Electronic Communications of the EASST
Volume 85 Year 2025

deRSE25 - Selected Contributions of the 5th Conference for Research Software Engineering in Germany

Edited by: René Caspart, Florian Goth, Oliver Karras, Jan Linxweiler, Florian Thiery, Joachim Wuttke

The Current State of CAQDAS is Insufficient for Open Science Qualitative Research

Jan Küster, Karsten D. Wolf

DOI: 10.14279/eceasst.v85.2709

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

The Current State of CAQDAS is Insufficient for Open Science Qualitative Research

Jan Küster[1] **and Karsten D. Wolf**[1]

ZeMKI, University of Bremen, Germany[1]

Abstract: Qualitative research, which relies on diverse methods and complex data, is increasingly utilizing qualitative data analysis software, known as Computer-Aided Qualitative Data Analysis (CAQDAS). While enhancing data management, coding and transparency, the current software landscape is dominated by proprietary software and presents significant challenges for open science principles. High costs, vendor lock-in and closed data formats hinder collaboration, data reuse and reproducibility. This study conducted a systematic review of 28 software tools for qualitative data analysis, which were identified through comprehensive searches. The review examined the tools' licensing, costs, platform availability, format support, interoperability, collaboration, AI integration, adherence to the FAIR principles for research software, auditability, security, and sustainability from the perspective of stakeholders. The findings reveal that, although proprietary tools offer broad format support and some collaboration features, they are expensive and often lack sufficient on-premises options for sensitive data, locking users into specific workflows. Open source alternatives exist, but often suffer from limited sustainable funding, smaller communities and less comprehensive features or documentation, which makes them less accessible, particularly for non-technical users. Furthermore, many tools lack the robust real-time collaboration and detailed audit trails that are necessary for transparency and intersubjective validation. AI integration is emerging, but it often operates as a black box, limiting user control and raising privacy concerns due to reliance on commercial vendors. Furthermore, support for FAIR principles, security policies and plug-in architectures remains marginal across the field. In conclusion, the current state of CAQDAS is inadequate for fully supporting open science qualitative research practices, such as collaborative, transparent and reproducible analysis, and long-term data archiving. There is a critical need to develop accessible, extensible and collaborative open source CAQDAS tools via community-driven approaches, in order to advance both research methods and teaching.

Keywords: caqdas,qda,qualitative research,research software,open science

1 Introduction

Qualitative research, informed by diverse theoretical perspectives such as symbolic interactionism, ethnomethodology or structuralism, fundamentally aims to analyze purposefully selected, often smaller, samples inductively, hermeneutically, or phenomenologically [Gup24]. This profound engagement with data, encompassing interviews, observations, memos, pictures, and multimodal sources such as video, aims to generate or refine theories at various scales through rigorous interpretation. Qualitative Data Analysis Software (QDAS) facilitates intricate data management and coding and enables the exploration of complex relationships within the data. This enhances the transparency and auditability of qualitative research. While less central, QDAS can also be employed strategically in quantitative research, particularly when managing and deductively coding large qualitative datasets, such as lesson transcripts, for subsequent statistical analysis.

Qualitative research therefore encompasses a variety of analysis methods that are increasingly being carried out with the help of software [KR20]. This is often referred to by the term "Computer Aided Qualitative Data Analysis" (CAQDAS) [FL93]. CAQDAS involves QDAS, but can also apply to qualitative data analysis using software not specifically developed for this purpose. To avoid confusion, this work will only use the acronym CAQDAS when referring to computer-aided qualitative data analysis. QDAS will be used to refer specifically to software and tools. The term "qualitative data analysis," commonly abbreviated as "QDA," will remain written out.

As research projects become larger and more collaborative, practices shift toward dividing labor across larger pools of data within or between teams. However, current QDAS does not adequately address this trend [Pau23]. The increasingly diverse range of methods (see [Fli22]) is not necessarily reflected in the commercially driven software landscape. It seems to be not profitable [O'K20, p.141]. Recently, there has been a push to integrate AI technologies into qualitative research, particularly for transcription, qualitative coding, summarization, and thematic analysis using large language models (LLMs) [CBW+23, preprint] [Chu23] [HEQ+23] [DP24] [TBX+24]. Despite the identification of potential applications of LLMs in qualitative research (see [BNP+24]), concerns persist regarding hallucinations, traceability, consistency, opaque training datasets and bias. Moreover, fundamental epistemological challenges remain inadequately addressed [Hay25].

1.1 Proprietary Domination

Over the last two decades, open source software (OSS) has firmly established itself in *quantitative* research with the emergence of the R project[1] as the de facto standard for computational statistics. However, there is no comparable OSS for qualitative research. The field-dominating QDAS, such as MaxQDA, Atlas.ti, and NVivo, are still distributed under proprietary licenses and with closed source. Some of these tools are direct continuations of their original versions from the late 1980s and 1990s [O'K20] [FFL98] [CM02]. While initially simple and compact, they evolved into large, complex, and cumbersome applications with sometimes confusing user interfaces [Wol18]. Some of these QDAS were founded by experts in the field, such as MaxQDA by Kuckartz and QCAMap by Mayring. A recent market consolidation occurred when direct competitors acquired established software tools.[2]. The high cost of licensing, vendor lock-in, and dependency on closed data formats impact stakeholders' decisions about selecting QDAS, which we will discuss in further sections.

[1] https://www.r-project.org/about.html
[2] https://atlasti.com/lumivero-acquires-atlas-ti

1.2 Marginal Open Source

From an open science perspective, there are several reasons why research software should be free[3] and open. Anzt et al. [ABD+21] assert that the provision of free research software is essential for facilitating validation, reproducibility, improvement and (re-) use of the software, reducing legal issues, meeting ethical obligations from public funding, and opening research software to the general public. In addition, there is a need for research software that produces data in accordance with the FAIR principles (findable, accessible, interoperable, reusable) [WDA+16] but also for the software itself being published under principles that support transparency, validity, and reproducibility. This led to the emergence of FAIR principles for research software (FAIR4RS; see [HCH+20] [LGK+20]), which are also seen as foundational competencies and responsibilities of a research software engineer [GAB+23]. In the field of CAQDAS these perspectives seems to play only a marginal role. Kuckartz and Rädiker even mention, that "open source software, in the true sense of a development project supported by many people with open source code, does not yet exist" [KR20, p.815, own translation]. Additionally, the effectiveness and capability of existing QDAS in meeting modern research needs, such as live collaboration, method diversity, and AI integration, is currently unknown.

1.3 The Missing Big Picture

Based on this initial situation, it seems clear that further improvements and innovations in the open source sector are necessary to advance the field of CAQDAS in line with current research and software engineering practices. But what is the general situation in CAQDAS today and what is holding back a more widespread adoption of open source QDAS? An extensive overview of the current state can serve as a foundation for future efforts. However, an extensive, up-to-date overview is currently lacking. This work aims to address the critical lack of information on the topic by providing an introductory overview of the field. This is accomplished by systematically collecting and examining data on available, relevant, and complete QDAS according to a defined set of criteria. The findings are then discussed from a perspective that is pro open science and pro open source, in order to promote further discussion and work. The authors want to emphasize at this point on their own authorship of an open source QDAS[4].

2 Methodology

We therefore designed a systematic software review that consisted of three steps: **1.** search for QDAS and related main resources (website, code repository); **2.** filter out results that are either unavailable, irrelevant or incomplete; **3.** examine remaining results for given criteria. In the following, we describe the proceeding for these three steps and the involved criteria in more detail.

[3] For a definition of free software, see https://www.gnu.org/philosophy/free-sw.html

[4] https://github.com/openqda

2.1 Finding Relevant QDAS

The search for QDAS was intended to provide a broad, yet purposeful range of results for both proprietary and free software. It was therefore decided to use a mix of relevant websites, as well as publicly available LLMs to conduct our search.

The CAQDAS Networking Project[5] represents one of the pioneering projects in the field of CAQDAS and was expected to list very relevant software. The Rotterdam Exchange Format Initiative (REFI), is a coalition of market-leading vendors of QDAS who defined a common interchange standard for interoperability [REF19] [ECNW20]. Their website[6] was selected to retrieve REFI-compliant QDAS. The Research Software Directory[7] was selected to retrieve QDAS, compliant to FAIR4RS, "while enabling a qualitative assessment of their impact" [HCH+20]. A search on Zenodo[8] was selected to find software that is more likely to (partially) meet the FAIR4RS principles. A search on GitHub[9] was selected to identify relevant open source repositories, independent of their current state of development. A search on Wikipedia[10] was selected for a broad range of results, suited for the general public. A SPARQL query on Wikidata[11] was selected to provide a structured result for instances or subclasses of CAQDAS, independent from their actual state of availability, relevance, or completeness. The list from the archived Social Science Software (SoSciSo) website[12] was selected to provide a legacy set of results (2017) that contain long-living software that is still available today.

In addition to these searches, a variety of LLMs were included to reveal results that had been placed in their training data and had gone unnoticed. Only resources and information that could be found publicly without registration or authentication were considered. To reduce potential fingerprinting affecting the results, a private browser session was expected to be used. GitLab was considered but not included because it does not allow public searches without registration.

2.2 Filtering the Results

A broad search was expected to produce results that were out of scope. This could be due to dead references (unavailable), software that is not suitable for qualitative data analysis (irrelevant), or a focus on a specific domain or analysis step, which would make a complete analysis impossible (incomplete). Starting from this position, we created the following inclusion criteria for filtering:

A result is considered **available** if it is associated with an operational website, a downloadable resource, or a repository. A brief description is provided, and the linked resource is clearly identified as research-related. The result is neither archived nor deleted. A result is seen as **relevant**, if it falls within the domain of qualitative research and is an executable piece of software. This excludes resources, such as learning materials, reports, guides, publications or documentation. To maintain an appropriate number of open source repositories, a repository must have at least

[5] https://www.surrey.ac.uk/computer-assisted-qualitative-data-analysis/resources/choosing-appropriate-caqdas-package
[6] https://www.qdasoftware.org/
[7] https://research-software-directory.org
[8] https://zenodo.org
[9] https://github.com
[10] https://wikipedia.org
[11] https://www.wikidata.org
[12] https://web.archive.org/web/20170706051957/http://www.sosciso.de/de/software/datenanalyse/qualitativ/qda/

two stars and be actively maintained. It should also have at least one significant contribution within the last few years or be marked as feature complete. A result is considered **complete** if it extends beyond a single method (e.g., transcription) and is not merely a script, plugin, or analogous extension of existing software. Finally, it must not be limited to a single domain.

2.3 Examination Criteria

The objective of examining the results was to use criteria that could be retrieved using only publicly available information, such as vendor websites, documentation, and README files, without the need for a deep inspection (i.e., purchasing, installing, running, and testing). With this in mind, we developed the criteria from the perspectives of stakeholders involved in CAQDAS: researchers, research groups, institutions, reviewers, and developers.

Firstly, the *researcher's perspective* is to be considered. It is represented by the individual who directly interacts with the QDAS. For them, it is desirable, that QDAS is available, relevant and complete, as previously incorporated into the filter criteria in Subsection 2.2. Relevant methods and features should be well described in a documentation. From a technical point of view, the individual is concerned with the executability of the software on their designated platform. Implications, by given license and costs, are of particular concern, especially for individuals covering expenses independently. Methodological aspects, such as supported media formats that can be coded, as well as supported import and export formats are fundamental to qualitative research, as it deals with a wide variety of content types. In qualitative data analysis, coding refers to annotating the content, embedded within the software, which requires type-specific capabilities, for example, a video player with annotation support for coding videos. Formats for coding include plain text, rich text, images and graphics, audio, and video, as well as documents. Finally, AI integration and custom scripting can be a useful addition in automating time consuming chore tasks (transcription, code, and category creation), programming workflows (i.e., see [Wan22]) and customizing output formats (visualizations, summaries, post-processing).

Research groups are another perspective to consider. In these groups, division of labor and collaboration drive a demand for interoperability and collaboration features. Especially, collaborative coding is seen as "[...] an established method for reducing bias during qualitative analysis" [RH18] via [BMJ+23, p.2]. Note, that collaboration itself implies proper platform support for individuals devices. In the event that collaboration is managed by a central server, then the involved servers should support on-premise hosting in order to obtain full control over data and workflows.

Thirdly, the *perspective of institutions* that intend to utilize a particular software for research and teaching purposes must be considered. An institutional implementation of software also raises the question of sustainability, in particular the continuation and maintenance of the software. Furthermore, legal factors, such as licensing, costs and compliance (privacy, security, ethics), as well as technical factors, such as on-premise deployment and technical support do play a role. For this study, we focused on whether a security policy was present and publicly accessible, whereas an assessment of privacy compliance and ethical implications was out of scope.

From the *perspective of a reviewer*, transparency is an inherent requirement for comprehending the analysis process. O'Kane et al. [OSL21] investigated existing CAQDAS techniques

Figure 1: SPARQL query to request instances and subclasses of “caqdas” from Wikidata

```
SELECT DISTINCT ?item
WHERE {
    ?item wdt:P31/wdt:P279* wd:Q4550939
}
```

that support transparency. However, presented techniques focused on outcomes of the analysis process, but scarcely reflect on the process itself. We therefore aimed to look for QDAS with audit functionality to document the analysis process and to further promote transparency and understanding.

Finally, the *developers' perspective* is a subject that is concerned with software engineering topics, including software architecture, technology stacks, maintenance, and code quality. We identified an extensible architecture as a key element for software sustainability, promoting code reusability and customization through plugins, enabling users to overcome limitations on their own. We therefore looked for information indicating support for plugins or extensions.

Further inquiries into various technology stacks, programming languages, databases, or software frameworks and dependencies were initially deemed important. However, we excluded this from our examination, because it would require a more thorough analysis. The same applies to the analysis of code quality, which was also beyond the scope of this work.

3 Data Collection

We collected the data between July/2024 and May/2025. All direct references from the REFI website, Soscio and the CAQDAS Networking project websites were manually collected. A search for the term “caqdas” on Wikipedia redirected to a related page[13], where all listed references were also manually collected. It should be noted that the page incorrectly uses the term CAQDAS, implying it being synonymous with QDAS.

A SPARQL query on Wikidata was prepared by a search for the main item that represented CAQDAS as class[14]. Then, the query (see Figure 1) was built to search for all instances or subclasses of CAQDAS on that platform[15], which returned 12 records that were manually reviewed and collected.

An initial search on GitHub for the term “caqdas” resulted in 13 entries overall. An additional search for term “qda software” resulted in exactly five results. In order to increase results, a third search was added using the rather generic term “qualitative data analysis”, leading to 343 results from which we collected 55 with two or more stars. On Zenodo a search for the term “caqdas” with filter for “software” returned two entries, which we were the authors of. Therefore, the search was also broadened by using the term “qualitative data analysis”, including the “software”

13 https://en.wikipedia.org/wiki/Computer-assisted_qualitative_data_analysis_software
14 https://www.wikidata.org/wiki/Q4550939
15 https://query.wikidata.org/

Table 1: Search results from the given platforms and LLMs

Platform	Inquiry Method	Raw Results	Cleaned
CAQDAS N.P.	Listed on page	18	18
Claude 3 Haiku	Prompt	8	8
GitHub	search: caqdas, qda software, qualitative data analysis	73	39
GPT-4o	Prompt	20	13
Llama 3.3	Prompt	88	26
Mistral small 3	Prompt	121	14
o3-mini	Prompt	13	13
REFI QDA	Listed on page	7	7
RSD	search: caqdas	0	0
socsio	Listed on page	37	37
wikidata	SPARQL query	13	13
wikipedia	search: caqdas	25	25
Zenodo	search: caqdas, qualitative data analysis	22	8
All			221
Unique			102

filter. This resulted in 50,038 results and it was decided to record the first page of 20 entries that Zenodo listed with highest relevancy. A search on the research software directory[16] for the term "caqdas" did not match any result.

Five AI models were consulted via Duckduckgo AI chat[17], where Duckduckgo acts as a proxy and enables an AI chat without registration or login. All LLMs were given the exact following input request text (prompt) to copmile and return a list of potential QDAS: "You are a researcher teaching qualitative research at the university. Compile a minimalist list of all known QDA Software (no limitation) where each entry consists of the name and an optional URL. Do not provide any further information and focus on list completeness." Outputs by LLMs were immediately copied and saved in a plain text. Due to the probabilistic nature of LLMs, we expected a high amount of false-positives, which turned out to be the case for "LLAMA 3.3 70b", reporting 88 references but only 26 were in any aspect related to qualitative research. A similar behavior was observed from "Mistral small 3" (121 entries, 16 after cleaning, no URLs provided), while the other three models ("Claude 3 Haiku"; GPT-4o" and "o3-mini") returned valid and usable results.

All queries, examinations and searches on all platforms resulted in 102 unique entries that remained after an initial cleaning of the raw results and are summarized in Table 1. All entries were then examined using available public information in order to match for inclusion (available, relevant, complete; see Table 2), resulting in 28 final entries, summarized in Table 3. The collected data, as well as the examination results have been published under a public domain license (CC-0 1.0) to Zenodo [KW25a] and Havard Dataverse [KW25b].

[16] https://research-software-directory.org/software?search=caqdas&order=mention_cnt&page=1&rows=12
[17] https://duckduckgo.com/?q=DuckDuckGo+AI+Chat&ia=chat&duckai=1&atb=v322-1

Table 2: Matches of inclusion criteria for 102 unique results. A = available; R = relevant; C = complete.

None	Available	Relevant	Complete	AR	AC	ARC
8	94	55	35	27	7	**28**

Table 3: Final 28 QDAS used for examination; sorted alphabetically; costs are in Euro and per user per year.

Name	License	Costs	URL
aquad	GPL-3.0		http://www.aquad.de/
ATLAS.ti	Proprietary	184,00	https://atlasti.com/de
DEDOOSE	Proprietary	156,00	https://www.dedoose.com/
Delve	Proprietary	200,00	https://delvetool.com/
dicto	AGPL-3.0		https://dictoapp.github.io/dicto/
ELAN	GPL-3.0		https://archive.mpi.nl/tla/elan/
f4analyse	Proprietary	111,00	https://www.audiotranskription.de/f4/
gQDA	Not found		https://github.com/jailop/gQDA/security
HyperResearch	Proprietary	199,00	http://www.researchware.com/products/hyperresearch.html#Power
lama	MIT		https://joss.theoj.org/papers/f91f03787429efa115f1e3aa144254bc
libreQDA	BSD-3		https://libreqda.org/
llm_qualitative_data_analysis	Not found		https://github.com/Gamma-Software/llm_qualitative_data_analysis/
LLMCode	MIT		https://github.com/PerttuHamalainen/LLMCode
MAXQDA	Proprietary	95,20	https://www.maxqda.com
Nvivo	Proprietary	430,00	https://lumivero.com/products/nvivo/
OpenQDA	AGPL-3.0		https://openqda.org/
qc	AGPL-3.0		https://github.com/cproctor/qualitative-coding/
QCAmap	Proprietary		https://www.qcamap.org/ui/de/home
QDA Miner	Proprietary	256,00	https://provalisresearch.com/products/qualitative-data-analysis-software/
QDAcity	Proprietary		https://qdacity.com/
Qualcoder	LGPL-3.0		https://qualcoder.wordpress.com/
Quirkos	Proprietary	96,00	https://www.quirkos.com
Requal	Custom open		https://www.requal.app/
Taguette	BSD-3		https://www.taguette.org
TamSys	GPL-2.0		https://tamsys.sourceforge.io
TigerDataCoding	Not found		https://github.com/paigerodeghero/TigerDataCoding/security
Transana	Proprietary	170,00	https://www.transana.com/
webQDA	Proprietary	151,00	https://www.webqda.net/

4 Findings

The following section summarizes the findings for the final data, which is attached as *03-examined-data-overview.csv*, as well as part of the published dataset. The final entries were examined using publicly available information, found in the given references (website, documentation, repository) and in regards to the criteria from Subsection 2.3. Proprietary vendors provided extensive documentation as separate documents, while open source projects either provided README files or or a Wiki within the same repository. Some of the open source tools were scarcely documented. Additional searches in the repository, using the repository-scoped search on GitHub, were undertaken to identify support for formats, collaboration, plugins, scripts or AI model support.

As shown in Table 3, we identified 25 licenses out of the 28 QDAS with 3 of them neither containing a license file nor a license statement. Nearly half of them were distributed under a proprietary license (13), while the other half was distributed under a free license (12). The minimal costs per user per year were recorded, monthly subscriptions were converted equivalently. The resulting costs fell within a range of €95.20 for the MaxQDA student license to €430 for the Nvivo annual subscription license. Two out of the 13 proprietary tools offer a complimentary license (QCAMap, financed via donations; QDACity, announced to potentially require payment in the future). It should be noted, that student license fees differ from the default user licenses. Some proprietary vendors offered additional licenses for certain features, such as AI integration, automatic transcription, or collaboration. These costs were not incorporated into the minimum cost calculation.

Native platform support was found for 13 tools. Primary build targets were Windows (13) and MacOS (12), while native Linux was supported by five open source tools and two proprietary vendors. Only six tools provided a platform-independent installation via Docker images. Native mobile builds were not found during the examination. A summary of platform availability is found in Table 4.

A great portion of QDAS (18) were available as web-based applications, where only a web browser is a requirement for using the software. Furthermore, the majority of tools (21) offered ways to run on a custom infrastructure, either because they are distributed as a local native application, or because they allow to run a local setup of involved server backend. For proprietary tools, this sometimes involved an additional on-premise license, including additional costs, while for some tools collaboration or AI features were not available on-premise at all (see Table 5).

We identified six major types of media that can be coded: text, images, PDF documents, spreadsheets, audio, and video. Text was clearly the dominating media type and was present in nearly all tools (26) with the exception of two audio/video-specific ones. Proprietary QDAS generally offered a wider range of formats, while two among them were the only ones that offered to code directly within spreadsheets. Audio and video were only supported by less than a half of the software. Additionally, three native desktop tools relied on a locally installed media player for their audio/video support. These dependencies are either system-level (Windows Media Player, Apple QuickTime) or require installation (VLC Player). One tool did not use local video files but connected via API to existing video platforms, such as YouTube or Vimeo in order to stream the videos. Table 6 provides an overview of the several formats.

Table 4: Platform availability for all 28 QDAS. Proprietary licenses are highlighted.

Name	Windows	MacOS	Linux	Web	Docker	On-Premise
aquad	x					
ATLAS.ti	x	x		x		off-premise
DEDOOSE	x	x	x	x		off-premise
Delve				x		
dicto				x		x
ELAN	x	x	x			x
f4analyse	x	x	x			x
gQDA	x	x	x			x
HyperResearch	x	x				x
lama				x	x	x
libreQDA				x	x	x
llm_qualitative_data_analysis				x		x
LLMCode				x		x
MAXQDA	x	x				x
Nvivo	x	x				x
openqda/openqda				x	x	x
qc				x	x	x
QCAmap				x		
QDA Miner	x					x
QDAcity				x		
Qualcoder	x	x	x			x
Quirkos				x		x
Requal				x	x	x
Taguette	x	x	x	x	x	x
TamSys		x	x			
TigerDataCoding				x		x
Transana	x	x		x		x
webQDA				x		
Proprietary	8	7	2	8	0	7
Open	5	5	5	10	6	13
All	13	12	7	18	6	20

Table 5: On Premise Availability for all 28 results; Off-premise server licenses were not included.

License Type	Server Involved	On-Prem Single	On-Prem Collab	On-Prem AI
Proprietary	8	2	1	0
Open	10	10	9	2

Table 6: Supported media formats that can be coded using the given 28 QDAS.

License Type	Text	Image	PDF	Spreadsheet	Audio	Video
Proprietary	13	9	4	2	7	7
Open	13	2	2	0	4	5
All	26	11	6	2	11	12

File formats were found to be unevenly supported, as shown Table 7. With the exception of one stream-based tool, all other 27 tools provided at least one way to import material from files. Few tools (such as MAXQDA or ATLAS.ti) provided extensive support for importing files in many formats, the rest offered only a very reduced set of formats. Text-based file formats (text, pdf, doc/x, rtf) are widely supported, followed by jpg images and mp3 audio. It should be noted, that importing a format does not imply coding support for these types of media. There were two QDAS that allowed for importing an entire database from a file.

Most supported export formats found were pdf (8), xls/x (8), doc/x (8), csv (6) and txt (5), while other formats, such as spss (3) and json (1) were also found. Interoperability via REFI is claimed to be supported by nine tools, which are mostly proprietary (7), and was partially supported by two tools via importing/exporting codebooks.

The majority of tools offered ways to collaborate, but only a minority (5) offered real-time collaboration, where contributions are proactively merged by the software. Among these, there was one tool (ATLAS.ti), that required an additional (proprietary) license for the usage of an additional web-based app in order to realize the collaboration. Other tools offered multi-user support for parallel collaboration via general multi-user and project features, but merging was required to be done manually. Finally, there were seven tools with no integrated collaboration support whatsoever and where users have to manually exchange data (if supported) in order to achieve collaboration. Audit features to improve reproducibility and to support collaboration were found in seven tools, though the level of detail of these logs has not been examined.

Explicit AI integration was found in nine results, while three additional results integrated AI only for their transcription. Only two tools allowed for the integration of local or custom models, while the other models were limited to a specific provider, such as OpenAI (5), the software vendor themselves (1) or were undisclosed (4). None of the QDAS that included AI models revealed, whether a system prompt was prepended, how it looked like or how users can alter it. Additionally there were no proprietary vendors that allowed to run the AI analysis on premise.

Only a minor set of results did include a citation metadata (such as .cff files; 3 out of 28), were archived on a platform like Zenodo (4 out of 28), supported project interoperability via REFI (9 out of 28; 2 support codebook exchange), provided a virtualization technique in form of a Docker setup (6 out of 28), supported code reusability via an extensible plugin architecture (1 out of 28; 1 provides partial support) or supported custom scripting (3 out of 28). There was no entry found that provided all criteria, while two entries provided four of these criteria (Taguette, provided cff file, Zenodo, REFI and Docker; OpenQDA, provided cff file, Zenodo, Docker and partial plugin support. A summary is given in Table 8.

While websites and repositories were searched for details on security, there were only one proprietary vendor website and one open source repository with an explicit security policy. Searching for security statements in licenses and privacy declarations revealed, that web-based services and collaboration servers of proprietary vendors may share data with several partners within the EU, US, and other countries, which users have to agree as part of the license.

There were eleven proprietary tools with licenses that generate a revenue stream that can be used to reinvest in development and maintenance. However, the team sizes were not disclosed. One proprietary project is funded by donations, but has apparently only one developer. Another proprietary project is free, but announced to be paid soon. Results for open source projects in contrast showed, that only one seems to have received a constant funding over the years[18], while the other projects did not claim to have received systematic funding yet. Additionally, there was only one project with a seemingly strong community of 32 contributors[19], while the rest of the projects were limited to much less or unknown contributors, as shown in Table 9.

Table 7: 20 most supported import file formats across all 28 QDAS.

Extension	Open	Proprietary	Combined
txt	9	10	19
pdf	4	10	14
doc/x	3	10	13
rtf	3	9	12
jpg	3	8	11
mp3	3	8	11
png	2	9	11
mp4	2	7	9
undisclosed	2	7	9
csv	5	3	8
wav	2	6	8
xls/x	1	7	8
gif	1	6	7
jpeg	1	4	5
mov	1	4	5
odt	3	2	5
avi	0	4	4
bmp	0	4	4
vlc	4	0	4
html	1	2	3

Table 8: Selected FAIR criteria found among all 28 results

License	Citation Metadata	On Zenodo	Full REFI	Docker	Plugins	Scripts
Proprietary	0	0	7	0	0	0
Open	3	4	2	6	1	0
All	3	4	9	6	1	3

[18] https://archive.mpi.nl/tla/elan/funding

[19] QualCoder, see https://github.com/ccbogel/QualCoder/graphs/contributors

Table 9: Contributors found in 15 open source repositories

Contributors	20+	10-20	5-10	2-5	1	Unknown
Repositories	1	2	1	6	3	2

5 Discussion

5.1 Licensing Costs and Proprietary Software Remain a Fundamental Problem

One of the core issues with CAQDAS is its vast domination by proprietary commercial software and its implications for teaching qualitative methods, as well as the research practice itself. This problem has been covered by recent literature: due to financial reasons, there is usually exactly one software product or service in use by research groups [WS17]. Staff has to learn the very tool that is in use [O'K20, p.146], which might differ a lot from the tools they used beforehand or were trained with.

Positively, "[c]ommercial research software often has revenue flows that can facilitate sustainable software development, maintenance, and documentation as well as the operation of adequate infrastructure." [ABD+21]. This argument is generally backed by the findings, as proprietary QDAS offer a wider range of supported formats and features, and enable a much richer coding experience, including live collaboration and AI integration. Additionally, the proprietary software vendors were leading in implementing an interchange standard between their QDAS. Nevertheless, the argument misses out the reality, that not all features are equally considered for being implemented, as already pointed out in Section 1 via [O'K20]. Open source solutions could potentially fill this need by implementing an open and extensible architecture to enable researchers and groups to implement their own solutions for edge cases and emerging methods.

For all found proprietary QDAS, users have to rely on statements about compliance of infrastructure and encryption. However, users cannot review code or analyze behavior of servers, which is especially problematic with servers involved in collaboration and AI tools, or servers outside of users' jurisdictions, given the sensitive quality of the data, such as interviews or observation protocols, including photographs or videos of participants. Furthermore, missing explicit security policies raise the impression of the topic being an afterthought. Neither users nor security researchers have a clear guidance on how to detect, report and fix vulnerabilities. Shockingly, we found this to be the case for proprietary and open source QDAS alike.

5.2 No AI model control

AI remains a blackbox in many QDAS, preventing users from access to internals such as system prompts, which is seen to have a negative influence on the quality of AI responses [GGL+24]. The current CAQDAS landscape – both proprietary or open source – shows a great dependency on commercial non-EU AI vendors, such as OpenAI. While this is highly problematic in regards to privacy and security, it seems to be not easily understandable for non-technical users. This is alarming, because it undermines critical reflection on the use of such technologies, a topic that is intensively discussed in the social sciences [ES24, preprint].

In order to systematically test the use of LLMs in qualitative data analysis, direct control of the

AI is required, e.g. to evaluate its validity through targeted repetition of AI-generated codings and comparison with human codings (see [GGL+24] [TBX+24]), but also to use optimization methods such as multi-shot and example-based prompting (see [XYL+23]) or chain-of- thought reasoning [Dun24, preprint]. Furthermore, it seems to be necessary to systematically assess methods such as Retrieval-Augmented Generation (see [LPLL24]), Fine-Tuning (see [SHR+24, preprint]) and multi agent systems (see [KPD25]). Such direct control can be realized by using local models, which is currently explicitly offered by only two QDAS. Summarized, the AI situation is even less sufficient than with collaboration, because inclusion of open source models or custom models is urgently needed.

5.3 Lack of Extendability and Collaboration by Design

Major contributing factors of successful OSS, besides healthy and large enough contributor communities (see [CAH03]), were identified in different ways of participation as well as documented software architecture build for collaborative development [NYN+02] [GCT+21]. In the field of quantitative research, the R project and its surrounding infrastructure are designed to be flexible, decoupled, and extensible through packages, lowering the barrier for contribution. While it might be a desired requirement to have an equivalent language ecosystem for qualitative research, the current state of CAQDAS does not provide any hints toward such an effort. This may be partly due to the fact, that qualitative research is not as well established as quantitative research in the field of (research-) software engineering. Also, qualitative researchers are more likely to lack the necessary skills to contribute to such an ecosystem.

In regards to existing QDAS, extendability can at least be achieved in a more narrow and restricted way using an architecture that is open for custom extensions. A proper plugin architecture could also be a potential solution to the issues, described in the prior subsections. While attempts toward a pluggable architecture are in progress[20], the majority of tools remains non-extensible and fixed toward their current features. Especially in established software products, implementing open interfaces for extendability would require a rethinking and redesign of their (sometimes legacy) architectures. In consequence, one proprietary vendor created a separate web-based extension for collaborative work, rather than redesigning the existing architecture[21].

Collaboration is an essential aspect of qualitative data analysis, because intersubjective replicability and communicative validation are two of the central quality and validity criteria in qualitative research [Kro05, pp.286-294] [CA14, pp.109-161] [Fli23, pp.489-494]. At the same time, proprietary vendors seem to lock in users into their infrastructure, as collaborative solutions are not, or only partially, available on-premise and require additional licenses (ATLAS.ti; MaxQDA). The tendency toward web-based applications (18 out of 28) can be seen as a first step for a better support of collaboration, because it is easier to implement collaboration on top of web based software architectures, as they often already include managing users, groups, projects and data exchange. Existing tools fail to fully and inherently support collaborative workflows with proper auditability. Tools with realtime representation of users' edits are a minority. Furthermore, attempts to collaborate across different QDAS were not found at all, although methods and data structures are likely to be similar (especially when considering existing interoperability

[20] See https://openqda.github.io/openqda/plugins/overview.html; affiliation to the project is addressed in Section 6.

[21] See https://atlasti.com/research-hub/live-collaboration-in-atlas-ti-web.

standards, such as REFI). While collaboration seems to be generally realizable using the majority of given tools, this is mostly done using parallel or even manual workflows. This situation is generally insufficient and there is a need for a generalized collaboration protocol across QDAS to support independent asynchronous realtime collaboration, may it be server-based or peer-to-peer.

5.4 It's just not FAIR

Research data and software in general received positive improvements toward sustainability and freedom by the establishment of FAIR and FAIR4RS principles. However, the findings indicate a tendency toward a highly un-FAIR software market of QDAS and their incorporated data formats. While the market dominating proprietary tools do support a wide range of formats, the closed nature of proprietary project formats can form a strong barrier for collaboration and result in data silos. A positive first step to solve this issue has already been made by publishing a common interchange standard in 2016 with the latest version being published in 20219 [REF19].

Despite this fact, the exchange between the proprietary QDAS is still lacking and not achievable without significant loss of data [ECNW20]. Additionally, the standard does not explicitly consider additional metadata being added, creating another barrier toward findability, accessibility, and archiving in general. This is justified in the REFI standard as database fields only being added, if at least two of the initiative's member group use them [REF19]. Since the developing member group currently consists of a few proprietary vendors, it is less likely to expand beyond their own supported fields in the near future. A growing and active community of researchers and RSEs around open and free QDAS could help to expand and improve the standard and the overall FAIRification of produced qualitative data.

Similar circumstances do exist for the software itself. Although there are some potential candidates for becoming a fully FAIR research software project, the majority of the tools are far away from such a state. Most open source QDAS are currently not a serious option, because installation and operation require technical skills beyond graphical user interfaces. Such skills are less expected in qualitative research groups, rendering most of the open source software inaccessible. Additionally, the majority of QDAS do not provide ways to audit or review the analysis process in a detailed way and the data loss during project export makes it impossible for reviewers and the general public to attempt any reproduction of published works.

Many of our arguments in this discussion indicate the need for a strong open source software community. However, there seems to be a lack of sustainable, long-term funding in the open source CAQDAS landscape and of developer collaboration to reach a more critical community size. This issue is known and has already been pointed out for open source in general [ABD+21]. Further work should attempt to investigate the reasons for this situation and suggest measures for improvement.

6 Critical Reflection

This work intends to reflect on the state of CAQDAS as a whole from a pro-open science, open data, open access, and open source perspective. However, due to its broad scope, which encompasses an entire field, it remains an introductory overview that highlights tendencies. A complete overview would require additional work involving stakeholders and in-depth reviews of QDAS. Furthermore, there is a clear bias and motivation by the authors against proprietary research software and for open science. The examination criteria can be seen as critically influenced by the works, cited in Section 1. This, in turn, may lead to more findings that support these arguments, than others. Additionally, we are actively developing an open-source QDAS, which is also listed in the results (OpenQDA).

Despite the extensive search, the sample is likely to be incomplete, and the examination was limited to publicly available information. At least two available, relevant and complete QDAS[22][23] were not among the results from any of the given platforms. Furthermore, open source tools may not be found on GitHub or Zenodo if they do not provide relevant search information in their README, description, or topic tags[24]. Consequently, a broader search using different search terms or refined prompts would have been desirable, but it was not feasible. Additionally, a systematic literature review could have revealed more relevant references, but it was beyond the scope of this work. The information used to examine the software tools may be incomplete because there was no objective indicator of documentation coverage. This could have led to inconsistent or false interpretations of the findings. Examining documentation coverage would imply a deep-level inspection of every QDAS but could be a good starting point for future work.

It is debatable, whether scripts, packages and plugins should have been excluded, since they also represent research software in a broader sense. However, analyzing the RQDA[25] package as an example of a large R package for qualitative data analysis, did not substantially add to the presented findings or the state of CAQDAS. The examination of the software did omit criteria, such as usability, code quality, test verification, and other best practices in software engineering. These criteria require not only a deep inspection of each software, but also a purchased license for some of the QDAS. At this point, we like to point at work done by others as a supplement [Tra25] [KR20]. Due to scope and legal complexity, the important topic of privacy has been omitted entirely, but it should be included as an important criterion in upcoming examinations.

7 Conclusion and Outlook

This work provides an initial overview of the current state of CAQDAS, though it remains incomplete. Proprietary software dominates the field. Not only is it expensive, but it also hinders the development of QDAS that can support open science principles in qualitative research. There seems to be a lack of extensibility, collaboration, AI model control, and auditability. Additionally, a generalized collaboration protocol is needed to support real-time collaboration across QDAS, whether server-based or peer-to-peer. Although the discussion was limited to 28 QDAS samples, the findings are likely representative of the broader CAQDAS field. Future work should build on this to provide a comprehensive overview.

[22] https://dokumet.de/, proprietary, web based

[23] https://www.mangold-international.com/de/produkte/software/interact-videographie-software.html, proprietary, focused on audio/video

[24] Examples include https://github.com/forTEXT/catma, https://zenodo.org/records/15000951 or https://textada.com/ which were not findable in relation to queries including "qualitative" but rather in relation to "text analysis".

[25] https://rqda.r-forge.r-project.org/, last released in 2018.

This study focused on QDAS, but the findings are also relevant to the teaching of qualitative methods in academia. The lack of open source QDAS that is both user-friendly and well-documented poses a significant challenge to its use in a classroom setting. At the same time, the high cost of proprietary software places a financial burden on university departments, creates vendor lock-in, and hinders students from using it for their personal research. This is problematic given the importance of qualitative research in many fields, including the social sciences, humanities, and health research. Therefore, it is crucial to raise awareness of the importance of open science principles in qualitative research and to encourage researchers to use open source tools.

From the perspective of methods development in qualitative research we propose a more open and collaborative approach to developing QDAS. This could entail forming a community of researchers and developers dedicated to creating open source QDAS that are accessible and extensible and that support collaboration. This approach would benefit researchers and contribute to the advancement of qualitative research methods. In our view, successful QDAS development communities require good example code, system architecture, and documentation, as well as different participation opportunities.

As a starting point for future examinations of the state of CAQDAS, we published the dataset and the documented methodology. For this purpose, we created the "State of CAQDAS" repository on GitHub[26]. The scientific community is encouraged to contribute to methodological improvements by suggesting additional sources for data inquiry and tools for automation, such as scraping, processing data, and generating reports.

Bibliography

[ABD+21] H. Anzt, F. Bach, S. Druskat, F. Löffler, A. Loewe, B. Y. Renard, G. Seemann, A. Struck, E. Achhammer, P. Aggarwal, F. Appel, M. Bader, L. Brusch, C. Busse, G. Chourdakis, P. W. Dabrowski, P. Ebert, B. Flemisch, S. Friedl, B. Fritzsch, M. D. Funk, V. Gast, F. Goth, J.-N. Grad, J. Hegewald, S. Hermann, F. Hohmann, S. Janosch, D. Kutra, J. Linxweiler, T. Muth, W. Peters-Kottig, F. Rack, F. H. Raters, S. Rave, G. Reina, M. Reißig, T. Ropinski, J. Schaarschmidt, H. Seibold, J. P. Thiele, B. Uekermann, S. Unger, R. Weeber. An environment for sustainable research software in Germany and beyond: current state, open challenges, and call for action. *F1000Research* 9:295, Jan. 2021.
doi:10.12688/f1000research.23224.2
https://f1000research.com/articles/9-295/v2

[BMJ+23] V. Bogachenkova, E. C. Martins, J. Jansen, A.-M. Olteniceanu, B. Henkemans, C. Lavin, L. Nguyen, T. Bradley, V. Fürst, H. M. Muctadir, M. Van Den Brand, L. Cleophas, A. Serebrenik. LaMa: a thematic labelling web application. *Journal of Open Source Software* 8(85):5135, May 2023.
doi:10.21105/joss.05135
https://joss.theoj.org/papers/10.21105/joss.05135

[26] https://github.com/zemki/state-of-caqdas

[BNP+24] A. Barany, N. Nasiar, C. Porter, A. F. Zambrano, A. L. Andres, D. Bright, M. Shah, X. Liu, S. Gao, J. Zhang, S. Mehta, J. Choi, C. Giordano, R. S. Baker. ChatGPT for Education Research: Exploring the Potential of Large Language Models for Qualitative Codebook Development. In Olney et al. (eds.), *Artificial Intelligence in Education*. Volume 14830, pp. 134–149. Springer Nature Switzerland, Cham, 2024.
doi:10.1007/978-3-031-64299-9_10
https://link.springer.com/10.1007/978-3-031-64299-9_10

[CA14] K. Charmaz, Askews & Holts Library Services. *Constructing grounded theory*. Introducing qualitative methods. SAGE, Los Angeles, 2nd edition. edition, 2014. OCLC: 892522517.

[CAH03] K. Crowston, H. Annabi, J. Howison. Defining Open Source Software Project Success. In *Proceedings of the International Conference on Information Systems (ICIS)*. Seattle, WA, USA, Dec. 2003.
doi:10.1287/mnsc.1060.0550

[CBW+23] R. Chew, J. Bollenbacher, M. Wenger, J. Speer, A. Kim. LLM-Assisted Content Analysis: Using Large Language Models to Support Deductive Coding. June 2023. arXiv:2306.14924.
doi:10.48550/arXiv.2306.14924
http://arxiv.org/abs/2306.14924

[Chu23] L. A. Chubb. Me and the Machines: Possibilities and Pitfalls of Using Artificial Intelligence for Qualitative Data Analysis. *International Journal of Qualitative Methods* 22:16094069231193593, Oct. 2023.
doi:10.1177/16094069231193593
https://journals.sagepub.com/doi/10.1177/16094069231193593

[CM02] J. W. Creswell, R. C. Maietta. Qualitative Research. In *Handbook of Research Design & Social Measurement*. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States of America, 2002.
doi:10.4135/9781412984386
https://methods.sagepub.com/book/handbook-of-research-design-social-measurement

[DP24] S. De Paoli. Performing an Inductive Thematic Analysis of Semi-Structured Interviews With a Large Language Model: An Exploration and Provocation on the Limits of the Approach. *Social Science Computer Review* 42(4):997–1019, Aug. 2024.
doi:10.1177/08944393231220483
https://journals.sagepub.com/doi/10.1177/08944393231220483

[Dun24] Z. O. Dunivin. Scalable Qualitative Coding with LLMs: Chain-of-Thought Reasoning Matches Human Performance in Some Hermeneutic Tasks. arXiv:2401.15170, Feb. 2024. arXiv:2401.15170.
doi:10.48550/arXiv.2401.15170
http://arxiv.org/abs/2401.15170

[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

[ES24] J. Eschrich, S. Sterman. A Framework For Discussing LLMs as Tools for Qualitative Analysis. July 2024. arXiv:2407.11198.
doi:10.48550/arXiv.2407.11198
http://arxiv.org/abs/2407.11198

[FFL98] N. Fielding, N. Fielding, R. M. Lee. *Computer analysis and qualitative research*. New technologies for social research. SAGE, London, 1. publ edition, 1998.

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

[Fli22] U. Flick. *The SAGE Handbook of Qualitative Research Design*. SAGE Publications, Limited, London, 1st ed edition, 2022.

[Fli23] U. Flick. *An introduction to qualitative research*. SAGE, Los Angeles London New Delhi Singapore Washington, DC Melbourne, 7th edition edition, 2023.

[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

[GCT+21] M. Guizani, A. Chatterjee, B. Trinkenreich, M. E. May, G. J. Noa-Guevara, L. J. Russell, G. G. Cuevas Zambrano, D. Izquierdo-Cortazar, I. Steinmacher, M. A. Gerosa, A. Sarma. The Long Road Ahead: Ongoing Challenges in Contributing to Large OSS Organizations and What to Do. *Proc. ACM Hum.-Comput. Interact.* 5(CSCW2):407:1–407:30, Oct. 2021.
doi:10.1145/3479551
https://dl.acm.org/doi/10.1145/3479551

[GGL+24] J. Gao, Y. Guo, G. Lim, T. Zhang, Z. Zhang, T. J.-J. Li, S. T. Perrault. CollabCoder: A Lower-barrier, Rigorous Workflow for Inductive Collaborative Qualitative Analysis with Large Language Models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. CHI '24. Association for Computing Machinery, New York, NY, USA, 2024. event-place: Honolulu, HI, USA.
doi:10.1145/3613904.3642002
https://doi.org/10.1145/3613904.3642002

[Gup24] A. Gupta. *Qualitative Methods and Data Analysis Using ATLAS. ti: A Comprehensive Researchers' Manual*. Springer Texts in Social Sciences Series. Springer International Publishing AG, Cham, 1st ed edition, 2024.

[Hay25] A. S. Hayes. "Conversing" With Qualitative Data: Enhancing Qualitative Research Through Large Language Models (LLMs). *International Journal of Qualitative Methods* 24:16094069251322346, Apr. 2025. Publisher: SAGE Publications Inc.
doi:10.1177/16094069251322346
https://doi.org/10.1177/16094069251322346

[HCH+20] W. Hasselbring, L. Carr, S. Hettrick, H. Packer, T. Tiropanis. From FAIR research data toward FAIR and open research software. *it - Information Technology* 62(1):39–47, Feb. 2020.
doi:10.1515/itit-2019-0040
https://www.degruyter.com/document/doi/10.1515/itit-2019-0040/html

[HEQ+23] L. Hamilton, D. Elliott, A. Quick, S. Smith, V. Choplin. Exploring the Use of AI in Qualitative Analysis: A Comparative Study of Guaranteed Income Data. *International Journal of Qualitative Methods* 22:16094069231201504, Oct. 2023.
doi:10.1177/16094069231201504
https://journals.sagepub.com/doi/10.1177/16094069231201504

[KPD25] U. Krähnke, T. Pehl, T. Dresing. *Hybride Interpretation textbasierter Daten mit dialogisch integrierten LLMs: Zur Nutzung generativer KI in der qualitativen Forschung*. 2025.
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-99389-7

[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

[Kro05] F. Krotz. *Neue Theorien entwickeln: eine Einführung in die Grounded Theory, die heuristische Sozialforschung und die Ethnographie anhand von Beispielen aus der Kommunikationsforschung*. Kommunikationswissenschaft. Herbert von Halem Verlag, Köln, 2005.

[KW25a] J. Küster, K. Wolf. State of CAQDAS 2025. Oct. 2025.
doi:10.5281/zenodo.17304035
https://doi.org/10.5281/zenodo.17304035

[KW25b] J. Küster, K. D. Wolf. State of CAQDAS 2025. 2025.
doi:10.7910/DVN/LKV8OT
https://doi.org/10.7910/DVN/LKV8OT

[LGK+20] A.-L. Lamprecht, L. Garcia, M. Kuzak, C. Martinez, R. Arcila, E. Martin Del Pico, V. Dominguez Del Angel, S. Van De Sandt, J. Ison, P. A. Martinez, P. McQuilton, A. Valencia, J. Harrow, F. Psomopoulos, J. L. Gelpi, N. Chue Hong, C. Goble, S. Capella-Gutierrez. Towards FAIR principles for research software. *Data Science* 3(1):37–59, June 2020.
doi:10.3233/DS-190026
https://journals.sagepub.com/doi/10.3233/DS-190026

[LPLL24] D. Lee, E. Park, H. Lee, H. Lim. Ask, Assess, and Refine: Rectifying Factual Consistency and Hallucination in LLMs with Metric-Guided Feedback Learning. In Graham and Purver (eds.), *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*. Pp. 2422–2433. Association for Computational Linguistics, St. Julian's, Malta, Mar. 2024.
https://aclanthology.org/2024.eacl-long.149/

[NYN+02] K. Nakakoji, Y. Yamamoto, Y. Nishinaka, K. Kishida, Y. Ye. Evolution patterns of open-source software systems and communities. In *Proceedings of the International Workshop on Principles of Software Evolution*. IWPSE '02, p. 76–85. Association for Computing Machinery, New York, NY, USA, 2002.
doi:10.1145/512035.512055
https://doi.org/10.1145/512035.512055

[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/S1479-838720200000012020/full/html

[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

[REF19] REFI-QDA - Exchange of processed data between qualitative data analysis software packages. Standard, The Rotterdam Exchange Format Initiative, Rotterdam, NL, Sept. 2019.
http://www.qdasoftware.org/

[RH18] K. A. R. Richards, M. A. Hemphill. A Practical Guide to Collaborative Qualitative Data Analysis. *Journal of Teaching in Physical Education* 37(2):225–231, Apr. 2018.
doi:10.1123/jtpe.2017-0084
https://journals.humankinetics.com/doi/10.1123/jtpe.2017-0084

[SHR+24] T. Susnjak, P. Hwang, N. H. Reyes, A. L. C. Barczak, T. R. McIntosh, S. Ranathunga. Automating Research Synthesis with Domain-Specific Large Language Model Fine-Tuning. Apr. 2024. arXiv:2404.08680.
doi:10.48550/arXiv.2404.08680
http://arxiv.org/abs/2404.08680

[TBX+24] R. H. Tai, L. R. Bentley, X. Xia, J. M. Sitt, S. C. Fankhauser, A. M. Chicas-Mosier, B. G. Monteith. An Examination of the Use of Large Language Models to Aid Analysis of Textual Data. *International Journal of Qualitative Methods* 23:16094069241231168, Jan. 2024.
doi:10.1177/16094069241231168
https://journals.sagepub.com/doi/10.1177/16094069241231168

[Tra25] V. B. Tran. How does QualCoder compare with NVivo, ATLAS.ti, MAXQDA? May 2025.
doi:10.5281/zenodo.15363745
https://doi.org/10.5281/zenodo.15363745

[Wan22] B. Wang. Programming for Qualitative Data Analysis: Towards a YAML Workflow. *ACIS 2022 Proceedings*, Dec. 2022.
https://aisel.aisnet.org/acis2022/17

[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

[Wol18] U. Wolski. The History of the Development and Propagation of QDA Software. *The Qualitative Report*, Mar. 2018.
doi:10.46743/2160-3715/2018.2984
https://nsuworks.nova.edu/tqr/vol23/iss13/2/

[WS17] C. WOOLF, N. H. SILVER. *Qualitative analysis using atlas. ti, nvivo and maxqda.* Routledge., 2017. OCLC: 987779390.

[XYL+23] Z. Xiao, X. Yuan, Q. V. Liao, R. Abdelghani, P.-Y. Oudeyer. Supporting Qualitative Analysis with Large Language Models: Combining Codebook with GPT-3 for Deductive Coding. In *28th International Conference on Intelligent User Interfaces*. Pp. 75–78. ACM, Sydney NSW Australia, Mar. 2023.
doi:10.1145/3581754.3584136
https://dl.acm.org/doi/10.1145/3581754.3584136