Title

Collaborative Ethnography at Scale: Reflections on 20 years of data integration

Authors
Borgman, Christine L.
Wofford, Morgan F.
Golshan, Milena S.
et al.

Publication Date
2020-02-13

Collaborative Ethnography at Scale: Reflections on 20 years of Acquiring Global Data and Making Data Global

Christine L. Borgman, Christine.Borgman@UCLA.edu, Corresponding Author
Morgan F. Wofford, mfwofford@ucla.edu
Milena S. Golshan, milenagolshan@ucla.edu
Peter T. Darch, ptdarch@illinois.edu
Michael J. Scroggins, mscroggins@ucla.edu
Center for Knowledge Infrastructures, UCLA

Invited paper for STS Special Issue on *Ethnographic data generation in STS collaboration*
https://sciencetechnologystudies.journal.fi/announcement/view/232

Original Submission, Sept 30, 2019

This Version, resubmitted January 24, 2020, in response to editors' comments

Abstract

A 5-year STS project in geography, starting in 1999, evolved into 20 years of data collection about scientific data practices in sensor networks, environmental sciences, biology, seismology, undersea science, biomedicine, astronomy, and other fields. By emulating the 'team science' approaches of the scientists studied, the UCLA Center for Knowledge Infrastructures accumulated a comprehensive collection of qualitative data about how scientists generate, manage, use, and reuse data across domains. Building upon Paul N. Edwards's model of 'making global data' – collecting signals via consistent methods, technologies, and policies – to 'make data global' – comparing and integrating those data, the research team has managed and exploited these data as a collaborative resource. This article reflects on the social, technical, organizational, economic, and policy challenges the team has encountered in creating new knowledge from data old and new. We reflect on continuity over generations of students and staff, transitions between grants, transfer of legacy data between software tools, research methods, and the role of professional data managers in the social sciences.

Collaborative Data Practices

This Special Issue on "Ethnographic data generation in STS collaboration" invited STS researchers to reflect upon ways in which their own research echoes the collaborative practices they study. Our team, under one PI (Borgman), has conducted collaborative qualitative studies of scientific practices since the late 1990s, accumulating a rich trove of interviews, ethnographic notes, documents, and publications on studies of data practices in the physical and life sciences. As our focus has evolved from active data collection to consolidation of our findings, we write this article to reflect on our research methods, in theory and in practice, to offer lessons learned and guidance for others who may embark on similar journeys. We write in the first person, using "the royal we," to represent the many members of the research team who have conducted this body of research over a 20-year period. The authors of this article are current or very recent members of the UCLA Center for Knowledge Infrastructures and its predecessors. Earlier members of the team, and our collaborators at UCLA and other universities, are represented by references to publications and projects in which they participated.

What is now apparent as a 20-year project on scientific data practices began as a five-year (1999-2004) effort to study the design and use of digital libraries in physical geography, conducted in collaboration with geographers and computer scientists. That project, known as the Alexandria Digital Earth Prototype (ADEPT), was not designed to frame a longitudinal study that would span many scientific domains, field sites, and research questions. As our data and our findings accumulated, their collective value become apparent. While our findings on uses of ADEPT in physical geography are reported in numerous publications (Borgman et al., 2000, 2005; Borgman, Leazer, et al., 2004; Borgman, Smart, et al., 2004), the ethnographic notes, documentation, and interviews on which those papers are based languish in boxes of paper, printouts, and legacy formats such as cassette tapes. The materials enshrine the body of work, but cannot be readily repurposed or integrated with data collected in subsequent studies.

Mid-way into the ADEPT project, we became founding members of the Center for Embedded Networked Sensing (CENS), a National Science Foundation Science and Technology Center from 2002 to 2012 (Borgman et al., 2012, 2015; Borgman, Golshan, et al., 2016; Mayernik et al., 2013; Wallis et al., 2013a). Because our information-studies-based team was studying how CENS scientific teams collected and managed their data, we became more deliberate in managing our own data. We have digital records of our CENS data, along with codebooks for documenting them.

As we expanded from geography, environmental sciences, biology, and seismology into astronomy and astrophysics, undersea science, and the biomedical sciences in later years, with other grants, more collaborators, and more staff, our methods became more systematic – and more problematic (Borgman, 2019b; Borgman, Darch, et al., 2016; Darch & Borgman, 2016; Pasquetto et al., 2017).

Conducting each of these projects individually, and starting anew with data collection each time, would have been far simpler than combining them into a long-term research program that requires continuous data management. We experienced many of the data-handling problems encountered by our research participants, such as tradeoffs between resources spent on data management vs. new data collection, protecting data of dissertation-stage students vs. sharing data with our faculty collaborators, discontinuities in research questions between projects and sites, maintaining continuity in our research program while proposing innovative new directions to obtain new funding, recoding data to make comparisons, migrating data to new platforms, dealing with software and hardware upgrades, handoffs between personnel, and so on.

We discuss these challenges and tradeoffs, comparing our experiences with those of the scientists we study. As Paul N. Edwards (2010) learned in the climate science community, "making global data" is a prerequisite to "making data global." Global data are those collected in consistent forms, usually based on agreements of methods and measures, so that they can be combined or compared. Investments in making global data span the entire research life cycle, from research design to data reuse. To make data global, which is the process of comparing, combining, and integrating data for scientific purposes, requires data science expertise. Whereas the importance of data science expertise is now being recognized in scientific domains, the prerequisite skills in curation and stewardship necessary to make global data rarely are part of graduate training in the sciences or social sciences.

Research data are scientific assets that can be mined, combined, and bartered, but they also are liabilities. Maintaining and servicing data are continuing challenges for scientists and social scientists alike. The payoff for investing in data management is the ability to integrate data across projects to address larger research questions. After nearly 20 years of investing in global data, we are achieving the payoffs of making data global in our own research, while facing similar challenges in data integration and stewardship as those of our scientific research participants. This special issue is a timely opportunity to offer our reflections on these challenges.

Open Science, Data Reuse, and Knowledge Infrastructures

Open science policy, which includes open access to publications, data management plans, and data release with publications, is based on arguments for the value of replication, reproducibility, transparency, and reuse of research data for education and innovation (Borgman, 2015a; European Commission High Level Expert Group on Scientific Data, 2010; Networking and

Information Technology Research and Development, 2009; Organisation for Economic Co-operation and Development, 2007; U.S. National Science Board, 2005). However, releasing scientific data often creates large burdens on researchers due to the labor and expertise involved in managing, curating, documenting, and providing access to those data (Borgman, 2015b; Mayernik, 2016a; Mayernik et al., 2013; Mello et al., 2020; Pasquetto et al., 2017, 2015; Wallis et al., 2013b). Many scientists view data sharing and release as unfunded mandates. Thus, the larger questions that drive our research agenda are to identify where the value lies in data acquisition and reuse, how costs and benefits are distributed among the many stakeholders in those data resources, and the practices by which scientists steward their data.

Disciplines and Data

Some disciplines invest heavily in maintaining data resources, such as astronomy, genomics, seismology, and certain areas of the environmental sciences. Other disciplines are characterized by local data management, sometimes keeping samples and digital data indefinitely and sometimes discarding them after associated publications are released. Our most consistent finding about data practices across disciplines is heterogeneity. Individuals keep some kinds of data and discard others. Disciplinary repositories acquire some kinds of data and reject others (Borgman et al., 2019). Scale is also a factor. Larger teams, especially those that generate larger volumes or varieties of data, are better able to invest in data management. Identifying these patterns, and theorizing relationships among them, is central to our agenda.

Another STS finding of our research is that data practices are embedded in complex social and technological contexts. The theoretical lens through which we view scientific data practices is knowledge infrastructures, a term first coined by Edwards (2010, p. 17) as "robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds." Data practices can be studied at spatial, disciplinary, and temporal scales (Bowker et al., 2009; Edwards, 2003; Edwards et al., 2013; Ribes & Baker, 2007; Star & Ruhleder, 1994).

Interdependencies of institutions also arise, as do relationships between software, code, data, and tools. Infrastructures that may appear durable often are fragile upon closer inspection (Borgman, Darch, et al., 2016). The very idea of "data" is problematized throughout our research (Borgman, 2015a, 2019b; Leonelli, 2016, 2019). Whereas science policies tend to imply that data are simply "facts," or otherwise static and bounded objects, they are more commonly malleable, mobile, and mutable (Edwards et al., 2011; Latour, 1987; Leonelli, 2016). The ability to generate, use, and reuse data in these collaborative and interdisciplinary environments often requires "interactional expertise" in addition to domain knowledge and technical skills (Collins & Evans, 2007; Pasquetto et al., 2019).

Research Agenda

Our research agenda lies in Pasteur's Quadrant, that of "use-inspired basic research" (Stokes, 1997). As members of a professional school, we are acutely aware of the benefits of engaging with the communities we serve (*ISchools*, 2019). We partner with the research groups we study, reporting back periodically on our findings, and offering guidance on their data practices upon request. We also publish and give talks in these scientific communities (Borgman, 2017, 2018, 2019a, 2019c; Borgman & Pasquetto, 2018; Darch, 2017, 2018b; Pasquetto, 2019; Wallis et al., 2013a; Wofford et al., 2019). Our studies of scientific data practices began as a subcontract to ADEPT, a five-year (1999-2004) digital libraries research project on the use of a digital collection of physical geography content for teaching undergraduate courses. Collaborators on ADEPT, which was funded by the U.S. National Science Foundation (NSF), spanned geography, earth sciences, computer science, education, and psychology. Our role was to address questions of what data were useful to physical geographers in their teaching and research and the degree to which the digital library would address those needs. Generally, we found that these geography faculty much preferred to draw data for teaching examples from their own research, rather than seeking external data resources. They were more interested in the digital library to manage their own research data than for its intended instructional purposes (Borgman, 2004a, 2006, 2004b; Borgman et al., 2000, 2005, 2001; Borgman, Smart, et al., 2004; Champeny et al., 2004; D'Avolio et al., 2005; Gazan et al., 2003; Leazer, Gilliland-Swetland, & Borgman, 2000; Leazer, Gilliland-Swetland, Borgman, et al., 2000; Mayer et al., 2002).

Our research design from the ADEPT project laid the foundation for studying data practices in Center for Embedded Networked Sensing (CENS), an NSF Science and Technology Center from 2002 to 2012. CENS, with five participating universities and 300 collaborators at its peak, spanned computer science, engineering, biology, environmental science, seismology, medicine and health, and other areas. Findings from ten years with CENS span four doctoral dissertations (Mayernik, 2011; Pepe, 2010; Shilton, 2011; Wallis, 2012), two masters theses, and approximately 100 publications. Overall, we identified a complex array of practices for data management, sharing, and reuse; mixes of incentives, disincentives, costs, and benefits of investing in data that varied by domain and team; vastly different concepts of "data" within and between collaborating research teams; and mapped social and authorial networks of CENS members and their external collaborators.

Starting in 2008, overlapping with CENS, we began to study data practices in astronomy as part of another large NSF center with collaborators from multiple physical and biological sciences, computer science, social sciences, and education. We partnered with the Sloan Digital

Sky Survey (SDSS) as they sought ways to curate, manage, and maintain access to a massive data resource as they neared the end of their funding for data collection. Astronomy differs greatly in scientific practices, infrastructure, scale, and other features from those of our partners in ADEPT and CENS. We leveraged our SDSS research to partner with other sites in astronomy and astrophysics. Overall, we find that astronomy has the most integrated knowledge infrastructure of any domain we have studied, spanning observational data, bibliographic records, archives, thesauri, software, and other resources (Borgman, Darch, et al., 2016). Yet, they too struggle with many aspects of data collection, processing, management, and reuse of data (Borgman et al., 2015; Boscoe, 2019; Darch, Sands, & Borgman, In Review; Darch, Sands, Borgman, et al., In Review; Darch & Sands, 2017; Sands, 2017; Wofford et al., 2019).

Concurrent with the astronomy research, we began studying two other large distributed collaborations, each at the invitation of their investigators. The first was in undersea science, where ocean drilling ships acquired core samples for physical and biological research. This body of work builds upon our ecological studies of CENS, given the scientific commonalities, and on the astronomy research, given the large-scale infrastructure required (Darch, 2016, 2018a; Darch & Borgman, 2016). The second collaboration is biomedicine, where a distributed and multidisciplinary array of labs, in a hub and spokes model, shares data about craniofacial abnormalities. While the biomedical collaboration is the farthest afield scientifically from our other sites, it has yielded striking comparisons in areas of data generation, reuse, and information policy (Pasquetto, 2018; Pasquetto et al., 2017, 2019, 2015, 2016).

In sum, we are studying multiple knowledge infrastructures, each of which has many components, and relationships among those infrastructures. In domains such as astronomy, the community has funding and critical mass to maintain sophisticated infrastructures that span decades and countries (Borgman, 2015a; Borgman, Darch, et al., 2016). In domains such as undersea science, where data are sparse and disciplines are emergent, the community relies upon multiple infrastructures that are maintained by other stakeholders (Darch & Borgman, 2016). Domains such as environmental sciences and biomedicine fall somewhere in between, each able to build some portions of their own infrastructures and to rely on multiple infrastructures that are controlled by other stakeholders.

Investing in Data Assets

Acquiring and managing data in ways that they can be kept 'alive' for future reuse is a far different process than collecting data for a single grant project or a single dissertation in which data can be abandoned shortly after the publication of results. Commitments to data preservation pervade the process, from team building, research design, data collection, data management, and publication, to stewardship.

Conducting each of these projects individually, and starting anew with data collection each time, would have been far simpler than combining them into a long-term research program that required continuous data management. However, by investing in our data management, we gained opportunities to reflect on the data handling challenges of our research participants, and to construct more nuanced interpretations by comparing new and old findings continuously. The overhead is considerable, but necessary to study multiple knowledge infrastructures across many domains over long periods of time.

Following the work of Edwards (2010), we distinguish between the process of making global data and making data global. In his framing example, making global data is the process of developing technical, social, governmental, and policy agreements by which weather services around the world could collect data in consistent forms that could be shared. Standards were lubricants in this century-long process, but friction remains a constant (Edwards, 2010; Edwards et al., 2011). Making data global is the process of integrating those data into computer models that could be used to model, predict, and theorize weather and meteorology. We have made similar investments in making global data, albeit on a significantly smaller scale. Subsequent generations of CKI researchers are now able to make these data global through comparisons over time and across projects.

Acquiring Global Data

In our case, the process of acquiring global data on scientific data practices can be grouped into several stages. To design research programs that produce reusable data, a first step is to take a team science approach and a second step is to build effective teams. Thereafter it becomes possible to pursue data reuse and integration across projects.

Team Science

The research groups we study reflect typical models of team science, with most researchers organized into groups of two to ten individuals. Team science, a much-studied research topic, offers benefits by assembling complementary expertise to address complex problems, balanced with the costs of communication and coordination (Bos et al., 2007; Cooke et al., 2015; Finholt, 2002; Gorman, 2010; Hackman, 2011; Jirotka et al., 2013; Majchrzak et al., 2012; Meyer, 2007; O'Leary & Mortensen, 2010; G. M. Olson et al., 2008; J. S. Olson et al., 1993; Pepe, 2010; Shrum et al., 2007; Wagner, 2018) Among the features of team science that create challenges for research, as identified in a recent National Academies of Science (NAS) study, are high diversity of membership in terms of age, gender, culture, religion, or ethnicity; deep knowledge integration; and high task interdependence. Teams in all of our studies exhibited various combinations of these characteristics. The largest and most distributed teams we studied also exhibited the features the NAS study identified for larger teams, such as goal misalignment, permeable boundaries, and geographic dispersion (Cooke et al., 2015).

Team Building

The UCLA Center for Knowledge Infrastructures (CKI), as our team is now known, grew out of a CENS science team of three: a professor, a post-doctoral fellow, and a graduate student. Following the model of CENS teams, this group became known as 'the Borgman lab,' and later the 'data practices team,' in partnership with the CENS statistics group (Borgman, Mayernik, et al., 2009). Over the 20 years of research work discussed in this article, more than 20 people were part of the CKI team, such as PhD students, postdocs, professors, graduate student researchers, volunteers, and staff. We sometimes had joint grants with faculty at other universities, creating a much larger science team.

Thus, members of the CKI are a social science team that functions as a science team, with shared goals and infrastructure, collaborative writing practices, standing meetings every week, and joint responsibility for the data and other knowledge products we produce. Every grant proposal, paper, and talk is developed as a team and workshopped iteratively. Our practices reflect the benefits of integrating diverse knowledge and the overhead of coordination and building infrastructure.

Data Reuse Practices

Our team science practice is central to acquiring global data. Among the challenges of collaborative ethnographic research is that ethnographic methods tend to be highly personal, with one person developing relationships with research participants over long periods of time. Individual ethnographers are often highly proprietary about their methods, notes, recordings, transcripts, and other field data collection. Sharing those data with others who have not participated in the research requires considerable explanation of context (Mannheimer et al., 2018; Mauthner et al., 1998; Medjedović, 2011; Pasquetto et al., 2019).

We address these challenges by setting data reuse as a common goal for the team. As each new student or other collaborator joins the team, we establish expectations about data sharing within the team, while maintaining the confidentiality of our research participants. These are largely informal agreements, encoded in meeting notes but not subject to formal contracts. When students reach the stage of developing dissertation proposals and conducting their data collection and writing, we give them a proprietary period for sole use of their data until the dissertation is filed. Thereafter, those data, which were acquired under grants to the university, become part of the CKI pooled resources for comparative research. In most cases, those who collected the data participate in writing joint publications which result, receiving authorship credit accordingly. Most of our publications are joint-authored, whether comparing data from multiple sites or addressing themes such as data reuse.

In one large collaboration involving faculty collaborators from multiple universities, each of whom were employing students and post-doctoral fellows on the project, we agreed that individuals who conducted interviews would always receive acknowledgements, but not necessarily receive co-authorship credit. The collaboration was too large and the number of papers too many to bring everyone ever involved into the writing process. That model has worked well. Our publications always acknowledge funding sources and include mention of anyone whose interviews were used, which might sometimes be a single quotation, but who did not participate in writing that paper.

Our research participants faced similar challenges in acquiring data that would remain useful for their teams, and in distributing authorship credit (Borgman et al., 2012; Wallis, 2012;

Wallis & Borgman, 2011). Team science was the norm in the scientific groups we studied, which we guided our collaborative approaches.

Research Designs for Collecting Global Data

Collecting global data requires agreements on research designs that balance the need for continuity across grant projects and individual work such as dissertations. By anchoring our research designs in common protocols and human participants consent forms, we found that we could vary other aspects of investigations while creating a pool of data resources that could be reused by the CKI team in the future

Common Protocols

Our protocols for ethnographic observation, interviews, and document analyses began organically, designed for our early grant projects. These were largely constructed, tested, and implemented by two students whose dissertations addressed data practices in CENS (Mayernik, 2011; Wallis, 2012). Core questions about data collection, analysis, sharing, reuse, and management provided continuity in other CENS studies. With that anchor, we could pursue new avenues and nuanced aspects of prior findings (Borgman, Bowker, et al., 2009; Borgman et al., 2012, 2014, 2006; Borgman, Wallis, & Enyedy, 2007; Borgman, Wallis, Mayernik, et al., 2007; Mayernik et al., 2013; Wallis et al., 2013a, 2007).

Core protocols that were developed for CENS proved reasonably robust for application to astronomy, undersea science, and biomedicine. We developed complementary questions to explore the specifics of these domains. Each grant proposal and dissertation pursued new research questions, with our overarching questions remaining at the core of our inquiries. Similarly, our scientific research participants often pursued common goals throughout their careers, carving out pieces of the larger problem for individual grants and dissertations.

Informed Consent

To maintain access to our interviews, observations, documents, and other data from each project, we needed consent from the participants of our research. Working with our university's Institutional Review Board (IRB), we developed consent forms that asked participants for permission to reuse interview recordings, transcripts, and notes in subsequent research by the team. Research participants could opt out of allowing reuse, opt out of recording, and withdraw from the project at any time, but almost all of our participants granted permission for us to reuse their data in later studies. We promised confidentiality, following the usual IRB rules, and did not request permission to contribute the data to a public repository. In later studies, we asked both for consent to participate in the research and a "deed of gift" for the interview recording and transcript to ensure that these documents could be reused. The alternative, which we encountered in our later studies in biomedicine, is to "reconsent" participants to reuse their data in other projects. Locating, contacting, and getting permission from participants interviewed or observed years earlier is untenable. Rather, a broader initial consent process enhances the ability to acquire global data.

Neither audio recordings nor transcripts can be anonymized. We are studying well known scientific projects; others in the field could readily identify individuals if these materials were to be released. Thus, we struck a middle ground that maintained confidentiality while allowing us to reuse data for subsequent projects on related topics.

The human subjects research permissions granted by our IRB must be renewed annually, with associated reports on data collection and analysis, both retrospective and planned. If we were to allow those permissions to lapse, we would not be able to analyze data from prior projects.

Making Data Global

Our efforts to maintain our data for reuse also began organically. Open science was in its ascendancy in the early days of our research program, and data management was a new topic in the field of Information Studies, known as 'iSchools' (*ISchools*, 2019). The CKI team has consisted largely of information studies students, a field that addresses the collection, selection, organization, curation, preservation, and accessibility of information. This educational background gives iSchools an advantage in acquiring global data and especially in making those data global. As individuals without a background in information studies later joined the team, we trained them in data management skills and in the ethics of data sharing. Even with expertise in data management and STS, making our data global required dedicated effort.

Curation

Curation activities are necessary to make data global, by which we mean processes of standardizing data in ways that they can be integrated with other data into larger models (Edwards, 2010, Chapter 10). In scientific contexts, these are local processes to add metadata, map variant terms to common forms, organize files, store and migrate files, preserve and steward records, and other ways to add value for future use of the data. In the scientific teams we study, most of these curation activities are ad hoc, falling to individual graduate students or post-doctoral fellows who may have minimal (or no) training in data curation. The smaller the team, the more ad hoc the processes tend to be, as the few people involved are able to share their knowledge locally (Mayernik, 2011, 2016b, 2017; Mayernik & Acker, 2017).

Of the domains we have studied, astronomy has the most formalized processes of metadata creation, data reduction pipelines, and standard sets of analytical tools. In large astronomy endeavors, such as the Sloan Digital Sky Survey and the Large Synoptic Survey

Telescope, as much as half of the overall budget may be devoted to data management. We also found that data curation in small astronomy teams tends to fall to graduate students and post-doctoral fellows (Borgman, Darch, et al., 2016; Boscoe, 2019; Goodman et al., 2014; Pepe et al., 2014; Scroggins & Boscoe, In Review; Wofford et al., 2019).

Ad Hoc Data Curation

When the 'Borgman Lab,' precursor to the CKI, consisted of two to four individuals, we too took an informal approach to data curation. We stored files on multiple local computers for redundancy, stored paper records in file cabinets, and documented records as necessary for each publication. Each interviewer had full responsibility for transcription and annotation of interview records. After several years of observation, and our first large round of interviews, we had sufficient material that we needed codebooks and metadata to annotate our files. We chose NVIVO for qualitative data analysis, importing Word files, and marking up our growing collection of interviews, notes, and other records as NVIVO files. At the time, this software had the best functionality available, despite its limitations in exporting files, as discussed further below.

Scaling Up: Professional Data Management

As grants grew larger, we included data management responsibilities in the job duties of an individual. At first, we hired masters students in information studies at about 25% time to maintain our data resources. They had skills in metadata and records management, could process interview transcripts, keep track of records, and correspond with research participants to set up interviews and send corrected transcripts. Delegating data curation to part-time masters students sufficed for a few years, but lacked continuity as the team and the data corpus grew in size.

When an opportunity arose in late 2013 to reorganize our staffing, we hired a full-time data manager with an MLIS degree and background in the sciences.

An essential, but often under-appreciated role a data manager can play in long-term research projects is to maintain bibliographies. Over the course of 20 years, we have accumulated more than 10,000 references in Zotero (*Zotero*, 2019) that represent the bibliographies of all of our publications, including dissertations and books; references to articles, documents, and books relevant to our research; and documents related to the teams and individuals we study. We mine this rich resource continuously as we write new papers, and as we assemble our annual reports to our funding agencies.

Data management is a growth area for individuals with graduate degrees in library, information, archives, and related areas of study. Data curators are "care givers" who preserve data for future work while maintaining policies, procedures, and promises associated with the context of the research (Baker & Karasti, 2018; Darch & Borgman, 2016; Jørn Nielsen & Hjørland, 2014; Pryor & Donnelly, 2009; Puig de la Bellacasa, 2011; Swan & Brown, 2008; Yakel, 2007).

Our investment in a professional data manager accelerated our ability to acquire global data and to make those data global. By delegating core data curation tasks, plus initial corrections to transcripts and coding, our intent is to give other team members more time for field research, data analysis, and writing. However, removing the responsibility for creating transcripts and routing access to the data corpus through a data manager can add distance between researchers and their participants.

We address this gap between qualitative researchers and their data in several ways. Individual researchers return to their data by listening to recordings, reading transcripts, adding more metadata, and writing interpretive memos to suit their inquiries. We involve the data curator in the research process by participating in selected interviews and observations and in writing. Our data curator, having listened to audio recordings of interviews, cleaned transcripts to correct scientific terminology and idioms, and conducted initial data coding passes, has intimate knowledge of our data resources that transcends that of any individual on the team.

Data Integration Practices

Integrating and reusing data globally requires long-term active management. Even with extensive experience in data management, our researchers find qualitative datasets difficult to integrate, frequently requiring additional cleaning and management prior to comparative analysis. Reflecting back on how we addressed these challenges, our approaches fall into two categories: crosswalks and software tools.

Metadata Crosswalks

Our efforts to make global data began organically, tested through iterative efforts to reuse data as our research on scientific data practices evolved. We began with a method widely used in library and archival practice, which is to build 'crosswalks' between the metadata in each of our protocols (Getty Research Institute, 2020; Library of Congress, 2020). By comparing questions in different interview protocols, we could integrate our data descriptions into a common codebook, as shown in Figure 1. This continual reintegration reflected observations in our own research sites where teams actively manage historical data to work with newer data collected using modern methods (Boscoe, 2019).

A	B	C	D	E	F
Number	Topic	Code	Scope	Follow the Data/Dissertation Questions	CENS Wind-down Questions
0:01:00	File Attributes	Researcher	Participant name		
0:02:00	File Attributes	Researcher role	This attribute should describe the employment status of the participant, using one of the following terms to describe them: Faculty, Student, Postdoc, or Staff.		
0:03:00	File Attributes	Project	This attribute should be used to record the project with which the participant identifies.	What is the main research project you are working on now?	What research project at CENS are you workin on right now?
0:04:00	File Attributes	Domain	This attribute should be used to capture the domain with which the participant identifies	What type of research of research do you do?	
0:05:00	File Attributes	Interview date	Date of interview in YYYYMMDD		
0:06:00	File Attributes	Interviewers	Interviewers present for the interview		
1:01:00	Project	Project description	General description of the project, this should be a much thicker description of the project than File Attributes - Project	What is the main research project you are working on now?	What research project at CENS are you workin on right now?
1:02:00	Project	Research questions	This code can also be used to capture other discussion about the theories behind the researcher's current work or what they are trying to learn in their current work, as well as the long term goals.		What are you research questions for this project?
1:03:00	Project	Research contribution	This code should be used to denote the form of the contributions or "findings", or the product of scholarly activities. This could include a theory, publication, dataset, equipment, system, program, etc.		
1:04:00	Project	Project initiation	Stories about how a project emerged and evolved - who, when, where, why.	When did you begin this project?	How and when did you come to join CENS? Why did you decide to join CENS? What data, tools, students, or staff did you bring with you to CENS?
1:05:00	Project	Project duration	Capture any short-term and long-term projections of how long the project or overall research trajectory will last. For example, the CENS grants will end this year, but the research trajectory will probably take the next 20 years.		What is the expected duration of this project?

Figure 1: Crosswalk of two early CENS protocols.

Following our initial exploration into data reintegration through crosswalks, we continued to modify and supplement our codebooks and protocols with each additional infrastructure. The gradual process, transpiring over almost twenty years, is mapped in Figure 2. Rather than abandoning previous research questions, codebooks, or protocols, we adjusted each iteration to incorporate new research topics thus simplifying comparative analysis. When feasible, we reanalyzed earlier data with new research questions in mind.

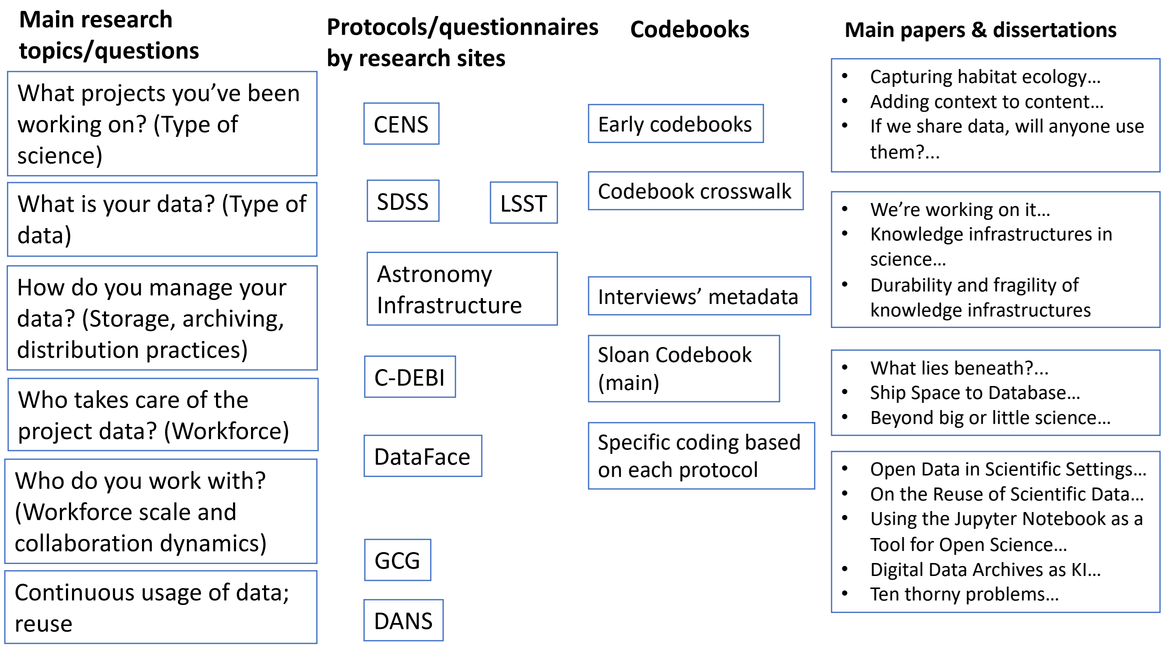

Figure 2: Mapping between research topics, protocols, codebooks, and papers.

Software Tools for Data Analysis

Building these crosswalks between protocols and mapping them to research questions, codebooks, and publications were essential steps in data integration, or making our data global. These steps were not sufficient, however, as our legacy files from NVIVO did not integrate easily. We encountered difficulties combining large files that were created in NVIVO versions spanning nearly 20 years, some on Apple and some on Windows computers. To be fair, NVIVO and most other qualitative software tools are designed for the canonical situation of one researcher on one project. Our files were large and heterogeneous, and our analytical goals were complex.

Our long-term solution to the legacy data problem was to start over, for the most part, with qualitative software that is better suited to collaborative research. As NVIVO did not have the capability to export our files with full analytical markup, we ingested our text files from

Word into Atlas.ti, creating a shared analysis file. Atlas.ti has two compelling features that facilitated making CKI data global. The first is a standardized XML format that makes sharing a single analytic file among CKI researchers feasible, and that allows us to export files to other analytical tools in the future. The second compelling feature of Atlas.ti is a tool derived from discourse analysis that supports more granular markup. The semantic linking features of Atlas.ti are proving helpful to explore commonalities and differences between activities like "maintenance," thus leveraging the entirety of the CKI data corpus (Scroggins & Pasquetto, 2020).

Keeping Data Alive

By keeping data "alive," or reusable over long periods of time, we are able to explore research questions at a much larger scale than would otherwise be possible. The ability to reuse data at scale is an inherent and underappreciated challenge of open science (Boscoe, 2019; Pasquetto et al., 2017, 2019). Here we summarize our lessons learned in collaborative ethnography at scale by identifying the analyses made possible by our approach and the challenges raised.

Curation and Continuity

Individual professors, as principal investigators, often have long-term research agendas that they pursue with cohorts of students and post-doctoral fellows. Maintaining continuity in the research agenda is difficult due to the high turnover of staff and the short term of research grants, typically one to three years in duration. Grant funding can be a precarious existence. Unless funding periods for staff overlap, one cohort leaves before the next arrives, leaving substantial gaps in team knowledge (Jackson et al., 2011).

The CKI team's investment in keeping our data alive has facilitated our continuity across cohorts and over long periods of time. In recent years, the data curator has trained each new graduate student and post-doctoral fellow in how to use our rich collection of data resources. As a consequence, new team members can build upon the work of prior staff. Alumni of the team also continue to collaborate, writing joint papers and helping to mine data they collected earlier.

Merging our dataset into a comprehensive database in Atlas.ti also has improved our data curation, because our data manager can focus on long-term preservation. Fewer intermediate steps are required than when each researcher managed a private analytic file in NVIVO to be merged later. Interview transcripts, memos, and other documents are more readily compared in our current analytical model. Working from a common analytical file also has created more robust conversations and more integrative analyses (Pasquetto et al., 2019).

Our challenges in continuity echo those of the teams we study. No matter how well documented, and how much knowledge is passed from one cohort to the next, the individual researchers who collected the data initially retain the deepest knowledge of context. As with our scientific teams, we contact our colleagues for further interpretation as needed, collaborating if substantial data integration will accomplish goals of mutual interest (Pasquetto et al., 2019).

New Grants, New Research Questions

Each new grant proposal must promise something new and innovative. Rarely can incremental funding be acquired successfully. Here the challenge is to propose new work that builds on the prior, without losing the continuity of the larger research agenda. We have focused on questions of data practices; research teams' abilities to share and reuse data; interactions between science policy and local practice; the concept of "data" as understood within and between scientific domains; and how knowledge infrastructures facilitate and constrain scientific work. This is a sufficiently broad agenda to allow us to ask new questions about how open science practices and policy play out in different domains, how old standards and new tools fit into knowledge infrastructures, and how infrastructures evolve and interact over time. We have sought funding from a wide range of sources, aligning our questions with their interests in individual sciences, in education, in infrastructure, in policy, and in scholarly communication.

As our curated data resources accumulate, we pitch them as a competitive advantage in seeking new funding. These resources also provide competitive advantage in hiring new graduate students and post-doctoral fellows. By joining our team, they have access to these data for use in constructing their own research agendas. We put graduate students into the field in their first year of study, which gives them at least two years of research experience by the time they begin their dissertations.

Diversity of Data Analysis

Maintaining a core set of research questions about data, data practices, and infrastructure across our projects and grants provides the continuity necessary to study knowledge infrastructures at scale. At the same time, we are careful not to be overly prescriptive in the details of study design or data analysis. Research questions in an individual grant are general enough to allow students considerable flexibility in pursuing their interests and following their instincts in data analysis. Some of our graduate students have relied more heavily on ethnographic observation and writing memos, some on interviews, and some on document analyses. The balance varies intentionally. Some do extensive coding in NVIVO or Atlas.ti, which provides the best records for further analyses. Others do basic coding with these tools, then print out sections for hand-coding with colored markers. The latter approach provides flexibility, but does not scale well beyond the space of a table or office floor.

As qualitative researchers steeped in grounded theory (Clarke, 2005; Glaser & Strauss, 1967), we encourage hypothesis building and testing, and iterative analyses. These can be done in many ways, some of which provide better record trails than others. For continuity purposes, we much prefer granular documentation. One of the factors distinguishing analytic practices is the amount of training in the information fields. Researchers with library and information science degrees (MLIS or equivalent) have spent more time on developing documentation, tables, protocols, and codebooks than have team members who came from technical or social science backgrounds. Librarians on the team more often ensured that their data were organized, well-described, stored in the secured CKI archive, and available for sharing with other team members. Before we hired a professional data curator, MLIS-trained researchers developed the Zotero library and maintained detailed bibliographies of CKI publications. CKI researchers without a background in the information field have required more encouragement to manage their data as a collective resource. We make sure to ingest their coded data and bibliographies before they leave our employ, usually when completing their degrees.

Our approach to managing our data falls between two extremes in the social sciences. One, more common to STS, is for individual students and post-doctoral fellows to maintain exclusive control over their data, not leaving copies behind for the supervisor or team. University practices vary widely in their expectations of students to have exclusive or non-exclusive rights over their knowledge products. The distinction may be a function of whether the research was conducted under grant funding or self-funded, and which open science policies may apply, whether by institution or government (Boulton et al., 2011, 2015).

At the other extreme are open science policies that promote transparency throughout the entire research process, from registering hypotheses to depositing datasets and code (Kidwell et al., 2016; Mancini et al., 2020; National Academies of Sciences, Engineering, and Medicine, 2018; Nosek et al., 2015; Nosek & Lakens, 2014; Schapira et al., 2019). These approaches are intended to increase reproducibility, accountability, and the ability to reuse data. Complete transparency of research is particularly problematic with qualitative human subjects research. Audio recordings and transcripts cannot be fully anonymized, especially for interviews conducted with well known science teams. Fully transparent approaches to open science also are controversial because of the resources required for investigators to comply with regulations, competitive advantages of researchers subject to different rules, and economic benefits that may accrue to external parties (Laine, 2017; Mirowski, 2018).

Conclusions

Twenty years of collaborative ethnography have enabled us to address big questions about how knowledge infrastructures develop, how they are used, when they are visible and when invisible, when they are robust, when they are fragile, and how they break down. We have learned that all infrastructures are fragile in the long run, no matter how robust they may appear in the present (Borgman, Darch, et al., 2016).

Similarly, "data" is the most complex concept in data science (Borgman, 2015a, 2019b). Throughout our research, we find that one person's signal is another's noise. Two researchers, working side by side, may not realize they have fundamentally different notions of essential variables such as "temperature," as we found in CENS (Borgman et al., 2012). An astronomy team that removed gas clouds from their images as part of their data reduction pipeline later hired a specialist in gas clouds. Instead of treating these gases as noise, they began to treat them as signals, offering new insights into their research program. Examples abound of nuanced interpretations of data and datasets; policies about openness, sharing, and reuse; and responsibilities for stewarding scientific knowledge products.

Data sharing and reuse, in turn, depend on the availability of knowledge infrastructures, on characteristics of the research domain, and on how competing stakeholders implement (or not) such policies. Throughout our research, we have attempted to identify factors that distinguish data practices, whether by domain, discipline, institution, career stage, scale, temporality, or public policies. We are often asked to compare practices by these or other factors, whether by funding agencies, policy makers, reviewers, or audiences at public talks. The larger the corpus of data practices and infrastructures we build, the more nuanced our conclusions become. Our efforts to distinguish data handling practices within a distributed interdisciplinary collaboration, for example, revealed that each individual researcher claimed multiple areas of disciplinary expertise, thus defying categorization. Of particular interest was how practices evolved as these people worked together, learning from each other. The disciplinary and methods training each person brought to the lab became part of an emerging set of practices. Any attempt to characterize how a discipline handles data may be a snapshot in time; generalizations are fraught. Focusing on the larger knowledge infrastructures in which these practices occur gives us a broader understanding of scientists' experiences, technologies, and access to the resources necessary to manage their data.

Acknowledgements

Research reported here was supported in part by grants from the National Science Foundation (NSF) and the Alfred P. Sloan Foundation (Sloan): (1) The Center for Embedded Networked Sensing (CENS) was funded by NSF Cooperative Agreement #CCR-0120778; (2) CENSNet: An Architecture for Authentic Web-based Science Inquiry in Middle and High School, NSF #ESI-

0352572; (3) Towards a Virtual Organization for Data Cyberinfrastructure, NSF #OCI-0750529; (4) Monitoring, Modeling & Memory: Dynamics of Data and Knowledge in Scientific Cyberinfrastructures, NSF #0827322; (5) If Data Sharing is the Answer, What is the Question?, Sloan #2015-14001. CENS interviews reported here were conducted by Jillian C. Wallis, Matthew S. Mayernik, and Christine L. Borgman, with additional data analysis by Elizabeth Rolando. Bernadette Boscoe provided helpful commentary on earlier drafts. Joshua Greenberg made constructive suggestions on methodological issues to address. Most of all, we acknowledge the generosity of the scientific research teams we study for welcoming us into their communities as observers and for their thoughtful discussions of data practices.

References

Baker, K. S., & Karasti, H. (2018). Data Care and Its Politics: Designing for Local Collective Data Management As a Neglected Thing. *Proceedings of the 15th Participatory Design Conference: Full Papers - Volume 1*, 10:1–10:12. https://doi.org/10.1145/3210586.3210587

Borgman, C. L. (2004a). Science, Cyberinfrastructure, and Knowledge Communities: Leveraging Scientific Data for Educational Applications. *Science and Cyberinfrastructure Session, Society for Social Studies of Science (4S)*.

Borgman, C. L. (2006). What can Studies of e-Learning Teach us about Collaboration in e-Research? Some Findings from Digital Library Studies. *Computer Supported Cooperative Work*, *15*(4), 359–383. https://doi.org/10.1007/s10606-006-9024-1

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

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

Borgman, C. L. (2017, June 7). *Big Data, Little Data, or No Data? Knowledge Infrastructures for the Earth Sciences* [Keynote]. EarthCube 2017 All Hands Meeting, Seattle, WA. https://works.bepress.com/borgman/410/

Borgman, C. L. (2018, October 3). *The invisible knowledge infrastructure of astronomy: A sharper focus on blurry data* [Colloquium]. Seamless Astronomy Colloquium, Harvard-Smithsonian Center for Astrophysics. https://seamlessastronomy.org/event/seamless-astronomy-colloquium-invisible

Borgman, C. L. (2019a, June 18). *Why Data Sharing and Reuse in Biomedicine are Hard to Do?* A UC-wide Conference on Artificial Intelligence in Biomedicine, UCLA Luskin Conference Center. https://escholarship.org/uc/item/0g6240dz

Borgman, C. L. (2019b). The Lives and After Lives of Data. *Harvard Data Science Review*, *1*(1). https://doi.org/10.1162/99608f92.9a36bdb6

Borgman, C. L. (2019c, July 25). *Whose Science? Whose Data? Whose Evidence?* [Keynote]. Open Science and the Role of Common Evidence Sage Bionetworks 10th Anniversary Assembly, Seattle, WA. https://escholarship.org/uc/item/9qc8c2qh

Borgman, C. L. (2004b). *The Interaction of Community and Individual Practices in the Design of a Digital Library* (S. Sugimoto, Ed.). University of Tsukuba. http://www.kc.tsukuba.ac.jp/dlkc/e-proceedings/papers/dlkc04pp9.pdf

Borgman, C. L., Bowker, G. C., Finholt, T. A., & Wallis, J. C. (2009). Towards a Virtual Organization for Data Cyberinfrastructure. *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries*, 353–356. https://doi.org/10.1145/1555400.1555459

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. 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

Borgman, C. L., Darch, P. T., Sands, A. E., Wallis, J. C., & Traweek, S. (2014). The Ups and Downs of Knowledge Infrastructures in Science: Implications for Data Management. *Proceedings of the 2014 IEEE/ACM Joint Conference on Digital Libraries (JCDL)*, 257–266. https://doi.org/10.1109/JCDL.2014.6970177

Borgman, C. L., Gilliland-Swetland, A. J., Leazer, G. H., Mayer, R., Gwynn, D., Gazan, R., & Mautone, P. (2000). Evaluating Digital Libraries for Teaching and Learning in Undergraduate Education: A Case Study of the Alexandria Digital Earth ProtoType (ADEPT). *Library Trends*, *49*(2), 228–250.

Borgman, C. L., Golshan, M. S., Sands, A. E., Wallis, J. C., Cummings, R. L., Darch, P. T., & Randles, B. M. (2016). Data Management in the Long Tail: Science, Software, and Service. *International Journal of Digital Curation*, *11*(1), 128–149. https://doi.org/10.2218/ijdc.v11i1.428

Borgman, C. L., Leazer, G. H., Gilliland-Swetland, A. J., & Gazan, R. (2001). Iterative design and evaluation of a geographic digital library for university students: A case study of the Alexandria Digital Earth Prototype (ADEPT). *Research and Advanced Technology for*

Digital Libraries: 5th European Conference, ECDL 2001, Darmstadt, Germany, September 2001 : Proceedings, *2163*, 390–401. http://dx.doi.org/10.1007/3-540-44796-2_33

Borgman, C. L., Leazer, G. H., Gilliland-Swetland, A. J., Millwood, K. A., Champeny, L., Finley, J. R., & Smart, L. J. (2004). How geography professors select materials for classroom lectures: Implications for the design of digital libraries. *JCDL '04: Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries (Tucson, AZ, June 7-11, 2004)*, 179–185.

Borgman, C. L., Mayernik, M., Mayoral, K., Lukac, M., & Hansen, M. (2009, May 12). *Tools for Dynamic Deployment and Data Management*. https://escholarship.org/uc/item/4n6669gf

Borgman, C. L., & Pasquetto, I. V. (2018, September 15). *Big Data, Little Data, or No Data? Systematic Reviews in an Age of Open Data* [Keynote]. Cochrane Colloquium, Edinburgh. https://colloquium.cochrane.org/event/opening-keynote

Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, *70*(8), 888–904. https://doi.org/10.1002/asi.24172

Borgman, C. L., Smart, L. J., Millwood, K. A., Finley, J. R., Champeny, L., Gilliland, A. J., & Leazer, G. H. (2005). Comparing faculty information seeking in teaching and research: Implications for the design of digital libraries. *Journal of the American Society for Information Science and Technology*, *56*(6), 636–657. https://doi.org/10.1002/asi.20154

Borgman, C. L., Smart, L. J., Millwood, K. A., Finley, J. R., Champeny, L., Gilliland-Swetland, A. J., & Leazer, G. H. (2004, April). Getting ready for class: The information seeking behavior of geography professors in support of teaching. *Annual Meeting of the American Educational Research Association*.

Borgman, C. L., Wallis, J. C., & Enyedy, N. (2007). Little science confronts the data deluge: Habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries*, *7*(1–2), 17–30. https://doi.org/10.1007/s00799-007-0022-9

Borgman, C. L., Wallis, J. C., & Enyedy, N. D. (2006). Building Digital Libraries for Scientific Data: An Exploratory Study of Data Practices in Habitat Ecology. In J. Gonzalo, C. Thanos, M. F. Verdejo, & R. C. Carrasco (Eds.), *Proceedings of the 10th European Conference on Research and Advanced Technology for Digital Libraries: Vol. LINCS 4172* (pp. 170–183). Springer Berlin Heidelberg. https://doi.org/10.1007/11863878_15

Borgman, C. L., Wallis, J. C., & Mayernik, M. S. (2012). Who's got the data? Interdependencies in science and technology collaborations. *Computer Supported Cooperative Work*, *21*(6), 485–523. https://doi.org/10.1007/s10606-012-9169-z

Borgman, C. L., Wallis, J. C., Mayernik, M. S., & Pepe, A. (2007). Drowning in Data: Digital library architecture to support scientific use of embedded sensor networks. *Joint Conference on Digital Libraries*, 269–277. https://doi.org/10.1145/1255175.1255228

Bos, N., Zimmerman, A. S., Olson, J., Yew, J., Yerkie, J., Dahl, E., & Olson, G. M. (2007). From shared databases to communities of practice: A taxonomy of collaboratories. *Journal of Computer-Mediated Communication*, *12*, 563–582.

Boscoe, B. M. (2019). *From Blurry Space to a Sharper Sky: Keeping Twenty-Three Years of Astronomical Data Alive* [Ph.D. Dissertation, University of California, Los Angeles]. https://escholarship.org/uc/item/2jv941sb

Boulton, G., Babini, D., Hodson, S., Li, J., Marwala, T., Musoke, M. G. N., Uhlir, P. F., & Wyatt, S. (2015). *Open Data in a Big Data World: An International Accord* [Outcome of Science International 2015 meeting]. ICSU, IAP, ISSC, TWAS. https://twas.org/sites/default/files/open-data-in-big-data-world_short_en.pdf

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

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2009). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International Handbook of Internet Research* (pp. 97–117). Springer Netherlands. http://link.springer.com/10.1007/978-1-4020-9789-8_5

Champeny, L., Mayer, R. E., Johnson, R. A., Borgman, C. L., Leazer, G. H., Gilliland-Swetland, A. J., Millwood, K. A., D'Avolio, L., Finley, J. R., Smart, L. J., & Mautone, P. D. (2004). Developing a digital learning environment: An evaluation of design and implementation processes. *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries*, 37–46.

Clarke, A. (2005). *Situational analysis: Grounded theory after the postmodern turn*. SAGE Publications, Inc.

Collins, H. M., & Evans, R. (2007). *Rethinking Expertise*. University of Chicago Press.

Cooke, N. J., Hilton, M. L., Committee on the Science of Team Science, Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education, & National Research Council (Eds.). (2015). *Enhancing the Effectiveness of*

Team Science. The National Academies Press. http://www.nap.edu/catalog/19007/enhancing-the-effectiveness-of-team-science

Darch, P. T. (2016, March 15). Many methods, many microbes: Methodological diversity and standardization in the deep subseafloor biosphere. *IConference 2016 Proceedings*. iConference 2016, Philadelphia, PA. https://doi.org/10.9776/16246

Darch, P. T. (2017, December 13). *Data Friction, Social Friction: Standardization Challenges in Emerging Domains of Science*. American Geophysical Union Fall Meeting 2017, New Orleans, LA.

Darch, P. T. (2018a, March 27). *Limits to the Pursuit of Reproducibility: Emergent Data-Scarce Domains of Science*. iConference 2018, Sheffield, UK. https://doi.org/10.1007/978-3-319-78105-1_21

Darch, P. T. (2018b, December 12). *The Core Issue: Researchers' Strategies for Establishing Trust in Data Derived From Physical Samples in Multidisciplinary Earth Science*. American Geophysical Union Annual Fall Meeting, Washington, D.C. http://adsabs.harvard.edu/abs/2018AGUFMIN34A..04D

Darch, P. T., & Borgman, C. L. (2016). Ship space to database: Emerging infrastructures for studies of the deep subseafloor biosphere. *PeerJ Computer Science*, *2*, e97. https://doi.org/10.7717/peerj-cs.97

Darch, P. T., & Sands, A. E. (2017). Uncertainty about the Long-Term: Digital Libraries, Astronomy Data, and Open Source Software. *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 1–4. https://doi.org/10.1109/JCDL.2017.7991584

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*.

Darch, P. T., Sands, A. E., Borgman, C. L., Golshan, M. S., & Traweek, S. (In Review). Do The Stars Align?: Infrastructural Tensions In The Transfer Of An Astronomy Dataset To Research Libraries For Curation. *Journal of the Association for Information Science and Technology*.

D'Avolio, L., Borgman, C. L., Champeny, L., Leazer, G. H., Gilliland-Swetland, A. J., & Millwood, K. A. (2005). From Prototype to Deployable System: Framing the Adoption of Digital Library Services. *Annual Meeting of the American Society for Information Science & Technology*.

Edwards, P. N. (2003). Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems. In T. J. Misa, P. Brey, & A. Feenberg (Eds.), *Modernity and technology*. MIT Press.

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

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

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science Friction: Data, Metadata, and Collaboration. *Social Studies of Science*, *41*(5), 667–690. https://doi.org/10.1177/0306312711413314

European Commission High Level Expert Group on Scientific Data. (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data* [Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission]. European Union. https://www.fosteropenscience.eu/content/riding-wave-how-europe-can-gain-rising-tide-scientific-data

Finholt, T. A. (2002). Collaboratories. In B. Cronin (Ed.), *Annual Review of Information Science and Technology* (Vol. 36, pp. 73–107). Information Today.

Gazan, R., Leazer, G. H., Borgman, C. L., Gilliland-Swetland, A. J., Smart, L. J., Ancona, D., & Nilsson, R. M. (2003). Use Scenarios in the Development of the Alexandria Digital Earth Prototype (ADEPT). *Annual Meeting of the American Society for Information Science & Technology*, *40*, 407–415.

Getty Research Institute. (2020). *Introduction to Metadata: Crosswalk (Getty Research Institute).* https://www.getty.edu/research/publications/electronic_publications/intrometadata/crosswalks.html

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

Goodman, A., Pepe, A., Blocker, A. W., Borgman, C. L., Cranmer, K., Crosas, M., Di Stefano, R., Gil, Y., Groth, P., Hedstrom, M., Hogg, D. W., Kashyap, V., Mahabal, A., Siemiginowska, A., & Slavkovic, A. (2014). Ten Simple Rules for the Care and Feeding of Scientific Data. *PLoS Computational Biology*, *10*(4), e1003542. https://doi.org/10.1371/journal.pcbi.1003542

Gorman, M. E. (Ed.). (2010). *Trading Zones and Interactional Expertise: Creating New Kinds of Collaboration*. The MIT Press.

Hackman, J. R. (2011). *Collaborative Intelligence: Using Teams to Solve Hard Problems*. Berrett-Koehler Publishers.

ISchools. (2019). https://ischools.org/

Jackson, S. J., Ribes, D., Buyuktur, A., & Bowker, G. C. (2011). Collaborative Rhythm: Temporal Dissonance and Alignment in Collaborative Scientific Work. *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, 245–254. https://doi.org/10.1145/1958824.1958861

Jirotka, M., Lee, C. P., & Olson, G. M. (2013). Supporting Scientific Collaboration: Methods, Tools and Concepts. *Computer Supported Cooperative Work (CSCW)*, *22*(4–6), 667–715. https://doi.org/10.1007/s10606-012-9184-0

Jørn Nielsen, H., & Hjørland, B. (2014). Curating research data: The potential roles of libraries and information professionals. *Journal of Documentation*, *70*(2), 221–240. https://doi.org/10.1108/JD-03-2013-0034

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLOS Biol*, *14*(5), e1002456. https://doi.org/10.1371/journal.pbio.1002456

Laine, H. (2017). Afraid of Scooping ? Case Study on Researcher Strategies against Fear of Scooping in the Context of Open Science. *Data Science Journal*, *16*. https://doi.org/10.5334/dsj-2017-029

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard University Press.

Leazer, G. H., Gilliland-Swetland, A. J., & Borgman, C. L. (2000). Evaluating the use of a geographic digital library in undergraduate classrooms: The Alexandria Digital Earth Prototype (ADEPT). *Proceedings of the Fifth ACM Conference on Digital Libraries*, 248–249.

Leazer, G. H., Gilliland-Swetland, A. J., Borgman, C. L., & Mayer, R. E. (2000). Classroom evaluation of the Alexandria Digital Earth Prototype (ADEPT). In *ASIS 2000 Proceedings of the 63rd ASIS Annual Meeting, Chicago, IL, November 12-16, 2000* (Vol. 37, pp. 334–340). Information Today.

Leonelli, S. (2016). *Data-Centric Biology: A Philosophical Study*. University of Chicago Press. http://www.press.uchicago.edu/ucp/books/book/chicago/D/bo24957334.html

Leonelli, S. (2019). Data Governance is Key to Interpretation: Reconceptualizing Data in Data Science. *Harvard Data Science Review*, *1*(1). https://doi.org/10.1162/99608f92.17405bb6

Library of Congress. (2020). *LC Linked Data Service: Authorities and Vocabularies (Library of Congress)* [Webpage]. http://id.loc.gov/authorities/subjects/sh2010013649.html

Majchrzak, A., More, P. H. B., & Faraj, S. (2012). Transcending Knowledge Differences in Cross-Functional Teams. *Organization Science*, *23*(4), 951–970. https://doi.org/10.1287/orsc.1110.0677

Mancini, D., Lardo, A., & De Angelis, M. (2020). Efforts Towards Openness and Transparency of Data: A Focus on Open Science Platforms. In A. Lazazzara, F. Ricciardi, & S. Za (Eds.), *Exploring Digital Ecosystems* (pp. 67–84). Springer International Publishing.

Mannheimer, S., Pienta, A., Kirilova, D., Elman, C., & Wutich, A. (2018). Qualitative Data Sharing: Data Repositories and Academic Libraries as Key Partners in Addressing Challenges. *American Behavioral Scientist*, 0002764218784991. https://doi.org/10.1177/0002764218784991

Mauthner, N. S., Parry, O., & Backett-Milburn, K. (1998). The data are out there, or are they? Implications for archiving and revisiting qualitative data. *Sociology*, *32*(4), 733–745. https://www.cambridge.org/core/journals/sociology/article/data-are-out-there-or-are-they-implications-for-archiving-and-revisiting-qualitative-data/21E3E723BF31F65E96675BC7E5892C6B

Mayer, R. E., Smith, T. R., Borgman, C. L., & Smart, L. J. (2002). Digital libraries as instructional aids for knowledge construction. *Educational Technology*, *42*, 38–42.

Mayernik, M. S. (2011). *Metadata Realities for Cyberinfrastructure: Data Authors as Metadata Creators* [PhD Dissertation, UCLA]. http://dx.doi.org/10.2139/ssrn.2042653

Mayernik, M. S. (2016a). Research data and metadata curation as institutional issues. *Journal of the Association for Information Science and Technology*, *67*(4), 973–993. https://doi.org/10.1002/asi.23425

Mayernik, M. S. (2016b). Research data and metadata curation as institutional issues. *Journal of the Association for Information Science and Technology*, *67*(4), 973–993. https://doi.org/10.1002/asi.23425

Mayernik, M. S. (2017). Open data: Accountability and transparency. *Big Data & Society*, *4*(2), 205395171771885. https://doi.org/10.1177/2053951717718853

Mayernik, M. S., & Acker, A. (2017). Tracing the traces: The critical role of metadata within networked communications. *Journal of the Association for Information Science and Technology*, *Opinion Paper*. https://doi.org/10.1002/asi.23927

Mayernik, M. S., Wallis, J. C., & Borgman, C. L. (2013). Unearthing the Infrastructure: Humans and Sensors in Field-Based Research. *Computer Supported Cooperative Work*, *22*(1), 65–101. https://doi.org/10.1007/s10606-012-9178-y

Medjedović, I. (2011). Secondary Analysis of Qualitative Interview Data: Objections and Experiences. Results of a German Feasibility Study. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *12*(3). http://www.qualitative-research.net/index.php/fqs/article/view/1742

Mello, M. M., Triantis, G., Stanton, R., Blumenkranz, E., & Studdert, D. M. (2020). Waiting for data: Barriers to executing data use agreements. *Science*, *367*(6474), 150–152. https://doi.org/10.1126/science.aaz7028

Meyer, E. T. (2007). Technological Change and the Form of Science Research Teams: Dealing with the Digitals. *Prometheus*, *25*(4), 345–361. https://doi.org/10.1080/08109020701689193

Mirowski, P. (2018). The future(s) of open science. *Social Studies of Science*, *48*(2), 171–203. https://doi.org/10.1177/0306312718772086

National Academies of Sciences, Engineering, and Medicine. (2018). *Open Science by Design: Realizing a Vision for 21st Century Research*. The National Academies Press. https://doi.org/10.17226/25116

Networking and Information Technology Research and Development. (2009). *Harnessing the power of digital data for science and society*. Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137–141. https://doi.org/10.1027/1864-9335/a000192

O'Leary, M. B., & Mortensen, M. (2010). Go (Con)figure: Subgroups, Imbalance, and Isolates in Geographically Dispersed Teams. *Organization Science*, *21*(1), 115–131. https://doi.org/10.1287/orsc.1090.0434

Olson, G. M., Zimmerman, A. S., & Bos, N. (Eds.). (2008). *Scientific Collaboration on the Internet*. MIT Press. http://ieeexplore.ieee.org/xpl/bkabstractplus.jsp?bkn=6267421

Olson, J. S., Card, S. K., Landauer, T. K., Olson, G. M., Malone, T., & Leggett, J. (1993). Computer-Supported Cooperative Work—Research Issues for the 90s. *Behaviour & Information Technology*, *12*, 115–129.

Organisation for Economic Co-operation and Development. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding* (p. 24). Organisation for Economic Co-Operation and Development. http://www.oecd.org/dataoecd/9/61/38500813.pdf

Pasquetto, I. V. (2018, May 23). *From Open Data to Knowledge Production: Biomedical Data Sharing and Unpredictable Data Reuses* [Ph.D. Defense].

Pasquetto, I. V. (2019, March 23). Do Scientists Reuse Open Data? [Blog Post]. *Sage Bionetworks Critical Assessment of Open Science (CAOS)*. http://sagebionetworks.org/in-the-news/do-scientists-reuse-open-data/

Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2019). Uses and Reuses of Scientific Data: The Data Creators' Advantage. *Harvard Data Science Review*, *1*(2). https://doi.org/10.1162/99608f92.fc14bf2d

Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017). On the Reuse of Scientific Data. *Data Science Journal*, *16*. https://doi.org/10.5334/dsj-2017-008

Pasquetto, I. V., Sands, A. E., & Borgman, C. L. (2015). Exploring openness in data and science: What is "open," to whom, when, and why? *Proceedings of the Association for Information Science and Technology*, *52*, 1–2. https://doi.org/10.1002/pra2.2015.1450520100141

Pasquetto, I. V., Sands, A. E., Darch, P. T., & Borgman, C. L. (2016). Open Data in Scientific Settings: From Policy to Practice. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 1585–1596. https://doi.org/10.1145/2858036.2858543

Pepe, A. (2010). *Structure and Evolution of Scientific Collaboration Networks in a Modern Research Collaboratory* (http://ssrn.com/abstract=1616935) [Ph.D. Dissertation, University of California, Los Angeles]. SSRN. http://dx.doi.org/10.2139/ssrn.1616935

Pepe, A., Goodman, A., Muench, A., Crosas, M., & Erdmann, C. (2014). How Do Astronomers Share Data? Reliability and Persistence of Datasets Linked in AAS Publications and a

Qualitative Study of Data Practices among US Astronomers. *PLOS ONE*, *9*(8), e104798. https://doi.org/10.1371/journal.pone.0104798

Pryor, G., & Donnelly, M. (2009). Skilling Up to Do Data: Whose Role, Whose Responsibility, Whose Career? | International Journal of Digital Curation. *International Journal of Digital Curation*, *4*(2). https://doi.org/10.2218/ijdc.v4i2.105

Puig de la Bellacasa, M. (2011). Matters of care in technoscience: Assembling neglected things. *Social Studies of Science*, *41*(1), 85–106. https://doi.org/10.1177/0306312710380301

Ribes, D., & Baker, K. S. (2007). Modes of Social Science Engagement in Community Infrastructure Design. In *Proceedings of Third International Conference on Communities and Technology* (pp. 107–130). Springer.

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]. http://escholarship.org/uc/item/80p1w0pm

Schapira, M., The Open Lab Notebook Consortium, & Harding, R. J. (2019). Open laboratory notebooks: Good for science, good for society, good for scientists. *F1000Research*, *8*, 87. https://doi.org/10.12688/f1000research.17710.1

Scroggins, M. J., & Boscoe, B. M. (In Review). Once FITS, always FITS? Astronomical infrastructure in transition. *IEEE Annals of the History of Computing, Governance in the History of Computing*. https://escholarship.org/uc/item/6zc599vp

Scroggins, M. J., & Pasquetto, I. V. (2020). Labor Out of Place: On the Varieties and Valences of (In)visible Labor in Data-Intensive Science. *Engaging Science, Technology, and Society*. https://escholarship.org/uc/item/8wp6t6nt

Shilton, K. (2011). *Building Values into the Design of Pervasive Mobile Technologies* [Ph.D. Dissertation, University of California, Los Angeles]. http://ssrn.com/paper=1866783

Shrum, W., Genuth, J., & Chompalov, I. (2007). *Structures of Scientific Collaboration*. MIT Press.

Star, S. L., & Ruhleder, K. (1994). Steps towards an ecology of infrastructure: Complex problems in design and access for large-scale collaborative systems. *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, 253–264. https://doi.org/10.1145/192844.193021

Stokes, D. (1997). *Pasteur's Quadrant: Basic Science and Technological Innovation*. Brookings Institution Press.

Swan, A., & Brown, S. (2008). *The skills, role and career structure of data scientists and curators: An assessment of current practice and future needs* [Report to the JISC]. JISC. https://eprints.soton.ac.uk/266675/

U.S. National Science Board. (2005). *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century* (US NSF-NSB-05-40). National Science Foundation. https://www.nsf.gov/pubs/2005/nsb0540/

Wagner, C. S. (2018). *The Collaborative Era in Science*. Springer.

Wallis, J. C. (2012). *The Distribution of Data Management Responsibility within Scientific Research Groups* [PhD dissertation, UCLA]. https://dx.doi.org/10.2139/ssrn.2269079

Wallis, J. C., & Borgman, C. L. (2011). Who is Responsible for Data? An Exploratory Study of Data Authorship, Ownership, and Responsibility. *Annual Meeting of the American Society for Information Science & Technology*, *48*, 1–10. https://doi.org/10.1002/meet.2011.14504801188

Wallis, J. C., Borgman, C. L., Mayernik, M. S., Pepe, A., Ramanathan, N., & Hansen, M. A. (2007). Know Thy Sensor: Trust, Data Quality, and Data Integrity in Scientific Digital Libraries. *Proceedings of the 11th European Conference on Research and Advanced Technology for Digital Libraries*, *LINCS 4675*, 380–391. https://doi.org/10.1007/978-3-540-74851-9_32

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

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

Wofford, M. F., Boscoe, B. M., Borgman, C. L., Pasquetto, I. V., & Golshan, M. S. (2019). Jupyter notebooks as discovery mechanisms for open science: Citation practices in the astronomy community. *Computing in Science & Engineering*, 1–1. https://doi.org/10.1109/MCSE.2019.2932067

Yakel, E. (2007). Digital curation. *OCLC Systems & Services: International Digital Library Perspectives*, *23*(4), 335–340. https://doi.org/10.1108/10650750710831466

Zotero. (2019). https://www.zotero.org/

tween

Borgman, C. L. (2004a). Science, Cyberinfrastructure, and Knowledge Communities: Leveraging Scientific Data for Educational Applications. *Science and Cyberinfrastructure Session, Society for Social Studies of Science (4S)*. Presented at the Paris. Paris.

Borgman, C. L. (2004b, March 2). *The Interaction of Community and Individual Practices in the Design of a Digital Library* (S. Sugimoto, Ed.). Retrieved from http://www.kc.tsukuba.ac.jp/dlkc/e-proceedings/papers/dlkc04pp9.pdf

Borgman, C. L. (2006). What can Studies of e-Learning Teach us about Collaboration in e-Research? Some Findings from Digital Library Studies. *Computer Supported Cooperative Work*, *15*(4), 359–383. https://doi.org/10.1007/s10606-006-9024-1

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

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

Borgman, C. L. (2017, June). *Big Data, Little Data, or No Data? Knowledge Infrastructures for the Earth Sciences*. Keynote presented at the EarthCube 2017 All Hands Meeting, Seattle, WA. Retrieved from https://works.bepress.com/borgman/410/

Borgman, C. L. (2018, October). *The invisible knowledge infrastructure of astronomy: A sharper focus on blurry data*. Colloquium presented at the Seamless Astronomy Colloquium, Harvard-Smithsonian Center for Astrophysics. Retrieved from https://seamlessastronomy.org/event/seamless-astronomy-colloquium-invisible

Borgman, C. L. (2019a). The Lives and After Lives of Data. *Harvard Data Science Review*, *1*(1). https://doi.org/10.1162/99608f92.9a36bdb6

Borgman, C. L. (2019b, June). *Why Data Sharing and Reuse in Biomedicine are Hard to Do?* Presented at the A UC-wide Conference on Artificial Intelligence in Biomedicine, UCLA Luskin Conference Center. Retrieved from https://escholarship.org/uc/item/0g6240dz

Borgman, C. L. (2019c, July). *Whose Science? Whose Data? Whose Evidence?* Keynote presented at the Open Science and the Role of Common Evidence Sage Bionetworks 10th Anniversary Assembly, Seattle, WA. Retrieved from https://escholarship.org/uc/item/9qc8c2qh

Borgman, C. L., Bowker, G. C., Finholt, T. A., & Wallis, J. C. (2009). Towards a Virtual Organization for Data Cyberinfrastructure. *Proceedings of the 9th ACM/IEEE-CS Joint Conference on Digital Libraries*, 353–356. https://doi.org/10.1145/1555400.1555459

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

Borgman, C. L., Darch, P. T., Sands, A. E., Wallis, J. C., & Traweek, S. (2014). The Ups and Downs of Knowledge Infrastructures in Science: Implications for Data Management. *Proceedings of the 2014 IEEE/ACM Joint Conference on Digital Libraries (JCDL)*, 257–266. https://doi.org/10.1109/JCDL.2014.6970177

Borgman, C. L., Gilliland-Swetland, A. J., Leazer, G. H., Mayer, R., Gwynn, D., Gazan, R., & Mautone, P. (2000). Evaluating Digital Libraries for Teaching and Learning in Undergraduate Education: A Case Study of the Alexandria Digital Earth ProtoType (ADEPT). *Library Trends*, *49*(2), 228–250.

Borgman, C. L., Golshan, M. S., Sands, A. E., Wallis, J. C., Cummings, R. L., Darch, P. T., & Randles, B. M. (2016). Data Management in the Long Tail: Science, Software, and Service. *International Journal of Digital Curation*, *11*(1), 128–149. https://doi.org/10.2218/ijdc.v11i1.428

Borgman, C. L., Leazer, G. H., Gilliland-Swetland, A. J., & Gazan, R. (2001). Iterative design and evaluation of a geographic digital library for university students: A case study of the Alexandria Digital Earth Prototype (ADEPT). *Research and Advanced Technology for Digital Libraries: 5th European Conference, ECDL 2001, Darmstadt, Germany, September 2001 : Proceedings*, *2163*, 390–401. Retrieved from http://dx.doi.org/10.1007/3-540-44796-2_33

Borgman, C. L., Leazer, G. H., Gilliland-Swetland, A. J., Millwood, K. A., Champeny, L., Finley, J. R., & Smart, L. J. (2004). How geography professors select materials for classroom lectures: Implications for the design of digital libraries. *JCDL '04: Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries (Tucson, AZ, June 7-11, 2004)*, 179–185. New York: Association for Computing Machinery.

Borgman, C. L., Mayernik, M., Mayoral, K., Lukac, M., & Hansen, M. (2009, May). *Tools for Dynamic Deployment and Data Management*. Retrieved from https://escholarship.org/uc/item/4n6669gf

Borgman, C. L., & Pasquetto, I. V. (2018, September). *Big Data, Little Data, or No Data? Systematic Reviews in an Age of Open Data*. Keynote presented at the Cochrane Colloquium, Edinburgh. Retrieved from https://colloquium.cochrane.org/event/opening-keynote

Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, *70*(8), 888–904. https://doi.org/10.1002/asi.24172

Borgman, C. L., Smart, L. J., Millwood, K. A., Finley, J. R., Champeny, L., Gilliland, A. J., & Leazer, G. H. (2005). Comparing faculty information seeking in teaching and research: Implications for the design of digital libraries. *Journal of the American Society for Information Science and Technology*, *56*(6), 636–657. https://doi.org/10.1002/asi.20154

Borgman, C. L., Smart, L. J., Millwood, K. A., Finley, J. R., Champeny, L., Gilliland-Swetland, A. J., & Leazer, G. H. (2004). Getting ready for class: The information seeking behavior of geography professors in support of teaching. *Annual Meeting of the American Educational Research Association*. Presented at the San Diego, CA. San Diego, CA: American Educational Research Association.

Borgman, C. L., Wallis, J. C., & Enyedy, N. (2007). Little science confronts the data deluge: Habitat ecology, embedded sensor networks, and digital libraries. *International Journal on Digital Libraries*, *7*(1–2), 17–30. https://doi.org/10.1007/s00799-007-0022-9

Borgman, C. L., Wallis, J. C., & Enyedy, N. D. (2006). Building Digital Libraries for Scientific Data: An Exploratory Study of Data Practices in Habitat Ecology. In J. Gonzalo, C. Thanos, M. F. Verdejo, & R. C. Carrasco (Eds.), *Proceedings of the 10th European Conference on Research and Advanced Technology for Digital Libraries*: *Vol. LINCS 4172* (pp. 170–183). https://doi.org/10.1007/11863878_15

Borgman, C. L., Wallis, J. C., & Mayernik, M. S. (2012). Who's got the data? Interdependencies in science and technology collaborations. *Computer Supported Cooperative Work*, *21*(6), 485–523. https://doi.org/10.1007/s10606-012-9169-z

Borgman, C. L., Wallis, J. C., Mayernik, M. S., & Pepe, A. (2007). Drowning in Data: Digital library architecture to support scientific use of embedded sensor networks. *Joint Conference on Digital Libraries*, 269–277. https://doi.org/10.1145/1255175.1255228

Bos, N., Zimmerman, A. S., Olson, J., Yew, J., Yerkie, J., Dahl, E., & Olson, G. M. (2007). From shared databases to communities of practice: A taxonomy of collaboratories. *Journal of Computer-Mediated Communication*, *12*, 563–582.

Boscoe, B. M. (2019). *From Blurry Space to a Sharper Sky: Keeping Twenty-Three Years of Astronomical Data Alive* (UCLA). Retrieved from https://escholarship.org/uc/item/2jv941sb

Boulton, G., Babini, D., Hodson, S., Li, J., Marwala, T., Musoke, M. G. N., … Wyatt, S. (2015). *Open Data in a Big Data World: An International Accord* [Outcome of Science International 2015 meeting]. Retrieved from ICSU, IAP, ISSC, TWAS website: https://twas.org/sites/default/files/open-data-in-big-data-world_short_en.pdf

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

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2009). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International Handbook of Internet Research* (pp. 97–117). Retrieved from http://link.springer.com/10.1007/978-1-4020-9789-8_5

Champeny, L., Mayer, R. E., Johnson, R. A., Borgman, C. L., Leazer, G. H., Gilliland-Swetland, A. J., … Mautone, P. D. (2004). Developing a digital learning environment: An evaluation of design and implementation processes. *Proceedings of the 4th ACM/IEEE-*

CS Joint Conference on Digital Libraries, 37–46. Tucson, AZ: Association for Computing Machinery.

Clarke, A. (2005). *Situational analysis: Grounded theory after the postmodern turn*. Thousand Oaks, Calif.: SAGE Publications, Inc.

Collins, H. M., & Evans, R. (2007). *Rethinking Expertise*. Chicago: University of Chicago Press.

Cooke, N. J., Hilton, M. L., Committee on the Science of Team Science, Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education, & National Research Council (Eds.). (2015). *Enhancing the Effectiveness of Team Science*. Retrieved from http://www.nap.edu/catalog/19007/enhancing-the-effectiveness-of-team-science

Darch, P. T. (2016). Many methods, many microbes: Methodological diversity and standardization in the deep subseafloor biosphere. *IConference 2016 Proceedings*. Presented at the iConference 2016, Philadelphia, PA. https://doi.org/10.9776/16246

Darch, P. T. (2017, December). *Data Friction, Social Friction: Standardization Challenges in Emerging Domains of Science*. Presented at the American Geophysical Union Fall Meeting 2017, New Orleans, LA.

Darch, P. T. (2018a, March 27). *Limits to the Pursuit of Reproducibility: Emergent Data-Scarce Domains of Science*. Presented at the iConference 2018, Sheffield, UK. Retrieved from https://doi.org/10.1007/978-3-319-78105-1_21

Darch, P. T. (2018b, December). *The Core Issue: Researchers' Strategies for Establishing Trust in Data Derived From Physical Samples in Multidisciplinary Earth Science*. Presented at the American Geophysical Union Annual Fall Meeting, Washington, D.C. Retrieved from http://adsabs.harvard.edu/abs/2018AGUFMIN34A..04D

Darch, P. T., & Borgman, C. L. (2016). Ship space to database: Emerging infrastructures for studies of the deep subseafloor biosphere. *PeerJ Computer Science*, *2*, e97. https://doi.org/10.7717/peerj-cs.97

Darch, P. T., & Sands, A. E. (2017). Uncertainty about the Long-Term: Digital Libraries, Astronomy Data, and Open Source Software. *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, 1–4. https://doi.org/10.1109/JCDL.2017.7991584

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*.

Darch, P. T., Sands, A. E., Borgman, C. L., & Golshan, M. S. (In Review). Library Cultures of Data Curation: Adventures in Astronomy. *Journal of the Association for Information Science and Technology*.

D'Avolio, L., Borgman, C. L., Champeny, L., Leazer, G. H., Gilliland-Swetland, A. J., & Millwood, K. A. (2005). From Prototype to Deployable System: Framing the Adoption of Digital Library Services. *Annual Meeting of the American Society for Information Science & Technology*. Presented at the Charlotte, NC. Charlotte, NC.

Edwards, P. N. (2003). Infrastructure and Modernity: Force, Time, and Social Organization in the History of Sociotechnical Systems. In T. J. Misa, P. Brey, & A. Feenberg (Eds.), *Modernity and technology*. Cambridge, Mass: MIT Press.

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

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science Friction: Data, Metadata, and Collaboration. *Social Studies of Science*, *41*(5), 667–690. https://doi.org/10.1177/0306312711413314

European Commission High Level Expert Group on Scientific Data. (2010). *Riding the wave: How Europe can gain from the rising tide of scientific data* [Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission]. Retrieved from European Union website: https://www.fosteropenscience.eu/content/riding-wave-how-europe-can-gain-rising-tide-scientific-data

Finholt, T. A. (2002). Collaboratories. In B. Cronin (Ed.), *Annual Review of Information Science and Technology* (Vol. 36, pp. 73–107). Medford, NJ: Information Today.

Gazan, R., Leazer, G. H., Borgman, C. L., Gilliland-Swetland, A. J., Smart, L. J., Ancona, D., & Nilsson, R. M. (2003). Use Scenarios in the Development of the Alexandria Digital Earth Prototype (ADEPT). *Annual Meeting of the American Society for Information Science & Technology*, *40*, 407–415. Long Beach, CA: Information Today.

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

Goodman, A., Pepe, A., Blocker, A. W., Borgman, C. L., Cranmer, K., Crosas, M., … Slavkovic, A. (2014). Ten Simple Rules for the Care and Feeding of Scientific Data. *PLoS Computational Biology*, *10*(4), e1003542. https://doi.org/10.1371/journal.pcbi.1003542

Gorman, M. E. (Ed.). (2010). *Trading Zones and Interactional Expertise: Creating New Kinds of Collaboration*. The MIT Press.

Hackman, J. R. (2011). *Collaborative Intelligence: Using Teams to Solve Hard Problems*. Berrett-Koehler Publishers.

ISchools. (2019). Retrieved September 5, 2019, from https://ischools.org/

Jackson, S. J., Ribes, D., Buyuktur, A., & Bowker, G. C. (2011). Collaborative Rhythm: Temporal Dissonance and Alignment in Collaborative Scientific Work. *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, 245–254. https://doi.org/10.1145/1958824.1958861

Jirotka, M., Lee, C. P., & Olson, G. M. (2013). Supporting Scientific Collaboration: Methods, Tools and Concepts. *Computer Supported Cooperative Work (CSCW)*, *22*(4–6), 667–715. https://doi.org/10.1007/s10606-012-9184-0

Jørn Nielsen, H., & Hjørland, B. (2014). Curating research data: The potential roles of libraries and information professionals. *Journal of Documentation*, *70*(2), 221–240. https://doi.org/10.1108/JD-03-2013-0034

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., … Nosek, B. A. (2016). Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLOS Biol*, *14*(5), e1002456. https://doi.org/10.1371/journal.pbio.1002456

Laine, H. (2017). Afraid of Scooping ? Case Study on Researcher Strategies against Fear of Scooping in the Context of Open Science. *Data Science Journal*, *16*. https://doi.org/10.5334/dsj-2017-029

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Leazer, G. H., Gilliland-Swetland, A. J., & Borgman, C. L. (2000). Evaluating the use of a geographic digital library in undergraduate classrooms: The Alexandria Digital Earth Prototype (ADEPT). *Proceedings of the Fifth ACM Conference on Digital Libraries*, 248–249. San Antonio, Texas, June 2-7, 2000: New York: Association for Computing Machinery.

Leazer, G. H., Gilliland-Swetland, A. J., Borgman, C. L., & Mayer, R. E. (2000). Classroom evaluation of the Alexandria Digital Earth Prototype (ADEPT). In *ASIS 2000 Proceedings of the 63rd ASIS Annual Meeting, Chicago, IL, November 12-16, 2000* (Vol. 37, pp. 334–340). Medford, NJ: Information Today.

Leonelli, S. (2016). *Data-Centric Biology: A Philosophical Study*. Retrieved from http://www.press.uchicago.edu/ucp/books/book/chicago/D/bo24957334.html

Majchrzak, A., More, P. H. B., & Faraj, S. (2012). Transcending Knowledge Differences in Cross-Functional Teams. *Organization Science*, *23*(4), 951–970. https://doi.org/10.1287/orsc.1110.0677

Mancini, D., Lardo, A., & De Angelis, M. (2020). Efforts Towards Openness and Transparency of Data: A Focus on Open Science Platforms. In A. Lazazzara, F. Ricciardi, & S. Za (Eds.), *Exploring Digital Ecosystems* (pp. 67–84). Springer International Publishing.

Mannheimer, S., Pienta, A., Kirilova, D., Elman, C., & Wutich, A. (2018). Qualitative Data Sharing: Data Repositories and Academic Libraries as Key Partners in Addressing Challenges. *American Behavioral Scientist*, 0002764218784991. https://doi.org/10.1177/0002764218784991

Mauthner, N. S., Parry, O., & Backett-Milburn, K. (1998). The data are out there, or are they? Implications for archiving and revisiting qualitative data. *Sociology*, *32*(4), 733–745.

Mayer, R. E., Smith, T. R., Borgman, C. L., & Smart, L. J. (2002). Digital libraries as instructional aids for knowledge construction. *Educational Technology*, *42*, 38–42.

Mayernik, M. S. (2011). *Metadata Realities for Cyberinfrastructure: Data Authors as Metadata Creators* (PhD Dissertation, UCLA). Retrieved from http://dx.doi.org/10.2139/ssrn.2042653

Mayernik, M. S. (2016a). Research data and metadata curation as institutional issues. *Journal of the Association for Information Science and Technology*, *67*(4), 973–993. https://doi.org/10.1002/asi.23425

Mayernik, M. S. (2016b). Research data and metadata curation as institutional issues. *Journal of the Association for Information Science and Technology*, *67*(4), 973–993. https://doi.org/10.1002/asi.23425

Mayernik, M. S. (2017). Open data: Accountability and transparency. *Big Data & Society*, *4*(2), 205395171771885. https://doi.org/10.1177/2053951717718853

Mayernik, M. S., & Acker, A. (2017). Tracing the traces: The critical role of metadata within networked communications. *Journal of the Association for Information Science and Technology*, (Opinion Paper). https://doi.org/10.1002/asi.23927

Mayernik, M. S., Wallis, J. C., & Borgman, C. L. (2013). Unearthing the Infrastructure: Humans and Sensors in Field-Based Research. *Computer Supported Cooperative Work*, *22*(1), 65–101. https://doi.org/10.1007/s10606-012-9178-y

Medjedović, I. (2011). Secondary Analysis of Qualitative Interview Data: Objections and Experiences. Results of a German Feasibility Study. *Forum Qualitative Sozialforschung /*

Forum: Qualitative Social Research, *12*(3). Retrieved from http://www.qualitative-research.net/index.php/fqs/article/view/1742

Meyer, E. T. (2007). Technological Change and the Form of Science Research Teams: Dealing with the Digitals. *Prometheus*, *25*(4), 345–361. https://doi.org/10.1080/08109020701689193

Mirowski, P. (2018). The future(s) of open science. *Social Studies of Science*, *48*(2), 171–203. https://doi.org/10.1177/0306312718772086

National Academies of Sciences, Engineering, and Medicine. (2018). *Open Science by Design: Realizing a Vision for 21st Century Research*. Retrieved from The National Academies Press website: https://doi.org/10.17226/25116

National Science Board (U.S.). (2005). *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century* (No. US NSF-NSB-05-40). Retrieved from National Science Foundation website: https://www.nsf.gov/pubs/2005/nsb0540/

Networking and Information Technology Research and Development. (2009). *Harnessing the power of digital data for science and society*. Washington, D.C.: Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., … Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137–141. https://doi.org/10.1027/1864-9335/a000192

O'Leary, M. B., & Mortensen, M. (2010). Go (Con)figure: Subgroups, Imbalance, and Isolates in Geographically Dispersed Teams. *Organization Science*, *21*(1), 115–131. https://doi.org/10.1287/orsc.1090.0434

Olson, G. M., Zimmerman, A. S., & Bos, N. (Eds.). (2008). *Scientific Collaboration on the Internet*. Retrieved from http://ieeexplore.ieee.org/xpl/bkabstractplus.jsp?bkn=6267421

Olson, J. S., Card, S. K., Landauer, T. K., Olson, G. M., Malone, T., & Leggett, J. (1993). Computer-Supported Cooperative Work—Research Issues for the 90s. *Behaviour & Information Technology*, *12*, 115–129. (ISI:A1993LB69800006).

Organisation for Economic Co-operation and Development. (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding* (p. 24). Retrieved from Organisation for Economic Co-Operation and Development website: http://www.oecd.org/dataoecd/9/61/38500813.pdf

Pasquetto, I. V. (2018, May). *From Open Data to Knowledge Production: Biomedical Data Sharing and Unpredictable Data Reuses*. Ph.D. Defense presented at the UCLA. UCLA.

Pasquetto, I. V. (2019, March 23). Do Scientists Reuse Open Data? [Blog Post]. Retrieved April 1, 2019, from Sage Bionetworks Critical Assessment of Open Science (CAOS) website: http://sagebionetworks.org/in-the-news/do-scientists-reuse-open-data/

Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2019). Uses and Reuses of Scientific Data: The Data Creators' Advantage. *Harvard Data Science Review*.

Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017). On the Reuse of Scientific Data. *Data Science Journal*, *16*. https://doi.org/10.5334/dsj-2017-008

Pasquetto, I. V., Sands, A. E., & Borgman, C. L. (2015). Exploring openness in data and science: What is "open," to whom, when, and why? *Proceedings of the Association for*

Information Science and Technology, *52*, 1–2. https://doi.org/10.1002/pra2.2015.1450520100141

Pasquetto, I. V., Sands, A. E., Darch, P. T., & Borgman, C. L. (2016). Open Data in Scientific Settings: From Policy to Practice. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems,* 1585–1596. https://doi.org/10.1145/2858036.2858543

Pepe, A. (2010). *Structure and Evolution of Scientific Collaboration Networks in a Modern Research Collaboratory* (Ph.D. Dissertation, University of California, Los Angeles). Retrieved from http://dx.doi.org/10.2139/ssrn.1616935

Pepe, A., Goodman, A., Muench, A., Crosas, M., & Erdmann, C. (2014). How Do Astronomers Share Data? Reliability and Persistence of Datasets Linked in AAS Publications and a Qualitative Study of Data Practices among US Astronomers. *PLOS ONE*, *9*(8), e104798. https://doi.org/10.1371/journal.pone.0104798

Pryor, G., & Donnelly, M. (2009). Skilling Up to Do Data: Whose Role, Whose Responsibility, Whose Career? | International Journal of Digital Curation. *International Journal of Digital Curation*, *4*(2). https://doi.org/10.2218/ijdc.v4i2.105

Puig de la Bellacasa, M. (2011). Matters of care in technoscience: Assembling neglected things. *Social Studies of Science*, *41*(1), 85–106. https://doi.org/10.1177/0306312710380301

Ribes, D., & Baker, K. S. (2007). Modes of Social Science Engagement in Community Infrastructure Design. In *Proceedings of Third International Conference on Communities and Technology* (pp. 107–130). London: Springer.

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

Schapira, M., The Open Lab Notebook Consortium, & Harding, R. J. (2019). Open laboratory notebooks: Good for science, good for society, good for scientists. *F1000Research*, *8*, 87. https://doi.org/10.12688/f1000research.17710.1

Scroggins, M. J., & Boscoe, B. M. (In Review). Once FITS, always FITS? Astronomical infrastructure in transition. *IEEE Annals of the History of Computing*, *Governance in the History of Computing*. Retrieved from https://escholarship.org/uc/item/6zc599vp

Shilton, K. (2011). *Building Values into the Design of Pervasive Mobile Technologies* (Ph.D. Dissertation, University of California, Los Angeles). Retrieved from http://ssrn.com/paper=1866783

Shrum, W., Genuth, J., & Chompalov, I. (2007). *Structures of Scientific Collaboration*. Cambridge, Mass.: MIT Press.

Star, S. L., & Ruhleder, K. (1994). Steps towards an ecology of infrastructure: Complex problems in design and access for large-scale collaborative systems. *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, 253–264. https://doi.org/10.1145/192844.193021

Stokes, D. (1997). *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington, D.C.: Brookings Institution Press.

Swan, A., & Brown, S. (2008). *The skills, role and career structure of data scientists and curators: An assessment of current practice and future needs* [Report to the JISC]. Retrieved from JISC website: https://eprints.soton.ac.uk/266675/

Wagner, C. S. (2018). *The Collaborative Era in Science*. Springer.

Wallis, J. C. (2012). *The Distribution of Data Management Responsibility within Scientific Research Groups* (Ph.D. Dissertation). University of California, Los Angeles, Los Angeles, CA.

Wallis, J. C., & Borgman, C. L. (2011). Who is Responsible for Data? An Exploratory Study of Data Authorship, Ownership, and Responsibility. *Annual Meeting of the American Society for Information Science & Technology*, *48*, 1–10. https://doi.org/10.1002/meet.2011.14504801188

Wallis, J. C., Borgman, C. L., Mayernik, M. S., Pepe, A., Ramanathan, N., & Hansen, M. A. (2007). Know Thy Sensor: Trust, Data Quality, and Data Integrity in Scientific Digital Libraries. *Proceedings of the 11th European Conference on Research and Advanced Technology for Digital Libraries*, *LINCS 4675*, 380–391. https://doi.org/10.1007/978-3-540-74851-9_32

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

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

Wofford, M. F., Boscoe, B. M., Borgman, C. L., Pasquetto, I. V., & Golshan, M. S. (2019). Jupyter notebooks as discovery mechanisms for open science: Citation practices in the astronomy community. *Computing in Science & Engineering*, 1–1. https://doi.org/10.1109/MCSE.2019.2932067

Yakel, E. (2007). Digital curation. *OCLC Systems & Services: International Digital Library Perspectives*, *23*(4), 335–340. https://doi.org/10.1108/10650750710831466

Zotero. (2019). Retrieved September 16, 2019, from https://www.zotero.org/