Curating QDAS Research Objects for Exchange and Reuse

The Qualitative Data Repository at Syracuse University (QDR) seeks a 3-year National Leadership Grant for Libraries to study the sharing, reuse, and preservation of research objects produced by and dependent upon qualitative data analysis software (QDAS). The recent development of a data exchange standard for QDAS provides an exciting opportunity to study these problems, as well as to develop open source tools that enable the reliable preservation and reuse of QDAS research objects in data repositories. Dr. Sebastian Karcher and Dr. Colin Elman at QDR, Dr. Nicholas Weber from the Information School at the University of Washington, and Dr. Nathaniel Porter from Virginia Tech University Library, will address: 1) Which components of QDAS-based research are most readily shareable and re-usable and for what purposes? 2) How can effective guidance for researchers improve the shareability and reuse of qualitative data? 3) How can repositories take advantage of new developments in QDAS formats to facilitate the deposit and allow the exploration of such data? 4) How can shared QDAS data be used to enhance instruction in qualitative methods and QDAS tools?

National Need: QDAS empowers researchers to systematically analyze and code qualitative data. QDAS applications such as NVivo and ATLAS.ti are used widely across the health and social sciences. Correspondingly, academic libraries spend considerable resources in purchasing QDAS licenses and training users. QDAS research objects (the combination of data, annotations or codes, and resulting analytic outputs) are critical to qualitative scholarship, but they are rarely shared.This is despite the fact that there is now a widely held consensus that open science provides better outcomes for individual researchers, scientific communities, and for society at large. Similar benefits would follow from making available QDAS research objects, including codebooks and, where possible, fully coded projects. Sharing these data and materials would significantly improve the transparency of the large amount of qualitative research that depends upon QDAS, and at the same time produce invaluable resources for instruction.

The sharing of QDAS outputs has been blocked, partially, by the fact that most QDAS is proprietary, and each different QDAS application depends upon a unique data model (Corti & Gregory, 2011). Incompatibilities between these data models made it impossible to reuse or combine outputs between the different tools. However, recently the developers of leading QDAS applications agreed on REFI-QDA, an open exchange standard that allows the outputs of any QDAS application to be ingested in, and reused by, other applications. This new standard has the potential to catalyze an extraordinary increase in qualitative data sharing.

The REFI-QDA standard alone, however, will not bring about this result. Three other needs must also be met: First, QDAS user communities need **concrete guidance** for how researchers can best exchange, reuse, and receive credit for their QDAS research objects. Second, barriers to the exchange and reuse of QDAS research objects can also be greatly eased by the **development of open-source tools** that are integrated with existing scholarly research workflows, and that include managing, depositing, and exploring such data within trusted storage environments. Such tools would allow for easy exclusion of specific files or codes from a package prior to deposit and would allow online exploration of QDAS data at the repository. Third, we need to encourage uptake by **developing training materials** that help librarians to employ real world examples of QDAS research objects as they teach students how to use QDAS applications.

As the foremost social science repository in the United States with a dedicated focus on qualitative data, QDR is uniquely positioned to build on the unprecedented opportunity offered by REFI-QDA, and to address the national need for sharing QDAS outputs. QDR is expert in the use of QDAS, and in sharing QDAS outputs. Moreover, QDR is the only US-based repository advising REFI on configuring the standard to maximize its potential to archive QDAS projects. In addition, as a leading participant in dialogues about openness, QDR has a comprehensive understanding of the ethical and practical challenges of qualitative data sharing.

Project Design: Work under this grant will take advantage of significant previous *exploratory* work by QDR on the topic, including a workshop with practitioners and software developers (Karcher & Pagé, 2017) and several existing data projects based on QDAS materials. The bulk of the grant work will consist in *piloting* approaches to better sharing QDAS materials and then working towards *scaling* their availability and use. The **first stage** in the project will complement prior work with a systematic survey of researchers on their current practices and views on sharing QDAS data, building on and extending existing surveys (Xiao et al., 2014). The survey will provide a broad and rigorous overview of QDAS users, as well as identify potential depositors for collaboration.

With the help of an external selection committee with representation from diverse backgrounds, we will then invite a group of researchers to deposit their QDAS data, and we will support their additional work required with a stipend. These piloteers will work closely with QDR curators and keep a structured log of their activities.

The **second phase** of the grant will build on the survey and deposited data to conduct four related activities:

1) We know from previous work on QDAS archiving that a tool will be necessary for existing data repositories to ingest, publish, and make QDAS research objects meaningfully accessible. We will develop a REFI-QDA repository tool that streamlines depositing QDAS data into any Dataverse repository and allows for online exploration of shared data within a secure environment that can protect sensitive research products.
2) We will analyze the deposit and curation of the pilot projects, and develop guidance in the form of manuals and FAQs. QDR will coordinate the joint efforts of researchers, archivists, and instructors as they develop guidance materials, so that QDAS projects can be safely and effectively shared.
3) We will use the deposited data and guidance materials for instruction in different formats and venues (including both one-off workshops, classroom and graduate seminars) at the three partner institutions. Based on student and instructor feedback, we will further evaluate the characteristics of a maximally re-usable data project, and update the guidance materials accordingly.
4) The initial work teaching with deposited data will also lead to the development of a set of pedagogical instructions based on QDAS data, including lesson plans.

Diversity, Equity, and Inclusion: The project seeks to support diversity, equity and inclusion by integrating and supporting traditionally underrepresented populations in three groups: piloteers, data subjects, and data users. A minimum of one third of pilot projects will be by researchers from traditionally underrepresented groups, and we will work directly with researchers to ensure participation provides not just financial support but substantive benefit in furthering their research and visibility. Likewise, projects whose subjects reflect diverse populations will be prioritized for selection, with instructional material developed in collaboration with researchers. Finally, we will actively seek out and collaborate with instructors at under-resourced universities in developing and promoting instructional materials using shared QDAS data.

National Impact: The **final year** of the grant will shift to fostering wide adoption of the guidelines, technical tools, and instructional resources. The wider availability of shared QDAS projects will present a significant step towards more transparent qualitative research, allowing for qualitative data re-use, and greatly improving instruction in qualitative methods. The guidance for creating shareable QDAS projects will be published as a CC-BY-licensed whitepaper as well as a publication in a journal targeting QDAS users and presented at disciplinary conferences. We will also work with other repositories archiving qualitative data to implement versions of our guidelines. Our proposal to author a REFI-QDA curation primer has been accepted by the Data Curation Network (DCN) and we have ensured participation of a DCN mentor in the creation of the primer.

Software tools developed as part of the grant will be released under a free/libre and open source license. The tools will be integrated with QDR's Dataverse instance, we will support their adoption in other Dataverse repositories, facilitate adoption by other repositories. We will present the tools at Open Repositories and the Dataverse Community Meeting and provide detailed implementation documentation. QDR's technical team has a significant record of adding features to the Dataverse codebase, and we have received statements of collaboration from the Dataverse team at Harvard as well as the Global Dataverse Community Consortium to make QDAS-related enhancements available to all (currently 60) Dataverse installations worldwide.

Improving instruction in qualitative methods and tools is a key benefit of shared QDAS materials. To empower data and instructional librarians, who frequently teach and support QDAS, we will hold workshops on teaching QDAS with shared data at the IASSIST annual meeting and the RDAP summit and openly publish lesson plans.

Budget: We anticipate the total cost of the project will be $340,132, which includes no cost sharing. PI salary and benefits totals $38,334. Support for a graduate research assistant at Syracuse during year 1 and 2 of the grant will be $53,347. Costs for technical development are $90,000. Virginia Tech Libraries and the University of Washington, Seattle, will receive subawards totalling $104,451. Incentives for survey participation and stipends for 15 selected pilot QDAS projects will be $34,500. Conference travel for team members, particularly during year 3 or the grant is budgeted at $9,000, and an additional $10,500 will cover expenses such as catering and room costs for workshops during year 3.