



Electronic Communications of the EASST
Volume 86 (1) Year 2026

deRSE26 - Selected Contributions of the 5th Conference for Research Software Engineering in Germany (Part 1)

Edited by Bernd Flemisch, René Caspart, Bernadette Fritzsch, Florian Goth, Jean-Noël Grad, Florian Thiery

CAQDAS Stakeholders and Their Relations to Open Science Practices

Jan Küster

DOI: 10.14279/eceasst.v86.1.2810

License: This article is licensed under a CC-BY 4.0 License.

Electronic Communications of the EASST (https://eceasst.org).
Published by **Berlin Universities Publishing**
(https://www.berlin-universities-publishing.de/)

CAQDAS Stakeholders and Their Relations to Open Science Practices

Jan Küster [1]

ZeMKI, University of Bremen, Germany[1]

Abstract: This study identifies potential stakeholders in CAQDAS and explores their relations to known Open Science practices. It builds on the premise that Open Science concepts are strongly influenced by quantitative disciplines and cannot be directly and unreflectively applied to qualitative research. An exploratory qualitative study using Applied Thematic Analysis was conducted on ten interviews with researchers, developers, and experts in qualitative research and/or Open Science. Findings revealed a complex stakeholder landscape in which actors may occupy multiple, context-dependent roles rather than mutually exclusive categories. Open Science was mainly reflected as process-oriented and contextually negotiated transparency, shaped by methodological, ethical, legal, and practical constraints. Collaborative qualitative research was reported to introduce data management concerns involving CAQDAS and external infrastructure. The findings largely align with existing literature and provide design implications for Research Software Engineers as precursors to future requirements engineering activities.

Keywords: rse,qualitative research,caqdas,open science,applied thematic analysis

1 Introduction

Qualitative research (QR) uses a wide variety of analytical methods, and software is increasingly used to support them [KR20]. This was originally described as "Computer Aided Qualitative Data Analysis" (CAQDAS) [FL93]. In the literature, as well as in this work, the term is often used to refer to *Computer Aided Qualitative Data Analysis Software*. These types of software help to manage and annotate (code) qualitative data, such as text, audio, or video, to uncover complex relationships, and to make the research process more transparent. In this context, they support digital workflows across larger parts of the QR process [Pau23]. However, their use also influences how QR activities are performed and documented, indicating the need for technological reflexivity [O'K20] [PL24]. Recent work, for example, demonstrates that differences between qualitative analysis tools can shape researchers' strategies and analytic decisions [Whe26]. CAQDAS therefore provides an important technological context for applied and transparent QR.

A transparent research process is also a foundational goal of Open Science (OS). Historically, the quantitative-dominated replication crisis underlined the necessity of systematic improvements in the reliability, transparency and rigor of research [Ope15]. Open Science includes the idea of making research and its various facets publicly accessible and thereby improving quality and reliability [UNE21]. For example, the FAIR principles can help to improve replication and archiving by making research data findable, accessible, interoperable and reusable

[WDA+16]. The ongoing establishment of these practices also drives increased expectations for transparency, accessibility, collaboration, and the responsible management of research outputs.

Applying OS practices to QR is particularly challenging because qualitative data can be highly contextual, sensitive, and difficult to anonymize without affecting their analytical meaning. There are benefits associated with opening QR, such as improved research quality, stimulation of method development, and reduced over-researching of particular fields by making data available for reuse [SFH+21, preprint]. However, openness in qualitative research cannot generally be understood as uncritical disclosure of research data, which is considered methodologically and ethically problematic [PKW20] [LRMS24]. Consequently, OS in QR requires a balance of transparency and reuse with methodological appropriateness, confidentiality, and participants' consent [CEB+20] [FVP+21, preprint] [SFH+21, preprint] [BCF+24, preprint] [LRMS24] [SMKP25] [SWS26]. This concerns not only eventual publication and data sharing, but also how research data and related materials are managed, documented, accessed, and exchanged throughout the research process [Moz26]. Since CAQDAS is considered to shape QR practices by its capabilities and limitations, it presents an important point of intersection between OS and QR. Existing research, summarized in Section 2, discusses the use and capabilities of CAQDAS, as well as the opportunities and tensions associated with OS in QR. At the same time, less attention has been paid to how OS practices intersect with CAQDAS-supported QR and how this intersection is experienced.

While CAQDAS tools support many aspects of the QR process, their specific potential for supporting OS practices remains insufficiently understood. Such an endeavor requires a better understanding of what is needed, who needs it, and the limitations involved. This study therefore addresses this intersection from a CAQDAS stakeholder-oriented perspective by exploring the needs, tensions, and constraints that might be encountered, rather than assuming that OS principles could be mapped directly into desirable software functionality. This includes perspectives from potential stakeholders who use, develop, provide, or support CAQDAS and QR processes, allowing the study to explore both the capabilities of these systems and the practices within which they are used. Requirements engineering (RE) can provide answers through methods using requirements, stakeholders and constraints [Poh10]. However, the stakeholders of this diverse domain have received limited attention. Furthermore, their needs and constraints for integrating OS in their QR practice are barely understood, because the literature mainly addresses methodological and ethical concerns or tool features. Such insufficient understanding can lead to fragmented tool support (see [KW25]) and unclear design priorities, making it difficult to integrate OS into CAQDAS without introducing new tensions or intensifying existing ones. Identifying potential CAQDAS stakeholders and how they perceive the intersection of OS and QR could therefore inform future efforts. With this perspective in mind, the study addresses the following research questions:

RQ1: Which stakeholder categories and roles relevant to CAQDAS and Open Science can be identified and refined from an exploratory stakeholder analysis?

RQ2: What recurring needs, constraints and tensions are reported when relating qualitative research and CAQDAS practices to Open Science?

By answering RQ1, this work aims to contribute an overview of potential stakeholders relevant to CAQDAS and Open Science. By answering RQ2, this work aims to contribute a thematic summary of recurring needs, tensions, and constraints that may have an influence on CAQDAS practice when OS is introduced to QR. Both questions are interrelated in the sense that emerging stakeholders can potentially be included as sources for answering RQ2. Relating these outcomes to existing literature and synthesizing them into design implications is intended to provide a structured foundation for future Research Software Engineering (RSE) efforts in CAQDAS development. In turn, this work is not expected to provide comprehensive stakeholder coverage or formal requirements across all CAQDAS tools and QR contexts.

2 Related Work

2.1 CAQDAS Capabilities and Stakeholders

At their core, CAQDAS support researchers in organizing, coding, retrieving, searching, comparing, and exploring qualitative data. However, functionality and architecture differ considerably between systems, regardless of whether they are proprietary or open source [KW25]. O'Kane [O'K20] demonstrates how basic and combined CAQDAS techniques can structure inductive analysis, support coder consistency, and make analytical procedures more transparent and traceable. Later, O'Kane et al. [OSL21] demonstrate a set of CAQDAS building blocks, combined techniques, and consistency checks through which researchers can make inductive analytical procedures more transparent and thereby enhance trustworthiness. Rather than functioning only as tools for the coding step, CAQDAS support broader digital research workflows ranging from literature review to the reporting of findings [Pau23]. Some CAQDAS also extend this to mixed methods workflows by combining material and codes with structured variables and enabling various forms of integration between qualitative and quantitative evidence [Fie12] [O'K20] [GJ23]. Interoperability between compatible CAQDAS has advanced through the open exchange standard and was demonstrated experimentally, with generally successful transfers but significant loss of application-specific data [vB19] [ECNW20]. Team-based coding and analysis can be supported through shared coding structures, division of coding work, coder comparison and consistency checks, as well as audit trails of changes and analytic decisions [OSL21]. While there are CAQDAS that enable a transparent integration of commercial and open LLMs[1], others keep this integration closed, creating further barriers for technological reflexivity [KW25]. Recent empirical research on developing qualitative research software also helps identify potential stakeholders. Hepp et al. [HHB23] distinguish between researched persons, researchers, and implementers and show how their respective practices and requirements can differ. Hirsbrunner et al. [HTM24] demonstrate how interdisciplinary research-software development involves multiple disciplinary and technical positionings that are negotiated throughout the research process. In summary, contemporary CAQDAS provide a comprehensive range of functions for qualitative and mixed-methods research, including capabilities that can enhance the transparency and trustworthiness of the analytical process. These capabilities can facilitate particular aspects associated with OS practices. However, the literature does not present OS support as an integrated

[1] https://github.com/ccbogel/QualCoder/wiki/2.3.-AI-Setup

aspect of CAQDAS functionality or systematically connect it to stakeholder perspectives.

2.2 Open Science and Qualitative Research

Open Science practices are still not widely established in QR because demands for openness can conflict with both epistemological grounds and ethical considerations. Pratt et al. [PKW20] argue that replication-oriented transparency standards can be inappropriate for interpretive qualitative work. Similarly, Lamb et al. [LRMS24] reject sharing full qualitative datasets merely to enable replication[2]. Transparency can instead be supported through reflexivity and explicit documentation of methodological and interpretive decisions [FVP+21, preprint] [LRMS24]. Steltenpohl et al. [SMKP25] argue that practices such as transparent documentation, data curation, and accessible materials can enhance credibility and collaborative potential when adapted to the research context through ethically sensitive protocols that preserve participant confidentiality and contextual richness. Schneijderberg et al. [SWS26] explicitly conceptualize this as appropriateness-transparency, a quality convention applicable to qualitative research, emphasizing transparency relative to the method and data rather than adherence to a uniform procedural standard.

Enhancing transparency through sharing qualitative data creates further tensions related to the nature of the data. Rich qualitative materials can be difficult to anonymize without either retaining disclosure risks or removing information relevant to interpretation, while consent to participate in a study does not automatically imply consent to unrestricted future sharing and reuse [CEB+20] [PHM+23] [WPM+25]. Data management plans (DMP), de-identification, and controlled or restricted access therefore provide alternatives to fully open publications of data [SFH+21, preprint]. Corti et al. [CEB+20] therefore treat consent, anonymization, access control, and contextual documentation as interconnected aspects of responsible data management rather than assuming that all research data should become fully open. More broadly, empirical research on research data management (RDM) shows that data sharing already creates difficulties within research collaborations, including technical difficulties, varying institutional or national conditions, and uncertainty about existing data [KL25]. In this context, Research Data Centers (RDC) can provide infrastructures that integrate data protection, research ethics, and interests in secondary use, rather than treating sharing as a binary choice between open publication and non-sharing [Moz26]. Preregistration may also be adapted to QR, making initial intentions and subsequent changes transparent [FVP+21, preprint]. Reuse likewise depends on sufficient documentation of origins and research context to permit meaningful secondary interpretation [CEB+20]. Corti et al. [CEB+20] explicitly treat providing context and assessing the opportunities and limitations of existing data as central parts of the research-data lifecycle.

Taken together, the literature suggests that OS in QR requires context-sensitive openness rather than maximal openness. This makes consent levels, access conditions, the origins of interpretations, and documentation relevant considerations for CAQDAS support.

[2] Sharing research data alone also does not guarantee replication of quantitative studies, as shown by [Ope15].

3 Research Design and Methodology

3.1 Study Type

The research questions are approached through an exploratory inductive qualitative study that is designed to derive insights from expert interviews. The goal is to generate structured themes that reflect the complex realities of stakeholders in relation to OS and CAQDAS. These themes will inform design implications that provide a foundation for subsequent software engineering work. The study follows an applied thematic analysis (ATA) approach by Guest et al., which "[...] is a rigorous, yet inductive, set of procedures designed to identify and examine themes from textual data in a way that is transparent and credible" [GMN12, p.15]. They explain that ATA adapts techniques and aspects from grounded theory, positivism, interpretivism, and phenomenology to applied research contexts and synthesizes them into a methodological framework. As in grounded theory, insights are grounded in data, but ATA does not necessarily require the development of formal theories. It broadens the analytical approach through "its pragmatic focus on using whatever tools might be appropriate to get the analytic job done in a transparent, efficient, and ethical manner" [GMN12, p.18]. While the identification of themes is indeed an interpretative process, the analytic process is biased toward a positivist view through an emphasis on always referring back to the raw data.

The study workflow comprises three main steps: 1. data collection, 2. data processing, and 3. data analysis. During data collection, experts and potential stakeholders in CAQDAS and/or OS are recruited and interviewed to collect qualitative data for analysis. The expected outcomes of this step are interview recordings, as well as collaboratively created artifacts, such as a stakeholder overview. Data processing prepares material for analysis by putting it into the correct format. This step also strengthens the author's familiarity with the data through active engagement with the transcripts. It involves correcting transcription errors and adding contextual information that might otherwise be lost. The analysis step is where the author is "locating meaning in the data" [GMN12, p.49] by identifying patterns, coding them, aggregating codes into codebooks, and identifying and evaluating candidate themes with the aid of data reduction techniques. Finally, the themes will be presented and discussed for their potential to address the research questions.

3.2 Ethics and Positionality

Ensuring the anonymity of participants and the confidentiality of sensitive information is a major ethical challenge at every step of the research workflow. This is because "[If] information were disclosed to others, or if data were not anonymized, then the privacy and security could be compromised[.]" [HHB20, p.71]. Therefore, the data processing step will include a systematic anonymization procedure in accordance with Hennink et al. [HHB20]. Furthermore, the collection, storage and retention of raw data must be carried out in accordance with a data management plan (DMP) that ensures no information is disclosed to any third party.

Recruitment has to be carried out in writing to allow participants time to respond and to carefully consider their decisions. A written consent form is sent to every participant during recruitment and before an interview is conducted. Any interviews for which consent has not been

obtained must not be included in the data analysis or any subsequent steps. Further details on the execution of the recruitment and the consent procedures can be found in Section 4.

Finally, the role of the researcher is another key aspect in qualitative research that needs explicit consideration and a statement of positionality. First, the author acts as the sole researcher throughout the study. Second, the author is also an active participant in the field. This is predominantly expressed through their main role as a CAQDAS developer in the OpenQDA project [KDH+26]. While this circumstance provides a fundamental motivation for this work, it also introduces a potential bias. Third, the author has been involved in open-source software development for over 15 years[3] and strongly supports the values of software freedom[4], especially in publicly funded contexts. With this background, the author supports OS as a development of scientific practice consistent with the open-source values that should be incorporated as a fundamental aspect of a humanistic society that values knowledge. The author therefore approaches the study as an OS advocate with prior exposure to qualitative research but who is not a qualitative researcher by profession. This position informs the author's active engagement in interviews, the focus on CAQDAS and OS, and collaboration on artifacts such as models and diagrams.

3.3 Sampling Strategy

Qualitative research "uses purposive sampling, which involves *purposefully* selecting participants with certain characteristics important to the study" [HHB20, p.92]. Based on the research questions, a central criterion for inclusion in the study population is a background in CAQDAS, qualitative research or Open Science or any combination of these. Second, participants should be part of the German academic system to help reduce cultural barriers and legal issues. The third criterion is to be able to discuss professional topics fluently[5] in either German or English to support detailed, context-rich interviews.

The main segmentation is achieved by the different roles that can relate to CAQDAS: user, developer, administrator, or legislator; and by level of experience: student, graduate researcher, postgraduate researcher, or senior researcher. The resulting sample size is intended to support a certain richness in data, because even smaller samples are seen as usable when purposefully selected [HHB20]. The goal of this approach is to capture a broad range of experiences within the German academic system. The study cannot cover all roles and experience levels, but it remains open to additional categories that may emerge during execution. A secondary factor for segmentation is a variation in research fields in order to capture different methodological peculiarities. However, interviewing multiple individuals from the same field is still expected to provide diverse experience as long as the main factors (role, experience) differ.

Recruitment targets can be identified through publications, software, and standards. Individuals involved with CAQDAS and its development can be found through code repository platforms and the related technical literature. Providing a reproducible description of the exact sources used to recruit participants would compromise their anonymity and is therefore omitted. A potential caveat of this strategy is that it is unlikely to reach individuals at student level, for whom an advertisement on a university public board is planned. Additionally, each recruited partici-

[3] https://github.com/jankapunkt
[4] https://www.gnu.org/philosophy/free-sw.html
[5] CEFR C1, C2 or native, see https://www.coe.int/en/web/common-european-framework-reference-languages

pant should be encouraged to suggest candidates who meet the inclusion criteria (snowballing, [HHB20, p.104]).

4 Study Execution

4.1 Recruitment

The active recruitment occurred from January to April of 2026 and was carried out proactively in writing by emailing individuals who matched the proposed criteria. Respondents who expressed interest in participating were provided with a document[6] that contained information about the purpose of the interview, a description of how their data would be processed, stored, and published, references to the GDPR and their individual rights, as well as a consent form that was required to be signed and sent back. A total of 16 individuals were contacted for this study. Of those contacted, 10 consented to participate in an interview, 2 declined due to self-reported mismatch with the topic, and 4 did not respond or were not reachable. Eight participants provided written consent prior to the interview, while two consented orally and later also in written form after they received and approved the anonymized transcript. Attempts were made to contact a research methods class. Additionally, advertisements were placed within the author's university. However, these efforts received no responses, meaning no students were recruited. No representative of a governing body (formerly legislator) could be recruited, as none of the reviewed literature on CAQDAS, OS, or QR involved any references to individuals in this role. The intended segmentation was not achieved because of the author's limited resources. The final group of participants comprised seven researchers, two developers, and one operator, totaling ten individuals (see Table 1). Overall, the sample was therefore considered an exploratory purposive sample, limited in breadth by the missing stakeholder groups.

4.2 Interview Procedure

The interviews were designed to be semi-structured and to explore experts' opinions while leaving room for collaboration and exchange of ideas. They were carried out from January to April 2026 using end-to-end encrypted video meetings, provided by the author's university Zoom X license. The interviews produced ten recordings, with durations ranging from 38 minutes to 1 hour and 15 minutes. All interviews combined yielded 517 minutes or about 8.5 hours of recording. The interview guide[7] did not represent a fully structured catalog of questions. Instead, it outlined topics and provided example questions for the interviewer. At the beginning of the interview, participants were welcomed and introduced to the topic. Then the recording was initiated upon agreement and the interview began, with the participants being encouraged to introduce themselves. The interviewer asked the participants about their background with CAQDAS, QR or OS as an introductory question, in order to better understand their methodological and epistemological background, and to build rapport [HHB20]. After that the interviewer actively engaged to move the discussion toward a specific topic. This resulted in not all categories being equally

[6] See appendix *interviews/materials/gdpr_form.pdf*
[7] See appendix *interviews/materials/guide.md*

Criterion 1	CAQDAS	Qualitative Research		Open Science
Criterion 2	Part of German academic system (10)			
Criterion 3	English (1)		German (9)	
Experience / Role	Researcher	Developer	Administrator*	Legislator*
Student	0	0	0	0
Graduate	1	2	1	0
Postgraduate	2	0	0	0
Senior Researcher	4	0	0	0
Total	7	2	1	0

Table 1: Overview of the final segmentation of the 10 interview participants. *Administrator and Legislator were subsequently renamed Operator and Governing Body, respectively.

covered across the ten interviews. Absence of evidence is therefore explicitly highlighted in the findings.

If participants agreed, work on the stakeholder overview began and was performed collaboratively. An initial version of the map[8], including the five core categories, was established before the interviews using relevant literature. In seven of ten interviews the author shared their screen and presented the current state of the overview diagram, followed by collaboration. Structurally, the categories were kept intact, while the names of three major categories were changed. The term "administrator" was deemed too focused on software and did not encompass individuals who administer data, such as data stewards and operators of research data centers. It was therefore replaced by the term "operator." The term "legislator" was changed to "governing bodies" in order to include entities that influence the research process but do not impose legal obligations. The term "data subjects" was proposed as a replacement for "third party" with the intent to formally represent actors and entities of this category. However, this change introduces legal implications, which were not properly reflected by the author at that time and will therefore receive further attention in Section 6. The term was therefore subsequently changed to the more neutral and inclusive label "Research Entities." The subcategories were collaboratively added, changed, moved or removed while being discussed with participants.

During the interview, the author sought to guide the discussion toward OS topics that either had been mentioned by the participants or that the author deemed to be related to their current roles and areas of expertise. In earlier interviews, the author asked more general questions about OS practices. Later interviews included more targeted questions about potential ideas and concepts that were noted during prior interviews[9]. At the end, the recording was stopped by the author to enable a quick informal exchange afterwards. The audio file was stored in private cloud storage at the author's university, while the video file was purged.

[8] See appendix *stakeholders/stakeholder-overview-old.pdf*
[9] See appendix *interviews/materials/notes.md*

4.3 Data Processing

Data processing started in April 2026, after all interviews had been recorded and stored, and aimed at preparing the interview data for analysis and publication. The transcripts were processed in their original languages: German for nine interviews and English for one. This helped maintain the richness and nuances of the data throughout the research process. Each recording was processed in four steps: 1. automatic transcription, 2. manual cleaning, 3. manual anonymization, and 4. an automated personally identifiable information (PII) check. To ensure confidentiality, each of these four steps was performed locally on the author's machine using locally installed tools. Transcription was done locally via aTrain 1.4.2 [HFKT24] configured for two-speaker diarization and German-language transcription (English for one interview). The "text with timestamps" format was chosen for easier navigation in a text editor. During cleaning, the recording was played locally while the transcript was read and errors and omissions were corrected using a simple local text editor. A verbatim transcription style based on Hennink et al. [HHB20, p.213] was chosen, preserving original phrasing and replacing filler words only when excessively used. Additionally, metadata and a legend were added at the beginning of the transcript (transcription label and transcription keys)[10].

Manual anonymization was applied by re-reading the cleaned transcript, optionally listening to the recording and replacing any PII with placeholders. Placeholders contained a generic contextual reference, such as [_somewhere_]. Once manual anonymization was complete, an automated scan checked for remaining PII by executing a script[11] that segmented the text and passed it to a locally installed Ollama with Llama 3.3 for inspection. This step was explicitly executed after the manual step in order to maximize human involvement and minimize the danger of incomplete anonymization due to false negatives. An additional manual cleaning step was applied, because the transcripts contained repeated omissions of words and sentences. After the cleaned and anonymized transcript had been finalized, the raw transcript was purged. However, the audio recording was retained to support analysis and will be deleted after publication.

4.4 Data Analysis

4.4.1 Codes and Codebook Development

The analysis workflow, shown in Figure 1, started with three randomly selected interviews that were reviewed to identify initial meanings and create codes. Structural codes, for example representing specific software names or platforms, were used to retain the context provided by interview questions and discussion topics. Content codes represented meanings identified in participants' responses and were provisionally grouped. Unrelated segments were coded as such, so that they did not overlap with any other codes. Code descriptions were added and progressively refined to specify the scope of codes and support consistent application. This resulted in 178 codes that were then structured into an initial codebook hierarchy[12]. The remaining transcripts were coded in random order, which iteratively extended and consolidated this list. Where codes were too specific to individual text segments, they were abstracted (scoping up). Conversely, new

[10] See appendix *interviews/materials/transcription_guide.md*
[11] See appendix *interviews/anoncheck/*
[12] See appendix *coding/Codebook1.qcd*

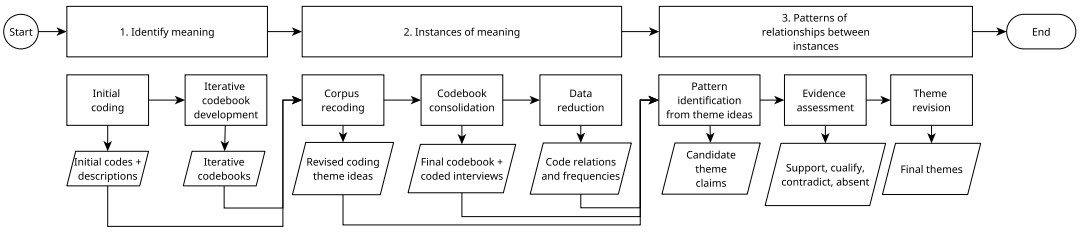


Figure 1: Data analysis workflow, inspired by [GMN12, pp.49-78]

codes were introduced when existing codes did not adequately represent a text segment. This iterative step produced several successive versions of the codebook[13] and preserved the connection between increasingly abstract codes and the underlying interview material. By doing so, the codebook progressively transformed from a rather flat collection of specific codes into a more hierarchical version in which related codes were grouped under increasingly abstract categories. During this step, a few domain- and gender-specific words and text passages were additionally anonymized, because neither manual nor automated anonymization had detected these in the data processing step.

During revision and after the inclusion of the related work section (see Section 2), the complete interview corpus was recoded. The related work further sensitized the author to concepts relevant to QR, OS, and CAQDAS, but did not provide a predefined deductive coding scheme. Existing codes remained provisional, and coding decisions continued to be revised in relation to the interview material. The latest version of the codebook, produced during the initial analysis, served as the provisional starting point rather than as a fixed coding scheme. During this iteration, codes could again be added, renamed, merged, reorganized, or removed when superseded by more appropriate codes. After all interviews had been recoded, the codebook underwent a final round of consolidation and abstraction. Overlapping and excessively narrow codes were merged or abstracted. Related codes were organized into code families with categories of meaning, and codes without remaining passages were removed. Finally, namespaces were added to child codes in order to make their relation to their top-level code family visible in non-hierarchical views in subsequent steps. The resulting final codebook[14] consisted of 219 codes, including 10 top-level code families (that should not be confused with themes): AI in QR and CAQDAS (AI), CAQDAS, Collaboration Practices (COLLAB), Open Science (OS), Qualitative Research (QR), Research Data Centers (RDC), Research Data Management (RDM), Research Software Engineering (RSE), Stakeholders (STAKE), Teaching and Learning (T/L).

4.4.2 Theme Development

Theme development followed the pattern-identification techniques described by Guest et al. [GMN12, p.66]. These included recurrence, domain-related categories and relationships, similarities and differences across coded material, and missing or contradictory evidence. Identified

[13] See appendix *coding/Codebook2.qcd* and *coding/Codebook3.qcd*
[14] See appendix *coding/Codebook5.qcd*

themes were initially noted in a separate document[15] which was later superseded by notes attached to code families during recoding. After recoding, the entire list of coded data, excluding interviewer codes, was exported as a list[16] for data reduction and visualization. Guest et al. [GMN12, p.129] describe data reduction, using quantification of the data (code frequencies), as an integral part of qualitative analysis that helps focus attention on specific themes. However, this quantification remains an analytic measure and is not used to draw conclusions about a wider population. For this purpose, a set of minimal tools[17] was created to clean, reduce, and visualize the dataset. The techniques involved were 1. repetition, to find recurrent codes and interview-specific concentrations, 2. distribution, to identify themes that span across participants and their backgrounds, 3. co-occurrence, to find relationships among codes that can inform candidate themes, 4. rarely used codes, to find opposing or contradictory opinions and rare subjects, and 5. overlaps, to find indicators of rich text segments that represent multiple codes. The reduction tables and corresponding visualizations are included in the supplementary materials[18].

Reduction data and visualizations were used to screen the emerging theme ideas from the notes and to decide whether theme ideas should receive more or less attention. This resulted in the following theme candidates: 1. Stakeholders can be represented in hybrid roles (HYBRID), 2. Open Science practices are individually negotiated as transparency in QR (TRANSP), and 3. Collaborative QR requires CAQDAS to integrate RDM on a workflow level (COLLAB). Other theme ideas were not further investigated or were incorporated into the three main themes. For example, AI was initially considered as a candidate theme but not retained because evidence was limited to two interviews. Because these interviews presented conflicting views of AI's role, AI was treated as a tentative aspect of the HYBRID theme. The learning and teaching of CAQDAS, QR and OS remained too fragmented to become a theme. RDM occurred across all interviews. However, its meaning depended strongly on context, so it was treated as an aspect of the main themes rather than a standalone theme.

For each candidate theme, the initial note was expanded into a theme development document, containing the theme description and the rules for evidence, as well as the classification of evidence. A text segment was classified as core evidence when it directly reflected the theme or a core aspect of the theme. Supporting evidence comprised segments that supported the theme at least partially. Counterevidence comprised text segments that directly challenged a theme claim. Based on these rules, an interview was classified as supporting if the evidence supported the claim of the theme. Qualifying interviews contained supporting evidence for a theme but either introduced a related topic or a variation of the theme or introduced limitations. A contradicting interview was used to test, revise, narrow, or, where necessary, reject candidate theme claims. A theme was considered absent if the interview did not contain any viable reference to the theme. During this step, the HYBRID theme was differentiated into two dimensions: pursuing multiple, context-dependent roles (ROLES) and the purpose of the stakeholder mapping (PURPOSE) presented in the findings. The classifications of evidence for the three candidates are summarized in Table 2 while theme-specific summaries are attached to the audit trail[19].

[15] See appendix *coding/theme-ideas.md*
[16] See appendix *coding/OpenQDA Open Science in CAQDAS Selections 8_15_2026_cleaned.csv*
[17] See appendix *themes/postprocessing/*
[18] See appendix *themes/reduction/*
[19] See appendix *themes/*

Interview	ROLES	PURPOSE	TRANSP	COLLAB
20260109-04016A	Absent	Absent	Qualify	Absent
20260109-04854A	Support	Absent	Qualify	Absent
20260204-03955A	Absent	Qualify	Support	Support
20260218-10038A	Qualify	Absent	Contradict	Qualify
20260224-05457A	Support	Support	Qualify	Absent
20260302-04448A	Absent	Absent	Support	Support
20260319-05705A	Support	Absent	Support	Qualify
20260326-05750A	Support	Support	Support	Support
20260408-10426A	Support	Support	Absent	Absent
20260413-04835A	Support	Support	Qualify	Support

Table 2: Evidence support for themes by interview

The retrieved material was then compared within interviews, across interviews, and across contributing codes with the aid of code-frequency data. Within-interview comparison examined participant consistency and variation. Between-interview comparison examined recurrence and differences between interviews. Between-code comparison examined how codes contributing to a candidate theme related to one another. Using the evidence classification across all interviews, the author revised candidate themes by examining what was shared across supporting cases and what boundaries were introduced by qualifying or contradictory cases. Theme claims were narrowed, restructured, or renamed where the evidence did not support their original formulation. This resulted in the final themes reported in Section 5. Member checking of codes, candidate themes, or final themes was not conducted. It had not been planned, and the remaining time and resources did not permit it. Theme confirmation therefore relied on comparison across the coded corpus, including the explicit examination of qualifying and contradictory evidence, rather than participant validation. The analysis was performed without collaborators, which is why no intercoder agreement methods were applied.

4.4.3 Software, Materials, and AI Usage

OpenQDA 1.0.7 [KDH+26], hosted at the University of Bremen[20], was used to organize and code the anonymized interviews and support theme development. The author used it to create and organize codes and scope descriptions, record emerging theme ideas in code notes, retrieve coded text selections, and export codebooks, selections, and notes in open formats. The author manually made all interpretive decisions, including code creation and revision and theme construction. Custom Python scripts were developed to clean, reduce, and visualize the coded data; Codex with ChatGPT 5.6 Sol assisted in writing them, and the author reviewed the code and generated the outputs. After manually constructing the themes, the author used Codex with ChatGPT 5.6 Sol to check established themes against anonymized selections and reduction data for overlooked supporting evidence or counterevidence (see Evidence assessment and Theme revision in Figure 1). Unlike the four locally executed transcript-processing steps (Subsection 4.3),

[20] Available at https://openqda.org.

this check involved submitting those materials to OpenAI-hosted ChatGPT. No full transcripts, personal data, or identifiable participant information were submitted. Neither the scripts nor ChatGPT autonomously assigned codes or constructed themes. Successive codebook versions, coded selections, notes, theme-development materials, software information, and prompts were included in an audit trail in open formats to support inspection.

5 Findings

Over the course of the study, a stakeholder map was iteratively developed and three main themes (see Figure 3) were identified through analysis of the interview material. The stakeholder map describes the major categories of actors identified in the QR and CAQDAS ecosystem. The first theme indicates that stakeholder roles in CAQDAS and QR are better understood as context-dependent roles rather than mutually exclusive actor categories. The second theme describes the compatibility of OS practices in QR, where OS can take the form of process-oriented, bounded, and contextually negotiated transparency. The third theme describes how collaborative QR turns team-level data sharing into an RDM requirement, spanning CAQDAS and external infrastructure. While the topical diversity of the three themes is likely influenced by answering two research questions, it also highlights the complexity and richness that arise when the domains of QR and OS intersect, with or without CAQDAS. The goal for this section is therefore to present these themes in sufficient detail while maintaining a close connection to the data, using direct quotes[21], as well as references to the sample and analysis results [GMN12].

5.1 Stakeholder Ecosystem

All ten interviews provided evidence that varied in form and richness that contributed to the stakeholder map, shown in Figure 2. Seven out of ten interview participants directly collaborated on the stakeholder map with the author by reflecting on its content and structure. In the resulting map, CAQDAS exists within an ecosystem of five categories of stakeholders: Users, Operators, Developers, Governing Bodies, and Research Entities. It differentiates these into more specific roles and actor groups hierarchically. It depicts a snapshot of the actors, roles, and relations that were identified as relevant to CAQDAS, QR and OS during this exploratory study, rather than providing an exhaustive enumeration of all actors potentially related to QR.

5.2 Context-dependent Stakeholder Roles

In contrast to the stakeholder map's structure, participants did not consistently express roles as distinct actor categories. Statements about individuals occupying multiple or overlapping roles occurred in five of the ten interviews. The combinations of roles and their circumstances varied across interviews. One explicit example involved a data scientist who also performed data-stewardship functions:

[21] The author translated quotations from German-language interviews into English while aiming to preserve the richness of the original as much as possible.

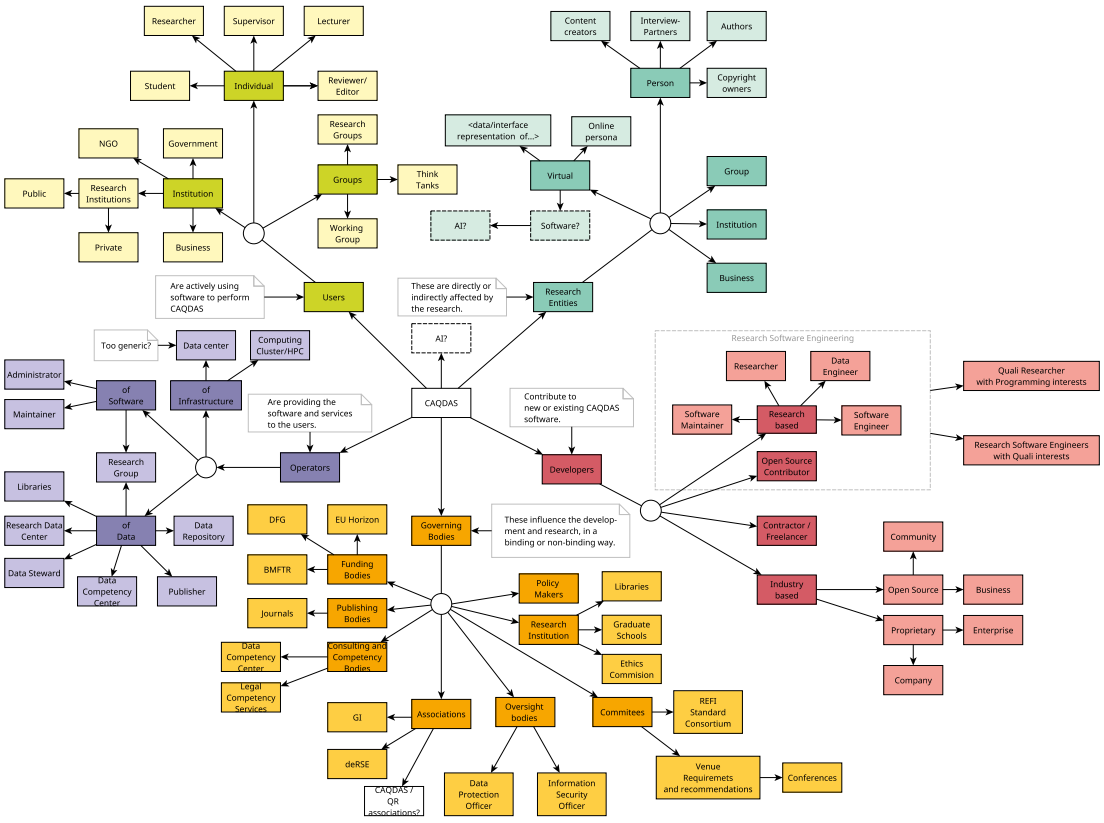


Figure 2: Overview of the QR and CAQDAS stakeholder ecosystem with five major categories and their corresponding subcategories after ten interviews.

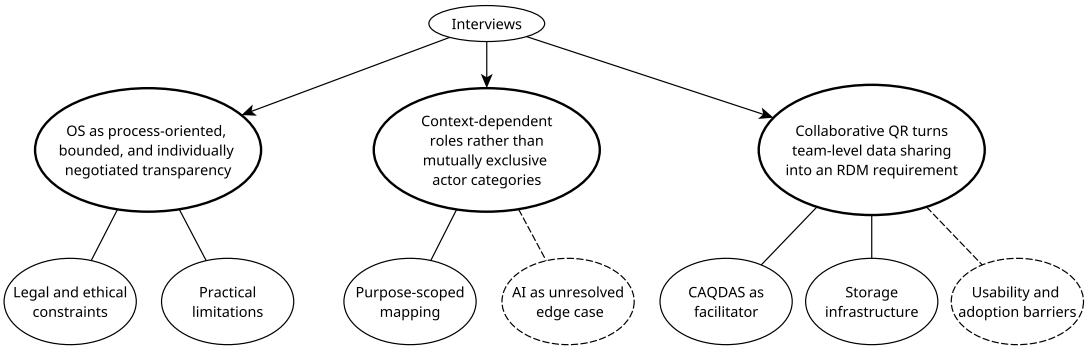


Figure 3: Overview of the three main themes, emerging subthemes and additional influential aspects; dashed outlines indicate tentative aspects.

> "Exactly, the way it is here is that we basically had a project before, where only data stewards worked. And now we're here as data scientists, but we're doing half data stewardship and half data science, I'd say. So now it's basically a combination of both, which I actually find quite useful from the perspective that we can already directly incorporate RDM elements into our data applications and all that. It all goes hand in hand, after all." (Data Consultant)

Other examples include researchers acting as software developers or vice versa, libraries acting as RDCs or ethics boards functioning as integral parts of research institutions. However, mixed roles were not limited to occupational status. Roles were also described as dependent on the setting in which an actor was currently operating. This was particularly explicit in reflections on RSE, where one participant questioned treating Researcher, Data Engineer, and Software Engineer as mutually exclusive top-level categories:

> "But I don't think you necessarily have to distinguish at the top level between researchers, data engineers, and software engineers. [...] I mean, those are just roles that you take on. Alright, I think it really comes down to the context in which you work, that's what determines the distinction." (Research Software Engineer)

In another example, someone who is primarily a researcher is switching into the role of a developer to build tools for a specific step in their research:

> "I just know that when I build a [...] tool myself, where I basically scrape or [...] collect data [...] from somewhere and then [...] process it further with code, then I do, so to speak, build in structures that allow me to trace which step [...] this data comes from [...]." (Graduate Researcher / PhD Student)

These statements indicate that the role overlap is broader than the initially identified "hybrid" professions. Rather, individuals can simultaneously occupy several stakeholder roles but also move between them depending on their research, development, support, organizational, or project context. This role-based perspective was not adequately captured by the actor-oriented structure of the map, which was also explicitly mentioned:

> "[...] well, stakeholders are never entities or actors, but rather roles. And that means, um, entities, people, institutions, and so on. They're often actually stakeholders in two or three ways, like, well, we are too: we're also data subjects, and we're developers." (Senior Researcher)

The claim that stakeholders are "never" entities or actors reflects one participant's position. However, the overlap of actor categories in this and other interviews supports representing the identified stakeholders as context-dependent roles rather than mutually exclusive actor classes. In this context, AI as a stakeholder was discussed only a few times, in two of the ten interviews. Interestingly, participants described it in contradictory ways, either as an actor or as software. While this indicated an unresolved classification in stakeholder modeling rather than a sufficiently grounded sub-theme, its potential relevance for future work should be acknowledged. Reflections on the stakeholder map also revealed views on an appropriate structure and purpose of the map. Comments about the map's structure occurred in five interviews. Explicit arguments

about its purpose and prioritization were concentrated in two. The importance of such mapping was appreciated. However, it was argued that such a mapping should not be understood as an exhaustive representation of an ecosystem, but as a selective representation of aspects relevant to its specific purpose:

> "Yeah, well, I actually think that's really important, and I also think it's justified to do it that way, and [...] I think that's just how it is, unfortunately, with these - with these stakeholder maps - it's not really about [...] mapping the whole world, but rather capturing certain aspects that are important for the software [...] right?" (Senior Researcher)

Others reached similar conclusions but emphasized usefulness in project contexts, rather than questioning the saturation itself:

> "[...] Speaking from practical experience, having too many roles and personas is usually bad for projects, in my experience. [...] What I always find important in situations like this [...] is: what is the smallest possible use case?" (Researcher and self-taught developer)

These findings suggest that a purpose-scoped approach to stakeholder mapping can be useful for CAQDAS development, if the goal is to identify those roles, practices, and relationships relevant to the software and its intended purpose.

5.3 OS as Process-oriented, Bounded, and Negotiated Transparency in QR

Participants did not generally frame OS in QR as requiring unrestricted access to all research data. Eight interviews described different ways in which QR processes could be made comprehensible. One reflected a substantially different understanding of transparency, and one contained no usable evidence. This process-oriented transparency was primarily associated with sharing methodological and analytical decisions. The related code occurred in six of ten interviews, although most instances were concentrated in three of them. One participant explicitly described this transparency as good QR practice and enumerated related materials:

> "But for qualitative research, greater transparency would be more appropriate, and to this day I still don't understand why qualitative articles, for example, don't include appendices. They simply don't exist. But you could include your guide, or you could include your categorization system in the appendix of an article. You could say, 'Okay, here's my research process, here's where I adjusted my research question again based on x, y, z.' That was basically the group I did the analysis with. Those were research hypotheses that I formulated and then discarded. So, actually, when you do good qualitative research, all of that is documented." (Senior Researcher)

Sharing qualitative data was presented as an optional part of this transparency. In the context of primary data publishing for reuse, this was subject to negotiation, taking into account legal and ethical aspects such as privacy, anonymity, consent, sensitivity, and practical constraints. However, these were not treated as reasons to reject OS:

> "But that doesn't mean that all strategies that can be implemented in Open Science are off the table. It just means that perhaps the raw data or even the anonymized data cannot be made available for reuse. But a data management plan, a workflow, a codebook - there are plenty of them - a metadata table that doesn't contain any PII, and things like that. In other words, a whole lot of supporting materials that researchers can still make use of, because they can then see how a particular workflow functioned and build their own research on that foundation." (Data Consultant)

Ethical considerations can place particularly strong constraints on this openness. Participants described these constraints in terms of potential consequences:

> "[...] I think it's very understandable that something like this might remain confined to a network [...] and not be made more widely available [...] and when it comes to [...] issues like, somewhat sensitive, qualitative data on [...] queer communities, I also find it very difficult [...], I think that, above all, the issue of safety [...] for the individuals [...] - if they haven't given their consent beforehand [...] - is more important than, so to speak, the [...] openness of science." (Graduate Researcher / PhD Student)

Practical feasibility was introduced as another constraint by a participant who reflected on previously preparing a qualitative dataset for reuse:

> "[...] but I wouldn't do that again [...]. [...] in the end, it took an unbelievable amount of time and effort to get this data into a format that [[_RDC_]] would approve. One of my student assistants basically spent twelve months working almost exclusively on making the data reusable." (Senior Researcher)

Together, these accounts indicate that decisions about openness do not simply revolve around sharing or withholding data. Instead, the scope and form of openness are negotiated in relation to methodological rigor, potential benefits of reuse, ethical and legal responsibilities, and the effort required. Methodologically, several participants explicitly distinguished intersubjective comprehensibility from exact replication. One participant explicitly warned against treating exact replication as the goal:

> "Well, I think the first thing is that we shouldn't create this illusion of replicability, because from a qualitative perspective, that's basically impossible. It's far too subjective for that. So, we can only try to make the subjective as comprehensible as possible, but it can never be the goal to somehow make something replicable." (Senior Researcher)

In contrast, one participant expressed a stronger replication-oriented view of transparency:

> "So [...] anyone who reads your paper has to be able to replicate exactly what you have done in order to make sure that that this work is solid, this work is transparent, and in order to progress even after that right? [...] We often don't share transcripts, depending again on what sort of study, how anonymized it is, how trackable they are to find the initial person and so on. [...] So for some studies, it does make sense to also share transcripts." (Postgraduate Researcher)

However, the participant's mixed-methods background and reported practice of not sharing transcripts suggest that even replication-oriented positions can remain cautious about opening primary qualitative data. In summary, these findings indicate a need for transparency in QR processes but without requiring unrestricted access, especially to primary data. Reported constraints included the ethical and legal implications of opening data, the effort required for compliance, and differing methodological views on transparency. The remaining tension concerns how transparency and potential reuse can be increased while participant protection, contextual information, methodological appropriateness, and practical feasibility are preserved.

5.4 Team-Level Data Sharing as an RDM Requirement in Collaborative QR

Collaboration was discussed in six of the ten interviews and described as a relevant practice in QR. Methodological aspects included collaborative interpretation, while practical workflows included separate coding, exchange of data, and requirements for collaborative software. One participant described a hermeneutic approach involving multiple researchers:

> "That's why all these group-based methods, such as all hermeneutically grounded methods, and even Mayring's qualitative content analysis, which focuses on intercoder reliability, are designed so that there are at least two people reviewing the material, in order to be able to reflect, so to speak, on where their own biases lie and where their own context and subjectivity come into play." (Senior Researcher)

This participant and others described collaborative analysis as a coordinated process. Four interviews directly connected collaboration with sharing research data in ongoing projects, where one participant described a failed collaboration caused by poorly coordinated version management:

> "[...] you end up with versions being sent back and forth, then someone keeps working on old versions, and that's just a disaster. [...] The way I'm currently handling this in my research practice is simply to have one person do the coding, and then another person does the next round of coding on top of that, or we just divide up the cases exactly [...]" (Senior Researcher)

They continued to describe workarounds such as sequential coding or dividing cases among team members. Another researcher reported a similar sequential workflow without difficulties, emphasizing their coordinating role as a supervisor. Other participants articulated the need for collaboration to be directly integrated into CAQDAS:

> "Yeah, well, if you were to translate that into a high-quality software application, you'd probably need something that syncs in real time, where the memos and comments appear, and most importantly where people could reply to each other's comments." (Researcher and self-taught developer)

Together, these experiences indicate that collaborative CAQDAS support concerns coordination of data, contributions, and versions, rather than just shared file access. Storage infrastructure formed a second aspect of these workflows. Institutional or cloud-based storage can provide shared access within and across organizations, but participants also described cases in which access restrictions encouraged technically convenient but ethically problematic alternatives:

> "We used to share stuff via Dropbox back then, which is highly problematic, so to speak, but of course the problem is often simply that other people can't access the university's own servers." (Senior Researcher)

The same participant highlighted their institutionally provided Nextcloud storage as a general solution for shared file access, yet they did not frame it as a solution to coordinated collaboration. This illustrates the tension between the need for frictionless collaboration and external requirements for privacy and reliable, secure storage. A further, but less broadly represented aspect concerned adoption. One participant described technically demanding security or encryption workflows as potential barriers for qualitative researchers. This suggests that secure infrastructure must also be simple enough to be used in QR practice.

In summary, these findings indicate a strong intersection with RDM when data are shared in collaborative settings. This includes software (such as CAQDAS), as well as infrastructure, because researchers need to coordinate project data (primary, as well as supplementary, see Subsection 5.3), versions, and analytical contributions while simultaneously managing storage, privacy, and organizational access constraints. However, the evidence does not support the claim that software-supported collaboration is inherently inadequate: One participant reported a case that challenged this interpretation: a positive experience using a purpose-built platform for collaborative hermeneutic interpretation "[...] that worked relatively well" (Senior researcher). Additionally, sequential collaboration was described not only as a workaround but also as an established, effective way of working. CAQDAS is not necessarily incapable of supporting collaboration and (external) infrastructures can support storage and access. However, in reported workflows, these remained separate and required manual coordination.

6 Discussion

6.1 Stakeholders

The stakeholder map provides an extensive, but non-exhaustive, representation of the stakeholder ecosystem in a hierarchical and actor-oriented way. It helped to identify stakeholders with different types of relationships to CAQDAS, QR and OS. However, its structure can limit the understanding of the contextual and situational reality of the stakeholders in this domain. Its current form does not adequately represent context-dependent and potentially overlapping roles of certain stakeholders. These overlapping roles are also described in the literature. Goth et al. [GAB+23, p.1] clearly state that "the roles of RSEs vary depending on the institutional context they work in", further describing "RSEs [often assuming] a multifaceted role at the junction of research, SE and data management". Practically, they organized their competency catalog around skills, rather than occupational definitions or actors. Hepp et al. [HHB23] also describe overlapping roles during research-software development, where researchers simultaneously act as users and infrastructure workers, depending on the specific development context. Additionally, the fixed terminology of categories may have its own implications. For example, the legal definition of "data subjects"[22] excludes non-human entities, such as data, software, and AI. Consequently,

[22] https://gdpr-info.eu/art-4-gdpr/

future efforts in developing CAQDAS may consider a more purpose-scoped approach to mapping stakeholders, rather than attempting to map the entire ecosystem. Scoping may also concern which types of software are included under the term CAQDAS, and whether this terminology or alternative definitions (see [BCK+22]) of (qualitative) research software are more appropriate.

6.2 OS as Bounded and Negotiated Transparency

The theme mainly confirms and contextualizes existing literature. Transparency is a central methodological need in QR to make research processes comprehensible. Guest et al. [GMN12, p.253] emphasize that "inadequate description of analytic procedures and reasoning is one of the weakest areas in published qualitative research". Recognizing that transparency in QR does not imply replicability is essential and is also discussed in the literature [PKW20] [LRMS24]. A further distinction is needed between transparency for primary qualitative data and supplemental data, such as documentation, methodological decisions, analytic materials, codebooks, and workflows. Based on how well they can be anonymized, they are affected by legal and ethical aspects that pose limitations to openness and transparency. This is well covered in the literature in various contexts and with varying proposed solutions [SFH+21, preprint] [BCF+24, preprint] [LRMS24] [SMKP25] [WPM+25] [Moz26]. These aspects should therefore be considered when developing CAQDAS solutions that aim to support transparency and openness. Additionally, the effort required to meet these constraints can pose a real challenge to researchers but also creates opportunities for CAQDAS to help manage this work. Negotiating these factors involves a central tension between OS and QR: openness and reuse versus ethical appropriateness, legal compliance and methodological adequacy. A balanced, appropriate transparency has also been widely discussed in the literature [CEB+20] [FVP+21, preprint] [SFH+21, preprint] [BCF+24, preprint] [LRMS24] [SMKP25] [SWS26]. In this regard, CAQDAS could act as a facilitator and mediator, supporting the decision-making process and implementation of individual solutions to achieve the best possible transparency in one's research.

6.3 Collaboration Needs RDM for Team-Level Data Sharing

Data sharing in QR does not begin with publication. Collaborative workflows already require researchers to share and manage data during ongoing research. Kralj and Landwehr state that "[s]haring data is a crucial element of collaboration [...]" [KL25, p.3]. In this context, data sharing refers to distributing data and project material among members of a research team, rather than to the OS practice of publicly sharing data for reuse. The reported workflows indicate RDM concerns around storage, access, privacy, version integrity, and coordination. Consequently, collaboration connects CAQDAS use with RDM during the ongoing research process, rather than only when research outputs are eventually prepared for publication. This interpretation is consistent with recent literature: "In the processing and analyzing phase, in addition to data protection, storage and organization emerged as the main areas in which researchers have needs to be able to collaborate successfully. The infrastructure for data storage represents one of the biggest needs" [KL25, p.6]. Because the shared material includes not only primary data but also projects, codes, memos, comments, versions, and other analytical contributions, managing collaboration requires attention to both the underlying data and the evolving analysis. CAQDAS

Sharing Context	Main purpose	Potential design concerns
Within ongoing research	Coordinate work and manage data within and between teams	Versions, contributions, work sequences, team access
Sharing materials to explain an analysis	Make methodological and analytical decisions comprehensible	Contextual documentation, selective export
Providing materials for secondary use	Enable appropriately governed reuse	Selection, documentation, access conditions

Table 3: Distinct contexts of qualitative data sharing, inspired by selected phases of the research lifecycle (1. Processing and Analyzing, 2. Sharing, 3. Reusing) [KL25].

tools support elements of collaborative workflows (see [OSL21]) and the findings indicate that software-based collaborative workflows can work. The tensions appear to arise from interactions between software, data management, and external research infrastructure, rather than from the analysis workflow alone. This infrastructure aspect is especially relevant because low-friction collaboration and secure RDM do not necessarily coincide. More real-time collaboration alone is insufficient for CAQDAS.

In summary, the findings in combination with the literature frame collaboration as an intersection of CAQDAS and RDM rather than as an isolated software feature. The main concern is not whether workflows are sequential or synchronized. Instead, the recurring concern is coordinating the involved data during the preparation phase and analytical work, such as coding, as well as after these sequences. This shifts the perspective for CAQDAS development from individual features toward how CAQDAS tools operate within broader data management endeavors.

6.4 Implications for CAQDAS and RSE

A synthesis of the themes in combination with the literature suggests that “data sharing” serves different purposes and audiences across research contexts (see Table 3). Team-level sharing addresses collaborative RDM needs, while preserving and documenting materials during ongoing research can also provide a foundation for OS practices involving transparency and appropriate reuse. CAQDAS design should therefore consider the sharing context, its purpose, and the associated needs, constraints, and tensions. The following implications offer candidate directions for CAQDAS design and subsequent requirements work.

6.4.1 Design Scope

CAQDAS design should be initially scoped by the intended transparency or sharing purpose, the targeted audience, and the responsibilities associated with the workflows. For example, during ongoing research collaboration and preparation of materials for release, these responsibilities could include preparing, coding, and analyzing data, administering access, advising on RDM, and approving provision and release. They may also overlap and vary with context. This scoping should be treated as a potential foundation for subsequent design work and not mistakenly interpreted as evidence that the software needs a fixed set of user roles.

6.4.2 Collaborative Integrity

Collaborative CAQDAS workflows in ongoing research should support sequential and simultaneous work where appropriate, while addressing tensions between ease of collaboration, version integrity, and controlled access. Practices to consider include versioning projects and materials, controlling visibility and access, organizing shared work, managing concurrent edits, preserving and discussing different interpretations, and communicating through shared notes. These practices can also help preserve the materials and analytical context needed for subsequent transparency and appropriate sharing beyond the team. Within these practices, CAQDAS should specifically focus on reducing friction and misunderstandings during handovers and concurrent tasks.

6.4.3 Storage Interoperability

Infrastructural barriers may encourage the use of questionable alternatives, creating tensions between convenient data exchange and privacy, security, and institutional requirements. CAQDAS design should therefore consider integration with local devices, standardized institutional storage solutions such as Nextcloud, and customized storage solutions suited to teams and institutions. Such integration can support responsible management and exchange of materials during collaboration and in preparation for appropriate OS practices. Because the software itself cannot establish cross-institutional access rules or guarantee legal compliance, integration support must remain assistive. Connections to external storage must still be assessed for privacy and security.

6.4.4 Support for Transparency

The findings on negotiated transparency indicate that supporting materials may contribute to OS even when primary data cannot be shared. CAQDAS design should therefore enable researchers to document relevant methodological and analytical decisions with enough context to make them understandable, while helping to manage the additional workload. Distinguishing between recording what changed and documenting why it changed can guide decisions about automatic logging and manual input. This supports the need for comprehensible research processes within practical constraints. However, the design should avoid implying that every action must be logged or that an activity log alone sufficiently explains the reasoning behind analytical decisions.

6.4.5 Multiple Levels of Access and Release

Supporting OS requires negotiating transparency and reuse within ethical, legal, and methodological constraints. Different materials may therefore require different levels of access and release, depending on the sharing context, consent, sensitivity, anonymity, effort, and methodological purpose. CAQDAS design should consider workflows for keeping primary data restricted while releasing suitable supporting materials, and for preparing selected material for controlled access or reuse. The design should avoid implying that materials will automatically be accessible to reviewers or eligible for publication and reuse. Release decisions should remain within applicable consent and institutional constraints, because a permission setting alone cannot establish

whether releasing particular material is ethical or lawful.

6.5 Limitations

The study's primary limitation stems from the fact that the author, constrained by limited resources, was the sole researcher. While advantageous during data collection, this limits the independence of the analysis and interpretation of findings. Even though the author already addressed their positionality and potential bias in Section 3, these issues remain relevant here. To address this limitation, the author included extensive supplementary analytical material and used several direct quotations to reveal the connection between the findings and the data. Second, more targeted recruitment of participants from diverse backgrounds could have resulted in a more balanced sample. The current sample appears to be skewed toward experienced qualitative researchers with generally positive attitudes toward OS. This bias is partially shown by the findings in the second theme, which largely reflects existing literature on OS in QR. Third, a narrower focus on a single research question (only stakeholders, only OS practices, only CAQDAS practices) could have helped to reduce the complexity of the analysis and the resulting themes. The findings and discussion revealed a complex interview structure that sometimes did not allow specific topics to be explored in more depth. Finally, the conclusions drawn from the findings represent no generalizations. Furthermore, they do not represent formal requirements or equivalent artifacts or products of a formal requirements engineering study. They should be treated as input for further studies, as a foundation for developing hypotheses, as well as considerations for future RSE endeavors in relation to CAQDAS and OS.

7 Conclusion

This work explored potential CAQDAS stakeholders in relation to OS in QR. It also explored needs, constraints, and tensions when these topics intersect. It produced a literature-informed and collaboratively extended stakeholder map whose hierarchical structure could not adequately represent context-dependent, overlapping, and changing roles. The findings largely support existing literature on negotiated transparency: researchers need to make research processes comprehensible while navigating ethical, legal, and methodological constraints. The challenge is to increase transparency within these constraints. The collaboration findings highlight related needs for managing and exchanging materials during ongoing research under human, technical, infrastructural, and legal constraints. These RDM concerns are relevant to OS because preserving and documenting materials within collaborative workflows can support subsequent transparency and appropriate sharing beyond the team. The resulting design implications address both this foundation and support for transparency and release, while retaining context-sensitive limits on openness. They offer directions for future CAQDAS development and requirements work rather than formal requirements.

Acknowledgments: The author thanks all interview participants for their valuable time, knowledge, and effort.

Data availability: The code and anonymized research materials supporting this study are publicly available for reuse at https://doi.org/10.5281/zenodo.23018877.

Declaration on the use of AI: AI tools supported selected tasks in local interview processing, code generation for post-processing coded data, and theme consistency checks. The tools, workflows, and human involvement are described in the relevant sections. Additionally, ChatGPT 5.6 and DeepL were used to critically review sections and citations, and to correct grammar and spelling. All AI-assisted outputs were critically reviewed by the author, who takes full responsibility for the content.

Bibliography

[BCF+24] C. Bonina, R. W. A. Clark, E. K. Farran, S. Gameiro, M. Hanchard, R. Higman, A. Korzeniowska, N. Sheppard, A. Marcu, R. Network, N. Tumelty, J. Uttley, E. Gowie, N. Mauthner, N. L. Cole. Open Qualitative Research: A Primer from UKRN. Dec. 2024.
doi:10.31219/osf.io/kje7r
https://osf.io/kje7r

[BCK+22] M. Barker, N. P. Chue Hong, D. S. Katz, A.-L. Lamprecht, C. Martinez-Ortiz, F. Psomopoulos, J. Harrow, L. J. Castro, M. Gruenpeter, P. A. Martinez, T. Honeyman. Introducing the FAIR Principles for research software. *Scientific Data* 9(1):622, Oct. 2022.
doi:10.1038/s41597-022-01710-x
https://www.nature.com/articles/s41597-022-01710-x

[vB19] F. van Blommestein. REFI-QDA - Exchange of processed data between qualitative data analysis software packages. 2019.
https://www.qdasoftware.org/

[CEB+20] L. Corti, V. Van den Eynden, L. Bishop, M. Woollard, M. Haaker, S. Summers. *Managing and sharing research data: a guide to good practice*. SAGE, Los Angeles London New Delhi Singapore Washington DC Melbourne, 2nd edition, 2020.

[ECNW20] J. Evers, M. U. Caprioli, S. Nöst, G. Wiedemann. What is the REFI-QDA Standard: Experimenting With the Transfer of Analyzed Research Projects Between QDA Software. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 21(2), May 2020. Number: 2.
doi:10.17169/fqs-21.2.3439
https://www.qualitative-research.net/index.php/fqs/article/view/3439

[Fie12] N. G. Fielding. Triangulation and Mixed Methods Designs: Data Integration With New Research Technologies. *Journal of Mixed Methods Research* 6(2):124–136, Apr. 2012.

doi:10.1177/1558689812437101
https://journals.sagepub.com/doi/10.1177/1558689812437101

[FL93] N. Fielding, R. M. Lee. *Using computers in qualitative research*. Sage, London, reprinted with an updated resources section as new edition, 1993.

[FVP+21] S. M. Field, D. Van Ravenzwaaij, M.-M. Pittelkow, J. M. Hoek, M. Derksen. Qualitative Open Science – Pain Points and Perspectives. Sept. 2021.
doi:10.31219/osf.io/e3cq4
https://osf.io/e3cq4_v1

[GAB+23] F. Goth, R. Alves, M. Braun, L. J. Castro, G. Chourdakis, S. Christ, J. Cohen, F. Erxleben, J.-N. Grad, M. Hagdorn, T. Hodges, G. Juckeland, D. Kempf, A.-L. Lamprecht, J. Linxweiler, M. Schwarzmeier, H. Seibold, J. P. Thiele, H. von Waldow, S. Wittke. Foundational Competencies and Responsibilities of a Research Software Engineer. Nov. 2023. arXiv:2311.11457 [physics].
doi:10.48550/arXiv.2311.11457
http://arxiv.org/abs/2311.11457

[GJ23] T. C. Guetterman, T. G. James. A software feature for mixed methods analysis: The MAXQDA Interactive Quote Matrix. *Methods in Psychology* 8:100116, Nov. 2023.
doi:10.1016/j.metip.2023.100116
https://linkinghub.elsevier.com/retrieve/pii/S2590260123000073

[GMN12] G. Guest, K. M. MacQueen, E. E. Namey. *Applied thematic analysis*. Sage, Los Angeles London New Delhi Singapore Washington DC, 2012.

[HFKT24] A. Haberl, J. Fleiß, D. Kowald, S. Thalmann. Take the aTrain. Introducing an interface for the Accessible Transcription of Interviews. *Journal of Behavioral and Experimental Finance* 41:100891, Mar. 2024.
doi:10.1016/j.jbef.2024.100891
https://linkinghub.elsevier.com/retrieve/pii/S2214635024000066

[HHB20] M. Hennink, I. Hutter, A. Bailey. *Qualitative research methods*. SAGE, Los Angeles London New Delhi Singapore Washington DC Melbourne, 2nd edition, 2020.

[HHB23] A. Hepp, F. Hohmann, A. Belli. Lasst uns unsere Tools gemeinsam bauen!: Möglichkeiten und Grenzen der Co-Creation von Forschungssoftware für die Kommunikations- und Medienwissenschaft. *Publizistik* 68(2–3):177–200, 2023.
doi:10.1007/s11616-023-00803-w
https://link.springer.com/10.1007/s11616-023-00803-w

[HTM24] S. D. Hirsbrunner, M. Tebbe, C. Müller-Birn. From critical technical practice to reflexive data science. *Convergence: The International Journal of Research into New Media Technologies* 30(1):190–215, Feb. 2024.
doi:10.1177/13548565221132243
https://journals.sagepub.com/doi/10.1177/13548565221132243

[KDH+26] J. Küster, S. S. Desha, A. Hepp, F. Hohmann, P. Sinner, K. Wolf. OpenQDA. 2026.
doi:10.5281/ZENODO.21456522
https://zenodo.org/doi/10.5281/zenodo.21456522

[KL25] S. N. Kralj, M. Landwehr. Research Data Management Practices Across the Research Data Lifecycle and Their Potential for Collaboration in an International Higher Education Alliance. *Data Science Journal* 24, Dec. 2025.
doi:10.5334/dsj-2025-036
https://datascience.codata.org/articles/10.5334/dsj-2025-036

[KR20] U. Kuckartz, S. Rädiker. Computergestützte Analyse qualitativer Daten (CAQDAS). In Mey and Mruck (eds.), *Handbuch Qualitative Forschung in der Psychologie*. Pp. 813–834. Springer Fachmedien Wiesbaden, Wiesbaden, 2020.
doi:10.1007/978-3-658-26887-9_55
http://link.springer.com/10.1007/978-3-658-26887-9_55

[KW25] J. Küster, K. D. Wolf. The Current State of CAQDAS is Insufficient for Open Science Qualitative Research. *Electronic Communications of the EASST* 85, Dec. 2025.
doi:10.14279/ECEASST.V85.2709
https://eceasst.org/index.php/eceasst/article/view/2709

[LRMS24] D. Lamb, A. Russell, N. Morant, F. Stevenson. The challenges of open data sharing for qualitative researchers. *Journal of Health Psychology* 29(7):659–664, June 2024.
doi:10.1177/13591053241237620
https://journals.sagepub.com/doi/10.1177/13591053241237620

[Moz26] K. Mozygemba. Managing, sharing and re-using qualitative data: Approaches that integrate data protection, research ethics and research interests: or: How researchers and data providers work together to make qualitative research data available for scientific re-use. *IASSIST Quarterly* 50(1), Mar. 2026.
doi:10.29173/iq1158
https://iassistquarterly.com/index.php/iassist/article/view/1158

[O'K20] P. O'Kane. Demystifying CAQDAS: A Series of Dilemmas. In Crook et al. (eds.), *Research Methodology in Strategy and Management*. Pp. 133–152. Emerald Publishing Limited, Oct. 2020.
doi:10.1108/S1479-838720200000012020
https://www.emerald.com/insight/content/doi/10.1108/S1479-83872020000012020/full/html

[Ope15] Open Science Collaboration. Estimating the reproducibility of psychological science. *Science* 349(6251):aac4716, Aug. 2015.
doi:10.1126/science.aac4716
https://www.science.org/doi/10.1126/science.aac4716

[OSL21] P. O'Kane, A. Smith, M. P. Lerman. Building Transparency and Trustworthiness in Inductive Research Through Computer-Aided Qualitative Data Analysis Software. *Organizational Research Methods* 24(1):104–139, Jan. 2021.
doi:10.1177/1094428119865016
https://journals.sagepub.com/doi/10.1177/1094428119865016

[Pau23] T. M. Paulus. Using Qualitative Data Analysis Software to Support Digital Research Workflows. *Human Resource Development Review* 22(1):139–148, Mar. 2023.
doi:10.1177/15344843221138381
https://journals.sagepub.com/doi/10.1177/15344843221138381

[PHM+23] A. M. B. Prosser, R. J. T. Hamshaw, J. Meyer, R. Bagnall, L. Blackwood, M. Huysamen, A. Jordan, K. Vasileiou, Z. Walter. When open data closes the door: A critical examination of the past, present and the potential future for open data guidelines in journals. *British Journal of Social Psychology* 62(4):1635–1653, Oct. 2023.
doi:10.1111/bjso.12576
https://bpspsychub.onlinelibrary.wiley.com/doi/10.1111/bjso.12576

[PKW20] M. G. Pratt, S. Kaplan, R. Whittington. Editorial Essay: The Tumult over Transparency: Decoupling Transparency from Replication in Establishing Trustworthy Qualitative Research. *Administrative Science Quarterly* 65(1):1–19, Mar. 2020.
doi:10.1177/0001839219887663
https://journals.sagepub.com/doi/10.1177/0001839219887663

[PL24] T. M. Paulus, J. N. Lester. Digital qualitative research workflows: a reflexivity framework for technological consequences. *International Journal of Social Research Methodology* 27(6):621–634, Nov. 2024.
doi:10.1080/13645579.2023.2237359
https://www.tandfonline.com/doi/full/10.1080/13645579.2023.2237359

[Poh10] K. Pohl. *Requirements engineering: fundamentals, principles, and techniques*. Springer, Berlin, 2010. XVII, 813 S. : Ill., graph. Darst.
https://suche.suub.uni-bremen.de/peid=B62305740

[SFH+21] I. Steinhardt, C. Fischer, M. Heimstädt, S. D. Hirsbrunner, D. Ikiz-Akinci, L. Kressin, S. Kretzer, A. Möllenkamp, M. Porzelt, R.-M. Rahal, S. Schimmler, R. Wilke, H. Wünsche. Opening up and Sharing Data from Qualitative Research: A Primer: Results of a workshop run by the research group "Digitalization and Science" at the Weizenbaum Institute in Berlin on January 17, 2020. *Weizenbaum Institute Publications*, 2021.
doi:10.34669/WI.WS/17
https://www.weizenbaum-library.de/handle/id/12

[SMKP25] C. N. Steltenpohl, B. Mawhinney, V.-M. Karhulahti, A. M. B. Prosser. Open science for qualitative research. *BMJ Evidence-Based Medicine* 30(6):413–419, Dec. 2025.

doi:10.1136/bmjebm-2024-113532
https://ebm.bmj.com/lookup/doi/10.1136/bmjebm-2024-113532

[SWS26] C. Schneijderberg, O. Wieczorek, I. Steinhardt. *The Handbook of Qualitative and Quantitative Content Analysis: Introduction to Classical, Digital, AI-supported, and Automated Data Analysis*. Routledge, London, 1st edition, Jan. 2026.
doi:10.4324/9781003496397
https://www.taylorfrancis.com/books/9781003496397

[UNE21] UNESCO. UNESCO Recommendation on Open Science. Technical report, UNESCO, 2021.
doi:10.54677/MNMH8546
https://unesdoc.unesco.org/ark:/48223/pf0000379949

[WDA+16] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. Da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 'T Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. Van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. Van Der Lei, E. Van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3(1):160018, Mar. 2016.
doi:10.1038/sdata.2016.18
https://www.nature.com/articles/sdata201618

[Whe26] K. Wheeler. Technological reflexivity in practice: how MAXQDA, NVivo, and ChatGPT shape qualitative survey analysis. *Qualitative Research in Psychology*, p. 1–27, Jan. 2026.
doi:10.1080/14780887.2025.2602820
https://www.tandfonline.com/doi/full/10.1080/14780887.2025.2602820

[WPM+25] H. A. Walsh, M. V. Parsons, J. Mozersky, A. Gupta, A. M. Lai, A. B. Friedrich, J. M. DuBois. Responsible Sharing of Qualitative Research Data: Insights From a Pioneering Project in the United States. *International Journal of Qualitative Methods* 24:16094069251329607, Apr. 2025.
doi:10.1177/16094069251329607
https://journals.sagepub.com/doi/10.1177/16094069251329607