Sharing your coding schemas: Developing a Platform to fit within the Qualitative Research Workflow

Julian Hocker
DIPF | Leibniz Institute for Research and Information in Education
Julian.Hocker@dipf.de

Taryn Bipat
Human Centered Design Engineering, University of Washington
tbipat@uw.edu

Mark Zachry
Human Centered Design Engineering, University of Washington
zachry@uw.edu

David W. McDonald
Human Centered Design Engineering, University of Washington
dwmc@uw.edu

ABSTRACT

Qualitative coding schemas are an essential part of qualitative research used in methods like Grounded Theory. To date, there is no platform to share these coding schemas. Sharing and exchanging these coding schemas has a great potential when it comes to the traceability of qualitative research and well as the re-use of coding schemas. Based on an interview study with qualitative researchers, we propose concepts for integrating a new platform for sharing qualitative coding schemas. Based on theoretical work by Birnholtz and Bietz (2002), it became clear that an easy-to-use system can foster the acceptance and the willingness of researchers to share their coding schemas. We identified three major points to focus for this on: the governance of the platform, the development of the ontology itself and integrating the sharing of qualitative coding schemas into the workflow of researchers by enabling direct upload from the qualitative coding software.

CCS CONCEPTS

• **Human-centered computing → Open source software**.

KEYWORDS

open science, qualitative research, interviews, data sharing, data reusing, ontology, coding schemas

ACM Reference Format:
Julian Hocker, Taryn Bipat, Mark Zachry, and David W. McDonald. 2020. Sharing your coding schemas: Developing a Platform to fit within the Qualitative Research Workflow. In *16th International Symposium on Open Collaboration (OpenSym 2020), August 25–27, 2020, Virtual conference, Spain.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3412569.3412574

OpenSym 2020, August 25–27, 2020, Virtual conference, Spain

ACM ISBN 978-1-4503-8779-8/20/08…$15.00
https://doi.org/10.1145/3412569.3412574

1 INTRODUCTION

The basic ideas behind Open Science are probably centuries old and are present in the earliest letters and establishment of journals for sharing research results. The expansion of science as a productive discipline, and concerns about junk science, have renewed the focus on the concept of Open Science. Since, roughly 2000, there has been a growing discussion of what constitutes Open Science [12]. Some clear components of Open Science include the sharing of scientific data, scientific methods, as well as research results.

The accessibility of data facilitates sharing data across multiple fields of knowledge, people, and institutions as well as the validation of results. According to the Open Science Collaboration (OSC), prior research "should not gain credence because of the status of authority of their originator but by the replicability of their supportive evidence" [7]. With the ongoing replication crisis it is important to be able to have access to prior data and clear methodologies for collecting and analyzing such data. Furthermore, sharing data allows researchers to build off the assumptions and efforts of past research. Access to open data enables the broadening of research scopes and the ability to diversify perspectives in science [10].

However, open data can be a double-edged sword. While some researchers can benefit from the sharing of data, others who might have more privilege might not see as much value. Some scientists believe that sharing data can lead to issues around commercialisation or have undue impact on their own reputation [18]. While making data open and available is important, these differing perspective make it extremely difficult [4]. Past researchers have explored and classified the problems related to open science across knowledge fields [10, 18].

In this study, we leverage a framework developed by Birnholtz and Bietz that classifies the type of difficulties researchers face when sharing their data [4]. To date, the majority of focus on data and methods sharing in open science has focused on quantitative type methods. In quantitative methods, the focus is on the replicability of research, whereas in qualitative research, traceability is the more important feature when it comes to open science. This creates special requirements in open science for qualitative research. In prior work an ontology to help describe qualitative research coding schemas was developed [14]. The ontology was embodied in a prototype semantic wiki system and populated with a small set of qualitative coding schemes that had been published in

peer-reviewed papers. Through a structured, task-based interview protocol, we had 20 participants interact with the prototype system. We sought to understand how this ontology might be used by qualitative researchers and how it may fit into their workflow. We analyzed the participants responses to understand how a qualitative coding schema ontology system can be used in qualitative coding processes. The findings from our analysis is then contextualized through the Birnholtz and Bietz framework.

Broadly, we believe the ontology for sharing qualitative coding schemas has the potential to make qualitative research more transparent as well as encouraging researchers to share more of their work, making their scientific process more visible. The ontology can make the reuse of qualitative schemas easier by providing a richer context for understanding the schema, how it can be applied, and for which types of data it is effective. However, as our analysis illustrates, for these types of systems to facilitate open science, the costs of participation need to be integrated into the workflow of researchers at a minimal cost to encourage sharing.

In the following we first outline prior work related to the way qualitative research schemas are shared. We then review the Birnholtz and Bietz [4] framework and state our research question. We explain our task driven data collection methods and present our findings in the context of the Birnholtz and Bietz framework. Finally, we close the paper by discussing the results and offer recommendations for what a future platform for sharing coding schemas might look like.

2 PRIOR WORK

Understanding coding schemas through prior research literature is challenging. In some cases the creation or development is the objective of the research. For example, for many who follow a a Grounded Theory approach, the coding schemas reflect an important part of the theory they will ultimately attempt to describe [20]. In other qualitative research, the coding schema is a means of getting to the real result. However, one commonality across most qualitative work is that there are few systematic ways of describing and cataloging the schema and how it was applied. This results in a paucity of literature detailing the systematic study of qualitative coding schemas.

Earlier research showed that documentation of data is an important factor in the reuse of that data. Kowalczyk, S., & Shankar, K. [16] define all this as context:"[...] context documents how datasets fit into their physical and technical environments (file formats and field descriptors) as well as into the scientific environment (experiment treatments and applications)." Faniel et al. also proved this when evaluating that metadata quality is important for the satisfaction of social scientists with data portals [9]. Therefore we infer that for the sharing of coding schemas the quality of the provided metadata is crucial. Based on this approach, the definition of metadata or ontologies as "a formal, explicit specification of a shared conceptualization" [21] and see this metadata, which is to describe the coding schemas as important information the field agreed on.

Research shows that when sharing data, researchers should describe the data since information experts often do not have adequate insight into the prior research. Therefore it is important to facilitate the sharing of coding schemas via easy-to-use platforms [4, 19]. Based on this, how data is input into such a system is important. It is necessary to encourage researchers to share their coding schemas, but there also needs to be governance of the platform. We also want to talk in this paper about how to ensure the government of a new system for the sharing of qualitative coding schemas.

An example of such a system is Dendro, which allows researchers to place data into a repository and also create ontologies in order to describe it in a meaningful way. This system also shows that the creation of ontologies is often a participatory process involving researchers and experts in the field [6].

In the sciences there is a huge amount of so-called dark data, which was created in small research projects often without data management plans and that is shared either on a personal level or not shared at all. Often this data is not well-documented [13].

Qualitative research often suffers this problem as well. Many projects can be considered small science [5] in which it is hard to find coding schemas in a meaningful way. Small science also means there are no large projects and great standardization, rather small projects, often done by single researchers and no large standards of platforms. Leaving this data as "dark data" makes it useless with no possibility to reuse it. Publishing the data gives other researchers the opportunity to use it and the researcher who created the data the possibility to be cited for their work.

One way to facilitate the sharing of data is the creation of platforms where researchers can expose how the data was created, which is also already done. For the sharing of qualitative coding schemas, there does not exist such a platform. There need to be more ways of attribution towards the sharing of coding schemas in order to motivate people sharing their data. Another way to enforce this type of sharing is for journals to enforce the sharing of coding schemas whenever an article is published [16]. Publishing data and coding schemas together with corresponding articles also gives the opportunity to easily find data, which helps in the sense of scientific knowledge graphs [3].

Prior research on barriers for the sharing of data claimed for the field of psychology that sharing of research data can be enforced via journals, funding agencies and institutions [15].

In recent years, there were several developments regarding the quality of the research data using the FAIR principles [23] as well as for the development of platforms with the CARE principles [2]. The FAIR principles mean findability, accessibility, interoperability and reuseability and formulate criteria for research data when these are shared within a platform. The CARE principles stand for collective benefit, authority to control, responsibility and ethics. These principles can be seen as blueprint for the creation of research data centers.

Bringing this together also shows that ontologies take a very important role in this process since they make the reuse of data in the sciences possible. Without well-described coding schemas, the re-use of coding schemas is not possible. But for the description of the coding schemas it is necessary that the researchers are incentivized to take the time to describe the data.

The group REFI[1] took one step towards sharing of qualitative coding schemes while creating the standard REFI-QDA Codebook[8], an exchange format with which it is possible to transfer a coding

[1] https://www.qdasoftware.org/

schema from one software to the next. The readability of file formats is a great problem in sharing data (data sharing in sciences), however this exchange format gives a start in order to archive and share data, still different software does use different notions of coding schemas, therefore not all information is shared within the format.

Summing up, there is currently no platform that allows the sharing of qualitative coding schemas. If such a system is to be developed, it has to fit into the workflow of the users in order to make it useful and to lower issues for uploading coding schemas to the system.

2.1 Ontology Description

We leverage one specific ontology designed to facilitate the sharing of qualitative coding schemas. These types of schemas are often used during systematic coding of qualitative data into descriptive categorical buckets. The ontology we use was developed through participatory design methods including interviews, observations and feedback rounds for prototypes [14].

The ontology consists of five main categories: publications, research data, study descriptions, coding schemas and codes. Each of these categories contains further information about these types of information. The following graphics shows the structure:

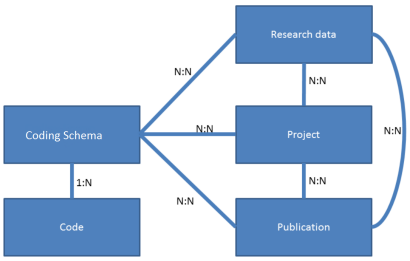

Figure 1: Structure of the ontology

The goal was to create an ontology that can be used in a standalone platform while integrating other types of information like publications and research data in the sense of a research graph, while also providing basic information about these links. The ontology therefore includes contextual information about project as well as research data. Naturally, not all projects will contain all potential artifacts, but the ontology was designed to represent key artifacts related to a schema. A prototype of the ontology can be found on Github[2].

A prototype implementation of the ontology was built using Semantic MediaWiki[3]. The prototype was populated with several coding schemas that our research participants were able to browse. We implemented the coding schemas as well as the connections to research data, projects and papers. The research participants were able to navigate freely in the system and choose the information they needed for the specific study tasks. The MediaWiki navigation was as follows: on the main page there were links to overview pages for publication, study, research data and coding schema. If the participant clicked on one of these main categories, they found a page with all content within these main categories, e.g. all coding schemas in the system. From this, participants could navigate to a specific item, e.g. coding schema. On these item pages, participants found all metadata as well as links to codes, publication, study and research data, which belong to this coding schema. This navigation largely reflects the links present in the ontology diagram of Figure 1.

[2] https://github.com/julianhocker/Quali-Codes-Ontology
[3] https://www.semantic-mediawiki.org

3 RESEARCH QUESTION

Our research question is.... How does a previously developed qualitative data ontology system fit within the existing work flow of qualitative researchers?

To better understand this question, we use a framework developed by Birnholtz and Bietz [4] that classified the difficulties centered around sharing data: (1) willingness to share, (2) locating shared data (3) using shared data. These authors noted that there is competition for reputation by scientists which in turn means that scientists may want exclusive rights to their own data. This exclusivity reduces their desire to share their methods or data. Additionally, scientific disciplines vary in the existence of shared data repositories, which hinders the ability find and use shared data. Further, without the context of a dataset's creation it can be difficult for researchers to recreate or appropriately use a shared dataset.

4 METHODOLOGY

Our goal was to understand potential user's experiences with this type of open data system. We wanted to understand how potential users might see this type of system as part of their qualitative research process. We structured our evaluation process around a pair of qualitative coding tasks that would require that participants in our study explore and use the various portions of the system.

We recruited participants for our study through flyers, targeted mailing lists, and personal contacts at the University of Washington in Seattle. All of our participants had prior experiences as a researcher or research assistant associated with a research project where they had performed a systematic qualitative coding process as part of their data analysis. While we recruited broadly at our university, all participants had a background in human-centered design or information science. Their qualitative experiences ranged from beginner to expert with respectively 1-5 years of experience. Most of our participants were undergraduate and graduate students working on qualitative research projects. We additionally had some university researchers who study student behavior and classroom feedback using qualitative methodologies. In total we had 20 participants complete our study.

4.1 Task-driven Exploration

Participants began with a 3-part introduction to the prototype with (i) a video tutorial, (ii) a system comprehension quiz and, (iii) two qualitative coding tasks that required them to use the system. In the first part, an 8 minute long video highlighted each section of the ontology, where one author described each section in detail and showed the participants how to navigate through the prototype. The goal of the overview video was to make sure every study

participant was systematically given the same introduction to the prototype. A link to our introductory video can be found in the appendix.

In the second part, the participant was given a prototype comprehension quiz to test their understanding of the system. Participants were allowed to freely explore the prototype before taking the quiz and could use it to answer questions on the quiz. After the participant completed the quiz, the researcher running the study went over the answers and clarified any answers that were wrong. The goal of the comprehension question was to make sure that the study participant understood and comprehended what they had been told during the video. The comprehension quiz served as a type of attention check on the information provided. The comprehension quiz can be found in the appendix.

In the final part, participants were given two different qualitative coding tasks. For each task, the participant was given 15 minutes to use the prototype to complete the task. The participant needed to identify and review one coding schema in the system and apply the schema codes to four sample texts. The participant was given a paper based coding sheet with the four text samples and had to circle or write down the code(s) that best fit the sample text, as the participant could determine based on the schema. Figure 2, shows one sample from one coding task. The coding schemas for both coding tasks were based on schemas from prior published work. The schemas put into the prototype were trimmed and simplified to facilitate a timely completion of the coding task. We were not trying to test the participants accuracy of qualitative coding. Instead, the goal was to motivate the participant to understand how the prototype represented prior methods and data so that the participant could more effectively reflect on how they might include systems like the prototype in their research process.

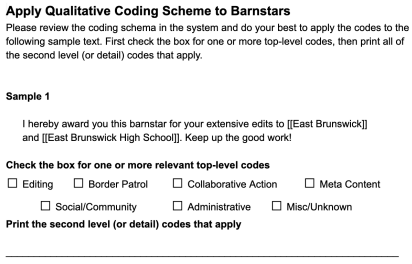

Figure 2: Example of a coding task. Participants were given instructions and two samples before beginning the task. They had the ability to freely use the system to complete all tasks.

4.2 Semi-structured Interviews

After the task-driven exploration, we conducted semi-structured interviews to obtain qualitative data about our participant's experiences with the prototype. Each interview lasted approximately 10-20 minutes. The semi-structured interview was structured around the following questions:

(1) Can you briefly describe your strategy for applying the codes?
(2) What was different about your strategy between the first coding scheme and the second?
(3) What questions did you have about the coding schemes that you could not answer with the system?
(4) What aspect of the system was most helpful to your work to apply the codes?
(5) What meta-data do you feel is missing from the system that would be helpful for these tasks?
(6) Is there anything else you would like to share about your experience with the system?
(7) Do you have any questions for us about this system or this study?

Audio recordings of the interviews were transcribed. We then conducted a thematic analysis of the transcripts. The research team met weekly to extract and refine the themes in the transcripts. The main findings from each participant were listed and then reviewed by all authors of the paper. The findings reflect common themes. However, in the process of identifying themes, we also realized that we were seeing some alignment with the Birnholtz and Bietz framework, as well as things that did not seem to fit the framework. In the following section we present our findings.

5 FINDINGS

We structure our presentation of the findings using the three broad categories from the Birnholtz and Bietz framework: (1) willingness to share, (2) locating shared data (3) using shared data. As we described above, all of our participants had prior experience with qualitative research. In particular, all met the condition that they had participated in a systematic qualitative coding where they had applied a schema. Many of our participants had also participated in the development of one or more coding schemas. Our analysis, consequently, reflects how experienced qualitative researchers view a system of data and schema sharing using the ontology.

5.1 Willingness to Share

Birnholtz and Bietz first noted that willingness to share is a problem for researchers. Scientists often have the potential to profit from their own data, so sharing might have implications for their reputation or other benefits such as grants and publications. Additionally, data sometimes has commercialization interests, which can lead to proprietary restrictions. In our analysis, participants did not mention issues such as this, probably because of the type of researchers we interviewed. All researcher informants were connected to the university system and publishing to a wide audience was key to their research having a broad impact. Sharing, however, was difficult to fit into their workflow. Sharing requires additional time and effort that is not always available for the type of researchers we interviewed.

5.1.1 The role of the system for use in qualitative methods. Participants compared their experience with our system to their own, prior qualitative coding experiences and hypothesized how our system would fit into that process. Most of the participants reported that they conduct coding tasks either on paper or using spreadsheets. A minority used commercial qualitative coding software such as Dedoose or Atlas.TI. In our tests, people were given the codes on a computer screen while they had to do the coding on

paper. Interviewees mentioned that it would make sense for them to have a better integration of the coding schemas into their working environment:

> *"I feel like if you can somehow include the things that I am doing on this paper, that would make it easier to do."* (P7)

This participant wanted the ability to apply codes to a given text in the system. They would, therefore, have both the ability to learn about the coding schema and then use the codes directly in the system to code and analyze their data.

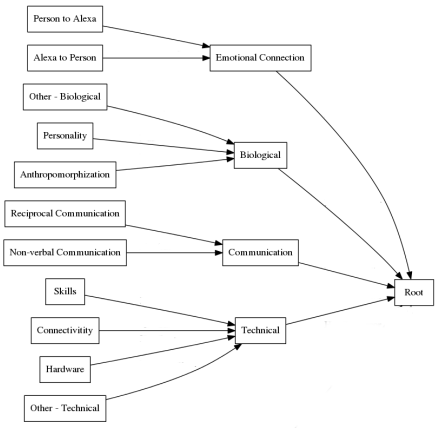

Figure 3: Coding schema presented as a graph visualization. The visualization allowed users to see the connections between codes and multi-level coding schemas.

Some interviewees also mentioned their own prior experiences with conducting coding in teams. Most of them described this as quite tough since they were working with multiple people who have different interpretations of the coding scheme. In their experience, coding schemas are not always described in a proper and consistent way. Participants mentioned using digital and "low tech" (P2) spreadsheets to organize their coding schemas, which did not allow for the entire schema to be visualized and required a lot of scrolling and maneuvering around the spreadsheet. In our ontology, we presented the coding schemas in a visualization as seen in Figure 3. Participants felt that this visualization was helpful because it allowed for less navigation than a spreadsheet and the entire coding schema could clearly be visualized on a single page:

> *"the visualization works better than a spreadsheet because a spreadsheet needs a lot of clicking and scrolling to navigate and find information"*(P15)

However, the prototype system we developed only allowed for a concrete representation of the final codes. Participants noted that creating a coding schema is a dynamic process and there can be shifts in the codes. Currently, our ontology does not accommodate that need:

> *"Once you have a project that is mature, you have a stable coding schema but it takes a great deal of work to get up to that point and what we really have not seen or tested is the extent which it can support a living document that has this kind of history of codes that get subdivided or merged or codes that definitions had.*(P10)

Being able to integrate more of the coding process into the prototype would be also helpful for qualitative researchers to share their codes across a team.

5.1.2 Time is an obstacle. Participants noted that while the prototype was useful to them, they could not see the system being integrated into their normal study schedule because it required additional work to input all of the metadata related to their coding schema. This problem is clearly connected with the need for strong documentation of coding schemas, but researchers often lack time to do the documentation. The lack of time leads qualitative researchers to be less willing to share. One participant noted that in a project he had created a memo system through email for his team. He would send out weekly email updates about the coding schema to the entire team; however, this method was not enacted by his team because it required too much overhead work. This participant suggested that adding another tool to the process such as a Wiki tool would be too difficult to integrate and required a high learning curve:

> *"I wonder if you raise the bar even further. If it is not just an email but a wiki tool. Even as simple as that is, to what extent you would have to lower the bar to realistically have that included in the scope of a project, time and effort. It strikes me that you have to be really aggressive. Everyone would say that this is really useful and these are things that people generally have. Everyone has a codebook but it is just the matter of doing that in your format. A change in format can be enough to discourage someone from using something."*(P10)

The addition of this type of system to an already existing workflow could lead to additional constraints for researchers. Participants noted that the addition of just a single new tool to their process would disrupt their information management practice:

> *"Here are a lot of fields that I manage in different places that I store like Zotero, archives.org. I have a workflow of information management. I do not know how useful it would be to fill out another form unless it attaches to these other services."* (P2)

5.2 Locating the Data

Researchers must become aware of who has the data they need and where the data are located [4]. Sharing data is a social exchange, which requires trust to gain access to the data [22].

We found that it would be easier for participants to locate data if such systems were already integrated into the tools that were part of their coding process.

5.2.1 Integrating the system into the workflow. Some participants discussed their use of software to do qualitative coding, but most of them mentioned using spreadsheets or coding on paper. After using our prototype ontology system, they liked the way in which the codes were presented, comparing that experience with it to their usage of other software:

> *"We used Dedoose for previous studies. That was not that great for looking up the codes and what it means"* (P7)

According to this participant, in Dedoose it was not easy to find a specific code with its definition. Participants noted that in our prototype, it was easier to explore codes, which were all presented on one screen:

> *"As well as that the presentation on the platform works better than a spreadsheet since it does not involve that much clicking."* (P15)

5.3 Using Shared Data

In our interviews, we found that presentation and documentation was key for users to locate and use the data. Participants noted that working with any given dataset requires additional contextual knowledge about the research to be able to understand how to use the data. We also found that integrating such system into the workflow of qualitative research was difficult because it added an additional learning curve for users.

5.3.1 The role of the system in open science in general. One part in the creation process of the prototype was to think about how to present coding schemas. In open science, documentation and presentation facilitates information retrieval [4]. In our ontology system, the coding schemas were presented in two ways: Using a table with a list of the codes and a visualization to show the relationship between the codes.

Overall, participants like the presentation of the codes. Participants mentioned that a coding schema was easier to apply with the use of the table and the visualization in the system. One participant noted:

> *"The visual representation of the codes was helpful. It was easier because of the structure, you can see the main code and then the codes under that category especially in the Alexa example."* (P6).

Some of them also mentioned that in their prior experience they used to work with coding schemes that were not very well documented, which lead to issues (P7, P10 and P16).

If coding schemes need to be reused, good documentation is essential to make them understandable for future researchers. Good documentation can also help make coding within a team easier. Documentation is important in both open science and qualitative methods for the traceability of research.

In open qualitative research good documentation helps support traceability or even replicability [1]. A platform that makes possible the collection and reuse of coding schemas together with metadata has the potential to make it easier for studies to become more traceable or even reproducible. It is important what metadata is collected because if researchers are able to fill out these metadata, they provide the information needed for traceability and are able to meet this quality criteria. Figure 3 and 4 show the coding schemas in the system.

5.3.2 Shared data allows for different perspectives of the research. The results of the interview showed that shared data can be used in different ways. One participant noted that their use of an ontology would allow them to gain a different perspective than just reading the original paper the coding schema was presented in:

> *"a different way of portraying a study. It actually is if you can see the coding schema and the examples of that codes. That would be a way to see how they made sense of the content rather than - because usually when you read papers, people will skip over the methods and just go to the research methods and finding. And I think actually looking at the codes is useful because - the conclusions they drew in the discussion might not be the only conclusion and the only significance in the coding that they did. So I think this provides a different way into the study than the paper itself."*(P18)

This participant noted that understanding the methodologies and the metadata was different through the use of this system.

6 DISCUSSION

Next we reflect on our findings and present design recommendations for the integration of new qualitative ontologies in open science. We saw the potential for such platforms to give researchers new perspectives of data, better documentation for re-use, as well as support for team coding and better traceability for qualitative research.

The main obstacle for people to share their data was time. It takes researchers extra time to document and upload their coding schemas into a new system. And many will not see much benefit from spending that extra time. In order to reduce this extra time, it is important to fit the system into the current workflows of the researchers. Due to the interpretative nature of qualitative research, the workflows can be quite diverse: Researchers use specialized software, or they use spreadsheets or even code their data on paper. Given the diversity of practice it is unlikely that any system could completely eliminate the extra time required to open up the research practice to sharing the wide variety of research artifacts. Therefore, we acknowledge that it is an open question as to how to create appropriate incentives to motivate this sharing while reducing the associated overhead.

We see three different levels that provide openings for resolving this sharing challenge: the governance of the system, the development of the ontologies and the integration of the actual sharing into the research process.

If the system is online, ultimately the researchers themselves provide the information about their coding schemas. We argue that it is more important to enable researchers to document their data in the way they see fit. The governance structures of the system will likely need to be settled by the researchers who participate. The system should facilitate and encourage representations of the research artifacts to help with re-use. We believe this means the ontology is easy to understand and researchers can integrate providing this information into their research workflows. Ultimately, in order for qualitative practice to participate in open science, researchers have to provide the information about their coding schemas, regardless of how the system in governed or which kind of system is used.

There are currently two governing models right now. One model followed by many research data centers is quite strict for people uploading data and requires a high quality level. While other services

	Definition	Example	Subcode of
Anthropomorphization	Are users attributing human characteristics to Alexa, especially in the way they address Alexa. Do they call the device her/she versus it/device/thing. (only what they are calling Alexa) Always have a code. Every line of code should have an antrophomorphization code.	Y: Alexa is overrated and she doesn't know anything you have to constantly repeat yourself. N: Gave this as a gift hopefully they are enjoying it. Y/N: I love playing with this device, she is so funny.	Biological
Personality	Are they describing the personality of Alexa? Is the device funny, sassy, etc..	I love speaking to Alexa, she is so funny. Sometimes she makes some snark remarks that I do not expect.	Biological
Other - Biological	Gives Alexa any other human traits that does not to seem to fit into any other human traits category		Biological
Reciprocal Communication	Any other issues or situations where the owners have conversations with Alexa.	We do have words; from time to time when I ask a simple question and get "I don't have an answer to that" or simple verbiage. After perhaps the third time of rephrasing the question, i may get the right answer.	Communication
Non-verbal Communication	The light at the top of the device, shutting down instead of giving a response.	Some times I know Alexa is responding to me or thinking about how to respond because the light at the top will be flashing.	Communication
Missing Alexa's company	The person using Alexa feels the absent of the device when they are not using it	When I am away from my home, I forget that Alexa is not there and I still try to talk to her	Companionship
Alexa as Family Member	Considers Alexa as part of the family	Had the Echo and Alexa in our lives almost a year now and she's like a member if the family.	Companionship
Alexa as Friend	Considers Alexa a friend	Alexa is my new bestie.	Companionship
Alexa to Person	Keywords: Love, Hate, etc… (human feelings that person feels toward alexa	I love Alexa. I do struggle with her understanding me sometimes.	Emotional Connection
Person to Alexa	Alexa is having these feelings or at least the human thinks alexa feels this way	Sometimes I think Alexa hates me.	Emotional Connection

Figure 4: Coding schema as presented in a table format. Participants could see the definition and example for each code.

like Zenodo, have few quality checks and make easy uploading of data possible. A third option might be an open portal where everyone can upload data, but for a paper publication it is mandatory to upload the coding schemas there and the quality is also checked by the journal. As claimed in prior research, forcing researchers to share their research data on publication is a strong incentive [15]. Another, fourth option, would be to implement the badges system by open science framework [11], which adds badges to a publication if material or data are open or if the study has been pre-registered.

The prototype of the system right now is built up as a single system, providing also basic information about research data and publications. In a future system, a system might also only provide information about the coding schemas or integrate the information about the coding schemas into a larger system.

Earlier work mentions as one obstacle the fear of scientists that their data might be misused [17]. This is a pervasive concern with regard to open science. We acknowledge this threat. However, we feel it is important to note that our interviewees did not mention this concern. We think that there is always the possibility to misinterpret coding schemas, but this risk can be decreased by having quality checks within the system so researchers provide high quality information. For example, high quality data that is rarely shared might include training materials used to train qualitative coders, and additional positive and negative examples of coding applications that were not included in a peer-reviewed publication. These kinds of additional research materials make it easier to validate results and make it easier to re-use a coding schema in a more accurate and trustworthy manner.

Ontologies should be developed and evaluated keeping in mind that information has to be provided as well as retrieved. Therefore the input of data should be kept as important as the retrieval. This also interconnects with the claim to provide easy to use interfaces for researchers to share data [4]. Following this, it is important to develop ontologies that consider the workflow of users. One concern was the amount of work researchers would have to do. Allowing for an easy integration of a system into the workflow of researchers with a smaller learning curve can foster the sharing of coding schemas. This also means that sharing should be part of the research cycle and planned from the beginning of the research. This can also make documentation easier if it is done from the start of the research.

The third point is making the actual sharing easier. This can done via an easy integration of the sharing of coding schemas into the QDA software. We see the development of export formats from the proprietary software. In a future version, it would make sense if these software would also support a direct upload of coding schema from the software, making it easy to fill in metadata. When researchers use spreadsheets for coding, an import of CSV-files might also make sense.

Some interviewees mentioned it would make sense to have branches and the functionality to upload certain snapshots of a coding schema, similar to Git. This also brings up interesting other ideas because it should be possible to see which coding schemas are reused by whom and how they changed. A simple solution is to create a type of reuse link in the ontology for coding schemas. Another possibility might be to have branches or hard forks in Git, making it possible to track how researchers reused a coding schema within other research.

7 CONCLUSION

Summing up, we can see that the researchers saw potential in the sharing of qualitative coding schemas and there is a potential for such a platform to make qualitative research more transparent and open in the sense of open science.

We used the framework found in Birnholtz and Bietz [4] to understand how it applies to sharing qualitative research. While the framework was developed based on studies of quantitative research we found some alignments with data from our study participants. We see three major obstacles that have to be tackled: the governance of the system should be organized making it easy to upload as well as easy to document the coding schemas. Also, the ontology for such a system should be developed considering both that information has to be put into the system by researchers as well as retrieved. The third method is to enable uploading from coding schemas and the metadata directly from qda software.

ACKNOWLEDGMENTS

We would like to thank Christa Womser-Hacker, Thomas Mandl, Marc Rittberger and Christoph Schindler for comments throughout the whole project. We would also like to thank our participants for taking part in the study.

REFERENCES

[1] Herman Aguinis and Angelo M Solarino. 2019. Transparency and replicability in qualitative research: The case of interviews with elite informants. *Strategic Management Journal* 40, 8 (2019), 1291–1315.

[2] Global Indigenous Data Alliance. 2019. CARE principles for Indigenous data governance. *GIDA https://www. gida-global. org/care* (2019). https://www.gida-global.org/care

[3] Sören Auer, Viktor Kovtun, Manuel Prinz, Anna Kasprzik, Markus Stocker, and Maria Esther Vidal. 2018. Towards a knowledge graph for science. In *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*. 1–6.

[4] Jeremy P Birnholtz and Matthew J Bietz. 2003. Data at work: supporting sharing in science and engineering. In *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*. 339–348.

[5] Christine L Borgman. 2015. *Big data, little data, no data: Scholarship in the networked world*. MIT press.

[6] João Aguiar Castro, Ricardo Carvalho Amorim, Rúbia Gattelli, Yulia Karimova, João Rocha da Silva, and Cristina Ribeiro. 2017. Involving data creators in an ontology-based design process for metadata models. In *Developing Metadata Application Profiles*. IGI Global, 181–214.

[7] Open Science Collaboration et al. 2015. Estimating the reproducibility of psychological science. *Science* 349, 6251 (2015), aac4716.

[8] Jeanine Evers, Mauro Ugo Caprioli, Stefan Nöst, and Gregor Wiedemann. 2020. What is the REFI-QDA Standard: Experimenting With the Transfer of Analyzed Research Projects Between QDA Software. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, Vol. 21.

[9] Ixchel M Faniel, Adam Kriesberg, and Elizabeth Yakel. 2016. Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology* 67, 6 (2016), 1404–1416.

[10] Beth A Fischer and Michael J Zigmond. 2010. The essential nature of sharing in science. *Science and engineering ethics* 16, 4 (2010), 783–799.

[11] Open Science Framework. [n.d.]. Badges to Acknowledge Open Practices. BadgestoAcknowledgeOpenPractices

[12] Google. 2020. Google Books Ngram Viewer. https://books.google.com/ngrams/graph?content=open+science&case_insensitive=on&year_start=1800&year_end=2008&corpus=15&smoothing=3&share=&direct_url=t4%3B%2Copen%20science%3B%2Cc0%3B%2Cs0%3B%3Bopen%20science%3B%2Cc0%3B%3BOpen%20Science%3B%2Cc0%3B%3Bopen%20Science%3B%2Cc0%3B%3BOpen%20science%3B%2Cc0#t4%3B%2Copen%20science%3B%2Cc0%3B%2Cs0%3B%3Bopen%20science%3B%2Cc0%3B%3BOpen%20Science%3B%2Cc0%3B%3Bopen%20Science%3B%2Cc0%3B%3BOpen%20science%3B%2Cc0

[13] P Bryan Heidorn. 2008. Shedding light on the dark data in the long tail of science. *Library trends* 57, 2 (2008), 280–299.

[14] Julian Hocker. 2020. Participatory design for ontologies. A case study of an open science ontology for qualitative coding schemas. *Aslib Journal of Information Management* (2020). in print.

[15] Bobby Lee Houtkoop, Chris Chambers, Malcolm Macleod, Dorothy VM Bishop, Thomas E Nichols, and Eric-Jan Wagenmakers. 2018. Data sharing in psychology: A survey on barriers and preconditions. *Advances in methods and practices in psychological science* 1, 1 (2018), 70–85.

[16] Stacy Kowalczyk and Kalpana Shankar. 2011. Data sharing in the sciences. *Annual review of information science and technology* 45, 1 (2011), 247–294.

[17] Harlan M Krumholz, Joseph S Ross, Cary P Gross, Ezekiel J Emanuel, Beth Hodshon, Jessica D Ritchie, Jeffrey B Low, and Richard Lehman. 2013. A Historic Moment for Open Science: The Yale University Open Data Access Project and Medtronic. *Annals of Internal Medicine* 158, 12 (2013), 910–911. https://doi.org/10.7326/0003-4819-158-12-201306180-00009 arXiv:https://www.acpjournals.org/doi/pdf/10.7326/0003-4819-158-12-201306180-00009 PMID: 23778908.

[18] Karen Seashore Louis, Lisa M Jones, and Eric G Campbell. 2002. Macroscope: sharing in science. *American Scientist* 90, 4 (2002), 304–307.

[19] Joana Rodrigues, João Aguiar Castro, João Rocha da Silva, and Cristina Ribeiro. 2019. Hands-On Data Publishing with Researchers: Five Experiments with Metadata in Multiple Domains. In *Italian Research Conference on Digital Libraries*. Springer, 274–288.

[20] Anselm Strauss and Juliet Corbin. 1994. Grounded theory methodology. *Handbook of qualitative research* 17 (1994), 273–85.

[21] Rudi Studer, V Richard Benjamins, and Dieter Fensel. 1998. Knowledge engineering: principles and methods. *Data & knowledge engineering* 25, 1-2 (1998), 161–197.

[22] Nancy A Van House, Mark H Butler, and Lisa R Schiff. 1998. Cooperative knowledge work and practices of trust: Sharing environmental planning data sets. In *Proceedings of the 1998 ACM conference on Computer supported cooperative work*. 335–343.

[23] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3, 1 (2016), 1–9.

A LINK TO VIDEO TUTORIAL

https://youtu.be/UuZ8HbtkZHI

B SYSTEM COMPREHENSION QUIZ

Please mark or circle the choice which you feel best answers the question based on the software orientation training you received today.

(1) What information can be found in the "Publication" page?
   (a) Information about the data collection methods.
   (b) Names of author(s) of at least one article related to this project.
   (c) The research questions.
   (d) Visualizations of the codes
(2) What information can be found in the "Research data" page?
   (a) Visualizations of the codes.
   (b) Information about sampling and creation of the data.
   (c) Names and affiliation information for authors of related article.
   (d) Participants who were involved in the study.
(3) What content can be found under the "Study" page?
   (a) Demographics or descriptions of the study participants.
   (b) Description of the data collection research team.
   (c) The coding schema exported from commercial qualitative analysis software.
   (d) Visualizations of the codes.
(4) What information can be found in the "Coding schema" page?
   (a) The research motivation.
   (b) The research questions.
   (c) Information about the data collection methods.

(d) Names and affiliation information for authors of related articles.
(5) What does the visualization in the middle of the "Coding schema" page show?
(a) The relationships between the most important people in the research field.
(b) The relationships between the authors of the related papers.
(c) The most frequent codes and how they relate to each other.
(d) The relation of this project in the broader research field.
(6) What information can be found in the "Code" page?
(a) Information about the data collection methods.
(b) A general description of the code, including examples.
(c) The author of the code.
(d) A list of negative code examples.

C EXAMPLE OF QUALITATIVE CODING EXERCISE

Figure 5

Ordering: ☐ First or ☐ Second Participant #: _Key_____

Apply Qualitative Coding Scheme to Barnstars

Please review the coding schema in the system and do your best to apply the codes to the following sample text. First check the box for one or more top-level codes, then print all of the second level (or detail) codes that apply.

Sample 1

I hereby award you this barnstar for your extensive edits to [[East Brunswick]] and [[East Brunswick High School]]. Keep up the good work!

Check the box for one or more relevant top-level codes

☑ Editing ☐ Border Patrol ☐ Collaborative Action ☐ Meta Content

☐ Social/Community ☐ Administrative ☐ Misc/Unknown

Print the second level (or detail) codes that apply

EMAJ

Sample 2

Another award for your mantlepiece, for your many vandalism reverts. Thanks for your hard work in making Wikipedia a better place.

Check the box for one or more relevant top-level codes

☐ Editing ☑ Border Patrol ☐ Collaborative Action ☐ Meta Content

☐ Social/Community ☐ Administrative ☐ Misc/Unknown

Print the second level (or detail) codes that apply

BVG

Sample 3

For outstanding copyediting and hard work improving Wikipedia articles, I award you this Barnstar.

Check the box for one or more relevant top-level codes

☑ Editing ☐ Border Patrol ☐ Collaborative Action ☐ Meta Content

☐ Social/Community ☐ Administrative ☐ Misc/Unknown

Print the second level (or detail) codes that apply

EMIN

Rev: 11.06.2019 page 1 of 3

Sample 4

For your hard work on making the Canadian discussion page more useful and effective, I award you a Working Man's Barnstar. Keep up the great work.

Check the box for one or more relevant top-level codes

☐ Editing ☐ Border Patrol ☑ Collaborative Action ☐ Meta Content

☐ Social/Community ☐ Administrative ☐ Misc/Unknown

Print the second level (or detail) codes that apply

CG

Sample 5

Thanks for helping with the team effort to improve the [[List of Andy Griffith Show episodes]].

Check the box for one or more relevant top-level codes

☑ Editing ☐ Border Patrol ☑ Collaborative Action ☐ Meta Content

☐ Social/Community ☐ Administrative ☐ Misc/Unknown

Print the second level (or detail) codes that apply

CG EG

Sample 6

I award you this Barnstar for reverting vandalism and tagging pages for cleanup, wikifying or deletion, like no one else on Earth. It's the first time I give anyone a barnstar and I'm absolutely sure I'll not regret giving it to such a hard worker as you.

Check the box for one or more relevant top-level codes

☑ Editing ☑ Border Patrol ☐ Collaborative Action ☐ Meta Content

☐ Social/Community ☐ Administrative ☐ Misc/Unknown

Print the second level (or detail) codes that apply

BDEL BVG ECLA

Rev: 11.06.2019 page 2 of 3

Sample 7

This Barstar for your tireless and endless work on improving and keeping the Wikiproject on Football and departments up and running.

Check the box for one or more relevant top-level codes

☐ Editing ☐ Border Patrol ☐ Collaborative Action ☐ Meta Content

☑ Social/Community ☐ Administrative ☐ Misc/Unknown

Print the second level (or detail) codes that apply

SL SC

Sample 8

I hereby award you this Barnstar as a reward for your tireless pursuit of [[Candidates for speedy deletion]], your endless welcoming of users, and all of your other great contributions, not to mention your glorious NPOV jpg!

Check the box for one or more relevant top-level codes

☑ Editing ☑ Border Patrol ☐ Collaborative Action ☐ Meta Content

☑ Social/Community ☑ Administrative ☐ Misc/Unknown

Print the second level (or detail) codes that apply

BDEL SW AB EMED SC

Sample 9

I do hereby award you your first Wikipedia Barnstar Award for unswerving dedication to expanding and fin-tuning the [[Ridge Route]] article, photographing its highlights and eliminating many red links by adding excellent articles to take their place. You da bomb!

Check the box for one or more relevant top-level codes

☑ Editing ☐ Border Patrol ☐ Collaborative Action ☐ Meta Content

☑ Social/Community ☐ Administrative ☐ Misc/Unknown

Print the second level (or detail) codes that apply

EINI EMED EMAJ SC

Rev: 11.06.2019 page 3 of 3