

TNA ACTIVITY REPORT

A sprint for FAIRer data: Development of features for data exchange and publication in research software for qualitative data analysis

Author: Radim Hladík

Current position: assistant researcher

Affiliation: Institute of Philosophy of the Czech Academy of Sciences

Host institution: TCDH, University of Trier

Mentor(s): Prof. Dr. Christof Schöch

Period of stay: 2024-04-14 / 2024-05-18

Introduction of the project

The project aimed to develop new features in a researcher software for computer-assisted qualitative data analysis (CAQDAS). The software, named Requal, is a free and open-source CAQDAS application engineered as an interface for annotating text and qualitative data. Its primary objective is to enhance reproducibility in qualitative research by promoting transparency, openness, reflexivity, and collaboration. The application operates cross-platform in web browsers and is built with the Shiny framework for the R language environment. The project focused on additional features that would increase Requal's capacity to handle data according to FAIR (Findable, Accessible, Interoperable, Reusable) principles.

The software's basic functionality consists of uploading documents, typically interview transcripts or field notes, which can then be "coded" – i.e. text segments are marked with distinctive labels – according to diverse methods employed in qualitative research. The coded segments can then be filtered and categorized, whereby the software aids researchers in making sense of the collected data and build theories of the phenomena of interest. Requal can operate either as a locally run application or as a self-hosted service that enables collaboration with adjustable permissions for team members. In the latter capacity, Requal has been used especially for teaching qualitative coding methods at universities, which can immensely benefit from using free and open source tools as opposed to the expensive proprietary solutions.



Project Goals

The project focused on new features that will enhance the software's mission to support and increase reproducibility of qualitative research. The access to the TCDH infrastructure and expertise enabled an extensive "sprint", i.e., a set period devoted to feature development for the software. Two specific features were proposed: 1) Import and Export of data based on REFI-QDA exchange format based on XML, which has been established as a standard in qualitative data analysis. 2) Anonymization of data to improve data sharing culture in qualitative data analysis.

1) Import and Export: By enabling users to exchange data between different software solutions, the open and reproducible character of Requal would be significantly strengthened. The export ability is particularly important to prevent users from getting locked in with their research projects. REFI-QDA exchange format based on XML has been established as a standard for data exchange in qualitative data analysis.

2) Anonymization of data: Although data sharing is becoming a mainstream practice in quantitative research, in qualitative research the culture of dataset publication remains weak. One of the reasons is that qualitative researchers often deal with sensitive data, such as interviews or diaries, where privacy is a serious concern. Providing researchers with tools to simplify anonymization of data can contribute to increased willingness to publish data, which in turn will increase the transparency in qualitative research.

Stay at the research infrastructure

The Trier Center for Digital Humanities (TCDH) was chosen as an optimal host for this project due to its commitment to open science principles and its extensive expertise in research software development. TCDH maintains and develops a portfolio of research software tools. One of its flagship software applications is FuD - a digital research environment for social sciences and humanities, which is a comprehensive system that allows researchers to annotate and analyse texts or prepare digital editions. TCDH's team includes a sizeable group of software engineers experienced in developing applications tailored for researchers' needs. This expertise ensures that the tools provided are both robust and user-friendly, facilitating advanced scholarly work across various disciplines.

The mentor for the project was Prof. Christof Schöch, one of the TCDH's directors. The mentor was chosen due to his previous work on the notion of reproducible research in the humanities.¹ His extensive scholarship and managerial experience enabled him to provide excellent mentoring. This guidance not only clarified concepts related to reproducibility but also offered valuable advice on developing a sustainable research software project. In addition to theoretical guidance, the mentor therefore provided practical advice on various aspects of software development specific to academic projects in the humanities. This included best practices for coding, documentation standards, and strategies for community engagement and collaboration.

¹ Schöch, Christof. 2023. "Repetitive Research: A Conceptual Space and Terminology of Replication, Reproduction, Revision, Reanalysis, Reinvestigation and Reuse in Digital Humanities." *International Journal of Digital Humanities*.

After settling on feasible dates with Prof. Schöch, the research visit at the CLS infrastructure lasted from April 14 to May 18, i.e., 5 full weeks. The duration was one week less than proposed, due to scheduling constraints, but sufficient to proceed with the project. Ahead of the visit as well as during the stay, TCDH provided me with necessary information and administrative support. Office space and IT equipment were made available as needed. The CLS Infra fellowship adequately covered all incurred costs.

Project implementation

The project implementation commenced with a thorough study of the REFI-QDA documentation, necessary for proper parsing of export files and mapping REFI-QDA onto Requal's data model. A sample file generated by proprietary software was obtained, followed by the development of parsing and import functions. Additionally, a user interface for importing projects was designed. Export functions required reverse mapping, revealing that REFI-QDA cannot fully capture Requal's collaborative features. Conversely, REFI-QDA encompasses objects not currently supported in Requal. Thus, non-technical aspects of the import/export functionality rely on identifying optimal two-way overlaps and planning future adjustments to Requal's data model to better align it with the REFI-QDA standard.

During the implementation, participation in TCDH's research colloquium was requested to present the project on May 2. A presentation was prepared, including a live demo of the Requal software and its import function under development. This event proved particularly useful, as it facilitated explaining the purpose of the research visit to the local research community. Following the presentation, extensive feedback was received from TCDH colleagues.

Discussions with TCDH members, notably Dr. Bremm, Dr. Petkov, and Dr. Hinzmann, steered the development work in new directions. Insights into the backend of TCDH's FuD tool revealed similarities in text tagging functions that inspired tasks for refactoring Requal's code. The highlight system was altered to better accommodate tag color importation and support future data anonymization features. The refactored highlight system now allows switching between background highlighting and underlining. This system is supported by newly developed JavaScript functions that enhance speed and efficiency for the server version of Requal. New CSS styles were also defined to improve the software's user interface further. Additional improvements from brainstorming sessions with TCDH staff included enabling a quick tag option based on the iframe method for maintaining text selection in the UI.

Finally, after consultation with Prof. Schoch, effort was dedicated to preparing a proposal for the OSCARS call for Open Science Projects. The application aims to sustain future Requal development and continue the work initiated during the CLS Infra fellowship. Particular emphasis was placed on anonymizing research data using Requal. An extensive literature review was conducted for this purpose, and the application was submitted on May 15. During this process, a redesigned static website for the Requal software was prepared and launched.

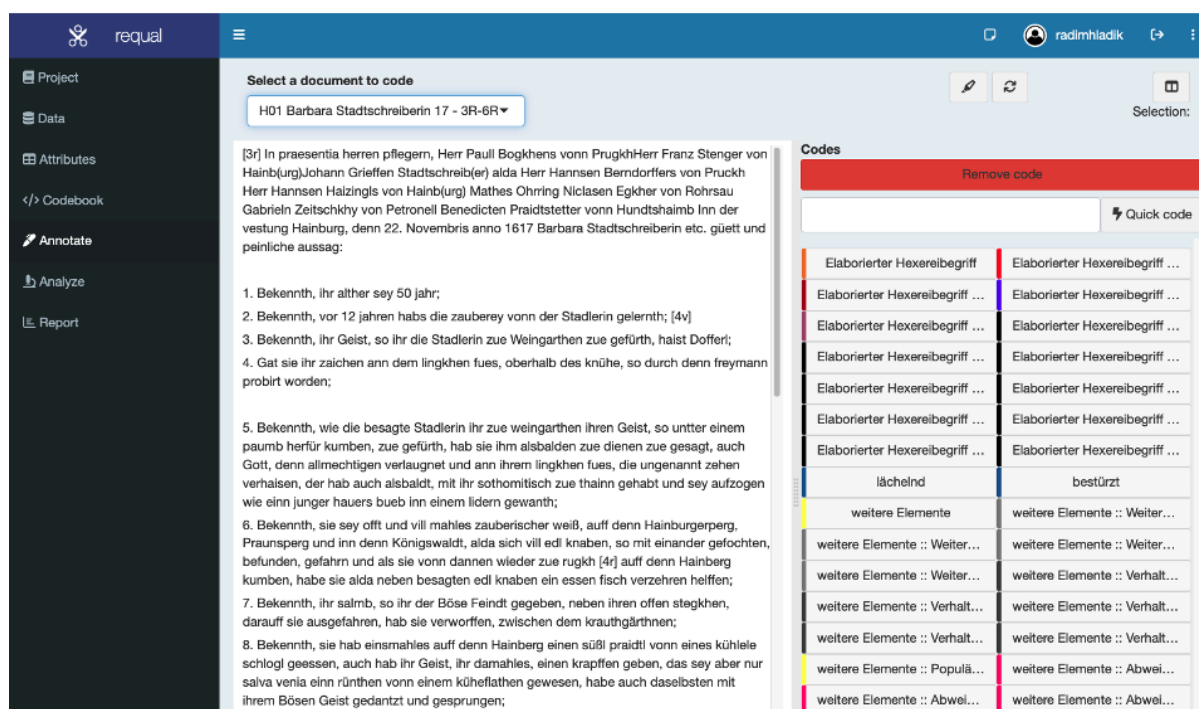


Figure 1: Screenshot of a document with a codebook imported into Requal software

Outcomes and future work

The main outcome of the project is a prototype import function for the REFI-QDA format in the Requal software. The feature is currently being tested before its inclusion in the official release. The development of export functionality remains a work in progress. Anonymization features were researched conceptually via a literature review and deemed outside the practical scope of the CLS Infra project due to technical and conceptual challenges, which will be better addressed in a separate, more extensive follow-up project.

Feedback from TCDH staff led to unforeseen tasks that resulted in important changes to both the user interface and the backend of the application. Consequently, the software eliminated several bugs and enhanced its efficiency. Requal is now a more robust tool thanks to the CLS Infra fellowship and TCDH personnel's assistance. Improvements immediately accessible to users were included in the official release, which was made available at the end of the fellowship.²

Future work entails completing the export functionality to ensure Requal's full compliance with the REFI-QDA format. Developing anonymization features will require additional time and resources. Securing these resources is a priority in the post-fellowship phase. Finally, the larger team behind Requal plans to publish a research paper to introduce the software to qualitative researchers. The valuable support provided by the CLS Infra fellowship will be duly acknowledged.

² <https://github.com/RE-QDA/requal/releases/tag/v1.1.2>