Title
What lies beneath?: Knowledge infrastructures in the subseafloor biosphere and beyond

Journal
International Journal on Digital Libraries, 16(2015)

ISSN
14325012, 14321300

Authors
Darch, Peter T. T.
Borgman, Christine L.
Traweek, Sharon
et al.

Publication Date
2015-02-15

Peer reviewed

Library Cultures of Data Curation: Adventures in Astronomy

AUTHORS:
Peter T. Darch (corresponding author)[1], ptdarch@illinois.edu
Ashley E. Sands[2], sands.ashley@gmail.com
Christine L. Borgman[2], borgman@gseis.ucla.edu
Milena S. Golshan[2], milenasvg@gmail.com

[1] School of Information Sciences, University of Illinois at Urbana-Champaign, 501 E Daniel Street, Urbana, IL 61821, tel: (217) 333-3280, fax: (217) 244-3302
[2]Department of Information Studies, University of California Los Angeles, GSE&IS Building, Bos 951520, Los Angeles, CA 90095, tel: (310) 825-8799, fax: (310) 206-4460

Abstract

University libraries are partnering with disciplinary data producers to provide long-term digital curation of research datasets. Managing dataset producer expectations and guiding future development of library services requires understanding the decisions libraries make about curatorial activities, why they make these decisions, and the effects on future data reuse. We present a study, comprising interviews (n=43) and ethnographic observation, of two university libraries who partnered with the Sloan Digital Sky Survey (SDSS) collaboration to curate a significant astronomy dataset. The two libraries made different choices of the materials to curate and associated services, which resulted in different reuse possibilities. Each of the libraries offered partial solutions to the SDSS leaders' objectives. The libraries' approaches to curation diverged due to contextual factors, notably the extant infrastructure at their disposal (including technical infrastructure, staff expertise, values and internal culture, and organizational structure). The Data Transfer Process case offers lessons in understanding how libraries choose curation paths and how these choices influence possibilities for data reuse. Outcomes may not match data producers' initial expectations but may create opportunities for reusing data in unexpected and beneficial ways.

Introduction

Research datasets can have enormous reuse potential many years after their initial production (Pasquetto, Randles, & Borgman, 2017), but only if the necessary infrastructure exists to ensure ongoing curation (Borgman, 2015). These datasets are typically produced by projects that receive funding on scales of months to years. As funding ends, valuable datasets are often irretrievably lost as dedicated staff disperse and computational systems cease to be maintained, with serious consequences for scientific progress. Loss of observational data taken at particular times and places, such as sky survey data or environmental data, cannot be repeated. Digital datasets are especially vulnerable, degrading quickly without regular maintenance (Edwards et al., 2013).

Researchers and collaborative research projects face increasing pressure to comply with data management policies from stakeholders such as research funding agencies (Bishoff & Johnston, 2015). Long-term data management is an institutional challenge; maintaining datasets is often well beyond the responsibility of researchers and research projects. With planning, datasets may be migrated to other institutions with the necessary workforces, tools, and technologies. One such institution is university libraries (Tenopir, Sandusky, Allard, & Birch, 2014). Many libraries are developing services centered on provision of digital infrastructure for data curation, but continue to face many decisions about how to implement these services (Cox, Kennan, Lyon, & Pinfield, 2017).

This paper is a study of how two research libraries responded to a request from a major astronomy collaboration, the *Sloan Digital Sky Survey* (*SDSS*), to curate a large-scale dataset. The data that this paper draws on were originally collected for a doctoral dissertation (Sands, 2017). This dissertation addressed broader questions about the meaning and process of data management in astronomy. Our paper asks different questions of the data, focusing on the institutional processes that affect data curation.

We refer to the collaborative process between SDSS and the two libraries, from initial approach and negotiations through to the libraries' curatorial work on the dataset, as the *Data Transfer Process.* The two libraries diverged significantly regarding what SDSS materials they curated, and in how they implemented associated services to support data access and use. The resulting possibilities to reuse the dataset – including by who and for what purposes – varied substantially between the libraries. Furthermore, the decisions taken by each library during the process were shaped by factors specific to that library. Our study particularly calls to attention the effects of values and culture. These deep-rooted factors both demonstrate the scale of the challenge in improving common understandings between dataset producers and libraries and reveal opportunities to enhance the value of libraries' data curation services.

Background

Data curation requires supporting infrastructure, and we apply an infrastructural perspective to understand development of libraries' digital data curation services (Bowker, 2005; Edwards et al., 2013). These services are an example of *knowledge infrastructures* (*KI*). Here, we review technical, social, and cultural dimensions of KI; how KI is built upon extant infrastructure; and infrastructural factors that shape how libraries develop data curation services.

Data Curation and Knowledge Infrastructures

The term data curation refers to actions taken on datasets at any stage of their existence (from initial conception of data collection, through collection, analysis, and beyond) that enhance their use or reuse value (Palmer, Weber, Munoz, & Renear, 2013). Research datasets produced for a particular purpose may be reused by scholars for multiple other purposes, including production of new scientific knowledge, reproducing existing results to ensure scientific integrity, and calibration of scientific instruments (Borgman, 2015; Pasquetto et al., 2017).

Reusing datasets may require considerable effort from a prospective reuser. This effort varies according to the resources available to the prospective reuser, and the purposes for which the dataset will be reused. Different curatorial actions performed on a dataset can affect this required effort, helping some prospective reusers and hindering others, and making some types of reuse easier and others more difficult (Chao, Cragin, & Palmer, 2015). Actions taken to preserve the dataset in a stable format are essential. Cleaning datasets can improve their quality and make it easier for prospective reusers to understand its contents (Arora, Esteva, & Trelogan, 2014). Providing accompanying metadata and other information about provenance can help prospective reusers to find and interpret the dataset, and to make judgments about its quality (Fear & Donaldson, 2012; Greenberg, 2017). Linking the dataset to other knowledge products such as code and journal articles can promote reproducibility (Brinckman et al., 2019).

These – and other – curatorial actions require supporting infrastructures. Infrastructures comprise not only technical dimensions, but also social, cultural, and organizational dimensions (Star & Ruhleder, 1996). Infrastructures for data curation include projects that generate data, repositories, databases, standards (e.g., for metadata), and staff (Borgman et al., 2015). Infrastructures for data curation are an example of *knowledge infrastructures* (KI) that, "generate, share, and maintain specific knowledge about the human and natural worlds" (Edwards 2010, p. 13).

Star and Ruhleder (1996) identified eight dimensions of infrastructures, of which two are particularly relevant to KI and data curation. One dimension is that an infrastructure is not built from scratch; instead, it is layered on existing infrastructure (or *installed base*). The form of the new infrastructure is shaped by, and thus "inherits strengths and limitations" from the installed base (Star & Ruhleder, 1996, p. 113). Another dimension is *links with convention of practice*, or the idea that an infrastructure is both shaped by, and shapes, the norms of a community of practice. This second dimension applies to the professional socialization and enculturation of those people involved in building and maintaining infrastructures – the *human infrastructure* of cyberinfrastructure (Lee, Dourish, & Mark, 2006). As they are socialized into this work environment, they inherit cultural practices and values already embedded in the extant infrastructure. Their values, in turn, shape subsequent development of the infrastructure (Shilton, 2015).

Research Data and University Libraries

University libraries are critical components of knowledge infrastructures, with an increasing number addressing research dataset curation challenges (Steinhart, 2014; Wittenberg & Elings, 2017). Half of the research libraries surveyed by Tenopir et al. (2017) either already provide, or plan to deliver, curatorial services that involve handling research datasets. Services can include acquiring datasets, such as selecting potential datasets as candidates for curation and selecting elements for inclusion in datasets; preparing datasets for deposition, such as transferring them from their original location; hosting datasets, including ingesting and storing datasets; and supporting access and use of datasets, such as providing search and retrieval tools, and implementing helpdesk services (Johnston et al., 2017). Cox et al. (2017) argue that libraries' services are progressing towards a "fully mature" state as they add new, and refine existing, features.

Although these services often require libraries to train existing staff in new skills or to hire new staff with requisite knowledge (Coates, 2014; Tenopir et al., 2015), libraries do not develop these services *de novo*. Per Star and Ruhleder's definition of infrastructure (1996), libraries must layer their new services on an installed base comprising already-available technologies, policies, personnel, and expertise. For instance, recounting experiences at Cornell University libraries, Steinhart (2014) identifies multiple precursors such as hiring data librarians and launching an institutional repository, that laid the ground for development of the library's data curation services. Other studies argue that the traditional expertise of librarians, such as information organization, may be adaptable to handling research datasets (Antell, Foote, Turner, & Shults, 2014).

Problem Statement and Research Questions

Once libraries have decided to curate research datasets, they must determine how to proceed, working out what resources they have available to support curation. Libraries build their digital dataset curation services on what they already have (*installed base*). This installed base has both affordances and constraints that shape how libraries can proceed. Building a data curation infrastructure on top of this installed base is not a trivial task for libraries: this installed base was typically built to support curation of bibliographic entities; in contrast to these entities, datasets are often "unruly and poorly-bounded" (Wynholds, 2011, p. 214).

In this paper, we address the following research questions:

1) What constitutes the infrastructure upon which libraries build digital services for research data curation?
2) What decisions do libraries make when building these services?
3) What factors shape these decisions?

Sloan Digital Sky Survey

The Sloan Digital Sky Survey (SDSS) is an international astronomy survey covering over a quarter of the night sky and providing optical and spectroscopic imaging data to astronomers. SDSS began collecting data in May 1998 and continues to this day through multiple project phases (I, II, III, and IV). SDSS data has supported more than 8000 peer-reviewed publications (SDSS Collaboration, 2016). The Alfred P. Sloan Foundation has been SDSS's primary funder. The U.S. National Science Foundation also provided substantial support. The first two phases of SDSS (henceforth, *SDSS-I/II*) are estimated to have cost between $70 and $85 million (Finkbeiner, 2010).

SDSS uses the Sloan Foundation Telescope to collect data. This telescope is located at Apache Point, New Mexico, and is one of three telescopes managed by the *Astrophysical Research Consortium* (*ARC*). ARC comprises eight research universities in the US and is governed by a body comprising senior leaders from these universities. ARC grants access to its telescopes to consortium members and, via leasing agreements, to researchers at other institutions.

SDSS-I/II contributed to a significant shift in the level of openness in astronomy data by providing annual large-scale public data releases. The SDSS collaboration chose to keep each data release intact over time to allow users to conduct long-term research with a stable and reliable data source. SDSS made the data available through its own online cyberinfrastructure operated at a National Laboratory (name withheld to ensure anonymity).

The final dataset of SDSS-I/II was released in October 2008 (Abazajian et al., 2009). By 2008, fiscal close of SDSS-II approached, and the datasets needed long-term management. The initial SDSS collaboration faced the loss of essential expertise and had no assurance of future funding. However, the collaboration remained committed to sustaining the value of SDSS-I/II data. SDSS leaders approached two university libraries to curate SDSS-I/II data. We refer to these two universities using pseudonyms, *Blue University* and *Red University*.

Methods

To address our research questions, we present findings from a longitudinal qualitative case study of the SDSS Data Transfer Process, conducted from 2011-2015. This study comprises semi-structured interviews, ethnographic observations, and document analysis, following standard research methods (Hammersley & Atkinson, 2007). We also draw upon knowledge gained from a broader UCLA study of SDSS since 2009.

We draw on 43 interviews with 39 participants across four sites. Some participants were interviewed more than once while others were interviewed in small groups of two to four persons. Two astronomers are affiliated with both the National Laboratory and Blue University. Table 1 indicates the number of participants for each site, and by workforce category. Interviews lasted between 30 and 90 minutes, averaging one hour. Interviews were audio-recorded and transcribed unless participants asked not to be recorded (two cases). Transcriptions total 818 pages. The interviews were semi-structured, enabling a balance between asking common questions of all interviewees (to enable comparability) and pursuing interesting lines of inquiry specific to individuals. Topics included interviewees' background, decision-making processes and collaborative practices, and challenges encountered during the Data Transfer Process. The second author conducted most of these interviews, with the remainder conducted by multiple other researchers including the third author.

Table 1. Interview participants by site and workforce.

	Blue University	National Laboratory	Red University	ARC/SDSS	**Totals**
Astronomers	3	7	5	7	20*
Library Staff	5	0	6	0	11
Administrators	1	1	1	2	5
Data Center Staff	0	3	0	0	3
Totals	9	11	12	9	**39**

To protect the confidentiality of our research subjects, individuals are referenced using unique identifiers. Interview quotations are concluded with a reference including the interviewee's identifier (randomly assigned), their affiliation, and their career path, for instance: (#28, Blue University, Astronomer).

The second author was embedded for two weeks in the National Laboratory, two weeks in the Blue University library, and five weeks in the Red University library. At each site, she observed team members participating in formal meetings and informal conversations, and completing work tasks. This author took detailed field notes (129 pages) about team members' interactions with each other and with people from other organizations, organizational hierarchies, work practices, and responses to unexpected challenges. The first and second authors also carried out observation at five major gatherings of the American Astronomical Society, affording a deeper understanding of how SDSS fits into the broader astronomy community, including uses of SDSS data and interdependencies with successor sky surveys.

The "UCLA Data Practices Interview Codebook" was used as the basis for data analysis of field notes, interview transcripts, and documents (SDSS operating documents and agreements between institutional actors). This codebook was developed iteratively from the interviews using principles of grounded theory (Glaser & Strauss, 1967). For the purposes of her doctoral dissertation, the second author coded many of the interviews using a subset of codes from the codebook relevant to her dissertation's research questions.

For the current paper, data were analyzed from scratch by the first author, given this paper's focus on new and different research questions. He read all data sources (field notes, interview transcripts, and documents such as SDSS operating documents and agreements between institutional actors) closely two times. He then coded all interviews and field notes, using a subset of codes from the codebook relevant to the current paper (these codes are presented in Table 2). This single-coder approach helped to ensure consistency and reliability of coding across the data corpus (McDonald, Schoenebeck, & Forte, 2019).

A number of measures helped to promote the validity of findings. The first author checked his interpretations of the data regularly with the other authors. The other authors were all familiar with the data, having conducted data collection, data processing, and their own analyses of the data. Using multiple types of data also helped to ensure validity by enabling triangulation (O'Donoghue & Punch, 2004), such as cross-checking interviewees' discussions of work practices with observations of these practices recorded in field notes. Other ways of validating findings included presentations of emerging findings at conferences including where informants were audience members (Darch & Sands, 2016; Darch, Sands, Borgman, Golshan, & Traweek, 2017; Sands & Darch, 2016), and five hours of conversations (audio-recorded, with transcripts available to authors) with five informants where emerging findings were discussed.

Table 2. Codes used in data analysis for this study.

Code category	Code
Personal attributes	Career and biography
Project information	Group members
	Project biography and funding
	Long-term goals
	Short-term goals
	State of technology
	Coordination work and tools
	Governance
Data infrastructure	Archival issues
	Data access
	Data tools
Data characteristics	Data definition
	Data uses
	Data types
	Data variety and scope
	Data volume
	Data standards and quality control
Collaboration	Positive collaboration experiences
	Division of labor
	Collaboration problems
	Collaborators and interconnections
Social and experiential factors	Place
	Policy and institutions
	Successes

Findings

At the start of the SDSS Data Transfer Process, both libraries took on an ostensibly common task: to archive and serve the SDSS-I/II dataset. However, as the process unfolded, the two libraries diverged substantially concerning their decisions about

- What units within the library were charged with carrying out Transfer Process activities;
- Which SDSS materials to include in curation activities; and
- What associated services to implement for data access and reuse.

This divergence occurred due to the extant infrastructure available to each library, including the culture and values embedded in this infrastructure. While the terms "archive and serve" appeared to have concrete meaning to the SDSS and Astrophysical Research Consortium (ARC) leaders, they were surprised by the extent to which the libraries differed in their decisions about how to interpret and implement the archiving and serving of SDSS-I/II data (e.g., "So there were some early discussions at [Red University IIDC]…I think they were sort of viewing this in a little different way" (#24, Blue University, Astronomer)).

Here, we first give an overview of the SDSS data, including the curation tasks undertaken by each library. Libraries were afforded a degree of choice in determining what tasks to take on, and we then explain how and why the libraries made their respective choices.

Overview of SDSS data

Leading figures in the SDSS-I/II collaboration were anxious to sustain open access to the data beyond fiscal close. However, in 2005, they realized the data were fragile. With the prospect of staff dispersing, and hardware and software ceasing to be maintained, these digital data could degrade quickly and the time-dependent observations they represent could never be repeated. As one of our interviewees explained:

> *"The value of the data we were collecting was enormously great, and we needed… some path to preserve it into the indefinite future, but ARC was not necessarily constituted to be the agent to do that, because for several reasons. One is that they had never anticipated that they would be a curator into the indefinite future. In other words, there wasn't a line item in the budget" (#24, Blue University, Astronomer)*

SDSS leaders were primarily concerned with protecting access for professional astronomers for four purposes in particular:

1) To produce new astronomy knowledge. SDSS data remained in active use by the international astronomy community;
2) To enable reproducibility of analyses involving the SDSS-I/II dataset, thereby enhancing trustworthiness of existing findings. As one of our interviewees explained, "There are, I don't know, 10,000 papers published on that [dataset]. So in a way, it's the ultimate reference for many of those" (#9, ARC/SDSS Astronomer);
3) To serve as a source to assist astronomers in choosing astronomical objects to study (e.g., "The imaging is used everywhere for target selection" (#9, ARC/SDSS Astronomer)); and

4) To continue to foster an ethos of open science within astronomy. SDSS leaders had long been committed to ensuring a "community resource for decades" (Szalay, Kunszt, Thakar, Gray, & Slutz, 2000, p. 405).

After two years of planning (2006-2008), ARC negotiated and signed Memoranda of Understanding (MOUs) with both libraries, with the aim that libraries would archive and make accessible (or, per the MOUs, "serve") the data until 2013 (Kron, Gunn, Weinberg, Boroski, & Evans, 2008). A small amount of funding accompanied each MOU.

The SDSS-I/II dataset is a complex aggregation of materials, 100 to 200 terabytes in size depending on what is included in its scope. These comprise the *SDSS Long-Term Scientific Data Archive* (*SDSS Archive*). Each library was able to pursue its visions and motivations when choosing what components of the archive to include in their activities and what associated services to implement to support data access and reuse (see Table 3). As an SDSS leader explained, "The two libraries had very different interests in what they wanted to do, and that's reflected in the MOUs…. So, [the key thing] is to engineer things so that everybody was doing what they wanted to do anyway" (#24, Blue University, Astronomer). Each library was also able to determine how to allocate the work of the Transfer Process within its organizational structure. While Blue library sought to integrate Transfer Process activities into multiple parts across its organization, Red library assigned its Data Transfer activities entirely to a semi-autonomous unit set apart from rest of the library's operations.

Table 3. SDSS long-term scientific data archive: Library task distribution.

SDSS Long-Term Scientific Data Archive Components	Blue University Library	Red University Library
Data Archive Server (DAS)	X	X
Catalog Archive Server (CAS)	X	X
Raw Data and Software	--	X
Administrative Archive	X	--
Help Desk	X	--

Components of the SDSS Archive include:

- *Data Archive Server (DAS)*, a complete set of all the processed data in a flat file format;
- *Catalog Archive Server (CAS)*, a collection of each version of the database released to the public during SDSS-I/II. The CAS provides a searchable and accessible version of the SDSS data;
- *Administrative Archive*, including internal communications, proposals, and operating documents;
- *Help Desk*, a service to answer queries from potential users of the SDSS dataset, for example, by answering questions about how to understand the dataset or how to assess information about data provenance;
- *Raw Data and Software*, data as they came off the SDSS telescope and software used to process the raw data and produce the DAS.

Blue University

Two factors drove Blue library's involvement in the Data Transfer Process. One is its library values of service to its university community. The second is that it wanted to expand its relevance to its university community. These drivers intersected with the affordances and limitations of the library's extant infrastructure (including organizational structure, services, and staff expertise) to guide the library's choices about what SDSS materials to curate and what associated services to implement.

Institutional Characteristics

Blue is a major private research university whose astronomy department was a member of the SDSS-I/II collaboration. In line with common university practice, this library has always emphasized service to its university community, including a commitment to long-term curation of the university's administrative records and faculty papers of historic interest.

Blue University library had developed infrastructure for curation of bibliometric objects over many years, and believed that elements of this infrastructure could prove useful for supporting the Transfer Process. One element was the library's Special Collections Unit, whose antecedents date to the late 19th century, that houses university and faculty archives. The unit's work is supported by more than a dozen specialized staff, many of long tenure and with expertise in the history of science. Library leaders anticipated that the expertise in this unit would prove suitable for curation of the Administrative Archive. Another example was the library's system for processing user queries about library activities. Library leaders envisioned this system being readily adaptable to operating the Help Desk.

Two rationales motivated Blue University library to undertake the Data Transfer Process. One is that the process aligned with their service mission. Because university faculty members were part of SDSS-I/II, the library felt a duty to take on the Data Transfer as a record of the university's participation in a prestigious collaboration. The second rationale is that the process promised strategic opportunities to enhance and promote the library's long-term sustainability. Library leaders were worried that, "A lot of people consider [libraries] sort of stodgy and moribund things" (#3, Blue University, Library Staff). Developing infrastructure and expertise to support faculty members' data management and curation needs was seen as an opportunity for the library to be more relevant to faculty members' scientific work practices. The following interview quotation explains how library staff hoped that the Transfer Process would contribute to developing this infrastructure:

> *"There is a strong future in data …[the transfer] was an opportunity for us to start looking at data and understand the library's role, and what we can and cannot do, and what kind of skill sets we need" (#35, Blue University, Library Staff)*

The library perceived the Data Transfer Process as an opportunity to develop infrastructure and expertise to serve data curation needs of researchers from a range of academic disciplines, and not just astronomy.

The Data Transfer Process was integrated into the main operations of the library at Blue University. Once the library decided to undertake this transfer, it assembled a team to perform the work. The library drew this team from multiple units, including the science library, the main research library, the digital library center, and the Special Collections Unit. Collectively, the team's skills and expertise included archival practices, software programming, library administration, and user services. Library leaders wanted the transfer to cause minimal disruption to the library's existing priorities and organizational structure, with Data Transfer process work being distributed across multiple existing units and absorbed into staff members' existing duties and roles (unrecorded interview with senior library leader).

Choosing What to Archive and Serve

Blue University library staff framed the SDSS Archive in terms of four components: DAS, CAS, Help Desk, and Administrative Archive. To illustrate the dynamics that resulted in this framing, we focus here on how the library chose to include the Administrative Archive in its curatorial activities. At the start of the Data Transfer Process, SDSS/ARC astronomers did not consider the Administrative Archive part of the Data Transfer Process. It was only during the MOU negotiation that preserving the archive emerged as an integral part of the process. The following quotation explains:

> *"[#24, Blue University Astronomer] and I went down to [Blue University], and we met with the library people and talked about the project. And what was interesting was that they were interested not only in the data, but in all of the project documents" (#17, National Laboratory, Astronomer).*

Library staff believed that preserving the Administrative Archive aligned well with the library's service mission. Primarily, library staff concluded that, as a service to prospective users of SDSS-I/II data and to the SDSS project itself, the archive could provide valuable insights into the dataset's provenance:

> *"What we learned as [the SDSS collaboration member] talked through how they did all of this work, was that there were so many changes to the data over time, and much of that was recorded not in the data itself, but in emails... To fully understand this data, you kind of needed all the surrounding information. That's what really led to us realizing this was far more than them saying, "We have 75 terabytes of data we want you to put somewhere." You know, they wanted to be able to actually keep a full record of how and why they'd made every different transformation" (#33 Blue University, Library Staff)*

The library also had a commitment to the university to archive faculty papers that documented important events in the university's history. As SDSS was a major astronomy project in which university faculty were involved, library staff also advocated curating the Administrative Archive ("We very much made the argument that we collect faculty papers" (#33 Blue University, Library Staff)).

The library also chose to handle the Administrative Archive due to existing infrastructure and expertise, namely the Special Collections Unit and its staff, which meant the library regarded preserving the Archive as a feasible goal:

> *"We brought in Special Collections…we also have an extremely strong history of science collection. It's one of our strengths. And so it also fit in there, especially with this idea of it not just being about the research data, but about actually archiving the history of this project" (#33 Blue University, Library Staff)*

Implementing Data Access and Reuse Services

The library developed services to support data access and reuse. To illustrate factors that shaped implementation of these services, we focus here on the Help Desk. Intended to assist prospective data users, the Help Desk aligned well with the library's overarching orientation toward user services. SDSS leaders hoped that the library would tailor Help Desk operations to the SDSS dataset, with library staff developing knowledge about astronomy so they could answer user queries directly. The library took a different approach, implementing a triage service that processed a user's question by directing the question to a suitable person within the SDSS collaboration. While they ran a well-organized ticketing system to handle queries, Help Desk personnel did not develop deep astronomy expertise.

Multiple factors shaped why the library implemented the Help Desk as a triage service. One factor is that the library was able to leverage its existing technical system for handling user queries. A staff member explained how this existing infrastructure worked:

> *"They were curious from us how we handled questions that were coming in, because we're able to queue them, distribute them to different people, we're able to track them. And so it was really a discussion about the system that we use, where questions can come in, everyone can see them, we can re-distribute them, we can track them, we can close them" (#16 Blue University, Library Staff)*

Operating the Help Desk as a triage service also suited library leaders' desire for the Data Transfer to remain minimally disruptive to existing library operations by not requiring staff to take time away from their existing duties to develop extensive astronomy expertise. Finally, the form of the SDSS Help Desk also aligned well with the library's strategic objective of using the Data Transfer to develop infrastructure suitable for a range of disciplines. Library leaders believed that a service focused on triaging queries would be easier to adapt and scale to a growing number and variety of datasets than a service tailored to the particular needs of SDSS data users.

Red University

A unit within the library, the *Institute for Innovation in Digital Curation* (*IIDC*, a pseudonym), handled the Data Transfer Process at Red University. Although some of its strategic objectives for undertaking this process overlapped with those of Blue University library, the IIDC differed from Blue library in terms of its mission, culture, organizational structure, and staff expertise.

Institutional Characteristics

IIDC's focus is on digital innovation. It is part of Red University's library but is autonomous within this library's organizational structure. IIDC's leadership considered the unit to have a "different culture" from the "dominant library culture" (#21 Red University, Library Staff). One driver of cultural difference is that IIDC has a research and development (R&D) orientation. An IIDC leader explained, "What we're doing is considered part of the research enterprise. It's not a library service" (#21 Red University, Library Staff).

A second driver of cultural difference is that the IIDC team mainly comprises software engineers employing agile software development methods to build information systems. Because agile methods are more flexible than traditional approaches to software development, enabling incorporation of new and changing user requirements into software, they are particularly suited to R&D environments. In common with other agile teams, the IIDC is largely self-managing, able to determine its priorities.

Two key rationales motivated IIDC to undertake the Data Transfer Process. One was the IIDC's R&D mission: the scale of the dataset promised IIDC staff an opportunity to push boundaries of the state-of-the-art in data curation. The second was strategic, providing potential opportunities for the IIDC to generate necessary financial resources, given its dependence on so-called "soft money." The Data Transfer Process brought funding to the IIDC, promised to make the IIDC a more attractive partner to potential future collaborators, and offered an opportunity to develop infrastructure that could subsequently be used to "sell services from the university to the outside world" (#27 Red University, Library Staff).

The IIDC undertook the Data Transfer Process alongside other projects. One key example is the *Guardianship of Datasets* (*GD*, a pseudonym) project. GD work involved external collaborators from a wide range of scientific disciplines at major research universities developing infrastructure to facilitate data management in multiple scientific disciplines. Like the Data Transfer Process, GD was an ambitious and ground-breaking project. The Data Transfer and GD had interdependencies that shaped the IIDC's Data Transfer work.

Choosing What to Archive and Serve

In the initial stages of MOU negotiation, SDSS leaders requested that IIDC take on the following components of the SDSS Archive: CAS, Raw Data, and Software. However, IIDC engaged in two moves that resulted in them not only taking on the DAS as well, but also prioritizing archiving the DAS ahead of archiving the CAS. The first move taken by IIDC leaders was to persuade SDSS leaders to allow them to take on DAS:

> *"They were interested in having us hold a copy of the part of the data that's called the CAS, the Catalog Access Server... I told them that that's great to have another copy somewhere, but we would be much more interested in actually preserving the data in a non-database form." (#7 Red University, Library Staff)*

The second move occurred within IIDC, where staff members prioritized archiving DAS as a core goal, to be pursued first before addressing the CAS as a "stretch goal" (#7 Red University, Library Staff).

One factors that contributed to the DAS becoming central to IIDC's efforts was the extant expertise and experience of IIDC members, for instance with handling the FITS data format. This expertise lent itself more readily to handling DAS than other elements of the SDSS Data Archive, as explained here:

> *"The database form is not really an easy form for us to manage because it's a particular technology that changes all the time whereas the data is in a format, the FITS format, that typically, with some other files floating around as well, that are flat files that can be ingested into an archive" (#7 Red University, Library Staff)*

Archiving DAS was attractive to the IIDC because it promised a relatively easy "win" compared with archiving other elements.

A second factor was that archiving DAS served IIDC's strategic objectives. The "archive" referred to in the quotation above is the GD infrastructure. IIDC envisioned that the Data Transfer Process would make a major contribution to the success of the coexisting project GD by helping to develop and showcase GD infrastructure:

> *"[The Data Transfer Process] became a natural basis upon which something like [Guardianship of Datasets] could be developed in order to understand the scientific data itself and what the system needed to be to curate and all that kind of stuff" (#27 Red University, Library Staff)*

Finally, the IIDC's organizational flexibility and its self-managing ethos also enabled it to prioritize archiving the DAS over the CAS. Although the MOU with SDSS assigned responsibility to IIDC for both DAS and CAS, the IIDC's organizational flexibility enabled it to choose to archive DAS first, before addressing CAS, reassigning staff from working on CAS to working on DAS as needed.

Implementing Data Access and Reuse Services

To illustrate the dynamics of how IIDC proceeded with developing services to support SDSS dataset users, this section focuses on the interface through which users would access data. IIDC decided that SDSS data would be accessible through the Guardianship of Datasets (GD) infrastructure's user interface and API, rather than through an interface tailored explicitly to SDSS data. This decision also reinforced IIDC's prioritization of the DAS over the CAS, as the GS's infrastructure would replace some of the functions of CAS for finding and accessing SDSS data.

Choosing to comingle Data Transfer activities and GD development served multiple purposes for the IIDC. Using GD infrastructure and resources for SDSS data access made economic sense for the IIDC. As one interviewee explained, "Most of what [SDSS] needs is encapsulated in the needs of [the Guardianship of Datasets] already" (#7 Red University, Library Staff). Leveraging GD infrastructure would avoid time-, labor-, and cost-intensive duplication of effort involved in building a separate infrastructure for SDSS data access.

This decision also promised IID the possibility of helping to secure grant money necessary to support its operations. IIDC leaders believed that integrating Data Transfer activities into GD project strengthened prospects of securing ongoing GD funding:

> *"The fact that it was such an important scientific dataset in the community…the SDSS dataset was really one of the things underpinning the proposal for [Guardianship of Datasets] itself in terms of pre-existing data that required curation" (#27 Red University, Library Staff)*

Another way in which IIDC leaders saw the transfer as potentially serving IIDC's strategic objectives was developing data infrastructure that could then be used to attract future "customers" (a term employed by several IIDC interviewees) for its data curation services. These customers were anticipated to come from a range of research fields, and even from outside academe. The IIDC saw the Data Transfer Process as an opportunity to develop infrastructure suitable for multiple domains and types of data. Hence, IIDC team members "went into [the Transfer Process] very much with the sense of we're not trying to build an astronomy data archive. We're trying to build a general data archive" (#21 Red University, Library Staff).

Finally, using the Data Transfer as an opportunity to support GD work aligned well with the R&D orientation and mission of the IIDC, given GD's status as a groundbreaking project at the bleeding edge of data infrastructure research. In this way, promoting the success of GD was central to the IIDC's mission.

Beyond the SDSS Data Transfer Process

As the end of Data Transfer Process approached in March 2013, both libraries had made progress towards fulfilling the terms of their respective Memoranda of Understanding (MOUs) with ARC/SDSS. Blue University library had fully archived the Administrative Archive and the Help Desk, and was serving the Catalog Archive Server (CAS). Red University library's IIDC had partially completed what they had set out to achieve. They had archived the raw data and software. However, they had not started to serve the DAS. To serve the DAS, they had become reliant on the Guardianship of Datasets (GD) project to provide the requisite infrastructure. However, GD was unexpectedly not funded to completion.

SDSS had secured funding for its third phase, SDSS-III, extending its lifespan beyond SDSS-I/II and enabling SDSS to continue serving its own data. (this perspective is discussed more fully in Darch et al. (In Review)). Following conclusion of the MOU, Blue University library stopped serving the CAS; instead astronomers now primarily access SDSS data via SDSS's own infrastructure. However, Blue University library committed to retaining a copy of the DAS, and to preserving the Administrative Archive in perpetuity. The library secured funding for this continuing SDSS work from a university foundation (#23, Blue University, Library Staff, unrecorded interview). Red University IIDC's engagement with the SDSS data is ongoing. The team secured funding from their university to continue archiving the raw data and software and the DAS until at least 2020 (#21, Red University, Library Staff). The outcomes of the Transfer Process, and the perspectives of the libraries and the astronomers on these outcomes, are developed in a separate paper (Darch et al., In Review).

Discussion

The study of the SDSS Data Transfer Process demonstrates that how libraries handle research datasets can differ markedly. First, even where libraries ostensibly set out to perform the same curation activity, they can interpret and implement this activity in very different ways. Second, how a library implements its curatorial activities, and the resultant outcomes, are shaped by multiple deep-rooted cultural, organizational, and infrastructural factors specific to that library. Third, libraries' curatorial actions open up some possibilities for future use and reuse, and restrict others.

Why the Libraries Diverged: An Infrastructural Perspective

Both libraries carried out similar activities, such as selection of materials for archiving and serving; transfer of data; storage; and development of associated services to support the discovery, retrieval, and use of data. However, differences emerged in how the libraries interpreted and carried out many of these activities. Each library had distinctive institutional characteristics that intersected to shape how each library interpreted and implemented the curatorial tasks. These institutional characteristics each generalize beyond the case presented here to other research libraries.

Before the transfer, each library had already developed extensive infrastructure to support its activities. This infrastructure is both technical and human (Lee et al., 2006). Each library both built upon, and worked within the constraints imposed by, this installed base (Star & Ruhleder, 1996) of infrastructure. Extant infrastructure, in turn, can be broken down into multiple components. The components described below each played a significant role in shaping the outcome of the Data Transfer Process. Other studies have already shown that libraries leverage and adapt existing technical and human infrastructure to develop services for research datasets (Antell et al., 2014; Steinhart, 2014). Our study has identified other components that also shape – for good and for ill – subsequent developments, notably a library's mission, values, and internal culture; and its organizational structure.

Technical Infrastructure and Staff Expertise

Other studies of libraries developing digital services for research data curation have emphasized how libraries build upon extant technical infrastructure, staff expertise, and precursor efforts (Antell et al., 2014; Steinhart, 2014). The SDSS Data Transfer Process demonstrates how the libraries' installed bases also imposes considerable constraints on development of data curation services.

For instance, Blue University's implementation of the Help Desk leveraged its extant infrastructure. The library was able to build upon its extant technical infrastructure (its system for handling library user queries) and human infrastructure (staff expertise in queuing and triaging library user queries). However, extant human infrastructure (e.g., the staff's lack of astronomy knowledge) also constrained the Help Desk's operations. Meanwhile, Red University leveraged other technical infrastructure (using Guardianship of Datasets infrastructure to ingest and provide accessibility to SDSS data), but also ran up against delays in building this infrastructure.

Organizational Structure

The university libraries' approaches also differed regarding how each allocated Transfer Process tasks internally and the effects of organizational structure on how these tasks were accomplished. Blue library distributed tasks across multiple units, and the Data Transfer Process was to proceed with minimal disruption to the library's organizational structure. At Red library, all Transfer Process activities were the purview of the IIDC, whose team had a flexible organizational structure that allowed members to select and prioritize the activities they performed.

Mission, Values, and Internal Culture

The units within each library charged with carrying out the Transfer Process represented two distinctive sets of missions, values, and culture. One set, in the units within Blue University library where Transfer Process work was carried out, was a service orientation emphasizing service to the needs of prospective SDSS data users and of the university in which the library was embedded. The other set, in Red library's IIDC, was a research and development orientation that emphasized pursuing novel solutions to digital curation challenges.

In general, the values embedded in the installed base shapes values in the infrastructure and services layered on top (Shilton, 2015). In the cases presented here, mission, values, and internal culture form part of the libraries' installed base and thus shaped each library's data curation activities. Star and Ruhleder (1996) identified *links with convention of practice* as a dimension of infrastructure. During recruitment processes, the library can choose staff whose practices already align with the library's existing values and mission. After recruitment, staff members become further socialized into the library's culture. For instance, in the case of Blue University, many staff members are librarians, graduating from degree programs that socialize students into library norms. On the job, these staff members become imbued with the library's service orientation to its users and the University. When these staff members came to work on the Data Transfer Process, they were already well-socialized into the library's culture, which accordingly shaped their decisions about how to implement the transfer.

Many university libraries display a mix of these values. All university libraries have an explicit service orientation. In addition, some libraries are ramping up their R&D activities (Steinhart, 2014). Differences between libraries in the mix of these values will give rise to differences in how they implement data curation services, with consequences for the outcomes of their curatorial activities.

Divergent Prospects for Dataset Reuse

Each library curated distinctive sets of materials, which afforded different possibilities for reuse. Datasets can be reused for multiple purposes (Pasquetto et al., 2017). Blue University's decision to curate the Administrative Archive affords historians and sociologists of science opportunities to use qualitative methods of textual analysis to better understand the social dynamics of topics such as scientific collaboration. Meanwhile, Red University IIDC's inclusion of raw SDSS data and processing software in its activities affords a very different group of scholars, e.g. astronomers who use computationally-intensive methods, the opportunity to reproduce the steps taken in processing raw data to produce the datasets released by the SDSS collaboration.

Conclusions

The study of the SDSS Data Transfer Process reveals some of the challenges of establishing and implementing data curation processes in university libraries. The two libraries' paths diverged from each other. The shift in context from the project where a dataset was initially curated to a library can change both the scholarly purposes that the dataset affords, and the scholarly communities who find the dataset useful. The effects of this shift in context on the perspectives of data producers and library staff are addressed in a separate paper (Darch et al., In Review).

These shifts in context, however, can also reveal new and unanticipated opportunities for dataset producers. In the case of SDSS, the decisions made by the libraries opened up new opportunities of potential benefit to the SDSS collaboration. One opportunity was facilitating greater transparency about the production of the SDSS-I/II data, both due to Blue University's curation of the Administrative Archive and Red University's curation of the raw data and pipeline processing software. Blue University's curation of the Administrative Archive also afforded historians of science the ability to study SDSS, thereby potentially helping to secure SDSS's position in the annals of astronomy history.

Cox et al. (2017) assessed the maturation of data curation services provided by different university libraries. The cases presented in this paper show that maturation does not necessarily imply convergence across libraries in terms of how they curate datasets and the resultant possibilities for dataset reuse. While the data curation services of the libraries involved in the Data Transfer were a long way from being mature, we saw how features of infrastructure – often locked-in many decades ago – profoundly shaped these libraries' subsequent decisions about how to develop infrastructure. Similarly, the maturation of other libraries' data curation infrastructure will involve building upon – and thus assuming the affordances and constraints of – their extant infrastructure.

The future of services provided by libraries that involve the handling and digital curation of research datasets is thus likely to involve a range of library infrastructures that differ in how they make datasets available and to whom. A straightforward solution to a dataset producer's long-term curation needs, but libraries with different approaches may complement each other well. The diversity of libraries' digital curation services poses many challenges, but this diversity can also be a source of value, ensuring the use and reuse of academic datasets in new and unexpected ways.

Acknowledgements

Funding from the National Science Foundation (award # 1145888) and the Alfred P. Sloan Foundation (awards #20113194, #201514001) supported this research. The UCLA Institutional Review Board, Study Protocol ID# 10-000909, provided oversight for this study. We thank L Wynholds and David S. Fearon for conducting early interviews. We thank Bernadette M. Boscoe, Irene V. Pasquetto, Michael J. Scroggins, and Sharon J. Traweek for comments on drafts. We also thank Sharon J. Traweek for mentorship and advice during the fieldwork. Finally, we are very grateful to those people we interviewed and observed in the institutions we studied for their time and efforts.

References

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., Allam, S. S., Prieto, C. A., An, D., … Zucker, D. B. (2009). The seventh data release of the Sloan Digital Sky Survey. *The Astrophysical Journal Supplement Series*, *182*(2), 543–558. https://doi.org/10.1088/0067-0049/182/2/543

Ahn, C. P., Alexandroff, R., Allende Prieto, C., Anderson, S. F., Anderton, T., Andrews, B. H., … Zinn, J. C. (2012). The Ninth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the SDSS-III Baryon Oscillation Spectroscopic Survey. *The Astrophysical Journal Supplement Series*, *203*(2), 21. https://doi.org/10.1088/0067-0049/203/2/21

Antell, K., Foote, J. B., Turner, J., & Shults, B. (2014). Dealing with data: Science librarians' participation in data management at Association of Research Libraries institutions. *College & Research Libraries*, *75*(4), 557–574.

Arora, R., Esteva, M., & Trelogan, J. (2014). Leveraging High Performance Computing for Managing Large and Evolving Data Collections. *International Journal of Digital Curation*, *9*(2), 17–27. https://doi.org/10.2218/ijdc.v9i2.331

Bishoff, C., & Johnston, L. (2015). Approaches to Data Sharing: An Analysis of NSF Data Management Plans from a Large Research University. *Journal of Librarianship and Scholarly Communication*, *3*(2), eP1231. https://doi.org/10.7710/2162-3309.1231

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.

Borgman, C. L., Darch, P. T., Sands, A. E., & Golshan, M. S. (2016). The durability and fragility of knowledge infrastructures: Lessons learned from astronomy. *Proceedings of the Association for Information Science and Technology*, *53*, 1–10. Retrieved from http://dx.doi.org/10.1002/pra2.2016.14505301057

Borgman, C. L., Darch, P. T., Sands, A. E., Pasquetto, I. V., Golshan, M. S., Wallis, J. C., & Traweek, S. (2015). Knowledge infrastructures in science: Data, diversity, and digital libraries. *International Journal on Digital Libraries*, *16*(3–4), 207–227. https://doi.org/10.1007/s00799-015-0157-z

Bowker, G. C. (2005). *Memory Practices in the Sciences*. Cambridge, Mass.: MIT Press. OCLC WorldCat FirstSearch (OCLC Number: 60776866).

Brinckman, A., Chard, K., Gaffney, N., Hategan, M., Jones, M. B., Kowalik, K., … Turner, K. (2019). Computing environments for reproducibility: Capturing the “Whole Tale.” *Future Generation Computer Systems*, *94*, 854–867. https://doi.org/10.1016/j.future.2017.12.029

Chao, T. C., Cragin, M. H., & Palmer, C. L. (2015). Data Practices and Curation Vocabulary (DPCVocab): An empirically derived framework of scientific data practices and curatorial processes. *Journal of the Association for Information Science and Technology*, *66*(3), 616–633. https://doi.org/10.1002/asi.23184

Coates, H. L. (2014). Building data services from the ground up: Strategies and resources. *Journal of EScience Librarianship*, *3*(1), 5.

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), 2182–2200. https://doi.org/10.1002/asi.23781

Darch, P. T., & Sands, A. E. (2016). It’s about time: How do sky surveys manage uncertainty about scientific needs many years into the future. *American Astronomical Society Meeting Abstracts# 228*, *228*.

Darch, P. T., Sands, A. E., Borgman, C., Golshan, M. S., & Traweek, S. (2017). From Sky to Archive: Long Term Management of Sky Survey Data. *American Astronomical Society Meeting Abstracts# 229*, *229*.

Darch, P. T., Sands, A. E., & Borgman, C. L. (In Review). Do the Stars Align?: Stakeholders and Strategies in Libraries’ Curation of an Astronomy Dataset. *Journal of the Association for Information Science and Technology*.

Edwards, P. N. (2010). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: The MIT Press.

Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., … Calvert, S. (2013). *Knowledge infrastructures: Intellectual frameworks and research challenges* (p. 40). Retrieved from University of Michigan website: http://hdl.handle.net/2027.42/97552

Fear, K., & Donaldson, D. R. (2012). Provenance and credibility in scientific data repositories. *Archival Science*, *12*(3), 319–339. https://doi.org/10.1007/s10502-012-9172-7

Finkbeiner, A. K. (2010, August 18). *The Sloan Digital Sky Survey: "A Grand and Bold Thing"* (N. Atkinson, Interviewer) [Universe Today]. Retrieved from http://www.universetoday.com/71445/the-sloan-digital-sky-survey-a-grand-and-bold-thing/

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago: Aldine Pub. Co.

Greenberg, J. (2017). Big Metadata, Smart Metadata, and Metadata Capital: Toward Greater Synergy Between Data Science and Metadata. *Journal of Data and Information Science*, *2*(3), 19–36. https://doi.org/10.1515/jdis-2017-0012

Hammersley, M., & Atkinson, P. (2007). *Ethnography: Principles in Practice* (3. ed., reprinted). Retrieved from https://www.routledge.com/products/9780415396059

Jackson, S. J. (2016). Speed, Time, Infrastructure. *The Sociology of Speed: Digital, Organizational, and Social Temporalities*, 169.

Johnston, L. R., Carlson, J. R., Hswe, P., Hudson-Vitale, C., Imker, H., Kozlowski, W., … Stewart, C. (2017). Data Curation Network: How Do We Compare? A Snapshot of Six Academic Library Institutions' Data Repository and Curation Services. *Journal of EScience Librarianship*, *6*(1), 3.

Kron, R. G., Gunn, J. E., Weinberg, D. H., Boroski, W. N., & Evans, M. L. (2008). *Final Report to the Alfred P. Sloan Foundation* (No. 2004-3–11).

Lee, C. P., Dourish, P., & Mark, G. (2006). The Human Infrastructure of Cyberinfrastructure. *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, 483–492. https://doi.org/10.1145/1180875.1180950

McDonald, N., Schoenebeck, S., & Forte, A. (2019). Reliability and Inter-rater Reliability in Qualitative Research: Norms and Guidelines for CSCW and HCI Practice. *Proceedings of the ACM on Human-Computer Interaction*, *3*(CSCW), 72.

O'Donoghue, T., & Punch, K. (2004). *Qualitative Educational Research in Action: Doing and Reflecting*. Retrieved from https://books.google.com/books?hl=en&lr=&id=2HBz4sWll8UC&oi=fnd&pg=PP1&dq=Qualitative+Educational+Research+in+Action:+Doing+and+Reflecting&ots=e4oZUFmsQu&sig=KOrVmI_9fE4heaZp5sZH1SRtVkM

Palmer, C. L., Weber, N. M., Munoz, T., & Renear, A. H. (2013). Foundations of Data Curation: The Pedagogy and Practice of "Purposeful Work" with Research Data. *Archive Journal*, *3*. Retrieved from http://www.archivejournal.net/issue/3/archives-remixed/foundations-of-data-curation-the-pedagogy-and-practice-of-purposeful-work-with-research-data/

Pasquetto, I., Randles, B., & Borgman, C. (2017). On the Reuse of Scientific Data. *Data Science Journal*, *16*. Retrieved from http://datascience.codata.org/articles/10.5334/dsj-2017-008/

Sands, A. E. (2017). *Managing Astronomy Research Data: Data Practices in the Sloan Digital Sky Survey and Large Synoptic Survey Telescope Projects* (Ph.D. Dissertation, UCLA). Retrieved from http://escholarship.org/uc/item/80p1w0pm

Sands, A. E., & Darch, P. T. (2016). What does it mean to manage sky survey data? A model to facilitate stakeholder conversations. *American Astronomical Society Meeting Abstracts# 228*, *228*.

SDSS Collaboration. (2016, February 3). Science Results | SDSS. Retrieved February 4, 2016, from http://www.sdss.org/science/

Shilton, K. (2015). Anticipatory ethics for a future Internet: Analyzing values during the design of an internet infrastructure. *Science and Engineering Ethics*, *21*(1), 1–18.

Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, *7*(1), 111–134. https://doi.org/10.1287/isre.7.1.111

Steinhart, G. (2014). An institutional perspective on data curation services. *Research Data Management: Practical Strategies for Information Professionals. Purdue University Press, West Lafayette, IN*, 303–323.

Szalay, A. S., Kunszt, P. Z., Thakar, A. R., Gray, J., & Slutz, D. (2000). The Sloan Digital Sky Survey and its archive. *Astronomical Data Analysis Software and Systems IX*, *216*, 405. San Francisco, Calif: Astronomical Society of the Pacific.

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), 4.

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.

Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., … Allard, S. (2017). Research data services in European academic research libraries. *Liber Quarterly*, *27*(1).

Wittenberg, J., & Elings, M. (2017). Building a Research Data Management Service at the University of California, Berkeley: A tale of collaboration. *IFLA Journal*, *43*(1), 89–97. https://doi.org/10.1177/0340035216686982

Wynholds, L. A. (2011). Linking to Scientific Data: Identity Problems of Unruly and Poorly Bounded Digital Objects. *International Journal of Digital Curation*, *6*(1), 214–225. https://doi.org/10.2218/ijdc.v6i1.183