

Processing Qualitative Interview Data - Development of a Software Platform to Support Open Data in the Humanities

Mollenhauer, Sabina

sabina.mollenhauer@uni-vechta.de
Universität Vechta, Deutschland

Introduction

Humanities research in Germany and Europe aligns with overarching strategies by NFDI and EOSC with technologies that support aims towards open science with digitally available shared open data (RfII, 2023). Despite this, humanities in Germany are not well represented digitally with only a small portion of researchers and data participating (Wuttke, 2022). Needs assessments in the humanities and social sciences underscore that a substantial portion of research data comprises textual and survey data (Imeri and Danciu, 2017, 10; Schopfel and Prost 2016, 100; Strunk 2018, 22–23). This dissertation thus delves into the complexities of digital humanities research data processing, particularly within the German context, with a focus on qualitative interview transcripts. It examines the challenges posed by limited software support for digital open data sharing.

While researchers acknowledge the importance of digital tools and value the reusability of research data, especially if tied to publications (Imeri and Danciu 2017, 19–20), barriers such as GDPR-compliance and resource-intensive de-identification processes hinder digital collection and preservation efforts (*ibid.*, 28).

In essence, the dissertation project explores the user and system requirements (Maiden, 2008) for humanities researchers willing to share qualitative interview data and evaluates the extent to which these requirements can be met by targeted software support with the aim of developing such software.

State of Research and Software

In current practice, qualitative interview data undergoes predominantly manual processing, often aided by personal consultation (Adena et al., 2020). This includes steps such as transcription, de-identification, and modeling before sharing or publication. The decision-making process for GDPR compliant processing can theoretically be aided by interactive virtual assistance BERD@NFDI (Herklotz et

al., 2021). Tools exist for de-identification through Named Entity Recognition or privacy-preserving data publishing, as well as for interoperability processing through metadata enrichment and data modeling (Pilán et al., 2022; Kuchma, 2018; Strathern et al., 2020; Prasser et al., 2020; Templ, 2008). However, these tools have not significantly increased the sharing of qualitative interview data in humanities research.

NFDI consortia aim to provide support and create technical infrastructure, yet an unfulfilled usability gap remains. NFDI4Culture provides user stories addressing digital data handling in the digital humanities, but it lacks coverage on the digitization of qualitative interview data (NFDI4Culture, 2023). NFDI4Memory has identified problem stories, including one concerning the anonymization of qualitative empirical data. However, this story lacks an assigned task area (Paulmann et al., 2022). Further research is needed to discover requirements that bridge this usability gap.

Research Project and Questions

The research project's overarching question focuses on identifying barriers faced by humanities researchers open to sharing their data and developing methods to mitigate these barriers through a technically guided workflow. An exploration of available infrastructure and services raises questions about existing technological solutions within the digital data pipeline and their adaptation for qualitative interview data processing.

Answering these questions will guide tool development within one specific aspect of the digital data pipeline. For instance, anonymization and automated data modeling are complex processes in and of themselves. At the same time, another central aspect of the digital data pipeline is the integration of each of these individual steps into a user-centrally designed user interface. Further research must therefore emphasize identifying the most feasible solution to bridge these gaps in the digital data pipeline of qualitative interview data generated in humanities research.

To address these overarching questions, the dissertation research adopts an iterative process. Initial exploratory steps involve literature reviews of existing needs assessments in the humanities, coupled with expert interviews. The methodology, inherently qualitative and based on grounded theory (Truschkat et al., 2011), ensures adaptability to the evolving landscape of digital data pipelines for open data.

The iterative process of software development begins with data collection through qualitative expert interviews (Helfferich, 2022) and contextual inquiry, utilizing qualitative methods from psychology, sociology, and anthropology (Raven and Flanders, 1996; DeBellis and Haapala, 1995). In conjunction with classic software requirement engineering (Bednar and Welch, 2009; ISO/IEC/IEEE 29148, 2011), this process is referred to as a context-sensitive software development process (Siadat and Song, 2012).

Initial evaluation of existing tools and interviews with technical experts is carried out as system requirement analysis. This analysis, focusing on the diverging architectural paradigms in open research data, aims to ensure that the existing open data pipeline is optimized. The goal is for the developed tool to be part of a structure that facilitates its persistence, remaining relevant to future developments in existing architectural paradigms (Diepenbroek et al., 2023; Machado, Costa, and Santos, 2022).

Bibliographie

- Adena, Maja, Christian Abmann, Doris Bambey, Katarina Blask, Andreas Blätte, Michael Bosnjak, Daniel Buck, et al.** "Consortium for the Social, Behavioural, Educational, and Economic Sciences (KonsortSWD)." Publisher: Zenodo, July 31, 2020. visited on 04/17/2023. <https://doi.org/10.5281/zenodo.3968372>. <https://zenodo.org/records/3968372>.
- Bednar, Peter M and Christine Welch.** 2008. Contextual inquiry and requirements shaping. In *Information Systems Development: Challenges in Practice, Theory, and Education Volume 1* (pp. 225-236). Boston, MA: Springer US.
- DeBellis, Michael, and Christine Haapala.** "User-centric software engineering." *IEEE Expert* 10, no. 1 (1995): 34-41.
- Diepenbroek, Michael, Ivaylo Kostadinov, Bernhard Seeger, Frank Oliver Glöckner, Marius Alfred Dieckmann, Alexander Goesmann, Barbara Ebert, Sonja Schimmler, and York Sure-Vetter.** "Towards a Research Data Commons in the German National Research Data Infrastructure NFDI: Vision, Governance, Architecture." In *Proceedings of the Conference on Research Data Infrastructure*, vol. 1. 2023.
- Helfferich, Cornelia.** "Leitfaden-und experteninterviews." In *Handbuch Methoden der empirischen Sozialforschung*, pp. 875-892. Wiesbaden: Springer Fachmedien Wiesbaden, 2022.
- Herklotz, Markus, Lars Oberländer, Irene Schumm, and Renat Shigapov.** 2021. iVA: Ein interaktiver Virtueller Assistent von BERD@ BW zur Aufbereitung von Rechtsfragen im Bereich Open Science. *Tage 2021*
- Imeri, Sabine and Ida Danciu.** 2017. Open Data. Forschungsdatenmanagement in den ethnologischen Fächern. Auswertung einer Umfrage des Fachinformationsdienstes Sozial- und Kulturanthropologie an der Universitätsbibliothek der Humboldt-Universität zu Berlin 2016. Teil I: Statistiken. http://www.evifa.de/cms/fileadmin/uploads/Umfrage_Bericht_Statistiken_1.0_14-06-2017.pdf (visited on 07/13/2023).
- ISO/IEC/IEEE 29148: 2011.** "ISO/IEC/IEEE International Standard—Systems and software engineering—Life cycle processes—Requirements engineering."
- Kuchma, Iryna.** 2018. "OpenAIRE Services for Open Science." *ITlib: Informacne Technologie a Kniznice* 4.
- Machado, Inês Araújo, Carlos Costa, and Maribel Yasmina Santos.** "Data mesh: concepts and principles of a paradigm shift in data architectures." *Procedia Computer Science* 196 (2022): 263-271.
- Maiden, Neil.** "User requirements and system requirements." *IEEE Software* 25, no. 2 (2008): 90-91.
- NFDI4Culture.** 2023. User Stories - NFDI4Culture. url: <https://nfdi4culture.de/de/ressourcen/user-stories.html> (visited on 07/13/2023).
- Paulmann, Johannes, John Wood, Klaus Ceynowa, Fabian Cremer, Silvia Daniel, Daniel Fähle, Barbara Fichtl, Peter Haslinger, Torsten Hiltmann, Rüdiger Hohls, Ursula Lehmkuhl, Marina Lemaire, Gerald Maier, Ole Meiners, Gisela Minn, Katrin Moeller, Andreas Neuburger, Claudia Prinz, Matthias Razum, Harald Sack, Johannes Sauter, Hildegard Schäffler, Eva Schlothuber, Stefan Schmunk, Arnost Stanzel, Helmuth Trischler and Thorsten Wübbena.** 2022. "NFDI4Memory. Consortium for the historically oriented humanities. Proposal for the National Research Data Infrastructure (NFDI)."
- Prasser, Fabian, Johanna Eicher, Helmut Spengler, Raffael Bild, and Klaus A. Kuhn.** 2020. "Flexible data anonymization using ARX—Current status and challenges ahead." *Software: Practice and Experience* 50, no. 7: 1277-1304.
- Raven, Mary Elizabeth, and Alicia Flanders.** 1996. "Using contextual inquiry to learn about your audiences." *ACM SIGDOC Asterisk Journal of Computer Documentation* 20, no. 1: 1-13.
- RfII.** 2023. RfII-Bericht „Föderierte Dateninfrastrukturen für die wissenschaftliche Nutzung“ – März 2023. url: <https://rfii.de/download/rfii-bericht-foederierte-dateninfrastrukturen-fuer-die-wissenschaftliche-nutzung-maerz-2023/> (visited on 04/17/2023).
- Siadat, Seyed Hossein, and Minseok Song.** "Understanding requirement engineering for context-aware service-based applications." (2012).
- Schopf, Joachim, and Hélène Prost.** 2016. "Research data management in social sciences and humanities: A survey at the University of Lille (France)."
- Strathern, Wienke, Moritz Issig, Kati Mozygema, and Jürgen Pfeffer.** 2020. QualiAnon - The Qualiservice tool for anonymizing text data. en. Tech. rep. TUM-I2087.
- Strunk, Sonja.** "Evaluationsbericht 2018: Bedarfserhebung Fachinformationsdienst (FID) Soziologie." (2018): 41.
- Truschkat, Inga, Manuela Kaiser-Belz, and Vera Volkmann.** "Theoretisches Sampling in Qualifikationsarbeiten: Die Grounded-Theory-Methodologie zwischen Programmatik und Forschungspraxis." *Grounded theory reader* (2011): 353-379.
- Wuttke, Ulrike.** 2022. "Wege bereiten, vermitteln und Denkräume schaffen! Reflexionen zu institutionellen und infrastrukturellen Erfolgsfaktoren für Digital Humanities an deutschen Universitäten auf Grundlage

von Expert*inneninterviews." *Zeitschrift für digitale Geisteswissenschaften* 2022, no. 7