Role of Communication in Data Reuse

Ayoung Yoon
Indiana University Purdue University Indianapolis (IUPUI)
535 W Michigan St. IT563
Indianapolis IN46202
ayyoon@iupui.edu

ABSTRACT

In acknowledging the potentials of existing data, researchers' interests in sharing and reusing data have recently emerged. However, sharing and reusing data is not a simple one-step process for researchers. Because data reusers build their work on other researchers' findings, the process of data reuse involves various interactions and communications with other relevant parties. Exploring the nature of communications around data is thus important to fully understand data reuse practices and to support smoother processes of data reuse. This study investigates communications occurring around data during data reusers' experiences through qualitative interview studies involving this group. This study's results show that the communications with different stakeholders mainly support data reuse in three areas: searching, learning, and problem solving. The findings provide valuable insights into the domain of scholarly communication, data reuse, and data services.

Keywords
data reuse, scholarly communication

INTRODUCTION

As data form the basis of scientific research, the importance of data and their roles in research has already been well recognized. Researchers conceive of the value of data as a way to address specific knowledge gaps, and valuable data help researchers answer their questions and/or test hypotheses (Akmon, 2014). In acknowledging the potentials of existing data, researchers' interests in sharing and reusing data have emerged. Existing research data can be used to not only validate current results but also to generate new findings built on the work of others (Birnholtz & Bietz, 2003; Borgman, 2011). Existing data can be

reanalyzed or combined with new data to verify published results or to arrive at new conclusions (Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age, 2009). Revolutions in the practices of science, known as data-intensive research or eScience, also support the need for sharing and reusing data (Kunze et al., 2011), as eScience promises discoveries and benefits that are not possible with more traditional methodologies.

However, sharing and reusing data is not a simple one-step process for researchers. Several studies have pointed out the difficulty of locating data for reuse, and understanding data is often a time-consuming process for data reusers, which sometimes lead to a failure in data reuse (e.g., Yoon, 2016a, 2017). Birnholtz and Bietz (2003) suggest that data are not simple carriers of meanings, and "converting raw data into scientific and social meaning is an active, context dependent process" (p. 341). Some studies have indicated the fundamental challenges of data reuse, showing the contextual nature of data from examinations of data reuse practices (e.g., Berg & Goorman, 1999; Zimmerman, 2007). Data reuse is not simply a transfer of numbers or images; it also entails the transfer of knowledge that is locally embedded during data creation. Additionally, documentation, which is one channel that is used to transfer contextual information about data, is often insufficient, which can make data reuse difficult (e.g., Fienberg, Martin, & Straf, 1985; McCall & Appelbaum, 1991; Niu & Hedstrom, 2008). Perhaps due to these challenges, the data reuse process is often not only an internal knowledge process for data reusers but also involves external communications and interactions with other researchers.

A dataset becomes an integral part of scholarly communication in data reuse, which is "a social process, in which scholars share ideas, data, resources, methodologies, innovations, and discoveries with their peers and with the public" (Martin, 2014, p. 53). Because data reusers build their work on other researchers' findings, the process of data reuse involves these interactions and communications with other relevant parties. Most commonly, previous research has found that data reusers contact data producers (the original investigators of the project in which the data for reuse were created) with questions (Niu, 2009; Yoon,

2016a, 2017). Sometimes mentors have been involved in locating and understanding data (Kriesberg, Frank, Faniel, & Yakel, 2013; Rolland & Lee, 2013).

These interactions and communications have been only sparsely examined in previous data reuse literature, as they naturally appeared as part of discussions about data reuse practices, but they have not been a major focus of past research. Understanding the nature of communications around data is important to fully understand data reuse practices and to support smoother processes of data reuse. This study investigates communications occurring around data during data reusers' experiences. The specific questions asked in this study are as follows:

- Who (what types of communities or individuals) is involved in communications with data reusers during their data reuse experiences?
- Why and how are data reusers interacting with those individuals or communities?

This study's findings provide valuable insights into the domain of scholarly communication and data reuse. Reusing data requires researchers' deep understanding of these data, and the process involves engaging in the different types of communications and interactions with relevant parties. By disclosing the nature of communications for reuse, this study offers insights into how to ease communications among researchers and how to support their reuse experiences.

LITERATURE REVIEW
Because data reuse practices are influenced by different disciplinary cultures and the processes for creating various types of data, the majority of this research has investigated data reuse practices and reusers' behaviors within specific disciplines, including cancer epidemiology research (Rolland & Lee, 2013), ecology (Borgman, Wallis, & Enyedy, 2007; Zimmerman, 2003), biological science (Chin & Lansing, 2004), environmental science (Van House, Butler, & Schiff, 1998), astronomy (Sands, Borgman, Wynholds, & Traweek, 2012), and earthquake engineering (Faniel & Jacobsen, 2010). Data reuse practices can be distinctive, depending on the types of data being reused, and some research has focused on certain types of data for reuse, such as quantitative (Faniel, Kriesberg, & Yakel, 2016) or qualitative data (e.g., Broom, Cheshire, & Emmison, 2009; Moore, 2007; Yoon, 2014). Several researchers have also conducted multidisciplinary studies, for instance, Birnholtz and Bietz's (2003) research on earthquakes, HIV/AIDS, and space physics, as well as Faniel, Barrera-Gomez, Kriesberg, and Yakel's (2013) investigation of quantitative social science, archaeology, and zoology, through the Dissemination Information Packages for Information Reuse (DIPIR) project from the University of Michigan and OCLC Research.

In addition to these efforts to understand data reuse practices in various disciplines, recent research has also investigated the criteria for data reuse from reusers' perspectives. Chin and Lansing (2004) present general data properties that are important to make data reusable, while Huang, Stvilia, and Jörgensen (2012) and Yoon (2016b) specifically address data quality dimensions. Faniel and Jacobsen (2010) discuss data reusability and point out relevancy, understandability, and trustworthiness as criteria. Adopting a theory of trust, Yoon's (2017) study also demonstrates users' criteria for judging the trustworthiness of data.

While different disciplines have various cultures and practices relating to data, existing research presents some common challenges in data reuse across disciplines. The major challenge is transferring information regarding the contexts of data. Many researchers have pointed out the contextual nature of data, as data are originally created for a specific purpose, with a local context embedded in them. To be reused beyond their original purpose, data are moved from the local sphere to the broader world (e.g., for interdisciplinary research), creating distance from their original context (Zimmerman, 2008). This distance poses problems when reusers try to understand how data were originally used (Cragin & Shankar, 2006; Faniel & Zimmerman, 2011; Jirotka et al., 2005; Zimmerman, 2008). Additionally, researchers often argue about the difficulty of deciding what contextual information is important for reusers beyond the original research purpose (Birnholtz & Bietz, 2003; Carlson & Anderson, 2007; Markus, 2001). Each reuser also has different technical skills and tacit knowledge about understanding data from other researchers and fields, which creates another layer of difficulty in understanding (Faniel & Zimmerman, 2011).

While documentation—which can include codebooks, related bibliographies, data-collection instruments, and metadata for resource discovery—is one channel for transferring contextual information to data reusers (Niu & Hedstrom, 2008), a number of researchers have also reported problems and challenges with it (Borgman, 2007; Corti, 2005; Fienberg et al., 1985; McCall & Appelbaum, 1991; Niu & Hedstrom, 2008; Zimmerman, 2003). Similar to the challenges of transferring contextual information embedded in data, documentation can be inherently insufficient because of the nature of tacit knowledge and communication reduction (Niu, 2009). People sometimes know things implicitly, and it is inevitable to miss information; not everything can be transferred (Carlson & Anderson, 2007; Niu, 2009; Niu & Hedstrom, 2008). On the other hand, documentation can be simply poor, as data producers might not want to prepare documentation well because of their lack of motivation or inadequate skills to do so (Baker & Yarmey, 2009).

Previous research findings also suggest active communications and interactions around data among various scholarly communities during the data reuse process. Besides the interactions with data producers, Zimmerman's (2007) study indicates that primary methods

of locating data for reuse include human interactions and communications from other scientists. Similarly, Sands and colleagues (2012) report informal or formal methods of communication involved in astronomers' data reuse processes. Communications regarding data appear important in transferring knowledge beyond locating the data. Niu (2009) points out interactive conversation as an important channel for transferring knowledge when data reuse is not supported by adequate documentation. Both Rolland and Lee (2013) and Kriesberg et al. (2013) discuss mentors' roles in data reuse, helping novice researchers understand data. While these studies imply rich communication and interaction occurring around data for reuse, such communication has not been a main focus of previous literature. Concentrating on the nature of communication around data and investigating it further would advance the understanding of data reuse practices and find better methods of supporting data reuse.

METHOD

This study employed a qualitative research method and conducted in-depth semi-structured interviews with quantitative data reusers in the fields of public health and social work. These two disciplines have a culture of data reuse and share similar characteristics, such as a professional orientation in research. Some datasets were also used by researchers in both disciplines, which helped in recruiting a homogeneous sample. The study participants were identified from searching major scholarly databases, such as EBSCOhost, SAGE Journals, ProQuest Social Science, and ERIC, using the keywords "secondary data" and "secondary analysis."

In total, 299 data reusers were identified from the search. An email invitation for an interview was sent to the potential participants. A total of 38 data reusers responded and were interviewed. Given the diverse geographic distribution of the study participants, all interviews were conducted by phone. The duration of the interviews ranged from 40 to 95 minutes, with an average of 60 minutes.

All interviews followed a predeveloped protocol. Nondirective, open-ended questions were asked about the participants' reuse experiences, as well as any interaction and communication they had during such experiences. They were encouraged to draw multiple experiences from their past if relevant, as their experiences using one dataset were often related to their past experiences or experiences in using other datasets. The interviews were recorded, fully transcribed, and analyzed using a qualitative data analysis tool, NVivo 10 for Mac.

The participants were researchers in various positions, consisting of 6 PhD students, 1 postdoctoral fellow, 12 assistant professors, 9 associate and full professors, and 9 research scientists. More women (25) participated in the study than men (13), with ages ranging from 20s to 70s. They had a minimum of 2 years to a maximum of 45 years of experience as researchers in their disciplines, averaging

about 15 years. Their years of experience in research data (both secondary and primary) also varied and were not always the same as their years of experience in their disciplines. Some participants had changed disciplines during their careers or worked with data only later in their research training. Their years of data experience in research ranged from 2 to 40, averaging about 16 years. With the exception of two participants, all researchers in this study had used more than three different datasets as secondary data for their research.

FINDINGS

The participants' data reuse experiences did not involve a linear process. Generally, their experiences consisted of the following multiple steps: discovering and searching for potential data for their research, selecting the data they wished to use, understanding the data through thorough examination, conducting analyses, and writing manuscripts. During this data reuse process, the participants communicated and interacted with different individuals and communities for various purposes.

Searching for new data and new opportunities

The data discovery and search process was an important first step in data reuse, but locating data was not always easy for the participants. A couple noted, "It's not like there is a unified library catalog for entire datasets" (IP17), and "It would be nice if there were more opportunities to find out what data is available. I know some people that will not necessarily post their data" (IS09). While the participants searched the published literature to see what data were produced and reused by other researchers, checked major data resources from government agencies, such as the Center for Disease Control or the National Institutes of Health, as well as looked in other data repositories, they also actively interacted with other scholarly communities to learn what data existed or to search for new data.

Interestingly, several participants reported having a "data talk" as part of their casual conversation with other researchers around them even when they were not in the process of actively searching for new data for their research. IS01 said, "It's just part of your research life; you hang out with your buddies, basically, after you work, coffee hour, lunch time, like that, and you talk about research […] and data." This casual data talk with other researchers could initiate a new project that would reuse existing data. IP05 said that this casual data talk led to collaboration with other colleagues as they "share[d] a lot of ideas about data pretty frequently […] and found a shared interest." The participants were interested in new ideas and opportunities that data could bring into their research. "I'm always willing to learn about [what] kind of data is available that I don't know about. I usually ask my colleagues about what their research is and what data they are using" (IS04).

When the participants more actively searched data for their research, they commonly reached out to different communities around them. Some participants reported being connected with the networks of their mentors or advisors. In addition to the participants' common practice to reuse data from their advisors' projects, several participants talked about their experience in being introduced to data created by a member of their advisor's network. The advisor's network served not only as a connection to the established network(s) in the field but also as a gateway to discover and obtain permission to reuse data. IP11 reused data produced by the "friend of my advisor, who was the director of [name] center" and recalled that "[the director] was willing to share her data with me" because of the relationship between the advisor and the director. For IP11, obtaining permission to reuse this data was "rather easy" compared to the case of "a total stranger, especially a student or a junior [researcher] asking to use your data." Similarly, IS13 used his advisor's network—a large one of past advisees, the advisor's colleagues, and the latter's own colleagues—to locate data for reuse. As a junior researcher, IS13 found this network a significant resource for identifying interesting data that he could reuse and for facilitating his collaborations with senior researchers. IS13 explained that his advisor's network was a community where people "just [hung out] together out of mutual scholarly interest and then interpersonal friendship." The community members helped one another, and data reuse among them was more open "even if I didn't know them personally" because of the scholarly lineage. "I always think of it as a family […]. I say to the people who are full professors and got their PhDs in the 70s, I'm like, 'So, I'm your littlest brother' […] in terms of the family lineage" (IS13).

The participants also communicated with their own scholarly social networks and particularly noted these networks' impact on discovering new data for their research. IP09 mentioned conferences as venues to network with other researchers and to gather new information about data. IS08 also shared a successful experience of actively searching for a particular dataset of interest through networking at a conference and a committee meeting:

> So, we actually didn't have any data related to the research questions that I was most interested in. And so, it was incumbent upon me to find somebody who was collecting data about this population [of interest] and to try and work with them in order to get my foot in the door, as someone who could become, hopefully, an expert in this area. So […], I went to conferences and networked that way. I also got on a planning committee for a specialty conference in my area of interest, and through that, I was able to meet a number of researchers interested

> in my [topic] [… and] had access to data in my area. […] [T]hrough that process, [i]f you make the right and the best connections, I was able to actually find quite a few people willing to work on data with me. At least two people came forward and said, 'Hey. You know, I think I might have what you're looking for in my dataset.'

Similar to the participants' experiences with their mentors' networks, being in the same network helped the data producers feel safe and willing to share data with the participants, which made data discovery and reuse easier for them. Although previous literature pointed out researchers' reluctance to share their data for various reasons, the participants found that locating data and contacting the data producers through their networks often lowered the barriers to data sharing. IS09 talked about his group's experience in gaining access (through networking) to "some mildly restricted data" from a data producer who was not very open about sharing her data. IS09 said, "People are like, '[…] I own my data. And you've got to give me a real[ly] good reason to allow me to [let you use it …].'" IS09 came to know the data producers from one of his network colleagues and believed that the reason for his being able to use data "that couldn't be networked [reused …] ha[d] to do with his social network."

Learning through communication
For many participants, communications and interactions around data with different individuals and communities comprised the process of learning data for reuse. IS11 argued for the importance of interactions and communications around data during the reuse process:

> So […] do these interactions do anything for me? Using secondary data is […] not just [gaining] access to it but [doing so …] as if you know it well. […] [I]t's like moving to a new neighborhood. If you don't make friends in [the] neighborhood, your life is gonna be more difficult because you need to be able to talk to somebody and [ask], 'Can you recommend a good plumber?' […] Talking to other people can make the experience easier because they increase your learning of the dataset more quickly. You can get to the real work. […] When you use secondary data, it's like a two-step process. The first thing you need to do is get to know the data or get to know the neighborhood, right? Then you can get down to real living. Or in this case, doing the actual research, right? So in that sense, I think being able to talk to

> people who use the data or have used the data, or official representatives who make the data available, that's very helpful in basically getting you to the real living things more quickly.

Learning and acquiring a deep understanding of data were important in data reuse, and other participants' experiences in communicating with different groups of people supported IS11's opinions on the importance of learning about data through various methods.

Learning through mentoring

Several participants recalled their experiences in being mentored during data reuse in their earlier careers as researchers, which occurred particularly when they acquired data through their advisors and their advisors' networks. Due to the existing mentor–mentee relationship, it was natural for both participants and data producers (their advisors or members of their advisors' networks) to have teaching–learning sessions about data in the participants' processes of examining and understanding data. IP11's regular meetings with the producer of the data he was using, his advisor, and his advisor's friend covered basic information and his questions about the data, issues he encountered during his analysis, and discussions for problem solving. IP02 enjoyed meeting with her mentor and the data producer, who was also a member of the mentor's network. According to IP02, it was extremely helpful because "it provide[d] a level of collaboration and mentoring that you [wouldn't] see when you just [got] a dataset." IS13 also pointed out the importance of this aspect, saying that he had "not just access to the data but more importantly, access to people. […] [I]t's very easy for me to go and talk to the [data producers] and ask what they're thinking about, […] what's acceptable to [them], what should be done, what shouldn't be done." This close interaction with data producers and mentoring regarding data were not always possible for the participants when they acquired and used data without personal connections.

Learning from data producers

Perhaps the most important communication around data reuse is with the data producers, who were the principal investigators of the original studies (or their equivalent). Many participants emphasized the significance of gaining access to the data producers. They expressed their preference for communicating with the data producers because sometimes, the data did not come with thorough and sufficient documentation, and speaking with the data producers would enhance the participants' understanding during the process of using and analyzing the data. The participants believed that direct communication with data producers would provide a behind-the-scenes examination of the study (IS08) and illuminate "potential problems, issues, or simply questions" (IP02). IP04 explained the greatest benefit of having access to data producers: "They

know what the challenges [associated with data] are and how to get around those challenges." Those challenges are often difficult to fully explain and document; the participants mentioned that additional help might not be possible without direct interaction with data producers. Interaction with the data producers was also essential for the participants when they found errors or inconsistencies in the data as only the data producers could explain what those were.

Problem solving and peer learning

Active communications with different communities occurred during the participants' processes of understanding data and their efforts to deal with any questions or issues with the data. For some participants, such contact involved an alternative way to handle data when they could not interact with data producers; others valued communicating with various communities regardless of their communications with data producers.

Scholarly community. As the participants were already members of larger scholarly communities, such as online forums, workshops, and conferences, they relied on these communities to solve data-related problems. These scholarly communities tended to be those that the participants already belonged to and engaged with regularly. Because these communities were not specifically formed around the use of certain data but coalesced around similar research topics and interests, the participants reported the frequent information exchange about data reuse. IP05 said, "It's pretty common in my field to go online and look for forums or discussion boards about [a] particular topic, which can be a data topic." IP08's community, a forum for members of research centers, was also "very active," with "a lot of talk about data in general."

Because these communities were not formed specifically around the use of certain data or data types, finding solutions or answers to data problems might not be guaranteed, and the chance of having data experts in such communities might be lower than in a data-driven community. However, IP08 also discussed getting help from a forum that was related to her research. The forum was "pretty active," and the members "share[d] a lot of things about research, ask[ed] each other questions, [and] answer[ed] the questions" (IP08). Data-related questions and discussions were just among the common themes in the forum. Because they already "talk[ed] about what they were using in terms of secondary analysis of existing data," it was not a "weird thing to do if I [brought] in a data question" (IP08). Interestingly, the participants did not simply communicate with these communities about searching for and discovering new data. Rather, they posed questions about the data they had found, raised issues about using these, and tried to obtain comments from the communities' members.

Data reuser groups. The participants often reached out to other data reusers. Data reuser groups were good sources of

information when the participants knew what data they wanted to work with or were already working on a specific dataset. The different ways to contact other reusers included a Listserv for specific data, data workshops at conferences, webinars, and user group meetings.

As data producers sometimes held workshops and webinars, attending those sessions was one valuable way to find other reusers. The workshops were also helpful in interacting with data producers. As IS04 said, "Not just the original investigators, the project manager, and [the] data manager but [also] the faculty members who worked directly with the data" attended the workshops. Similarly, on the Listserv regarding the data that IP03 used, there was "a mix of university faculty, […] university researchers, and federal researchers." Those data reusers were not necessarily researchers in the same fields or disciplines as those of the participants but shared similar interests in the data. Neither IS04 nor IP03 recognized all of the people in the workshops or the Listserv and had no established relationship with them; nonetheless, they expressed appreciation for the other data reusers' help and the experiences that the latter shared with them. IP03 said, "It was [the] more experienced researchers who answered the most […] questions," which she found "super helpful" because "they'd know how to [solve the issue and] […] understand why [I had it]." For the same reason, IS04 noted "knowing other people who were closely working with the data," and "talking among ourselves" gave us "confidence" about solving the issue.

Local research community. The participants already belonged to their local research communities through research meetings, laboratory meetings, and even casual conversations with their colleagues at their universities or research institutions. Local research communities were often the most accessible, reliable, trustworthy, and easiest sources for discussing any research-relevant topic, including data reuse. Thus, the participants used these communities for problem solving. IP04 attended a weekly team meeting where the group members discussed their research progress, shared research ideas, and presented their challenges with the data. Although the group meetings were not specifically about the data that IP04 was using, IP04 said, "We'[d] discuss [the data]; we'[d] bring [these] to the group and see what the group ha[d] to say because everyone ha[d] a different experience and expertise about things." The participants also showed respect for those who were members of local research communities. For IP04, the "collaborative effort to solve issues" was helpful when the members needed advice on dealing with errors and when one member had already worked on the data. By the same token, IP14 tried to find "people around me who were more familiar, who could help me with understanding the data and the whole process." IS06 felt "very lucky to have [a peer's] expertise on the dataset while I [was] working on it" when she was unable to obtain assistance from the data producers or others. IS14 said, "[My colleague] has worked [on] the data quite a bit. He has a lot of insider knowledge,

which is very helpful. And he knows a few people who've worked [on] the dataset as well." Some participants' colleagues also introduced them to someone who could help. IS13's colleague said, "Oh, I have a friend who knows a bunch about it" and initiated an informal e-mail exchange.

Working with data professionals
Not all participants had extensive experiences in working with data professionals, but some discussed their experiences in communicating with statisticians and data repository staff during their data reuse processes. The participants who approached statisticians or statistical programmers during their data reuse experiences argued for the need for and the importance of statistical consultations in data reuse. IP07 said, "We talked with the statistician if we had a problem or commiserated," and the statistician could let them know "if the problem [was] really me or the data." When it was "more like a programming issue of why my merge[d dataset] wasn't working," the participant was able to get help from the statisticians. When the issue was "really about the actual data, like if the variable didn't seem to be reflecting what it was supposed to be" (IP13), the participant directly approached the data producers or others.

The participants also interacted with the staff at data repositories for help with questions if they had acquired the data from those repositories. Those participants initially contacted the repositories rather than the data producers. Although the participants perceived the repositories as "neutral" places, often as "warehouse[s]" (IS09), and thought of the data producers as primarily responsible for any issues with the data, the participants still expressed appreciation for the repositories' help with their questions. IS09 contacted one data repository when he found that some of his results did not make sense. The repository staff first responded, "We're not really sure," which made IS09 think of the repository as just a warehouse. However, the staff later offered to examine the data and helped him solve the problem.

> IS09: The people who [were] warehousing it were totally different people. And those people could sometimes be more helpful because they felt like that was their job. It's like, I was not talking to the PI but was [instead] talking to people [who] had […] energy and expertise about the data.

Limitations of communication around data
Despite many participants reporting the importance of communicating with different individuals and communities about data reuse, several participants also mentioned some challenges and limitations of these communications. One was the challenge of being junior researchers, causing some participants to feel the barriers to becoming part of an established network to obtain data. IS09 said, "My problem is, being junior, I don't have a great social network. I'm not

strongly affiliated with a lab or [a] large research outfit." IS09 was aware that certain data were shared among researchers but not with him because sharing was done through networks, and "people don't know me, and [in] a lot of the cases, I don't even know what [other researchers have] got." IS09 recognized the opportunities that networks could bring him, but without an established network, it was difficult to gain access to other existing networks and discover what data would be available.

Another difficulty involved communication and interaction with data producers, which was not always successful. A few participants experienced difficulty in receiving responses. Sometimes it took too long for them to do so, or they received only "partial answers," which they found "very frustrating" (IS09). IP12 found the interaction "tricky, as it [would] just depend [on] get[ting] the right person who ha[d] the knowledge [of the questions]," particularly when working on a large amount of institutional data. Usually the institutional data provided a contact person, who might or might not have been directly involved in the data collection or the original analysis. Although the participants believed that the contact person would be an expert on the data and the right person to speak to, the negative experiences made IP12 wonder whether "all the staff [members] were really busy or they [might] not exactly know."

DISCUSSION

This study has examined the role of communication during the data reuse process, as well as the nature of the communication. This study's findings reveal various individuals and communities engaging in data reuse to ease the process. Communication is not always necessary for all data reusers. Several participants experienced working on data without any external communication because the dataset was well documented, well prepared, and easily understandable. However, this was not consistently reported by the participants, as they worked on multiple datasets from different data producers with varying conditions for reuse.

This study's results show that the communications with different stakeholders mainly support data reuse in three areas: searching, learning, and problem solving. While data reusers usually have their own strategies and practices for finding and understanding data and figure out any issue encountered during the process of reusing data, communications with various parties provide often complementary and sometimes the only means of successfully reusing data and overcoming challenges associated with reuse.

Data discovery and search is the most important first step in data reuse, but difficulties in finding data for reuse are among the well-known challenges. As reported by the participants, datasets are stored in various locations without any unified catalog or search system. Additionally, existing research on data sharing points out various factors that inhibit data producers' sharing behaviors (e.g., Cragin,

Palmer, Carlson, & Witt, 2010; Tenopir et al., 2011), meaning that not all data are shared. Using scholarly lineage or scholarly networks is an effective alternative way to learn about new data that are not publicly available or deposited in repositories, and having in-person or casual conversations often lowers the barriers to data access. The participants reported being connected to an established community as a huge benefit in terms of learning about data and obtaining permission to reuse the data. On the contrary, not being part of this established community can be a disadvantage for other participants, usually junior faculty members, by not gaining access to data. Due to the person-to-person data exchange or frequent sharing, building relationships among scholarly communities and having formal or informal data talks are significant for data searching and access.

Communications with others also add a level of collaboration and learning. Data reuse was not always solo work for the participants, who learned more about the data from others' experiences. As IS11 argued, communication with others might be the core component of data reuse because in that process, researchers would work with unfamiliar data that they did not create. The participants had to rely on the given information about the data through different mechanisms, which were not always thorough or sufficient. As much of existing research has already pointed out, understanding data is not always easy because of their contextual nature (e.g., Berg & Goorman, 1999; Bishop, 1999; Sandusky & Tenopir, 2007; Stewart, 1996; Zimmerman, 2007). External communications and interactions with other researchers, mentors, and data producers play an important role in sufficiently transferring embedded knowledge in the data.

Finally and most importantly, communications around data are central in problem solving, which help researchers overcome challenges and difficulties in reusing data. The participants experienced different challenges and difficulties in the various phases of data reuse, from data discovery and search, understanding data, and seeking answers to questions, to solving issues and problems that emerged during data reuse. Interacting with different individuals and communities and sharing information with one another were common ways for the participants to solve the problems they encountered. This finding suggests that effective communications can be the key to support successful data reuse. Existing research has argued that the lack of support in data reuse is one of the critical components of failed reuse (Yoon, 2016a). Communications with peers, data professionals, and relevant research communities are ways to provide support for problem solving, which can even turn unsuccessful data reuse experiences to successful ones.

While active communications around data produce rich conversations and collaborations among scholars, they also reveal the need to provide more formal support for data reuse. Communication with others was often the only way

to overcome the challenges of data reuse—in this study, searching, learning, and problem solving—because data reusers did not know where else to get help. IS02 said that she was just "wandering around to see who [would] know this." Sometimes, early in their careers, data reusers found it even difficult to engage in data talk without an established network. While academic libraries have initiated efforts to support scholars' data work, their data services usually aid in scholars' data sharing by providing data management and documentation guidelines; not many offer services relevant to data reuse (Yoon & Schultz, 2017). Active support for scholars' communications around data would be necessary, such as developing a tool to foster effective communications among scholars and their information sharing for data reuse.

CONCLUSION

As the findings demonstrated, data talk was an aspect of scholarly conversation for the participants. It was natural for them to use or rely on existing scholarly communities and connections for their data reuse. Even if contacting such communities might not always guarantee answers to the participants' questions regarding the data or help with their data reuse experiences when those communities were not specifically built around data, the participants naturally discussed data issues and found the communities helpful. New communities also formed around the data, such as data reuser groups, which could support reusers' experiences more effectively.

As shown by previous studies, data reuse is not a simple process for researchers. Finding and understanding data require certain efforts on the part of data reusers. This study suggests that communications with and among other researchers and/or data reusers play an important role in searching, learning, and problem solving during the data reuse process. Considering the roles and the importance of communications occurring around data, ways to support these communications among scholars should be investigated further.

REFERENCES

Akmon, D. R. (2014). *The role of conceptions of value in data practices: A multi-case study of three small teams of ecological scientists*. (Doctoral dissertation). Retrieved from http://deepblue.lib.umich.edu/handle/2027.42/107162

Baker, K. S., & Yarmey, L. (2009). Data stewardship: Environmental data curation and a web-of-repositories. *International Journal of Digital Curation, 4*(2), 12-27. doi:10.2218/ijdc.v4i2.90

Birnholtz, J. P., & Bietz, M. (2003). Data at work: Supporting sharing in science and engineering. In *Proceedings of the ACM Conference on Supporting Group Work* (pp. 339–348). Sanibel Island, FL. New York, NY: ACM.

Berg, M., & Goorman, E. (1999). The contextual nature of medical information. International Journal of Medical Informatics, 56(51-6), 51-60. doi:10.1016/S1386-5056(99)00041-6

Borgman, C. L. (2007). *Scholarship in the digital age: Information, infrastructure, and the Internet*. Boston, MA: MIT Press.

Borgman, C. L. (2011). *The conundrum of sharing research data*. Retrieved from SSRN eLibrary: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1869155

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. doi:10.1007/s00799-007-0022-9

Broom, A., Cheshire, L., & Emmison, M. (2009). Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sociology*, *43*(6), 1163–1180. http://doi.org/10.1177/0038038509345704

Carlson, S., & Anderson, B. (2007). What are data? The many kinds of data and their implications for data re-use. *Journal of Computer-Mediated Communication*, *12*(2). 635–651.

Chin, G., & Lansing, C. S. (2004). Capturing and supporting contexts for scientific data sharing via the biological sciences collaboratory. In *Proceedings of CSCW'04, ACM Conference on Computer Supported Cooperative Work* (pp. 409–418). Chicago, IL: ACM.

Committee on Ensuring the Utility and Integrity of Research Data in a Digital Age. (2009). *Ensuring the integrity, accessibility, and stewardship of research data in the digital age*. Retrieved from National Academy of Sciences: http://www.nap.edu/catalog.php?record_id=12615#toc

Corti, L. (2005). Qualitative archiving and data sharing: Extending the reach and impact of qualitative data. *IASSIST Quarterly*, *2005*(Fall), 8-13.

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

Cragin, M. H., & Shankar, K. (2006). Scientific data collections and distributed collective practice. *Computer Supported Cooperative Work*, *15*(2–3), 185–204. http://doi.org/10.1007/s10606-006-9018-z

Faniel, I. M., Barrera-Gomez, J., Kriesberg, A., & Yakel, E. (2013). A Comparative Study of Data Reuse A mong Quantitative Social Scientists and Archaeologists.

iConference 2013 Proceedings (pp. 797-800). .doi: 10.9776/13391

Faniel, I. M., & Jacobsen, T. E. (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work*, *19*(3–4), 355–375. http://doi.org/10.1007/s10606-010-9117-8

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

Faniel, I. M., & Zimmerman, A. (2011). Beyond the data deluge: A research agenda for large-scale data sharing and reuse. *International Journal of Digital Curation*, *6*(1), 58-69. doi:10.2218/ijdc.v6i1.172

Fienberg, S., Martin, M., & Straf, M. (1985). *Sharing research data*. Retrieved from National Research Council: http://www.nap.edu/openbook.php?isbn=030903499X

Huang, H., Stvilia, B., & Jörgensen, C. (2012). Curation roles and perceived priorities for data quality dimensions and skills in genome curation work. *Proceedings of the American Society for Information Science and Technology*, *49*(1), 1–3. http://doi.org/10.1002/meet.14504901312

Jirotka, M., Procter, R., Hartswood, M., Slack, R., Simpson, A., Coopmans, C., … Voss, A. (2005). Collaboration and trust in healthcare innovation: The eDiaMoND case study. *Computer Supported Cooperative Work, 14*, 369–398.

Kriesberg, A., Frank, A., Faniel, I. M., & Yakel, E. (2013). The role of data reuse in the apprenticeship process. In *Proceedings of the Association of Information Science and Technology, 50*(1), 1-10. Montréal, QC. doi: 10.1002/meet.14505001051

Kunze, J., Cruse, P., Hu, R., Abrams, S., Hastings, K., Mitchell, C., & Schiff, L. R. (2011). *eScholarship: Practices, trends, and recommendations in technical appendix usage for selected data-intensive disciplines*. Retrieved from http://escholarship.org/uc/item/9jw4964t

Markus, M. L. (2001). Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success. *Journal of Management Information Systems*, *18*, 57–93.

Martin, V. (2014). *Demystifying eResearch: A primer for librarians*. Santa Barbara, CA: ABC-CLIO.

McCall, R. B., & Appelbaum, M. I. (1991). Some issues of conducting secondary analyses. *Developmental Psychology, 27*(6), 911–917. http://doi.org/10.1037/0012-1649.27.6.911

Moore, N. (2007). (Re)using qualitative data. *Sociological Research Online*, *12*(3).

Niu, J. (2009). Overcoming inadequate documentation. In *Proceedings of the Annual Meeting of the American Society for Information Science & Technology*. Vancouver, BC.

Niu, J., & Hedstrom, M. (2008). Documentation evaluation model for social science data. In *Proceedings of the Annual Meeting of the American Society for Information Science & Technology, 45*(1), 11-11. Columbus, OH. doi: 10.1002/meet.2008.1450450223

Rolland, B., & Lee, C. P. (2013). Beyond trust and reliability: Reusing data in collaborative cancer epidemiology research. In *Proceedings of the 2013 Conference on Computer-supported Cooperative Work* (pp. 435–444). New York: ACM. Retrieved from http://doi.org/10.1145/2441776.2441826

Sands, A., Borgman, C. L., Wynholds, L., & Traweek, S. (2012). Follow the data: How astronomers use and reuse data. *Proceedings of the American Society for Information Science and Technology*, *49*(1), 1–3. http://doi.org/10.1002/meet.14504901341

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., … (2011). Data sharing by scientists: Practices and perceptions. *Plos One, 6*(6). Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101

Van House, N. A., Butler, M. H., & Schiff, L. R. (1998). Cooperative knowledge work and practices of trust: Sharing environmental planning data sets. In *Proceedings of the ACM Conference on Computer-supported Cooperative Work* (pp. 335–343). Seattle, WA.

Yoon, A. (2014). Making a square fit into a circle: Researchers' experiences reusing qualitative data. In *Proceedings of the Association of Information Science and Technology, 51*(1), 1–4. Seattle, WA. doi: 10.1002/meet.2014.14505101140

Yoon, A. (2016a). Red flags in data: Learning from failed data reuse experiences. In *Proceedings of the 79th Association for Information Science and Technology Annual Meeting, 53*(1), 1–6. Copenhagen, Denmark. doi: 10.1002/pra2.2016.14505301126

Yoon, A. (2016b). Visible evidence of invisible quality dimensions and the role of data management. In *Proceedings of iConference 2016*. Philadelphia, PA.

Yoon, A. (2017). Data reusers' trust development. *Journal of Association for Information Science and Technology*, *64*(8), 946–956. doi: 10.1002/asi.23730

Yoon, A., & Schultz, T. (2017). Research data management services in academic libraries in the US: A content analysis of libraries' websites. *College & Research*

Libraries [pre-print is available at: http://crl.acrl.org/index.php/crl/article/view/16620/0].

Zimmerman, A. S. (2008). New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology & Human Values, 33*(5), 631–652. http://doi.org/10.1177/0162243907306704

Zimmerman, A. S. (2007). Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse. International Journal on Digital Libraries, 7(1-2), 5–16. doi:10.1007/s00799-007-0015-8

Zimmerman, A. S. (2003). *Data sharing and secondary use of scientific data: Experiences of ecologists* (Doctoral dissertation). Retrieved from http://search.proquest.com.libproxy.lib.unc.edu/docview/287907131/abstract?accountid=14244