

# Unpacking the “Black Box”: A Preliminary Study of Visualizing Humanists and Social Science Scholars’ Data and Research Processes

Wei Jeng<sup>1</sup>, Eleanor Mattern<sup>1, 2</sup>, Daqing He<sup>1</sup>, and Liz Lyon<sup>1</sup>

<sup>1</sup>School of Information Sciences, University of Pittsburgh

<sup>2</sup>University Library System, University of Pittsburgh

## Abstract

Due to the significant variations in their needs, attitudes, and practices toward their research, different scholars could have different research processes and interact with different kinds of data. We report the preliminary results of a qualitative study that examines how humanities and social science (H&SS) scholars interact with their data throughout the course of their research process. By conducting two focus group sessions with eight H&SS scholars, our results show that scholars indeed took different research patterns and thus interacted with distinctive data. The sketches produced in this study provide insights on future larger-scale focus group design, and also identify opportunities for an academic library or data service provider to support H&SS scholars’ research activities.

**Keywords:** Humanities and social science data; data practices; research process; research data services; visual methods

**Citation:** Jeng, W., Mattern, E., He, D., & Lyon, L. (2016). Unpacking the “black box”: A preliminary study of visualizing humanists and social science scholars’ data and research processes. In iConference 2016 Proceedings.

**doi:** 10.9776/16173

**Copyright:** Copyright is held by the authors.

**Contact:** wej9@pitt.edu; emm100@pitt.edu; dah44@pitt.edu; elyon@pitt.edu

## 1 Introduction

Research data commonly refers to the raw material obtained or generated during the research process, and it is “commonly accepted in the scientific community as necessary to validate research findings” (OMB Circular A-110). There are three critical attributes of research data: when and where data are obtained (situation), what they consist of (content), and why they are used (context) (Martin, 2014).

Most comprehensive and systematic studies on data practices focus on research in the STEM fields. Data service providers such as academic libraries or discipline repositories have already gained much insights about possible barriers, motivations, and concerns among individual STEM researchers (Kim, 2012; Lage et al., 2011). Less attention has been paid to how H&SS scholars conduct their research from the perspective of “data”: what data these scholars collect (content), how they interact with such data on what stage of their research (situation), and what research methods they adopt (context). More importantly, variations in H&SS scholars’ research needs, attitudes, and practices could further complicate the understanding of data and services supporting their work.

Through a preliminary qualitative study, this article aims to advance the understanding of how H&SS scholars collect, process, and interact with data at each stage of their research process, and thus opens the “black box” on how they conceptualize their research processes and data in their research.

Our study involved two focus group sessions with early career H&SS scholars. We employed a methodology involving visual sketching and narration to probe the following questions:

- RQ1: How do H&SS scholars conceptualize the research process through visual sketching?
- RQ2: How do H&SS scholars interact with research data throughout their research process?

Recently, researchers and practitioners in academic libraries have joined iSchool researchers in demonstrating an increasing interest in research data management and curation. These libraries are currently building or considering adding support for research data management to their faculty and students and mapping such services to a visual representation of the research process (e.g. University of California Irvine Libraries, University of Virginia Library). Our library system at University of Pittsburgh is enhancing their research support services. Thus the results of this study aim to directly benefit their services by building an understanding how scholars on campus approach their research process and data interaction.

## 2 Background

*Data practices for H&SS scholars.* There are few existing work on data practices for H&SS scholars. Among the limited literature on the data practices of humanities and social science scholars, the most noticeable research project is Data Curation Profiles (DCP). Cragin et al. (2010) presented the DCP tool, which supports assessment and analysis of data characteristics and scholars’ discipline data practices. Each completed profile is a resource that helps other professionals quickly capture how specific data was generated, used, and reused in a certain research area. Their project revealed common data practices for

H&SS domains, such as standard file exchange formats, file size, and data's embargo time. However, Cragin et al. (2010) also found that although developing services for data curation and sharing is considered to be very important, no field-wide norms have been established, and very few scholars routinely deposit their research data into data repositories.

*Common research activities for H&SS scholars.* There are a host of texts on research methodologies that provide insight on common activities that form the research process in H&SS domains. Although the sequence of research activities presented vary in different research paradigms (e.g., quantitative studies versus qualitative studies), a set of activities that generally cut across disciplinary work include (Carole & Julian, 2004): planning the research; surveying related work; identifying research questions; gathering data; analyzing data and interpreting the research outcomes; and communicating the research findings. The literature tends to illustrate activities in a research process as a journey in which scholars travel from their research questions and motivations to the research outcomes (Malins & Gray, 2013). Such a journey is typically visualized as a linear model (Bhattacharjee, 2012; Fraenkel & Wallen, 2003). However, some researchers described research process as a lifecycle in H&SS (Johnson & Christensen, 2008) where there is no explicit startpoint and endpoint. For any node, one can go back to the same node by moving along the directed links. The structures of the research process in H&SS can even be more complex than linear or cyclical ones. For example, the process of research can form a "daisy" (or a spoke-hub pattern) (Mackey, 2009), with the central idea placed in the center, connecting to neighboring components via links that are often bidirectional. These neighboring components form a cycle among themselves, too. This structure allows high flexibility at visualizing the course of research or only one stage of research.

### 3 Research Design

#### 3.1 Visual Method

Visual method is an ethnographic methodology deriving from visual anthropology (Pink, 2003). Today, the interests in employing visual method is increasing in cultural studies, queer studies, and consumer research (Pink, 2003). Compared with purely talk-based approaches (e.g., interviews) or text based ones (e.g., diaries, social media content analysis), scholars who employ this method usually ask their participants to create video diaries, photography, or drawings as research materials (Buckingham, 2009). These visual materials usually represent participants' life stories and their points of view.

Our research methodology drew upon visual narrative inquiry, a qualitative methodology in which participants communicate their experiences visually (Connelly and Clandinin, 1990; Bach, 2008; Bowler et al., 2014). This study was rooted in the observation that there are a host of visual representations of research that academic libraries use for mapping and communicating services. We were driven by an interest to learn directly from H&SS scholars about how they interact with their research data by visualizing their own experiences.

#### 3.2 Data Collection and Analysis

We used convenience sampling for our data collection, recruiting eight H&SS scholars who completed or were nearing completion of their doctoral degree. In December 2014 and January 2015, two focus group sessions of four participants each were held at the iSchool at University of Pittsburgh. Seven out of eight participants held a PhD degree at the time of the study, and we had an equal number of male and female participants. The participants are from five different research programs: information sciences, library and information science, history, anthropology, and philosophy.

We first introduced the study objectives to the participants and asked for permission to create an audio recording of the session. Each participant then spent 15 minutes sketching their research process and later verbally described their sketches. To minimize our influence on the participants' drawings, we intentionally avoided giving instructions about how they might approach the visualizations (i.e. whether they might draw their process as a lifecycle or a flowchart or as comic book panels). We finally asked the participants to talk about their sketch of research process and data. The average time for each pilot focus group session was an hour and forty-five minutes. A more detailed discussion of this particular research method, including the overall protocol and implications about the visual method, was reported in Mattern et al. (2015). This article focuses on researchers' self-reported data challenges and their recommendations regarding library's research data service.



Figure 1-8. Participant #1-8 (from left to right and from top to bottom)

## 4 Preliminary Results

### 4.1 Visualizing the Research Process in Humanities and Social Sciences

As shown in Figure 1-8, each participant in the pilot focus groups created their own sketch that visually narrates their research process and how they interact with their research data.

Table 1 on the next page summarizes and compares the characteristics that we observed in the research processes. We assigned participants IDs (P01 to P08), which are referenced in this paper. We found that the majority of participants (i.e., P01, P02, P05, P07, & P08) depicted a linear process, displaying their work as sequential steps. In particular, P01’s linear process involved a decision about the “kind of data” the research targets, with two options in a flowchart style: simulation versus actual data (see Figure 1). Other H&SS scholars visualized their research as a cyclical process. For P05 and P06 (Figures 5 and 6), when their research reaches an endpoint, they return back to the start. Their representations are reminiscent of the research processes described in Johnson and Christensen (2008).

P03 (Figure 3), on the other hand, constructed a “daisy” shape (spoke-hub) like that described in Mackey (2009). Her research activities were each connected to this central worldview and the resultant drawing is messier, less sequential than others produced during the sessions. For example, conducting a literature review does not occur at one point in the research process. Instead, as she prepares her research questions, selects her methods, and designs the study, she returns to this node consistently.

ID	Status	Research Focus	Institutions	Research Method/Data collection method	Observed type of visualization
P01	Doctoral Candidate	Data Mining	iSchool	Quantitative/ epidemiology datasets in a historical data repository	Linear-flowchart
P02	Visiting Scholar	Public Library Management	iSchool	Mixed/ Business documents and datasets	Linear
P03	Visiting Assistant Professor	Public Library Management	iSchool	Qualitative/ Diaries kept by participants	Daisy
P04	Postdoctoral Researcher	Archival Studies	History	Qualitative/Archival sources and interviews	Linear
P05	Postdoctoral Researcher	Education Technology	iSchool	Mixed/ Survey responses, interviews	Cycle
P06	Research Fellow	History and Philosophy of Science	Philosophy	Qualitative/ Published research articles	Cycle
P07	Postdoctoral Researcher	Medical Anthropology	Anthropology	Qualitative/ Field notes by the researcher, interviews from informants	Linear
P08	Postdoctoral Researcher	World History	History	Qualitative/Archival sources and Published research articles	Linear

Table 1. An Overview of Participants

While the participants’ research process might vary, we looked into participants’ research activities through the lens of four main themes articulated in Table 2: conceptualization, plan, execution,

and reporting. Note that these four components are common to the research process (Gray & Malins, 2013) but are not required, and there is no specific or necessary chronological order among them. There were variations in the activities that the participants described. For example, data preparation, processing, and analysis seemed to dominate participant P01's research, whereas P03 and P08 put more emphasis on the conceptual design of the research. Interestingly, two participants (P01 and P05) did not mention the stages of conceptualization and reporting in their sketches and narratives.

ID	Methods	Conceptualization		Design		Execution		Reporting		
		Developing research questions	Literature review	Selecting research method	Defining variables /samples	Data gathering	Data analysis	Results Interpretation	Authoring	Giving access to others
P01	QUANT			X	X	X	X			
P02	Mixed	X	X	X		X	X		X	X
P03	QUAL	X	X	X	X	X	X	X	X	
P04	QUAL	X	X	X	X	X	X	X	X	
P05	Mixed				X	X	X			
P06	QUAL		X			X	X		X	X
P07	QUAL	X	X		X	X	X		X	
P08	QUAL	X	X	X		X	X	X	X	X

Table 2. Common Elements of Research Process

We observed several interesting results. First, for P06 and P08, the literature review step includes data collection and data analysis because their main source of research data is published texts. Second, in the focus group sessions, almost every H&SS scholar addressed the data gathering and data analysis steps (i.e. the execution stage of their research.) Third, only three out of eight participants mentioned giving access to others, while the rest of the participants at most considered the authoring step in the reporting stage. Despite the variations in research activities, we discovered a common execution stage that appeared in every sketch.

#### 4.2 Research Data Practices in H&SS

Even if participants' research patterns are similar (e.g. both P06 and P07 illustrated a linear research pattern), the emphases of their research vary significantly due to the difference in their research data. In this section, we take a deep dive into P01, P06, and P07's sketches and explore their data practices, based on their self-report responses.

*When primary data are from literature.* P06 begins his research journey by browsing literature, represented as "Stage A" in Figure 6. This is unique to the sketches, as most of the participants (e.g. P02, P03, P04, and P08) mentioned that they would start with constructing a research idea. The cloud in "Stage B," a metaphor for a collection of published useful and irrelevant articles, became P06's primary material for his study. "Stage C" in his process illustrates the submission process to an academic journal, including the interactions with potential anonymous reviewers that requires a return to his data. P06 explained, "I have to go back and find all those things that were in the cloud (Stage B)... And then I incorporate all those and send it back." In P06's view, the articles, or "this big cloud of stuff" are the primary materials for his study. Once his article is published, his work will go back and become other humanists' research materials. It is worth noting that while P06 did not explicitly mention the concept of data reuse, his sketch and narrative implicitly suggested that his published works act as data that other scholars in his community can draw upon.

*When primary data are derived from informants.* P07 is an anthropologist who visualized his dissertation research process in Figure 7. Interestingly, he referred the research methodology that he was trained in as "a black box": "it confuses a lot of students who are in the process of doing their first fieldwork." In his sketch and narratives, he describes how he (the figure with purple hair) gets closer to the informants throughout the time: "there's a big gap, metaphorically and everything... So then after some time has passed, this gap isn't as wide and you're able to kind of leap over it and have a closer advantage with the people you're studying and interacting with (P07)." Understanding and successfully interpreting his informants becomes the end task of his drawing. The structure of P07's research process reveals the dependency between components: A linear process implies that a step cannot begin until the previous step has finished. A "daisy" shaped process, such as P03's visualization (Figure 3), shows that the scholar might go back to the same step iteratively or might work on multiple steps concurrently.

*When primary data are from a third-party repository.* Although the sketch also shows a linear structure, P01's research focuses on numerical data which are obtained from third-party repositories, and therefore involves different activities from others who are also depicted a linear pattern. The primary data source for P01 is tabular epidemiology datasets from a historical data center. As shown in Figure 1, the participant starts her research journey by defining the data type, splitting datasets into so-called actual data and simulation data. P01's research process is heavily driven by data and resembles a common research paradigm in recent data-intensive studies: researchers often start with data gathering,

processing, cleansing, and then try to find patterns or relationships in the data. The discovered patterns are then used to help make decisions about where to take the research.

## 5 Implications and Future Work

In this paper, we report on the preliminary results on how humanities and social science scholars visualize their research processes and data practices throughout these processes. Eight participants from five different disciplines revealed different research patterns: linear, cyclical, and a daisy; as well as different focus on their research data: from literature, from informants, and from a third-party repository.

Our preliminary study confirms that there are diverse research patterns in the humanities and social sciences that have been previously relatively overlooked in the literature. In particular, we also observed different ways of conducting research when participants dealt with different kinds of primary data. The preliminary results also suggest that in future work, it will be worthwhile to collect basic information about their research patterns and primary data types while or before investigating their data practices.

For the library system at University of Pittsburgh, the insights gleaned from these sketches and from the sketches produced by future focus groups will help to identify opportunities to support research activities. For example, because each participant in this study referenced the execution stage in their sketches, university libraries or data service providers may consider focusing their attention to the development and enhancement of their services that assist researchers at this stage. We also expect that our final results will inform researchers in the realm of sociology of social scientific knowledge, who treat social science itself as a sociological inquiry and discussed how social scientists embody value and use their tacit knowledge when conducting social research (e.g. Maynard & Schaeffer, 2000).

There are two directions of future work. First, we would like to recruit more H&SS researchers for additional focus group sessions. Since the current sample participants worked primarily with qualitative data, we expect to pay some more attention to quantitative scholars. For a further investigation for possible factors which shape researchers' research patterns, a quantitative survey at a larger scale is needed to fulfill this query.

## 6 References

- Bhattacharjee, A. (2012). *Social science research: principles, methods, and practices*. CreateSpace Independent Publishing Platform.
- Bowler, L., Mattern, E., & Knobel, C. (2014). Developing design interventions for cyberbullying: A narrative-based participatory approach. *iConference 2014 Proceedings*.
- Buckingham, D. (2009). "Creative" visual methods in media research: possibilities, problems and proposals. *Sage Publications*, 31, 4-652.
- Connolly, F. M., & Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational researcher*, 19(5), 2-14.
- Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, 368(1926), 4023-4038.
- Fraenkel, J. R. & Wallen, N. E. (2003). *How to design and evaluate research in education* (5th ed.). Boston: McGraw-Hill.
- Gray, C. & Malins, J. (2013). *Visualizing research: A guide to the research process in art and design*. Ashgate Publishing, Ltd.
- Israel, M. (2014). *Research ethics and integrity for social scientists: Beyond regulatory compliance*. Sage.
- Johnson, B., & Christensen, L. (2008). *Educational research: Quantitative, qualitative, and mixed approaches*. Sage.
- Lage, K., Losoff, B., & Maness, J. (2011). Receptivity to library involvement in scientific data curation: a case study at the University of Colorado Boulder. *portal: Libraries and the Academy*, 11(4), 915-937.
- Mackey, K. (2009). *Research Process*. Retrieved from [http://www.clark.edu/Library/iris/types/research\\_process/research\\_process\\_p3.shtml](http://www.clark.edu/Library/iris/types/research_process/research_process_p3.shtml).
- Mattern, E, Jeng, W., He, D., Lyon, L., & Brenner, A. (2015). Using participatory design and visual narrative inquiry to investigate researchers' data challenges and recommendations for library research data services. *Program: electronic library and information systems*. 49(4): 408-423.
- Maynard, D., & Schaeffer, N. (2000). Toward a Sociology of Social Scientific Knowledge Survey Research and Ethnomethodology's Asymmetric Alternates. *Social Studies of Science*, 30(3), 323-370.
- Martin, V. (2014). *Demystifying eResearch: A Primer for Librarians*. Libraries Unlimited.
- Pink, S. (2006). *The future of visual anthropology: Engaging the senses*. Taylor & Francis.
- University of California Irvine Libraries (n.d.). *Digital Scholarship Services*. Retrieved from <http://www.lib.uci.edu/dss/>.

University of Virginia Library Research Data Services (n.d.). Steps in the Data Lifecycle. Retrieved from <http://data.library.virginia.edu/data-management/lifecycle/>