International Journal of Child and Adolescent Resilience

Brief Report: Data Sharing and Resilience: Turning Lemons into Lemonade

Laura Schwab-Reese[1], and Scottye Cash[2]

1 Department of Health & Kinesiology, Public Health Graduate Program, Purdue University, West Lafayette, IN, USA.

2 College of Social Work, The Ohio State University, Columbus, OH, USA.

Corresponding author: Laura M Schwab-Reese, MA, PhD, Assistant Professor, Department of Health & Kinesiology, Public Health Graduate Program, College of Health & Human Sciences, Purdue University. 765.496.6723, lschwabr@purdue.edu

Abstract:

Sharing and reusing data is an important aspect of research in the United States. When done well, it has the potential to improve research by reducing duplicate data collection, improving statistical analysis techniques and software implementation through low- or no-cost opportunities, and increasing researchers' incentive to minimize errors by encouraging no-cost replication of analyses. Sharing data also allows students and trainees the opportunity to pursue research that would not be otherwise feasible. However, sharing data that was not intended for research purposes is a complex undertaking. The authors recently engaged in a data sharing agreement that, while useful for both research and future research management purposes, was challenging in many ways. This article describes their experiences with a data sharing agreement process, the issues with the agreement, and lessons learned by the authors.

Introduction

Sharing and reusing research datasets has become an important part of research in the United States since the National Institutes of Health started requiring sharing for data that resulted from projects with more than $500,000 in direct costs per year (National Institutes of Health, 2003). Data sharing has many potential benefits, including reduced burden on agencies/organizations to collect or recollect data (Tenopir et al, 2011) and improved coordination across agencies for shared clients, while reducing the data collection burden for individuals served by the agencies (Kingsley & Goldsmith, 2013). For research, data sharing may lead to more efficient use of funding through reduced duplicate data collection, improved methods and statistical analysis techniques, and software implementation through low- or no-cost opportunities to pursue these activities, and increased incentive to minimize errors in research by encouraging no-cost replication of analyses (Piwowar & Chapman, 2010). In addition, data sharing creates additional opportunities for substantive and methods training for students, trainees, and professions (Piwowar & Chapman, 2010; Tenopir et al., 2011).

Several data archives exist to facilitate data sharing between researchers. One of the largest data archives, the Inter-university Consortium for Political and Social Research, houses more than 250,000 research files and 21 discipline-specific collections (Institute for Social Research, 2018). Some large studies also self-archive and manage the data sharing process, including The National Longitudinal Study of Adolescent to Adult Health. This nationally-representative cohort study has resulted in more than 6,000 journal articles, presentations, books, books chapters, and dissertations on many aspects of social, behavioral, mental, and physical health for adolescents (Carolina Population Center, 2017). However, in this type of data sharing, the data were created for research and managed by organizations with extensive experience with sharing, which reduces the complications associated with data sharing. Some data created for non-research purposes, such as data available from the National Child Abuse and Neglect Data System (NCANDS), has been used for many years and so specific data security and ethics processes have been established (National Data Archive on Child Abuse and Neglect (NDACAN), 2018).

Other non-research sources of data, such as social media data, have not been used extensively for child maltreatment studies (Schwab-Reese, Hovdestad, Tonmyr, and Fluke, 2018). As such, there are several ethical and practical concerns that must be considered when exploring novel data sharing agreements. First, data sharing requires careful consideration of the expected privacy and confidentiality by participants and legal and institutional regulations around privacy and confidentiality, which may differ across disciplines and institutions (UK Data Service, 2017). Second, different computer science and statistical analysis skills and expertise are often needed to construct and use data from these different platforms, which increases potential for data management and analysis errors if not conducted by individuals with adequate expertise and skills (boyd & Crawford, 2012). Finally, data sharing processes and agreements that develop without specific, well-conceived guidelines may cause difficulties for both the original owner of the data and the recipient of the data.

A Real-World Example of Data Sharing and Resilience

The authors recently engaged in a data sharing agreement with a technology-based organization that engages adolescents and young adults. While the research findings that resulted from this data sharing agreement are worthwhile, the data sharing process substantially complicated the research process.

Several years ago, one of the authors realized young victims of child maltreatment were likely to seek the support provided by the organization as the platform provided support while allowing users to remain anonymous. Subsequently, she contacted the organization to determine if they collected information on child maltreatment. At the time, they were not collecting or aggregating information on child maltreatment disclosures, but they concluded it would be possible to add a "child abuse/neglect" tag to summaries completed by the workers after the conversation, which would allow identification of trends across time and geographical location. Unexpectedly, they contacted the authors many months later indicating that they would like to share deidentified message-level data. The organization was in the process of hiring an individual who would manage the data agreements, related protocols, and data sharing process, with the intent of piloting the process with a small number of researchers, then expending the sharing process to screened and qualified researchers. Soon after, we sought IRB approval for the analyses, which were determined to be non-human subjects research by the local institutional review board because they were deidentified in such a way that identification of the participants was impossible.

Over the course of nine months, many aspects of the initial data sharing arrangement were altered from the initial agreement. For example, data were initially to be downloaded by the authors through a secure process but were ultimately moved to a secure server with multi-step security protocols and processes. Researchers were not allowed to download or use the data outside of the secure server environment. This change posed a challenge to the researchers who had intended to conduct analyses through statistical and qualitative software, which was not available on the secure server environment. In addition, proficiency in computer language tools, which the authors did not possess, became necessary for data manipulation.

The data agreement process was finalized approximately nine months after the initial data sharing discussion, however the authors experienced ongoing difficulties accessing the new platform. Approximately four months after finalizing the data sharing agreement, one author gained consistent access to the platform and data analysis began. For the next three months, the researchers worked with the open data agreement manager and other personnel on creating data coding dictionaries, understanding the data from both the texter and the crisis counselor, and developing an analytical strategy. Three potential manuscripts were outlined with preliminary data analyses and each paper was discussed with the open data manager. Approximately three months into data analysis, the organization announced that they were terminating all data sharing agreements effective in sixty days. The data sharing agreement was a pilot project to assess if the open data process was feasible so it was possible that the organization determined continued open data sharing was not feasible. The reasons for terminating the data sharing agreement program were not shared.

Lessons Learned: Lemons to Lemonade

The cancellation of the data sharing agreement was not anticipated and posed numerous challenges. First, the analyses and data checks for reliability had to be completed quickly, and additional data analyses often completed during the publication revision process were not possible. Setting up reliability and validity checks throughout the analysis process was instrumental to being able to complete rigorous analyses. Second, the end of the data sharing agreement required all data-related information be destroyed, except for completed results and tables. As a result, when the authors were finalizing the coding and analysis framework, they created a content analysis coding and analysis framework that included specific text examples. Prior to the data sharing agreement termination, the authors complied with the request to destroy all data-related information, but the information available in the content analysis framework may reduce the negative impact on publishing the papers.

Although the data sharing agreement had a disappointing end, the process and outcome of the agreement were important to future research projects with data sharing arrangements and agreements. From a research perspective, the authors developed a coding scheme for analyzing text-based data and wrote several papers based on this work for publication. From a management perspective, the authors learned to include additional assurances in any signed data sharing agreements to minimize the disruptive nature of shifting organizational priorities. While every organization has its reasons for changing agreements from time-to-time, researchers may find it useful to include specific agreement language that protects their ability to conduct and disseminate high quality research, including an agreed upon process for terminating the data sharing agreement and a clearly defined process for organization input in dissemination efforts.

This data sharing agreement was valuable, from both a research and process improvement perspective. Overall, future research may be improved by having an established coding and analysis framework. In addition, the authors will have more informed discussions with future partners on their data sharing agreements, expectations, and commitments. Finally, the experiences from this project, both the positive and negative, may be helpful to other researchers who are embarking on data sharing agreements.

References

boyd, d., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society, 15*(5), 662-679.

Carolina Population Center. (2017). *Publications database*. Retrieved from http://www.cpc.unc.edu/projects/addhealth/publications/database.

Institute for Social Research. (2018). *About ICPSR*. Retrieved from https://www.icpsr.umich.edu/icpsrweb/content/about/.

Kingsley, C. & Goldsmith, S. (2013). *Getting data to the good guys*. Retrieved from https://datasmart.ash.harvard.edu/news/article/getting-big-data-to-the-good-guys-140.

National Data Archive on Child Abuse and Neglect. (2018). *Datasets*. Retrieved from https://www.ndacan.cornell.edu/datasets/datasets-list.cfm.

National Institutes of Health. (2003). *Final NIH statement on sharing research data*. Retrieved from https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html.

Piwowar, H.A., & Chapman, W.W. (2010). Public sharing of research datasets: a pilot study of associations. *Journal of Informetrics, 4*(2), 148-156.

Schwab-Reese, L.M., Hovdestad, W., Tonmyr, L., & Fluke, J. (2018). The potential use of social media and other internet-related data and communications for child maltreatment surveillance and epidemiological research: Scoping review and recommendations. *Child Abuse & Neglect*. DOI: 10.1016/j.chiabu.2018.01.014

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wi, L., Read, E., … Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLOS One, 6*(6), e21101.

UK Data Service. (2017). *Big data and data sharing: Ethical issues*. Retrieved from https://www.ukdataservice.ac.uk/media/604711/big-data-and-data-sharing_ethical-issues.pdf.