White Paper: Building and Supporting the Use of Digital Research Infrastructure Among Qualitative Researchers

December 14, 2020

***Katherine A. Tamminen,** Faculty of Kinesiology & Physical Education, University of Toronto

Heidi Bohaker, Department of History, University of Toronto; Co-Director of the *Great Lakes Research Alliance*

Andrea Bundon, School of Kinesiology, The University of British Columbia

Denise Gastaldo, Bloomberg Faculty of Nursing, University of Toronto

Brenda M. Gladstone, Dalla Lana School of Public Health, University of Toronto; Director, *Centre for Critical Qualitative Health Research*

Cara Krmpotich, Faculty of Information, University of Toronto; Co-Director of the *Great Lakes Research Alliance*

Meghan McDonough, Faculty of Kinesiology, University of Calgary

Brett Smith, Durham University (UK), and on behalf of the *International Society of Qualitative Research in Sport and Exercise*

* The contact person for this White Paper is Katherine Tamminen (katherine.tamminen@utoronto.ca); other authors are listed in alphabetical order by last name.

1. Digital research infrastructure for qualitative researchers: Background and context

The advancement of open science initiatives and corresponding investments in data repositories has proceeded rapidly within such fields as the physical sciences and survey-based, quantitative social sciences (McKiernan et al., 2916). While the existing repositories and infrastructure may accommodate qualitative data files, the engagement of qualitative researchers with open science practices more broadly, and data sharing in particular, is considerably less than that of researchers working with quantitative data - the scope and volume of qualitative data available in open repositories "pales in comparison to the volume of archived quantitative data" available for re-use (Jones et al., 2018). There are several reasons why this might be the case: First, there are several ethical and epistemological concerns that researchers face when deciding whether sharing data from qualitative projects is possible and advisable. The resources and training that currently exist do not sufficiently address the concerns associated with qualitative research; many questions, dilemmas and ethical problems remain. Second, researchers lack guidance, training, and support about the options available to them to support open data sharing: the landscape of data repositories and support is complex and fragmented, and there is a lack of guidance on where and how researchers may store and share qualitative data. Third, the work associated with preparing qualitative data for storage and future use creates additional burden and demands for researchers and participants, for which there is currently little, if any, support (e.g., time, training, compensation).

In developing the digital research infrastructure for qualitative researchers, it is imperative that the infrastructure is built, implemented, and supported in a manner that is accessible by researchers across a variety of disciplines. Critical to supporting researchers' engagement with digital research infrastructure is the development and provision of training and information to make informed decisions about whether it is epistemologically and ethically possible to openly store and share their data. This training and information should be respectful of disciplinary differences and should accommodate the needs and commitments of qualitative researchers to their participants and research sites.

1.1. What do we mean by digital research infrastructure?

Digital research infrastructure refers to the tools and services used across the research process; four elements of this infrastructure are: 1) a digital network for research and education, 2) data management processes, 3) research software, and 4) advanced computing for managing large amounts of data. In the current paper, the issues and challenges raised herein concern primarily data storage in repositories or databases, and associated data management processes (Government of Canada, 2020).

1.2. What is qualitative inquiry and what are forms of qualitative data?

Qualitative inquiry explains social phenomena using high quality, rich information, to address how and why individuals and social groups understand, feel and interact in specific contexts (Facey et al., 2018). Qualitative inquiry, and the production, analysis and interpretation of qualitative data, may help to generate new problems and questions, promote the development of theories that explain social processes and behaviours, and may also be used to evaluate policies, practices, and innovations. Qualitative inquiry is not 'preliminary' or 'ancillary' to 'real research' (Sandelowski, 1997). Qualitative inquiry produces knowledge that is idiographic and naturalistic and which may help us to generate ideas, to 'see differently, and to act differently' (Peshkin, 1993): "the goal of qualitative inquiry is not the mere accumulation of information, but rather the transformation of understanding" (Sandelowski, 1997, p.128). A feature of qualitative inquiry is to be flexible and have an emergent design (Facey et al., 2018); this is a strength as it allows researchers to be attuned and responsive to topics, issues, and new lines of questioning that are identified as important as the study unfolds, adapting their research plan to what is being learnt about the questions, participants, and methods used.

Qualitative projects include various forms of data, including demographic information, interviews (audio recordings and transcripts), field notes and observations of people and their interactions, photos and

videos of research sites and participants, documents, and other media (e.g., public and private social media posts). Qualitative researchers often combine multiple types of data to inform their analyses and interpretations of the phenomenon or topic under investigation. Because the focus of qualitative research is idiographic and naturalistic and attends to individuals' subjective experiences and to context(s), the data collected in qualitative studies can be highly personal and sensitive, and it can be difficult to de-identify participants and settings within a dataset. This is especially so when conducting research with small populations, or those with distinct, unique experiences or expertise. Furthermore, it can be analytically problematic to remove contextual information from datasets in which experiences and interactions are understood.

1.3. Benefits of enhancing digital research infrastructure for qualitative researchers

There are several benefits of enhancing the digital research infrastructure for qualitative researchers. Enhanced infrastructure and engagement in data storage and sharing can promote new research questions and encourage diversity in analysis (Chauvette et al., 2019), it can enable other researchers to build upon, expand, and critique existing analyses, and it can be used for teaching and learning purposes, to support trainees' development and exposure to diverse forms of data and approaches to analysis (Haaker & Morgan-Brett, 2017; McKiernan et al., 2016; Tamminen et al., under review).

2. Current issues: Challenges with digital research infrastructure for qualitative researchers

There are several challenges and fundamental issues facing qualitative researchers engaging with digital research infrastructure to potentially store and share data openly for future use (e.g., for secondary analysis, re-analysis, teaching purposes, etc.). The two core concerns to draw attention to in this paper are the epistemological issues and ethical issues related to qualitative inquiry; a discussion of these issues is necessary as they precede the challenges that many researchers face when considering whether and how to engage with digital research infrastructure.

2.1. Epistemological issues and re-use of qualitative data: Qualitative inquiry is not the same as postpositivist forms of inquiry

The epistemological commitments of many qualitative researchers are founded in interpretivism and subjectivism, which emphasizes the idea that reality and knowledge is relational and gained by accessing individuals' subjective perspectives and interpretations (Jones et al., 2018; Tamminen & Poucher, 2020). Some researchers working from these positions maintain that data are context-dependent and "cannot be re-used by anyone other than the original researcher" (Jones et al., 2018), and that any attempts to re-interpret the data by other researchers for purposes other than the original project are impossible. Alternatively, some qualitative researchers maintain that data re-use and secondary analysis may be possible, with sufficient consideration of the original purposes, context, and conditions under which the data were originally produced (Kuula, 2011).

Replication is a concept and value associated with postpositivist, quantitative research wherein the push toward open science and having access to data is seen as valuable to enable others to verify the analyses and findings of published research (Asendorpf, 2013; Nosek, Spies, & Motyl, 2012). However, the notion of replication and reliability are not relevant for (most) qualitative researchers because these concepts neither fit the logic of qualitative research or work when doing qualitative research (e.g., when interviewing people over time they do naturally change their view and have different experiences). As such, creating a digital research infrastructure for the purposes of 'auditing' qualitative analyses or to 'verify' the validity of interpretations is incompatible with the epistemological positions of many qualitative researchers.

Creating digital research infrastructure can, however, be useful to support researchers in documenting their research process. Storing qualitative data, methods, and materials can also support the sharing of data which can (in some cases) be used for secondary analyses with the goal of producing new

interpretations or knowledge of a phenomenon. We support the value of encouraging a plurality of perspectives in the production of knowledge, rather than resting simply on the idea of storing and sharing data in order to replicate or verify previous research findings.

2.2. Ethical issues and sensitive data: Some data should not be made available for re-use

In some cases, the sensitivity of the research topic and the data is heightened, such that there are risks to participants if data were made available to other researchers for re-use. This concern is exacerbated in much qualitative research, as the data often situate participants in context, providing great detail about individuals and communities. Even in cases where stringent restrictions could be put in place regarding the access and future uses of the data, there are some forms of data and topics of inquiry where the benefits of storing and sharing the data do not outweigh the risks to participants. Researchers must weigh their ethical responsibilities and potential risks to participants against the potential benefits of other researchers having access to the data in the future.

Ethical concerns may arise within research on sensitive topics where participants may face psychological, social, or legal risks if they can be re-identified through subsequent researchers' use and reconstruction or representation of the data – known in ethics as deductive disclosure. For example, individuals who participate in research on topics such as discrimination in the workplace, persecution, illegal behaviours, immigration, and health status (e.g., disclosures of mental health or physical health conditions) could face future harm if their identity could be disclosed. Researchers conducting inquiry with some populations, such as children and youth, may decide that it is too great a risk to ask individuals to consent to an undetermined future use of their data, particularly if these individuals might be identified. Furthermore, it may be an ethical violation of the child's confidentiality to seek parental consent to store the child's data for future use, particularly if the parent wishes to review the child's data before it is stored. Another example arises in ethnographic research or focus group interviews where multiple members of a community or group know one another and where some individuals may wish to have their data stored for re-use, while others do not. In this case, participants who have not consented to the use of their data beyond the original project could risk being identified in the data provided by other participants, and researchers must decide whether they can share any of the data from the project without compromising each individual participants' confidentiality and right to refuse to have their data shared for future re-use. These are only a few examples of cases where some data should not be made open and accessible for re-use; these examples fit the 'extreme risk' classification (Portage Network, 2020) and "should not be deposited anywhere, beyond the direct storage and access needs of the research team" (p.8).

Data ownership and sovereignty are also important considerations for researchers who generate and/or steward Indigenous knowledge and data. How and where data are digitized, stored, and accessed is indicative of attitudes toward intellectual, political, and territorial sovereignty (Kukutai & Taylor 2016; Rainie et al., 2017; Wemigwans, 2018). Researchers in these contexts need access to infrastructures where protocols for digitization, storage, and access can be co-created and negotiated, not dictated.

2.3. Are digital research infrastructure tools and resources being used?

At the moment, there are multiple databases and infrastructure that are available to researchers, but the landscape of the digital research infrastructure is complex and difficult to navigate. A review of the literature in one research field indicates that there is very limited engagement in open science practices, and researchers conducting qualitative studies are using a digital research infrastructure in very limited ways (Tamminen & Poucher, 2018). Challenges for qualitative researchers include:

- Ethical considerations may preclude the possibility of storing data for future re-use.
- Lack of information and training about using digital research infrastructures in an ethical manner in qualitative inquiry.
- Data stewardship decisions are unclear: Most repositories are researcher-controlled, which provides flexibility in what to upload, how to store it, and control over access to research data and resources;

however, researcher-controlled access to data and materials does not address long-term storage and stewardship issues (e.g., after a researcher retires; moves from one institution to another).

- The issue of data ownership varies for qualitative researchers across institutions, and within different jurisdictions. Information regarding the ownership of qualitative data between researchers and participants requires thoughtful consideration. Data ownership has implications for who has the right to grant access to the data in the future and what controls are required.
- The time and burden to prepare materials for storage is high for qualitative researchers and participants. Time must be spent de-identifying data (e.g., redacting information from transcripts, de-identifying people in photos/videos, anonymizing digital media), which may also require advanced technological skills and software. There is considerable participant burden beyond the end of the original study that may be required for participants to make an informed decision about how their data is to be stored (e.g., reviewing redacted transcripts, photos, videos, to ensure that they are comfortable with information that is being stored indefinitely).

3. Future digital research infrastructure for qualitative researchers

Our vision for a cohesive Canadian digital research infrastructure ecosystem is one that provides a flexible system for researchers who choose to store and share qualitative data and project materials while respecting the sensitivity of the data, the ethical commitments to their participants, and the epistemological integrity of their research. Below we outline key considerations for a digital research infrastructure to support qualitative researchers who may wish to engage in open storage and sharing of qualitative data and materials. Some of these are already possible within various data repositories, although not all options are available – for example, Dataverse is a repository which provides researchers with a great deal of control over access to their data; however, this does not provide an option for long-term control and stewardship by a qualified third party. Alternatively, the Qualitative Data Repository (QDR) is an example of a repository where researchers can deposit their data and access is controlled by the host (according to the security/access requirements stipulated by the researchers depositing the data); however, this data repository is located in the U.S. and Canadian researchers may not be able to deposit and store sensitive data on U.S. servers.

Digital research infrastructure that supports the needs of qualitative researchers should:

3.1. Be flexible and optional

Digital research infrastructure for qualitative researchers should provide researchers with the ability to store and share some or all parts of their data and project materials as they and their participants deem appropriate. Respecting the ethical and epistemological issues underpinning qualitative inquiry, engaging in digital research infrastructure should be an optional system to support qualitative inquiry.

3.2. Be capable of storing and cross-linking multiple file types

Given that qualitative researchers often combine multiple sources of data and types of data to inform their interpretations and analysis, it is also important that the digital research infrastructure be capable of storing and maintaining the datasets as a ‘whole.’

3.3. Provide options for researcher-controlled and third-party stewardship

Researchers should have the option to control access to data in repositories (e.g., requests to access data are reviewed and granted by the researcher), and there should also be options for researchers to stipulate conditions of access and security for data to be controlled by a qualified third party (e.g., requests to access data are reviewed and granted by a third party/data steward organization).

3.4. Provide options for archiving and long-term stewardship of data beyond the lifetime of the project and the researcher's career

Options should be available to store data where access is controlled by the researcher, as well as options to archive materials or transfer them to a third-party data steward if appropriate.

3.5. Provide various options for data security and access depending on the sensitivity of the data

Different types of data and project materials may have different levels of sensitivity and require different levels of restriction. Guidelines currently exist for researchers to determine the sensitivity of their data and strategies for de-identifying and storing sensitive data; these guidelines and training should be provided to researchers with specific emphasis on qualitative data.

3.6. Provide resource and training that are specific to qualitative inquiry

Materials and training are currently targeted at primarily quantitative, postpositivist forms of inquiry; practical 'how-to' resources are insufficient to address the complex ethical and epistemological issues facing qualitative researchers. To support qualitative researchers in making use of the digital research infrastructure in Canada, it is imperative to provide training on these issues. Strategies for this training can include modules and webinars similar to the TCPS-2 training required by all researchers at institutions in Canada, as well as training embedded in graduate student education. Topics to be addressed in digital research infrastructure training include:

- **Ethical issues with qualitative data** (confidentiality, anonymity, downstream risks of being identified from qualitative data)
- **Data ownership and sovereignty** (e.g., participant informed consent and decision-making about their data ownership; data collected and produced by Indigenous communities)
- **Preparing data for storage** (from raw, unprocessed data to redacted, summarized, processed data)
- **Determining the degree of sensitivity of their data**
- **Determining the level of restriction required when storing and sharing data**
- **Data attribution, terms of use and re-use, copyright**

4. How to Bridge the Gap

4.1. Because qualitative researchers may have specific concerns and hesitation about engaging in open data sharing, NDRIO should aim to target qualitative researchers for contribution and participation during the upcoming town hall meetings as part of their needs assessment process to consult on issues and concerns related to the development of digital research infrastructure.

4.2. Education and training are paramount to support researchers in making informed decisions about engaging in digital research infrastructure. NDRIO can support this initiative by providing a platform for disseminating information and education to researchers about the digital research infrastructure options available to them as well as training on the ethical and epistemological implications of using digital research infrastructure.

4.3. As one of the key tasks in developing digital research infrastructure in Canada concerns the training and retention of highly qualified personnel, resources should be allocated to the training and retention of HQP with expertise in qualitative research.

4.4. Representation of qualitative researchers should be supported through membership on the NDRIO Researcher Committee and/or consultations with qualitative researchers working from various epistemological positions and across a variety of disciplines. This representation should be supported to contribute to discussions about the long-term stewardship of qualitative data that requires careful controls and restrictions, particularly with sensitive, contextual data that could potentially identify participants if it were shared with other researchers for secondary analysis and re-use.

References

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J., Fiedler, K., ... & Perugini, M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality, 27*(2), 108-119. https://doi.org/10.1002/per.1919

Chauvette, A., Schick-Makaroff, K., & Molzahn, A. E. (2019). Open data in qualitative research. *International Journal of Qualitative Methods, 18,* 1-6. https://doi.org/10.1177/1609406918823863

Facey, M., Gastaldo, D., Gladstone, B., & Gagnon, M. (2018) *Centre for Critical Qualitative Health Research - Learning and teaching qualitative health research in Ontario: A resource guide.* Toronto: E-Campus Publication http://qualitativeresearchontario.openetext.utoronto.ca/

Government of Canada. (2020). *Digital research infrastructure.* https://www.ic.gc.ca/eic/site/136.nsf/eng/home

Haaker, M., & Morgan-Brett, B. (2017). Developing research-led teaching: Two cases of practical data reuse in the classroom. *Sage Open, 7,* 1–9. http://dx.doi.org/10.1177/2158244017701800

Jones, K., Alexander, S. M., et al. (2018). *Qualitative data sharing and re-use for socio-environmental systems research: A synthesis of opportunities, challenges, resources and approaches.* SESYNC White Paper. https://doi.org/10.13016/M2WH2DG59

Kukutai, T. & Taylor, J. (Eds.) (2016). *Indigenous data sovereignty: Toward an agenda*. ANU Press.

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., ... & Spies, J. R. (2016). Point of view: How open science helps researchers succeed. *elife, 5,* e16800. https://dx.doi.org/10.7554/eLife.16800

Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7*(6), 615-631. https://doi.org/10.1177/1745691612459058

Portage Network. (2020). *Sensitive data toolkit for researchers. Part 2: Human participant research data risk matrix*. Retrieved from http://doi.org/10.5281/zenodo.4060448

Rainie, S. C., Schultz, J. L., Briggs, E., Riggs, P., & Palmanteer-Holder, N. L. (2017). Data as a strategic resource: Self-determination, governance, and the data challenge for Indigenous nations in the United States. *International Indigenous Policy Journal, 8*(2). https://doi.org/10.18584/iipj.2017.8.2.1

Sandelowski, M. (1997). "To be of use": Enhancing the utility of qualitative research. *Nursing Outlook, 45*(3), 125-132. https://doi.org/10.1016/S0029-6554(97)90043-9

Sparkes, A. C., & Smith, B. (2009). Judging the quality of qualitative inquiry: Criteriology and relativism in action. *Psychology of Sport and Exercise, 10,* 491–497. https://dx.doi.org/10.1016/j.psychsport.2009.02.006

Tamminen, K. A. & Poucher, Z. A. (2020). Research philosophies. In D. Hackfort & R. Schinke (Eds.), *The Routledge international encyclopedia of sport and exercise psychology* (vol.1: Theoretical and methodological concepts). Routledge.

Tamminen, K. A., Poucher, Z. A. (2018). Open science in sport and exercise psychology: Review of current approaches and considerations for qualitative inquiry. *Psychology of Sport and Exercise 36,* 17-28. https://doi.org/10.1016/j.psychsport.2017.12.010

Wemigwans, J. (2018). *A digital bundle: Protecting and promoting indigenous knowledge online.* University of Regina Press.