?” And those things might not even include data management, sharing or preservation, or at least not on the surface, as far as the researcher is concerned. In acknowledging the myriad of complexities that surround the requirements being made by funding agencies, publishers and others are making, the temptation is to try and address these problems at scale. However, our experiences with the Data Curation Profiles Toolkit and exposure to the Value Proposition Canvas demonstrate the value of thinking iteratively in the short term as a way of eventually realizing longer-term gains.

Conclusion

We set out to explore how we could build the next iteration of the Data Curation Profiles Toolkit to try to address the low rate of deposit into data repositories. We saw the DCPT serving as a foundation for informing librarians and other information professionals how to better prepare their data over the course of the research data lifecycle for eventual deposit and for informing repositories how they could structure their submission process to best connect with researchers. We still believe that the DCPT is an excellent means of gathering information about the practices and needs of individual researchers, however we have learned from these workshops that it is not a suitable instrument for sparking broad based culture change. The richness and depth of the DCP comes at the price of a significant investment of time and effort. As we learned from Professor Lynall in our second workshop, agility and the ability to gather quick responses from a lot of the potential pool of customers is a key facet in developing innovative products that are more likely to succeed. To understand researcher needs we might ask questions about what would make their life easier, what might be a useful solution to help them. Typically these are not the kinds of questions we ask directly in DMP consultations or Data Curation Profile interviews and perhaps we lose out on useful information that would provide insight into services.

Approaching the challenge of data sharing from an entrepreneurial standpoint can help jump start efforts to increase data deposits. The cultures of practice surrounding data management, sharing and preservation in many research fields are still evolving. The direction and speed in which they take shape will be determined less by abstract ideals and more by how data sharing can aid researchers in accomplishing what they set out to do. Though we have learned a great deal from the surveys, interviews and other information gathering efforts that have been done by librarians and others in the past decade or so, we have not yet been able to develop practical tools that address the on the ground issues that facilitate or hinder deposit into data repositories.

We know that making research data widely available can benefit the public, the research community and the individual researcher him or herself. The challenge is in finding ways in which data sharing and data deposits will become a normative part of the research process in all fields rather than an exception. Making progress will likely require use to move away from relying solely on surveys and other cumbersome information gathering approaches towards more lightweight and rapid approaches that can be used to fashion prototypes of tools and resources that can be brought out and tested. Recasting our thinking and approaches on instantiating data sharing by grounding them on local scale needs offers another promising path forward.

Lastly, it is not clear whether the DCPT can or should be adapted to fit this approach. The goal of providing a profile of data management and use is different than identifying better ways to encourage data deposit. As shown by the number of downloads of Data Curation Profiles (9434 downloads since October 2012), there seems to be use for detailed profiles. But while a DCPT "lite" might cut down on problems of format and time needed to gather information, it wouldn't likely provide a quick and easy solution for connecting researchers to repositories. Further research on what such a tool or method would look like is to be pursued in future research.

Acknowledgements

The authors would like to thank their institutions for providing time and resources to complete this work, as well as the Institute of Museum and Library Services for the Planning Grant that made it possible. They would also like to thank Abigail Goben, Information Services Librarian, University of Illinois-Chicago Health Sciences Library, and Lisa Johnston, Research Data Management/Curation Lead and Co-Director of the University Digital Conservancy, University of Minnesota Libraries, for providing input to and feedback on this white paper.

Appendix 1 Workshop 1 participants and goals

Workshop 1 was held June 15-16, 2015. As noted, this Workshop brought together data management and curation experts with a wide range of backgrounds and experiences in university settings. Participants included:

Project partners:
Scott Brandt-PI (Purdue University)
Jake Carlson-PI (University of Michigan)
Suzie Allard-co-PI (University of Tennessee, Knoxville)
Sherry Lake (University of Virginia)
Angus Whyte (DCC, University of Edinburgh)
Sarah Jones (HATII, University of Glasgow)
Todd Vision (Dryad) [unable to attend]

Invited experts:
Elizabeth Hull (Dryad)
Lisa Johnston (University of Minnesota)
Wendy Kozlowski (Cornell University)
Joan Starr (California Digital Library)
Thomas Padilla (Michigan State University)
Limor Peer (Yale University)
Lizzy Rolando (Georgia Institute of Technology)
Karen Baker (University of Illinois, Urbana Champaign)
Abigail Goben (University of Illinois, Chicago)

The initial goals for the workshop were to:

1. Review the literature on issues related to research data to provide context and help define the problem
2. Discuss connecting researchers and repositories from the data's point of view to identify problems that inhibit or prevent the transfer of data
3. Describe possible ways to address problems identified, determine feasibility of approaches
4. Apply business model paradigm to articulate value propositions for specific approaches identified through discussion with experts (personas/scenarios, questions database, best practice recommendations)
5. Provide suggestions to bring this all together to create outline and framework for White Paper

Appendix 2 - Activities to support the value propositions (from workshop 1)

<u>Value propositions</u>
Providing the data in a meaningful way
Save time
Do new science - ask new questions
Enabling them to produce products of commercial value
Discover evidence / compliance track best practices through access to the data - (usage data)
Learning and education
More informed policy making / discovery
Enabling evidence based discussion / research / actions,

<u>Activities to support value propositions</u>
Make data available will full documentation / context
Promote clarity and understandability of the documentation - transparency - generic / non-discipline specific - understandable
Delineate relationships between data sets to enable interoperability
Consider design of the product - imagining the re-use - speculate the utility of the data
Consider the tools needed to use the data in meaningful ways (i.e., design tools for use)
Define service workflows to derive a particular result/outcome and accommodate free and open exploration (product design) - modules/tools of processing - curriculum
Promote discovery tools (UI for humans and API for machines, OAI-PMH for indexers) - that connect to accessibility standards and are tested
Visualization and Analysis Tools - embedded or linked
Foster interoperable formatting (open and migration)
Make data machine readable (ready for automated consumption - building apps on top of)
Ensure IP - Licensing and rights - what are consumers allowed to do with the data to promote good data governance
Provide a means of maintaining the data (through a repository or other means - ours or another)
Develop a brand that generates trust
Define customer support system and how it operates
Define service levels - from the consumer's vantage point / standard of practice
Form User communities

Appendix 3 Workshop 2 participants and goals

Workshop 2 was held at Purdue University June 6-7th, 2016 to further investigate adoption of research data management as an innovation from a startup perspective. Participants included:

Scott Brandt-PI (Purdue University)
Jake Carlson-PI (University of Michigan)
Suzie Allard-co-PI (University of Tennessee, Knoxville)
Sherry Lake (University of Virginia)
Lisa Johnston (University of Minnesota)
Wendy Kozlowski (Cornell University)
Abigail Goben (University of Illinois, Chicago)

Goals of the Workshop were:

1. Review our “customer focus” on researchers who have requirements to share data (funders, publishers, or peers) but do not have a natural workflow for depositing data
2. Review the problems associated with connecting researchers to repositories
3. What are the problems, and what would alleviate them?
4. What are possible options to remedy problem? How to find out what researchers want? Can more focus on Customer Segments and Value Proposition help?
5. Review structure for white paper

REFERENCES

AGU Publications Data Policy. (2017). American Geophysical Union website. Retrieved July 27, 2017: https://publications.agu.org/author-resource-center/publication-policies/data-policy/.

Akers, K. G., Sferdean, F. C., Nicholls, N. H., & Green, J. A. (2014). Building Support for Research Data Management: Biographies of Eight Research Universities. *International Journal of Digital Curation*, 9(2), 171–191. http://doi.org/10.2218/ijdc.v9i2.327.

Almeida, N. (2013). “A New Polemic: Libraries, MOOCs, and the Pedagogical Landscape.” In the Library with the Lead Pipe. http://www.inthelibrarywiththeleadpipe.org/2013/a-new-polemic-libraries-moocs-and-the-pedagogical-landscape/.

Andrews, C., Young, S., Ochs, M., Shea, A., & Morris-Knower, J. (2016). Research Practices and Support Needs of Scholars in the Field of Agriculture at Cornell University. http://hdl.handle.net/1813/45090.

American Political Science Association. Data Access & Research Transparency. Retrieved July 27, 2017: https://www.dartstatement.org.

Assante, M., Candela, L., Castelli, D., & Tani, A. (2016). Are scientific data repositories coping with research data publishing? *Data Science Journal*, 15. http://datascience.codata.org/article/10.5334/dsj-2016-006/.

Blank, S. (2005). Five Steps to Epiphany. Retrieved May 3, 2016: http://web.stanford.edu/group/e145/cgi-bin/winter/drupal/upload/handouts/Four_Steps.pdf.

Blank, S. (2012a). Innovation Corps: A Review Of A New National Science Foundation Program To Leverage Research Investments. Retrieved December 20, 2016: http://science.house.gov/sites/republicans.science.house.gov/files/documents/HHRG-112-SY21-WState-Sblank-20120716_0.pdf.

Blank, S. (2012b) The Lean LaunchPad Online. Retrieved December 20, 2016: https://steveblank.com/2012/09/06/the-lean-launchpad-online/.

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

Brandt, D. S.,Witt, M., Carlson, C. Palmerr, & C. Cragin, M. (2007) “Investigating Data Curation Profiles Across Research Domains.” (IMLS NLG-06-07-0032-07) Unpublished grant application.

Briney, K., Goben, A., & Zilinski, L. (2015). Do You Have an Institutional Data Policy? A Review of the Current Landscape of Library Data Services and Institutional Data Policies. *Journal of Librarianship and Scholarly Communication*, 3(2). http://doi.org/10.7710/2162-3309.1232.

Carlson, J. (2010) "The Data Curation Profiles Toolkit: Interviewer's Manual" *Data Curation Profiles Toolkit.* Paper 2. http://dx.doi.org/10.5703/1288284315651.

Carlson, J. (2014). The Use of Life Cycle Models in Developing and Supporting Data Services. In: Ray, Joyce M., editor. *Research Data Management: Practical Strategies for Information Professionals*. West Lafayette, Indiana: Purdue University Press. p. 63-86.

Chang, H. F., & Milligan, S. (2016). ITHAKA S&R research report services: Prospectus for the field of agriculture: Oklahoma State University Library local report. Retrieved December 20, 2016: http://hdl.handle.net/11244/47160.

Charbonneau, D. H., & Beaudoin, J. E. (2016). State of Data Guidance in Journal Policies: A Case Study in Oncology. *International Journal of Digital Curation*, 10(2), 136-156. http://dx.doi.org/10.2218/ijdc.v10i2.375.

Center for Informatics Research in Science and Scholarship. Data Curation Education Program. Retrieved July 27, 2017: http://cirss.ischool.illinois.edu/Project/project-details.php?id=19.

Collins, F. S., & Tabak, L. A. (2014). NIH plans to enhance reproducibility. *Nature*, 505(7485), 612–613. http://www.nature.com/news/policy-nih-plans-to-enhance-reproducibility-1.14586.

Corrall, S., Kennan, M. A., & Afzal, W. (2013). Bibliometrics and Research Data Management Services: Emerging Trends in Library Support for Research. *Library Trends*, 61(3), 636–674. http://doi.org/10.1353/lib.2013.0005.

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 A: Mathematical, Physical and Engineering Sciences*, 368(1926), 4023–4038. http://doi.org/10.1098/rsta.2010.0165.

Crowston, K. (2015). “Personas” to Support Development of Cyberinfrastructure for Scientific Data Sharing. *Journal of eScience Librarianship* 4(2): e1082. http://dx.doi.org/10.7191/jeslib.2015.1082.

Curty, R. G. (2015). "Beyond “Data Thrifting”: An Investigation of Factors Influencing Research Data Reuse in the Social Sciences". Dissertation: Syracuse University. http://surface.syr.edu/etd/266/.

Delaney, G., & Bates, J. (2015). Envisioning the academic library: a reflection on roles, relevancy and relationships. *New review of academic librarianship*, *21*(1), 30-51. http://dx.doi.org/10.1080/13614533.2014.911194.

Faniel, I. M., & Jacobsen, T. E. (2010). Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues’ Data. *Computer Supported Cooperative Work (CSCW)*, 19(3-4), 355–375. http://doi.org/10.1007/s10606-010-9117-8.

Fearon, D. Jr., Gunia, B., Lake, S., Pralle, B. E., Sallans, A. L. (July 2013) SPEC Kit 334: Research Data Management Services. http://publications.arl.org/Research-Data-Management-Services-SPEC-Kit-334/.

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

Fielt, Erwin (2011) Business service management: understanding business models. Smart Services CRC, Eveleigh, NSW. [Working Paper] Retrieved December 20, 2016: https://eprints.qut.edu.au/41609/.

Friesike, S. (2015). Reputation Instead of Obligation: Why We Need to Forge New Policies to Motivate Academic Data Sharing. Alexander von Humboldt Institute for Internet and Society. http://www.hiig.de/reputation-instead-of-obligation-why-we-need-to-forge-new-policies-to-motivate-academic-data-sharing.

Heidorn, P. B. (2011). The Emerging Role of Libraries in Data Curation and E-science. *Journal of Library Administration*, 51(7-8), 662–672. http://doi.org/10.1080/01930826.2011.601269.

Herold, P. (2015). Data Sharing Among Ecology, Evolution, and Natural Resources Scientists: An Analysis of Selected Publications. *Journal of Librarianship and Scholarly Communication.* 3(2), p.eP1244. http://doi.org/10.7710/2162-3309.1244.

Holdren, J.P. (2013). Increasing Access to the Results of Federally Funded Scientific Research. “Public access memorandum from the Office of Science and Technology Policy”. http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

Hudson-Vitale, C., Imker, H., Johnston, L., Carlson, J., Kozlowski, W., Olendorf, R. Stewart, C (2017) SPEC Kit 354: Data Curation. http://publications.arl.org/Data-Curation-SPEC-Kit-354/.

IDCC10. "Why Is It Taking So Long?" Panel moderated by Carly Strasser. Retrieved July 27, 2017: https://www.youtube.com/watch?v=2M6v7d2VdYo.

Jason, V. (2013). “Defining Technological Innovation.” *Library Technology Reports*, *49*(7), 10-46. https://journals.ala.org/index.php/ltr/article/view/4593.

Johnston, Lisa R. (2014). A Workflow Model for Curating Research Data in the University of Minnesota Libraries: Report from the 2013 Data Curation Pilot. University Digital of Minnesota Conservancy. http://hdl.handle.net/11299/162338.

Kaplan, F. (2015). A map for Big Data research in Digital Humanities. *Frontiers in Digital Humanities*, 2(1). https://doi.org/10.3389/fdigh.2015.00001.

Kim, Y., & Adler, M. (2015). Social scientists’ data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. International *Journal of Information Management*, 35(4), 408–418. http://doi.org/10.1016/j.ijinfomgt.2015.04.007.

Kim, Y., & Zhang, P. (2015). Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories. *Library & Information Science Research*, 37(3), 189-200. https://doi.org/10.1016/j.lisr.2015.04.006.

Klump, J. (2017). Data as Social Capital and the Gift Culture in Research. *Data Science Journal,* 16, p 14. http://doi.org/10.5334/dsj-2017-014.

Kutay, S. (2014). Advancing Digital Repository Services for Faculty Primary Research Assets: An Exploratory Study. *The Journal of Academic Librarianship*, 40(6), 642–649. https://doi.org/10.1016/j.acalib.2014.08.006.

Margolis, R., Derr, L., Dunn, M., Huerta, M., Larkin, J., Sheehan, J., ... & Green, E. D. (2014). The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: capitalizing on biomedical big data. *Journal of the American Medical Informatics Association*, 21(6), 957-958. https://dx.doi.org/10.1136/amiajnl-2014-002974.

McLure, M., Level, A. V., Cranston, C. L., Oehlerts, B., & Culbertson, M. (2014). Data Curation: A Study of Researcher Practices and Needs. *Portal: Libraries and the Academy*, 14(2), 139–164. http://doi.org/10.1353/pla.2014.0009.

McNutt, M., Lehnert, K., Hanson, B., Nosek, B. A., Ellison, A. M., & King, J. L. (2016). Liberating field science samples and data. *Science*, 351(6277), 1024-1026. http://doi.org/10.1126/science.aad7048.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., ... & Laitin, D. (2014). Promoting transparency in social science research. *Science*, 343(6166), 30. http://doi.org/10.1126/science.1245317.

National Institutes of Health. (2007). NIH Data Sharing Policy. https://grants.nih.gov/grants/policy/data_sharing/.

National Science Board. (2005). Long-lived digital data collections: Enabling research and education in the 21st century. http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf.

National Science Foundation. (2011) Program Solicitation NSF 11-560: Innovation Corps Program (I-Corps). Retrieved May 12, 2016: https://www.nsf.gov/pubs/2011/nsf11560/nsf11560.htm?org=IIP.

Nature.com. Got data? (2007). *Nature Neuroscience*, 10(8), 931–931. http://doi.org/10.1038/nn0807-931.

Norman, H. (2014). Mandating data archiving: experiences from the frontline. *Learned Publishing*, 27(5), 35–38. http://doi.org/10.1087/20140507.

Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers*. John Wiley & Sons.

Osterwalder, A., Pigneur, Y., Bernarda, G., & Smith, A. (2014). *Value proposition design: how to create products and services customers want*. John Wiley & Sons.

Parham, S. W., Bodnar, J., Fuchs, S. (2012). Supporting tomorrow’s research: Assessing faculty data curation needs at Georgia Tech. *College & Research Libraries News*. 73(1), 10-13. http://crln.acrl.org/content/73/1/10.full.

Peters, C. and Dryden, A.R. (2011). Assessing the Academic Library's Role in Campus-Wide Research Data Management: A First Step at the University of Houston. *Science & Technology Libraries* 30(4). http://dx.doi.org/10.1080/0194262X.2011.626340.

Pinfield, S., Cox, A. M., & Smith, J. (2014). Research Data Management and Libraries: Relationships, Activities, Drivers and Influences. *PLoS ONE*, 9(12), e114734. http://doi.org/10.1371/journal.pone.0114734.

Piwowar, H. A. (2011). Who Shares? Who Doesn't? Factors Associated with Openly Archiving Raw Research Data. *PLoS ONE*, 6(7), e18657. http://doi.org/10.1371/journal.pone.0018657.

PLoS ONE. (2015). Data Availability. http://journals.plos.org/plosone/s/data-availability.

PLoS ONE Editors. (2017). Expression of Concern: Adaptive Pacing, Cognitive Behaviour Therapy, Graded Exercise, and Specialist Medical Care for Chronic Fatigue Syndrome: A Cost-Effectiveness Analysis. *PLoS ONE*, May 2, 2017 https://doi.org/10.1371/journal.pone.0177037

Pryor, G. (2013) Options and approaches to RDM service provision. In Pryor, G., Jones, S., & Whyte, A. (Eds.). *Delivering research data management services: Fundamentals of good practice*. London: Facet Publishing. p. 21-40.

Purdue University. (2015) Purdue NSF I-Corps Site. Retrieved May 3, 2016: http://www.krannert.purdue.edu/centers/nsf/home.php.

Ray, J. M. (Ed.). (2014). *Research data management: practical strategies for information professionals*. West Lafayette, Indiana: Purdue University Press.

Roche, D. G., Lanfear, R., Binning, S. A., Haff, T. M., Schwanz, L. E., Cain, K. E., … Kruuk, L. E. B. (2014). Troubleshooting Public Data Archiving: Suggestions to Increase Participation. *PLoS Biol*, 12(1), e1001779. http://doi.org/10.1371/journal.pbio.1001779.

Rogers, E. M. (2010). *Diffusion of innovations*. New York, NY: Simon and Schuster.

Shen, Y. (2016). Research Data Sharing and Reuse Practices of Academic Faculty Researchers: A Study of the Virginia Tech Data Landscape. I*nternational Journal of Digital Curation*, 10(2), 157–175. http://doi.org/10.2218/ijdc.v10i2.359.

Soehner, C., Steeves, C., & Ward, J. (2010). E-science and data support services: A study of ARL member institutions. Retrieved from Association of Research Libraries website: http://www.arl.org/storage/documents/publications/escience-report-2010.pdf.

Sturges, P., Bamkin, M., Anders, J. H., Hubbard, B., Hussain, A., & Heeley, M. (2015). Research data sharing: Developing a stakeholder-driven model for journal policies. *Journal of the Association for Information Science and Technology*, 66(12), 2445-2455. http://doi.org/10.1002/asi.23336.

Sutton, S., Barber, D.; Whitmire, A. L. (2013). Oregon State University Libraries and Press Strategic Agenda for Research Data Services. Oregon State University Libraries. http://hdl.handle.net/1957/38794.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., ... & Frame, M. (2011). Data sharing by scientists: practices and perceptions. *PloS ONE*, 6(6), e21101. http://doi.org/10.1371/journal.pone.0021101.

Tenopir, C., Birch, B., & Allard, S. (2012). Academic Libraries and Research Data Services: Current Practices and Plans for the Future. An ACRL White paper. http://www.ala.org/acrl/sites/ala.org.acrl/files/content/publications/whitepapers/Tenopir_Birch_Allard.pdf.

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., … Dorsett, K. (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLoS ONE*, 10(8), e0134826. http://doi.org/10.1371/journal.pone.0134826.

Tenopir, C., Hughes, D., Allard, S., Frame, M., Birch, B., Baird, L., … Lundeen, A. (2015). Research Data Services in Academic Libraries: Data Intensive Roles for the Future? *Journal of eScience Librarianship*, 4(2). https://doi.org/10.7191/jeslib.2015.1085.

Van den Eynden, V. and Bishop, L. (2014). Sowing the Seed: Incentives and motivations for sharing research data, a researcher's perspective. A Knowledge Exchange Report. Retrieved June 2016:: http://repository.jisc.ac.uk/5662/1/KE_report-incentives-for-sharing-researchdata.pdf.

Vasilevsky, N. A., Minnier, J., Haendel, M. A., & Champieux, R. E. (2017). Reproducible and reusable research: are journal data sharing policies meeting the mark?. *PeerJ*, 5, e3208. https://doi.org/10.7717/peerj.3208.

Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLoS ONE*, 8(7), e67332. http://doi.org/10.1371/journal.pone.0067332.

Westra B. (2014) Developing Data Management Services for Researchers at the University of Oregon. In: Ray JM, editor. *Research Data Management: Practical Strategies for Information Professionals*. West Lafayette, Indiana: Purdue University Press. p. 375–91.

Whitmire, A. L., Boock, M., & Sutton, S. C. (2015). Variability in academic research data management practices: implications for data services development from a faculty survey. Program. http://doi.org/10.1108/PROG-02-2015-0017.

Witt, M., Carlson, J. R., Cragin, M. R., & Brandt, D. S. (2009). Constructing Data Curation Profiles. *International Journal of Digital Curation*, 4(3). http://dx.doi.org/10.2218/ijdc.v4i3.117.

York, J., Gutmann, M., Berman, F., (2016) "Will Today's Data Be Here Tomorrow? Measuring the Stewardship Gap," Proceedings of the 2016 International Conference on Digital Preservation (IPRES 16), 102-111. https://ipr16.organizers-congress.org/frontend/organizers/media/iPRES2016/_PDF/IPR16.Proceedings_4_Web_Broschuere_Link.pdf.

Zhang, T., Zilinski, L., Brandt, D. S., & Carlson, J. (2015). Assessing Perceived Usability of the Data Curation Profile Toolkit Using the Technology Acceptance Model. *International Journal of Digital Curation*, 10(1), 48-67. http://dx.doi.org/10.2218/ijdc.v10i1.344.

BIBLIOGRAPHY: CONNECTING RESEARCHERS TO REPOSITORIES

Akmon, D., Zimmerman, A., Daniels, M., & Hedstrom, M. (2011). The application of archival concepts to a data-intensive environment: working with scientists to understand data management and preservation needs. *Archival Science*, 11(3-4), 329–348. https://doi.org/10.1007/s10502-011-9151-4.

Almas, B., Bicarregui, J., RTI, A. B., Hill, S., Lannom, L., Pennington, R., ... & CAS, Z. Y. (2015). Data Management Trends, Principles and Components–What Needs to be Done Next?. Report from the Research Data Alliance Data Fabric Interest Group, draft version https://www.rd-alliance.org/system/files/documents/paris-doc-v6-1.pdf from September.

Anagnostou, P., Capocasa, M., Milia, N., Sanna, E., Battaggia, C., Luzi, D., & Destro Bisol, G. (2015). When Data Sharing Gets Close to 100%: What Human Paleogenetics Can Teach the Open Science Movement. *PLoS ONE*, *10*(3), e0121409. http://doi.org/10.1371/journal.pone.0121409.

Anderson, W. L. (2004). Some challenges and issues in managing, and preserving access to, long-lived collections of digital scientific and technical data. *Data Science Journal*, 3, 191-201. https://www.jstage.jst.go.jp/article/dsj/3/0/3_0_191/_article.

Antell, K., Foote, J. B., Turner, J., & Shults, B. (2013). Dealing with Data: Science Librarians' Participation in Data Management at Association of Research Libraries Institutions. *College & Research Libraries*, 75(4), 557–574. http://doi.org/10.5860/crl.75.4.557.

Baker, K. S., & Yarmey, L. (2009). Data stewardship: Environmental data curation and a web-of-repositories. *International Journal of Digital Curation*, 4(2). http://www.ijdc.net/index.php/ijdc/article/view/115.

Beardsley, T. M. (2015). Notes on Changing Practices in Data Publication. *BioScience*. http://bioscience.oxfordjournals.org/content/65/7/645.full.

Berman, F. (2008). Got Data?: A Guide to Data Preservation in the Information Age. *Communications of the ACM*, *51*(12), 50–56. http://doi.org/10.1145/1409360.1409376.

Berman, F., & Cerf, V. (2013). Who Will Pay for Public Access to Research Data? *Science*, *341*(6146), 616–617. http://doi.org/10.1126/science.1241625.

Borgman, C. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, Massachusetts: The MIT Press.

Bolukbasi, B., Berente, N., Cutcher-Gershenfeld, J., Dechurch, L., Flint, C., . . . Walker, D. (2013). Open Data: Crediting a Culture of Cooperation. *Science*, Letters to the Editor, 342, 1041-1042. http://doi.org/10.1126/science.342.6162.1041-b.

Brandt, D. S. (2011). Disambiguating the role of data lifecycle gatekeeper. In: Workshop on Research Data Lifecycle Management (RDLM 2011), Princeton University, Princeton, NJ. July

18-20, 2011. Retrieved December 20, 2016: http://www.columbia.edu/~rb2568/rdlm/Brandt_Purdue_RDLM2011.pdf.

Bruna, E. M. (2014). The opportunity cost of my #OpenScience was 36 hours + $690 (UPDATED). *The Bruna Lab | UF.* [Posted September 4, 2014]. http://brunalab.org/blog/2014/09/04/the-opportunity-cost-of-my-openscience-was-35-hours-690/.

Callaghan, S., Tedds, J., Kunze, J., Khodiyar, V., Lawrence, R., Mayernik, M. S., Murphy F., Roberts T., & Whyte, A. (2014). Guidelines on Recommending Data Repositories as Partners in Publishing Research Data. *International Journal of Digital Curation*, *9*(1), 152–163. http://doi.org/10.2218/ijdc.v9i1.309.

Callaghan, S., Donegan, S., Pepler, S., Thorley, M., Cunningham, N., Kirsch, P., et al. . (2012). Making Data a First Class Scientific Output: Data Citation and Publication by NERC's Environmental Data Centres. *International Journal of Digital Curation*, *7*(1), 107–113. http://doi.org/10.2218/ijdc.v7i1.218.

Christensen-Dalsgaard, B. (2012). Ten recommendations for libraries to get started with research data management. Final Report of the LIBER working group on E-Science / Research Data Management. http://libereurope.eu/blog/2012/08/24/ten-recommendations-for-libraries-to-get-started-with-research-data-management.

Cox, A. M., Kennan, M. A., Lyon, L., & Pinfield, S. (2017). Developments in research data management in academic libraries: Towards an understanding of research data service maturity. *Journal of the Association for Information Science and Technology*, 68(9). http://doi.org/10.1002/asi.23781.

Cragin, M. H., & Shankar, K. (2006). Scientific data collections and distributed collective practice. *Computer Supported Cooperative Work* (CSCW), 15(2), 185-204. https://doi.org/10.1007/s10606-006-9018-z.

Curty, R. G., & Qin, J. (2014). Towards a model for research data reuse behavior. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1–4. http://doi.org/10.1002/meet.2014.14505101072.

Dafoe, A. (2014). Science Deserves Better: The Imperative to Share Complete Replication Files. *PS: Political Science & Politics*, *47*(01), 60–66. http://doi.org/10.1017/S104909651300173X.

DCC (2014). "Five steps to decide what data to keep: a checklist for appraising research data, v.1". Edinburgh: Digital Curation Centre. http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep.

Douglass, K., Allard, S., Tenopir, C., Wu, L., & Frame, M. (2014). Managing scientific data as public assets: Data sharing practices and policies among full-time government employees. *Journal of the Association for Information Science and Technology*, 65(2), 251-262. doi.org/10.1002/asi.22988.

Duerr, R. (2013). Data Archives and Repositories. In E. G. Nijoku (Ed.), *Encyclopedia of Remote Sensing: Springer Science+Business Media*. New York, NY: Springer-Verlag.

Edinburgh University DataShare. (2015). “Checklist for deposit”. Retrieved November 2, 2015: http://www.ed.ac.uk/information-services/research-support/data-library/data-repository/checklist.

Ember, C., & Hanisch, R. (2013). “Sustaining Domain Repositories for Digital Data: A White Paper: Sustaining Domain Repositories for Digital Data Workshop,” June 24-25, 2013, University of Michigan, http://datacommunity.icpsr.umich.edu/sites/default/files/WhitePaper_ICPSR_SDRDD_121113.pdf.

Fary, M., & Owen, K. (2013). “Developing an Institutional Research Data Management Plan Service”. EDUCAUSE, ACTI DMWG – Advanced Core Technologies Initiative Data Management Working Group. http://www.educause.edu/library/resources/developing-institutional-research-data-management-plan-service.

Federer, L. (2016). Research data management in the age of big data: Roles and opportunities for librarians. *Information Services & Use*, 36(1-2), 35-43. doi.org/10.3233/ISU-160797.

Flanders, J., & Muñoz, T. (2011). An Introduction to Humanities Data Curation. *DH Curation Guide*. http://guide.dhcuration.org/intro/.

Flores, J. R., Brodeur, J. J., Daniels, M. G., Nicholls, N., & Turnator, E. (2015). Libraries and the research data management landscape. In Maclachlan, J. C., Waraksa, E.A. and Williford, C. (Eds.) The Process of Discovery: The CLIR Postdoctoral Fellowship Program and the Future of the Academy. (p. 92). https://www.clir.org/pubs/reports/pub167.

Franke, M., Heinzel, S., Mauer, R., Neumann, J., Neuroth, H., Pfeiffenberger, H., Senst, H., … Winkler-Nees, S. (2015): "Research Data at Your Fingertips": A Position Paper by the Research Data Working Group, Potsdam : Deutsches GeoForschungsZentrum GFZ. http://doi.org/10.2312/allianzfd.002.

Henderson, M. E., & Miller, H. (2016). Compliance: Data Management Plans and Public Access to Data, University of Massachusetts and New England Area Librarian e-Science Symposium, Amherst, MA, April 6, 2016. http://escholarship.umassmed.edu/escience_symposium/2016/program/8/.

Hickson, S., Poulton, K. A., Connor, M., Richardson, J., & Wolski, M. (2016). Modifying researchers’ data management practices: A behavioural framework for library practitioners. *IFLA Journal*, 42(4), 253-265. https://doi.org/10.1177%2F0340035216673856.

Islam, A., Agarwal, N. K., & Ikeda, M. (2015). How do academic libraries work with their users to co-create value for service innovation?: A qualitative survey. Qualitative and Quantitative Methods in Libraries, 4(3), 637-658. http://www.qqml.net/papers/September_2015_Issue/4313QQML_Journal_2015_IkedaandAgarwal_637-658.pdf

Kelly, J., & Eells, L. (2015). Global scholarship: The role of subject repositories in advancing research from the developing world. *College & Research Libraries News*, 76(5), 265-268. https://doi.org/10.5860/crln.76.5.9313.

Keralis S.D., Stark S., Halbert M., Moen W.E. (2013). Research Data Management in Policy and Practice: The DataRes Project. In *Research Data Management: Principles, Practices and Prospects*. CLIR publication: p16-38. https://www.clir.org/pubs/reports/pub160.

Kervin, K., Michener, W., & Cook, R. (2013). Common Errors in Ecological Data Sharing. *Journal of eScience Librarianship*, *2*(2). http://dx.doi.org/10.7191/jeslib.2013.1024.

Kindling, M., Pampel H., van de Sandt, S., Rucknagel, J. Vierkant, P. Kloska, G., Witt, M.…. Scholze, F. (2017). The Landscape of Research Data Repositories in 2015: A re3data Analysis. *D-Lib Magazine*, *23*(3/4). https://doi.org/10.1045/march2017-kindling.

Kollen, C., Kouper, I., Ishida, M., Williams, S., & Fear, K. (2017). Research Data Services Maturity in Academic Libraries. American Library Association, Association of College and Research Libraries. http://hdl.handle.net/10150/622168.

Kratz, J. E., & Strasser, C. (2015). Researcher Perspectives on Publication and Peer Review of Data. *PLoS ONE*, *10*(2). http://doi.org/10.1371/journal.pone.0117619.

Leeper, Thomas (2015). What's in a Name? The Concepts and Language of replication and Reproducibility, Thomas J. Leeper Blog. [Posted May 12, 2015]. http://thomasleeper.com/2015/05/open-science-language.

Lyon, L., & Brenner, A. (2015). Bridging the data talent gap: Positioning the iSchool as an agent for change. *International Journal of Digital Curation*, 10(1), 111-122. http://dx.doi.org/10.2218/ijdc.v10i1.349.

Marx, V. (2012). My data are your data. *Nature Biotechnology*, *30*(6), 509–511. http://doi.org/10.1038/nbt.2243.

McNutt, M. (2014). Reproducibility. *Science*, *343*(6168), 229–229. http://doi.org/10.1126/science.1250475.

Michener, W. K., Vieglais, D. A., Vision, T., Kunze, J., Cruse, P., & Janée, G. (2011). "DataONE: Data Observation Network for Earth — Preserving Data and Enabling Innovation in the Biological and Environmental Sciences". *D-Lib Magazine,* 17(1/2). http://www.dlib.org/dlib/january11/michener/01michener.html.

Mills, J.A., Teplitsky, C., Arroyo, B., Charmantier, A, Becker, P.H. et al. (2015). Archiving Primary Data: Solutions for Long-Term Studies. *Trends in Ecology & Evolution, 30*(10), 581-589. http://dx.doi.org/10.1016/j.tree.2015.07.006

National Science Foundation, Office of Cyberinfrastructure, Directorate for Computer & Information Science & Engineering, "Sustainable Digital Data Preservation and Access Network Partners (DataNet)." http://www.nsf.gov/pubs/2007/nsf07601/nsf07601.htm.

Naughton, L. & Kernohan, D., (2016). Making sense of journal research data policies. *Insights*. 29(1), pp.84–89. http://doi.org/10.1629/uksg.284.

Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry*, *23*(3), 217-243. http://dx.doi.org/10.1080/1047840X.2012.692215.

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science*, *7*(6), 615-631. https://doi.org/10.1177/1745691612459058.

Ogburn, J. L. (2010). The Imperative for Data Curation. *Portal: Libraries and the Academy*, *10*(2), 241–246. http://doi.org/10.1353/pla.0.0100.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). http://doi.org/10.1126/science.aac4716.

Palumbo, Laura B., Ron Jantz, Yu-Hung Lin, Aletia Morgan, Minglu Wang, Krista White, Ryan Womack, Yingting Zhang, and Yini Zhu. (2015). Preparing to Accept Research Data: Creating Guidelines for Librarians. *Journal of eScience Librarianship* 4(2): e1080. http://dx.doi.org/10.7191/jeslib.2015.1080

Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., … Dierolf, U. (2013). Making Research Data Repositories Visible: The re3data.org Registry. *PLoS ONE*, *8*(11). http://doi.org/10.1371/journal.pone.0078080.

Peer, L., Green, A., & Stephenson, E. (2014). Committing to Data Quality Review. *International Journal of Digital Curation*, *9*(1), 263–291. http://doi.org/10.2218/ijdc.v9i1.317.

Peng, G., Ritchey, N., Casey, K., & Kearns, E. (2016). Scientific Stewardship in the Open Data and Big Data Era — Roles and Responsibilities of Stewards and Other Major Product Stakeholders. *D-Lib Magazine*, *22*(5/6). http://doi.org/10.1045/may2016-peng

Peng, R. D. (2011). Reproducible Research in Computational Science. *Science*. 334(6060), 1226–1227. http://doi.org/10.1126/science.1213847.

Piwowar H.A., Day R.S., Fridsma D.B. (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLoS ONE*. 2(3). http://doi.org/10.1371/journal.pone.0000308.

Poole, A. H. (2014). How has your science data grown? Digital curation and the human factor: a critical literature review. *Archival Science*, 15(2), 101–139. http://doi.org/10.1007/s10502-014-9236-y.

Raboin R, Reznik-Zellen R, Salo D. (2112). Forging New Service Paths: Institutional Approaches to Providing Research Data Management Services. *Journal of eScience Librarianship*. http://escholarship.umassmed.edu/jeslib/vol1/iss3/2/.

Research Information. (2015). “What makes researchers willing to share their data?”. http://www.researchinformation.info/features/feature.php?feature_id=511.

Reilly, S. K. (2014). Rounding up the data: libraries pushing new frontiers. *Learned Publishing*, *27*(5), 33–34. http://doi.org/10.1087/20140506.

Sacchi, S., Wickett, K., Renear, A., & Dubin, D. (2011). A framework for applying the concept of significant properties to datasets. *Proceedings of the American Society for Information Science and Technology*, *48*(1), 1–10. http://doi.org/10.1002/meet.2011.14504801148.

Salo, D. (2013). “Data Curation’s Dirty Little Secret | Peer to Peer Review”. Retrieved November 2, 2015: http://lj.libraryjournal.com/2013/06/industry-news/data-curations-dirty-little-secret-peer-to-peer-review.

Share alike: Research communities need to agree on standard etiquette for data-sharing. (2014). *Nature*, *507*(7491), 140–140. http://doi.org/10.1038/507140a.

Smith M. (2014). Data Governance: Where Technology and Policy Collide. In: Ray, Joyce M., editor. *Research Data Management: Practical Strategies for Information Professionals*. West Lafayette, Indiana: Purdue University Press. p. 45–59.

Starr J, Castro E, Crosas M, Dumontier M, Downs RR, Duerr R, et al. . (2015) Achieving human and machine accessibility of cited data in scholarly publications. *PeerJ Computer Science*. https://dx.doi.org/10.7717/peerj-cs.1.

Stodden, V., Guo, P., & Ma, Z. (2013). Toward Reproducible Computational Research: An Empirical Analysis of Data and Code Policy Adoption by Journals. *PLoS ONE*, *8*(6), e67111. http://doi.org/10.1371/journal.pone.0067111.

Strasser, C. (2012). “Data Management: A Scientist's Perspective”. Presentation given at the Data Management Training Workshop, University of Florida Health Sciences Library. http://www.slideshare.net/carlystrasser/data-mgmt-scientist-perspective.

Strasser, C. A., Cook, R., Michener, W., Budden, A., & Koskela, R. (2011). Promoting Data Stewardship Through Best Practices. In *Proceedings of the Environmental Information Management Conference 2011*. Santa Barbara, CA: University of California. Retrieved from https://eim.ecoinformatics.org/eim2011/eim-proceedings-2011/at_download/file.

Swauger, S., & Vision, T. J. (2015). What Factors Influence Where Researchers Deposit their Data? A Survey of Researchers Submitting to Data Repositories

. *International Journal of Digital Curation*, *10*(1), 68–81. http://doi.org/10.2218/ijdc.v10i1.289.

Tenopir, C., Sandusky, R., Allard, S., & Birch, B. (2013). Academic librarians and research data services: preparation and attitudes. *IFLA Journal*, 39(1), 70–78. http://doi.org/10.1177/0340035212473089.

Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, 36(2), 84-90. https://doi.org/10.1016/j.lisr.2013.11.003

University of Edinburgh.(2011). Research Data Management Policy. http://www.ed.ac.uk/information-services/about/policies-and-regulations/research-data-policy.

University of Minnesota Libraries. (2015). The Supporting Documentation for Implementing the Data Repository for the University of Minnesota (DRUM): A Business Model, Functional Requirements, and Metadata Schema. Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/171761.

Van der Graaf, M. (2009). *The European Repository Landscape 2008 : Inventory of Digital Repositories for Research Output*. Amsterdam: Amsterdam University Press. http://dare.uva.nl/aup/nl/record/316871.

Vines, T. H., Albert, A. Y. K., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., … Rennison, D. J. (2014). The Availability of Research Data Declines Rapidly with Article Age. *Current Biology*, *24*(1), 94–97. http://doi.org/10.1016/j.cub.2013.11.014.

Wallis, J. (2014). Data Producers Courting Data Reusers: Two Cases from Modeling Communities. *International Journal of Digital Curation*, *9*(1), 98–109. http://doi.org/10.2218/ijdc.v9i1.304

Ward, P. (2015). *Repositories Unleashing Data: Ideas: Ideas and comments gathered from RepoFringe 2015 participants* . Poster session presented at RepoFringe 2015 (Repositories Fringe), Edinburgh, United Kingdom. http://www.research.ed.ac.uk/portal/files/21378016/Repositories_Unleashing_Data_Ideas.pdf.

Weber, N. M., Chao, T. C., Palmer, C. L., & Varvel, V. E. (2011). *Report on the Data Curation Research Summit*. Center for Informatics Research in Science and Scholarship, University of Illinois. Available: https://www.ideals.illinois.edu/handle/2142/28355.

White, E. P., Baldridge, E., Brym, Z. T., Locey, K. J., McGlinn, D. J., & Supp, S. R. (2013). Nine simple ways to make it easier to (re)use your data. *Ideas in Ecology and Evolution*, *6*(2). http://library.queensu.ca/ojs/index.php/IEE/article/view/4608.

Whyte, A. (2015). “Where to keep research data: DCC checklist for evaluating data repositories”, v.1. Edinburgh: Digital Curation Centre. http://www.dcc.ac.uk/resources/how-guides-checklists/where-keep-research-data.

Wickett, K. M., Isaac, A., Doerr, M., Fenlon, K., Meghini, C., & Palmer, C. L. (2014). Representing Cultural Collections in Digital Aggregation and Exchange Environments. *D-Lib Magazine*, 20(5). http://www.dlib.org/dlib/may14/wickett/05wickett.html.

Witt, M. (2008). Institutional Repositories and Research Data Curation in a Distributed Environment. *Library Trends*, *57*(2), 191–201. http://doi.org/10.1353/lib.0.0029.

Wuchty, S. Jones, B. F., Uzzi, B. (2007). The Increasing Dominance of Teams in Production of Knowledge. *Science*. 316, 1036–1039. http://doi.org/10.1126/science.1136099.

Zaslavsky, I. (2012). EarthCube Roadmap: Cross-Domain Interoperability Test Bed Group, Version 1.1. https://www.earthcube.org/sites/default/files/doc-repository/EarthCube%2520Cross-Domain%2520Interoperability%2520Roadmap.pdf.

Zimmerman, A. S. (2008). New Knowledge from Old Data The Role of Standards in the Sharing and Reuse of Ecological Data. *Science, Technology & Human Values*, *33*(5), 631–652. http://doi.org/10.1177/0162243907306704.