Rhetoric of Health & Medicine
Vol. 3, No. 4, pp. 483–493
DOI: 10.5744/rhm.2020.4008

Pivoting Toward Rhetorical Ethics by Sharing and Using Existing Data and Creating an RHM Databank: An Ethical Research Practice for the Rhetoric of Health and Medicine

Kristin Marie Bivens and Candice A. Welhausen

We argue that by using existing data and sharing research in a databank, RHM scholars can practice a research habit that conserves and optimizes intellectual and institutional resources. When possible, by using existing datasets, scholars avoid *data waste*, that is ignoring or bypassing existing data. The data distinctions that we call attention to—derived, compiled, and designed—account for various ethical and rhetorical concerns regarding privacy and confidentiality, expected context, and consent. Equally important to the aforementioned data deliberations we explore, collecting and managing shared RHM data in a databank, while possible, are not without ethical, logistical, and rhetorical difficulties.

KEYWORDS: compiled, designed, derived data

The National Endowment for the Humanities (NEH, 2018) defines data as "materials generated or collected during the course of conducting research" (p. 1). Data might be quantitative (for example, numerically continuous or discrete) or qualitative (descriptive or narrative). Qualitative data—like much of the data used in rhetoric of health and medicine (RHM)-oriented research—might be designed (new and original—that is, collected via observations, focus groups, and/or interviews, for instance), derived (existing or found), and/or compiled (curated or made).

Because most research in scientific fields is funded through federal grants, it is common practice for this information to be shared and/or otherwise made publicly available. For instance, large quantitative (continuous and discrete) datasets are collected by many government agencies (such as the United States Bureau of Labor Statistics and Data, 2020; CDC, 2019) and are centralized for researchers to redeploy and analyze in various research contexts. Further, when researchers in scientific and medical fields have created their own datasets, academic journals (such as *BMJ,* 2020; *JAMA*, 2020; *Nature,* 2020) have often encouraged authors to share this information.

In many cases, researchers have created data repositories. For example, scholars developed a collaborative system for integrating biomedical data for analysis, anonymization, and sharing or iDASH (see Ohno-Machado et al., 2012). According to the group, "iDASH was designed to allow as many researchers as possible to leverage other researchers' work and accelerate discoveries" (para. 1). Furthermore, Theresa L. Harris and Jessica M. Wyndham (2015) claimed that "The enormous potential benefits of sharing data, only now becoming possible as a result of new technologies, are driving demands for researchers to make their data openly accessible" (p. 334).

These trends, too, have surfaced in the digital humanities through Cheryl Ball's (2013) statements in a Blog Carnival post for the University of Michigan's Sweetland Digital Rhetoric Collaborative, which states, "humanities researchers collect and publish their data for others to potentially use" (para. 1). In fact, although Ball's call to share data may have been novel for writing and rhetoric scholars, it was not for humanities researchers. As she noted, writing studies projects such as Research Exchange Index (Rex or WritingPro) that provide opportunities for writing and rhetoric scholars to upload datasets exist. In this way, there is momentum toward "[approaching] our research as scientists do, understanding that data is as important as our analysis of it and others may—and will—find value in it" (Ball, 2013, para. 7).

As a field, RHM is growing and maturing, and as it expands, so do our systematic investigations, our research materials, and the affiliations and relationships that we create with researchers outside of the field (Baldwinson, 2018, pp. 217–218). In this ethical exposure, our simplest assertion is to be frugal scholars and researchers who share and use shared datasets. We argue that by being prudent with RHM data and widely leveraging existing derived, compiled, and designed data, we can avoid *data waste* (that is, ignoring or bypassing existing derived or compiled datasets to collect and create more designed data), which we present as a response to Raquel Baldwinson's (2018) rallying call for rhetorical ethics to be tethered to research practices.

We call for RHM scholars to embrace analyzing existing data, while also considering the implicit and potential ethical implications of doing so. In kind, we endorse a "distinctively ethical" (Baldwinson, 2018, p. 213) frame for sharing and storing RHM data that is rooted in human rights. The thesis of the documentary *The Great Hack* contends that "data rights are human rights," which is the optic—human rights—we use to frame our argument. By viewing RHM through a human rights lens, we connect to similar discourses in human research ethics, as well as use human rights as a framework for considering the ethical responsibilities related to sharing and storing RHM data. Furthermore, data rights as a human right should include individual privacy rights regarding data—a point we take up later in our discussion of storing shared data.

Rather than endeavoring to address the full range of ethical considerations involved in data sharing in RHM, we seek to begin the conversation about how RHM scholars might responsibly and ethically account for the idea of data waste as we have defined it. Specifically, we suggest that using derived data, which we illustrate using one of our research projects, contributes to conserving valuable intellectual (for example, advisors', colleagues') and institutional (for example, IRB) resources. In making this argument, we discuss the ethical, logistical, and rhetorical intricacies involved in creating a databank where datasets can be responsibly shared with the RHM community. We also recognize that data sharing and storage are not without their nuanced, contextual, ethical difficulties. To exemplify these considerations, we turn next to our project that focuses on the emergency intervention app, PulsePoint Respond, to illustrate how RHM researchers can use existing data and to highlight several ethical considerations of doing so before contemplating the complexities of an RHM databank.

The PulsePoint Project and Considerations for Using Existing Data

Curious about the volunteer-based networks of lay trained cardiopulmonary resuscitation (CPR) responders, the individuals who have experienced sudden cardiac arrest (SCA) and received CPR from bystanders (Hansen et al., 2015; Ringh et al., 2015), and the mapping technologies that enable these practices, our PulsePoint Project[1] used content analysis to examine review comments posted about the PulsePoint Respond mHealth app on the iOS and Android platforms, respectively (Welhausen & Bivens, forthcoming). PulsePoint Respond is a crowdsourced, emergency response app that connects people experiencing SCA, which is often fatal, with nearby responders (who are frequently off-duty, trained healthcare providers) (see Dainty, Vaid, & Brooks, 2017).

This study emanated from our preoccupation with using existing user comments to explore users' practices with this tool, while also avoiding leaving data on the table (and preventing data waste). More specifically, when we began conceptualizing this project, we discovered there were nearly 30,000 review comments posted online (on the combined iOS and Android platforms) about the app, which we believed could lend insight into users' unstructured, self-reported experiences with the app, as well as their self-perceptions of these experiences.

In what follows, we examine several logistical and temporal cogitations related to this project. Whether quantitative or qualitative, designing a rhetorical research study (whether ethnographic or computational) and participating in research activities is a time-intensive process that draws upon a multitude of resources. Some of those resources are personal and professional (for example, having a conversation with a colleague at a conference), intellectual (for example, synthesizing information or troubleshooting research logistics), or institutional (for example, IRB oversight). When it comes to ethical or "distinctively ethical" RHM research (Baldwinson, 2018, p. 213), we are not suggesting elements of the research process can be ignored. However, we argue that since designing RHM research activities is an essential, yet also time- and resource- intensive processes, by sharing and using derived and compiled datasets, we can widely leverage the prior research activities of other researchers, as well as conserve and optimize our resources. For example, by using derived data for the PulsePoint Project (specifically, existing review comments), we estimate saving close to 40 hours of work spread across study design, IRB approvals, and recruiting participants.

---

[1] An Association of Computing Machinery's Special Interest Group on the Design of Communication (ACM SIGDOC) research grant provided funding for this research.

Although using derived data provided certain temporal and institutional resource affordances, there were, at the same time, other ethical particularities regarding our project context that we needed to consider: privacy and confidentiality, removal and disconnection from the expected context, and informed consent (see National Academies of Sciences, Engineering, and Medicine, 2015). Although our methodology for this project included analyzing publicly available comments, users who posted this information might not have intended for their words and the events they described to be used for research. Users' intended (and arguably expected) audience were likely the PulsePoint design team and app administrators. For example, one user wrote:

> . . . In May our neighbor crashed his plane at a local airport. I heard the sirens, turned on PulsePoint and feared it was him. I was able to gather my stuff, go get his wife and take her to the scene . . .

It is fair to assume that this user did not consider the confidentiality—the responsibility to keep health information private—of this user's neighbor, nor did the user protect the privacy of their neighbor. In fact, based on the information provided (such as local airport and PulsePoint Respond connected community areas), it would not be difficult to discover who this person might be, even if the commenter's username is a pseudonym. Since users comment in a particular context (such as posting feedback about an app online) for a specific audience (such as an app design team and/or administrator), analyzing these comments must also meaningfully consider the commenter's expectations (for example, a testimony to app efficacy). However, when taken outside of that context, the comment might be used in a manner that does not account for the user's intentions. In this way, protecting the privacy and the identity of the person described in the comment becomes our ethical responsibility (as the researchers who conducted this study), as well as the responsibility of anyone who uses this derived data or our compiled dataset (in our case, we eliminated all usernames in the Excel sheet we created to manage our datasets).

As our example illustrates, accounting for confidentiality and privacy of information, as well as the expected context, are ethical considerations for using existing (i.e., derived) data such as the information in our PulsePoint dataset. For instance, researchers can acknowledge users' expectations regarding who would read their comments and for what purpose in the methodology section of any publications that result from the study. In addition, consent should be central to decision-making when researchers consider using derived, compiled, and/or designed data. In qualitative studies that include human participants and their words, behaviors, and actions, researchers are required to obtain participants' consent. For our PulsePoint project, the majority of users posted comments using pseudonyms, so it would not be possible (in most cases) to seek consent. However, researchers need to carefully consider these decisions on a case-by-case basis. More specifically, scholars who use derived and/or compiled data must reasonably ascertain if IRB-approval is needed to engage in ethical research practices.

These ethical considerations regarding privacy and confidentiality, expected and original context, and consent are not easily ascertained without rhetorical consideration. In fact, discerning between derived, compiled, and designed data adds intricate ethical layers that complicate the process of sharing data. For example, the data we used in the PulsePoint project is derived from existing user comments, and in our content analysis, we compiled and curated this information to serve our particular research purpose.

At the same time, it is also important to point out that in sharing our dataset, subsequent researchers will need to determine how to address comments like the one we include above that divulge potentially sensitive personal information. Further, if we intentionally designed a study that elicited information from identifiable user comments, we would need to submit our research for IRB review. Clearly, the data distinctions that we call attention to in this article—derived, compiled, and designed—account for various ethical and rhetorical concerns regarding privacy and confidentiality, expected context, and consent.

Collecting and Managing Data for an RHM Databank

Equally important to the aforementioned data deliberations, collecting and managing shared RHM data (which might be housed in a databank, for instance) are not without ethical, logistical, and rhetorical difficulties. For example, we propose that designed (intentional and new) data must include specific statements in recruiting materials and informed consent documents that contain language, such as "data collected during this study might be used and analyzed by other researchers in additional studies." These decisions would also take heroic honesty on the part of the researcher explaining their designed study because once the data are shared, realistically how they are used cannot be controlled; researchers would need to make statements akin to the above in their recruiting materials and informed consent documents. For this reason, not all datasets, such as those that are multimodal like video recordings of focus groups and interviews, are ideal prospects for data sharing. However, in such situations it might be possible for researchers to share redacted transcripts and/or observational notes.

Creating a databank for RHM data also requires coordinating human research protection review boards. Since, technically, using derived and compiled data might involve multiple research sites, these kinds of collaborations must agree to Reliance Agreements or IRB Authorization Agreements, which allow one IRB to serve as the formal IRB of record. We suggest, prior to embarking on a designed research study or using derived or compiled data, that RHM researchers broach this possibility with their deans of research and other institutional IRB officials. Scholars working in RHM might consider encouraging IRB administrators to pursue IRB coordination certification or, at the very least, becoming familiar with it. Although typically acquired for clinical trials, IRB coordination certification will likely be useful in thinking through how cooperating IRBs can guide RHM databank contributions. Nonetheless, the role IRBs play in using databank data, whether they are derived, compiled, or designed, should also be carefully considered.

Additionally, there are a host of ethical issues related to informed consent that need to be deliberated for RHM data sharing to even be a possibility. For example, Harris and Wyndham (2015) add that, depending on the kind of information collected (e.g., genomic), informed consent for communities must also be considered (p. 335). If RHM researchers do not bring up the possibility of publicly sharing de-identified data with people and communities during the consent negotiation, it would be an abnegation of their duties as researchers. In other words, if sharing an anonymized dataset is not agreed to by participants during the informed consent negotiation, then that dataset simply cannot be shared.

Furthermore, Michael Parker noted, “Consent alone does not make research ethical . . . [and that] responsible conduct in data sharing requires protections around discrimination, security standards, and standards of confidentiality and privacy . . .” (as cited in National Academies of Sciences, Engineering, and Medicine, 2015, para. 12). Governance and oversight of shared data, whether derived, compiled, or designed, is trickier to ensure. For example, information for the Child Health and Development Studies (2020) datasets states ". . . researchers using the data . . . are required to sign [and are legally bound by] a Confidential Data Disclosure Agreement . . ." (para. 1). However, there is not an oversight agency that specifically ensures the confidentiality of these data. In fact, we wonder if such oversight is possible or if it is just safest to assume that once the data are shared, they are publicly available and virtually uncontrollable.

Logistically, the prospect of an RHM databank involves factors such as cost, usability, accessibility, and security. Such considerations are typical and likely unoriginal considering the wide range of data-sharing databases in science and medicine, such as those managed by the CDC (2020). However, there is murkiness around ownership of compiled and designed data contributions. Some have suggested that custodianship might be a frame for any data collected and shared, while others have advocated for "an approach to data sharing that involves setting the stage for reasonable behaviors at the outset which, in turn, requires reasonable oversight and governance, and a fair exchange" (as cited in National Academies of Sciences, Engineering, and Medicine, 2015, para. 17).

However, these logistical concerns regarding ownership do not end there. As the field of RHM continues to grow across the globe, we will need to consider how nations other than the United States legally and ethically treat data ownership, custodianship, or stewardship. We call those in RHM who hold scholarly spaces in other disciplines to provide direction regarding ownership and identifying (or creating) those guidelines that align with RHM research methodologies and research practices. Furthermore, sharing data across international borders presents concerns, in addition to those of ownership, especially for the kinds of qualitative research prevalent in RHM scholarship. Local or global human research protection boards might provide advice along these lines. For example, if data are collected in Kenya (Mara & Mara, 2018), can they be included in an RHM databank in the United States? Or, if parents of premature infants are interviewed and observed in Denmark (Bivens, 2018), can the transcripts and observation notes be shared in a Canadian databank? Do research participants in Denmark, the researcher, or the databank own the data? The coordination of human research protection boards might assist decision-making in this regard;

however, the logistical deliberations cannot be meaningfully mulled over without attending to rhetorical considerations, as well.

The rhetorical considerations of a databank present an equally complex provocation. Overall, we wonder what is lost (or gained) when we share and use derived, compiled, and designed data. To illustrate, consider notes taken during observations of emergency medical services (EMS; see Angeli, 2018). Since EMS provides critical, timely care, what kinds of re-uses could reflect the original context and expectations regarding that context? To wit, even if observation notes are paired with interview transcripts, can compiled or shared designed data be expected to convey enough information about the context for a different and perhaps uninitiated or novice researcher? For these reasons, we suggest that sharing designed (new and original) textual, computational, and/or visual datasets (collections of images) might provide fewer rhetorical complexities than compiled and designed data involving humans.

For some research projects, it may not be feasible for researchers to use derived and/or compiled data. Yet, as we have endeavored to propose, by sharing designed datasets (such as those from interviews, observations, focus groups, etc.), we can leverage the original research collected by RHM scholars to avoid data waste. Furthermore, by using derived and compiled data—like the PulsePoint project we describe—we can pivot toward rhetorical ethics practices (Baldwinson, 2018) to encourage habits that conserve and optimize intellectual and institutional resources. Although sharing data in certain contexts is impossible, when data can ethically and legally be shared with other researchers or derived or compiled data can be analyzed within the parameters of a project, we suggest doing so. It is possible, and dare we say likely, that such practices, too, might accelerate RHM discoveries—perhaps in surprising and unexpected ways.

Kristin Marie Bivens (@kmbivens) is an associate professor of English at Harold Washington College—one of the City Colleges of Chicago. Her scholarship examines the circulation of health information from expert to non-expert audiences in critical care contexts.

Candice A. Welhausen is an assistant professor of English at Auburn University where she specializes in technical and professional communication. Her research focuses on the ways that data visualizations of public health threats construct knowledge about disease, illness, and health.

References

Amer, Karim, & Noujaim, Jehane. (Directors). (2019). *The Great Hack* [Documentary Film]. Netflix.

Angeli, Elizabeth L. (2018). *Rhetorical work in emergency medical services: Communicating in the unpredictable workplace*. Routledge.

Bivens, Kristin Marie. (2018). Rhetorically listening for microwithdrawals of consent in research practice. In Lisa Melonçon & J. Blake Scott (Eds.), *Methodologies for the rhetoric of health & medicine* (pp. 138–156). Routledge.

Baldwinson, Raquel. (2018). Ethics for rhetoric, the rhetoric of ethics, and rhetorical ethics in health and medicine. *Rhetoric of Health & Medicine, 1*(3–4), 213–238.

Ball, Cheryl. (2013). From big data to boutique data. *Digital Rhetoric Collective.* https://www.digitalrhetoriccollaborative.org/2013/11/12/from-big-data-to-boutique-data/

British Medical Journal (BMJ). (2020). Authors. https://mh.bmj.com/pages/authors/#editorial_policy

Bureau of Labor Statistics (2020). Data. United States Department of Labor. https://www.bls.gov/data/

Centers for Disease Control and Prevention (CDC). (2019). Resources data. *Centers for Disease Control.* https://www.cdc.gov/nceh/tracking/resources_data.htm

Centers for Disease Control and Prevention (CDC). (2020). CDC Wonder. *Centers for Disease Control.* https://wonder.cdc.gov/.

Child Health and Development Studies (CHDS). (2020). Information for researchers. http://www.chdstudies.org/research/information_for_researchers.php

Dainty, Katie N., Vaid, Harris, & Brooks, Steven C. (2017). North American public opinion survey on the acceptability of crowdsourcing basic life support for out-of-hospital cardiac arrest with the PulsePoint mobile phone app. *JMIR mHealth and uHealth*, *5*(5), e63.

Hansen, Carolina Malta, Kragholm, Kristian, Pearson, David A., Tyson, Clark, Monk, Lisa, Myers, Brent, . . . & Granger, Christopher B. (2015). Association of bystander and first-responder intervention with survival after out-of-hospital cardiac arrest in North Carolina, 2010–2013. *Journal of the American Medical Association, 314*(3), 255–264.

Harris, Theresa L. & Wyndham, Jessica M. (2015). Data rights and responsibilities: A human rights perspective on data sharing. *Journal of Empirical Research on Human Research Ethics, 10*(3), 334–337.

JAMA. (2020). JAMA Network. Instructions for authors. https://jamanetwork.com/journals/jama/pages/instructions-for-authors#SecDataSharingStatement

Mara, Miriam, & Mara, Andrew. (2018). Blending humanistic and rhetorical analysis to locate gendered dimensions of Kenyan medical practitioner attitudes about cancer. *Technical Communication Quarterly*, *27*(1), 93–107.

National Academies of Sciences, Engineering, and Medicine. (2015). Committee on Population. *Sharing research data to improve public health in Africa: a workshop summary*. National Academies Press. https://www.ncbi.nlm.nih.gov/books/NBK321546/

National Endowment for the Humanities. (2018). Data management plans for NEH Office of Digital Humanities: Proposals and awards. https://www.neh.gov/sites/default/files/2018-06/data_management_plans_2018.pdf

Nature. (2020). Springer Nature. Research data. https://www.springernature.com/gp/authors/research-data

Ohno-Machado, Lucila, Bafna, Vineet, Boxwala, Aziz A., Chapman, Brian E., Chapman, Wendy W., Chaudhuri, Kamalika, . . . & the iDash team. (2012). iDash: Integrating data for analysis, anonymization, and sharing. *Journal of the American Medical Informatics Association, 19*(2), 196–201.

Ringh, Mattias, Rosenqvist, Mårten, Hollenberg, Jacob, Jonsson, Martin, Fredmen, David, Nordberg, Per, . . . & Svensson, Lief. (2015). Mobile-phone dispatch of laypersons for CPR in out-of-hospital cardiac arrest. *New England Journal of Medicine, 372*(24), 2316–2325.

Welhausen, C.A., & Bivens, K.M. (2021; forthcoming). “mHealth apps and usability: Using user-generated content to explore users’ experiences with a civilian first responder app.” *Technical Communication*.