



Article scientifique

Article

2021

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

Towards Open Science for the Qualitative Researcher: From a Positivist to an Open Interpretation

Class, Barbara; De Bruyne, Miguel Angel; Willemin, Claire; Donzé, Dimitri; Claivaz, Jean-Blaise

How to cite

CLASS, Barbara et al. Towards Open Science for the Qualitative Researcher: From a Positivist to an Open Interpretation. In: International journal of qualitative methods, 2021, vol. 20, p. 1–15. doi: 10.1177/16094069211034641

This publication URL: <https://archive-ouverte.unige.ch/unige:155794>

Publication DOI: [10.1177/16094069211034641](https://doi.org/10.1177/16094069211034641)

Towards Open Science for the Qualitative Researcher: From a Positivist to an Open Interpretation

International Journal of Qualitative Methods
Volume 20: 1–15
© The Author(s) 2021
DOI: 10.1177/16094069211034641
journals.sagepub.com/home/ijq
 SAGE

Barbara Class¹ , Miguel de Bruyne², Claire Wullemmin³ , Dimitri Donzé³ ,
Jean-Blaise Claivaz³

Abstract

This reflection by a qualitative researcher stems from a concrete experience with data handling in a funded research project. The researcher followed Open Research Data guidelines and found optimal solutions to pseudonymise data, but this later evolved into a deep epistemological questioning on praxis. During the first phase of the project, a tailor-made software was developed with help from librarians and an IT professional to automate the pseudonymisation of the 150 data chunks generated by 16 students, 3 tutors and 3 decision makers. In the second phase of the project, this experience sparked questions about the meaning of such data handling and interpretations of Open Science, which led the researcher to suggest a framework for the professional development of qualitative researchers in their understanding of Open Science. The article raises awareness of normative frameworks in institutional data handling practices and calls for active contributions to defining qualitative research in an Open Science perspective, particularly taking as a reference the recent draft recommendation by [UNESCO \(2020\)](#).

Keywords

open science, data handling, pseudonymisation, epistemological questioning

Introduction

Open Science is a new approach to the entire scientific process based on transparency, collaboration, new ways of disseminating knowledge and evolving interactions between science and society ([de la Fuente, 2017-19](#); [Ramjoué, 2015](#)). This approach requires stakeholders to develop new practices, institutions to create new infrastructures and researchers to explore new epistemologies for conceptualising research (e.g. [Aguinis et al., 2020](#); [Banks et al., 2019](#)). Open Science represents a major paradigm shift that completely remodels the scientific landscape, for instance by providing free access to research data. Open Research Data (ORD), which is one specific type of Open Data, refers to providing unrestricted access to the observations that lead to scientific results ([European Commission, n.d. b](#); [WikipediaContributors, 2020](#)). For qualitative researchers, this requires an extra measure of creativity given the specificities of their research, such as a naturalistic setting, contextual and iterative studies and the type of data produced.

One of the major epistemological arguments against qualitative data sharing concerns the unique context of qualitative data creation (e.g. [Broom et al., 2009](#); [Chauvette et al., 2019](#); [Feldman & Shaw, 2019](#); [Hammersley, 2010](#); [Tsai et al., 2016](#)). In this view, the use of qualitative data outside of the study in which they originated unavoidably alters that data ([Feldman & Shaw, 2019](#)). Ethical arguments against qualitative data sharing centre around the risk of causing harm to research participants by disclosing data that

¹University of Geneva, Faculty of Psychology and Educational Sciences, TECFA (Educational Technology Unit), Geneva, Switzerland

²University of Geneva, CUI (Computer Science Centre), Geneva, Switzerland

³University of Geneva, Geneva, Switzerland

Corresponding Author:

Barbara Class, Faculty of Psychology and Educational Sciences, TECFA (Educational Technology Unit), University of Geneva, 40, boulevard du Pont-d'Arve, Geneva 4 CH-1211, Switzerland.
Email: Barbara.Class@unige.ch



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

concern them (Tsai et al., 2016). Because informed consent is generally requested at the onset of a study for a specific purpose, some argue that it might not be relevant for potential future use of the data (Jones et al. 2018; Chauvette et al. 2019). In opposition, some scholars hold that data context does not preclude data reuse, and that qualitative data are fit for secondary analysis whether the context of the new study is akin to that of the original study or not. Others argue that attaching detailed metadata to datasets (Bishop, 2009; Jones et al., 2018) would allow subsequent researchers to treat each dataset, or dataset chunk, as a unique case and decide the extent to which each one could be shared and reused, depending on the nature of the data (Chauvette et al., 2019; Roulston, 2019), data-protection regulations and deeper methodological questions (Antonio et al., 2020). Despite this active and evolving debate on qualitative data sharing, however, no studies have yet suggested a framework for the development of qualitative researchers' understanding of Open Science and how this translates into their praxis.

The purpose of this study, which takes the form of a reflexive report conducted in a Scholarship of Teaching and Learning (SoTL) approach (Boyer, 1990), is to document one qualitative researcher's praxis (Freire, 1994 cited by Evans, 2007) as it relates to Open Science, specifically ORD. Although the study was initially undertaken in order to meet institutional requirements – that is, finding solutions to share the qualitative dataset – it evolved into a much deeper epistemological reflection on the development of qualitative researchers' interpretations of Open Science beyond the framework of institutional positivism (Piron, 2019).

The article starts with a dialectic presentation of the Open Science approach and data protection laws which illustrates the lack of consensus on the definition of Open Science. It describes issues involved with data handling and discusses some challenges of Open Data for qualitative researchers, providing some practical recommendations in this regard. Next, the article explains how the SoTL approach was used to guide the researcher's reflections. It then uses the case study of a Swiss National Science Foundation (SNSF) project on Open Education research methodology teaching to illustrate how qualitative data were handled and a tailor-made software designed and developed to automate pseudonymisation. Finally, the article opens a new line of epistemological questioning, namely, beyond the focus on data sharing (Antonio et al., 2020), what could the praxis of qualitative researchers look like in an Open Science perspective (UNESCO, 2020)?

Open Science

While Open Science has been gaining momentum in recent years, it is still an emerging movement (Vicente-Saez & Martinez-Fuentes, 2018) that entails profound changes in research culture and practices. Open Science itself can be

considered as a part of broader initiatives such as Open Knowledge (García-Peñalvo Francisco et al., 2010) or Open Innovation in science (Beck et al., 2020). Several studies have examined in detail what is meant by the concept of Open Science. Depending on the perspective used (stakeholder, domain, etc.), these studies have produced a remarkable variety of definitions. One, for example, states that the components of Open Science are Open Access,¹ Data, Reproducible Research, Science Definition, Science Evaluation, Science Guidelines, Science Policies, Science Projects and Science Tools (Pontika et al., 2015). Another describes Open Science as a complex that entails at least Open Access, Data, Source/Software and Hardware, Science Infrastructures, Evaluation, Educational Resources, Engagement of Societal Actors and Openness to Diversity of Knowledge, including Indigenous Knowledge Systems and Scholarly Knowledge & Inquiry (UNESCO, 2020).

From a historical point of view, Open Science is said to have its roots in the late sixteenth and early seventeenth centuries, when research societies made science public for the first time. Previously, the scientific profession had been veiled in secrecy and reserved for a select few researchers who were funded by royal, aristocratic or ecclesiastic patrons (David, 2013; Eamon, 1985). In the 1990s, the invention of the World Wide Web represented a major milestone for the Open Science movement. Its specific goal was to share information rapidly amongst academics worldwide (CERN, n.d.), and it essentially allowed scientific communities to re-establish the tradition of sharing their work publicly that had been compromised by copyright laws passed earlier in the 1900s (Dulong de Rosnay & Langlais, 2017; Langlais, 2015). In 2002, the Budapest Open Access Initiative set the stage for the free circulation of academic research findings (Budapest Open Access Initiative, 2002), further highlighting the technological possibilities of the web. The web has also made possible new ways to carry out research projects and distribute results (García-Peñalvo Francisco et al., 2010) and driven a re-evaluation of various issues surrounding scientific research, such as research management, outputs, assessment (San-Francisco-Declaration-on-Research-Assessment, 2012) and economics (David, 2013; Partha & David, 1994). These discussions are ongoing and continue to infuse the Open Science movement with new energy.

One notable issue currently under discussion concerns the absence of a formal definition for Open Science (Fecher & Friesike, 2014; Kraker et al., 2011; OECD, 2015; Vicente-Saez & Martinez-Fuentes, 2018). Open Science essentially serves as an umbrella term encompassing various conceptions and understandings of knowledge creation and dissemination that are highly dependent on one's context, point of view and interests (Fecher & Friesike, 2014). Open Science is frequently defined as the sum of its parts – for example, accessible, transparent and inclusive² (Beck et al., 2020). Table 1 (below) presents one tentative description of Open Science in terms of five different schools of thought, each based on different

Table 1. Variation of Open Science schools of thought, adapted from (Fecher & Friesike, 2014, p. 19).

	Infrastructure school	Public school	Measurement school	Democratic school	Pragmatic school
Assumption	Efficient research depends on available tools and applications	Science needs to be made accessible to the public	Scientific contributions need alternative impact measurements	Access to knowledge is unequally distributed	Knowledge creation could be more efficient if scientists worked together
Goal	Create openly available platforms, tools and services for scientists	Make science accessible for citizens	Develop an alternative metric system for scientific impact	Make knowledge freely available for everyone	Make the process of knowledge creation more efficient and goal-oriented
Keywords	Collaboration platforms and tools	Citizen science	Altmetrics, peer review, citation and impact factor	Open Access, intellectual property rights, Open Research Data and Open Code	Wisdom of the crowds, network effects, open research data, open code

assumptions and pursuing different goals (Fecher & Friesike, 2014).

Data Management, Anonymisation and Pseudonymisation

To operationalise Open Science in the ‘infrastructure school’ perspective and in line with one definition (Pontika et al., 2015) of Open Science, tools such as Data Management Plans (DMPs) were introduced in order to encourage better data management and stewardship throughout the data life cycle. DMPs require researchers to state at the outset of a research project how they will handle their data during the different stages of the project, including after it is completed (Burgi et al., 2017). The FAIR (Findable Accessible Interoperable Reusable) principles, published in 2016 by a consortium of researchers (Wilkinson et al., 2016), are a set of 10 principles covering data, metadata and infrastructure issues. Closely interconnected with DMPs (Jones et al., 2020), they are meant to strengthen the framework for the sharing and reusability of scientific digital assets and help create a valuable base for good practice in data management and stewardship, thereby ensuring more a sustainable and systematic dissemination of knowledge. The research community benefits from policies and requirements such as DMPs and the FAIR principles to guide stakeholders in the common effort to establish a more open research environment (Ayrís et al., 2018).

In qualitative research, however, these tools can be problematic for two reasons. Firstly, DMPs were designed for quantitative studies and, as such, are not adapted to support the emergent and participatory aspects of qualitative research (Antonio et al., 2020). Secondly, sharing research outputs comes with a number of difficulties and contradictions. For example, for research participants’ identities to be protected, ORD guidelines must be made to align with laws on personal data protection (Hardy et al., 2016). Not all types of data lend themselves to public sharing – especially sensitive or personal data, which qualitative data often are. The motto of ORD is ‘as open as possible, as closed as necessary’

(European Commission, 2016). If researchers are to share their data to the best of their ability, they must process data containing personal or sensitive information in such a way that it is impossible to identify the participants in a study.

With the introduction in Europe of the General Data Protection Regulation (GDPR) on top of existing national data protection laws, the close monitoring of data and the preservation of private life have taken centre stage in many fields. Although protecting the privacy of research participants was already a top concern, researchers have begun to pursue new ways of automating the challenging and time-consuming activity of data anonymisation. Several ways of rendering data untraceable to its original subject and/or context already exist or are being investigated in order to comply with data-sharing regulations. However, some datasets may become meaningless if they are altered, rendering moot the question of their reusability. In general, researchers must make their data-sharing and data anonymisation decisions taking into account the interests of all stakeholders, data-protection regulations and mandates of their institutions and/or funders while safeguarding the fundamental principles of qualitative data (Chauvette et al., 2019).

The introduction of data sharing has sparked persisting debates amongst qualitative researchers. These debates revolve around various types of issues, including methodological, infrastructural, technological, ethical and epistemological issues. In this study, we focus on ethical and epistemological issues that stem from the differences between quantitative and qualitative data, as well as data-protection practices in the context of qualitative studies – including the question of whether qualitative data can and should be shared.

As stated in the introduction to this article, epistemological arguments against qualitative data sharing relate to the situatedness of qualitative data (see, e.g., Broom et al., 2009; Chauvette et al., 2019; Feldman & Shaw, 2019; Hammersley, 2010; Tsai et al., 2016). Indeed, qualitative data are often ‘co-created’ through the unique relationship between researchers and their participants (Broom et al., 2009 citing Moore, 2007). Additionally, in contrast to

quantitative datasets, qualitative datasets are usually not shared for reproduction purposes.

Ethical arguments made against qualitative data sharing centre around the risk of causing harm to participants if the data are disclosed (Tsai et al., 2016). Qualitative data sharing bears the same risk as quantitative data sharing, meaning that it similarly can be mitigated and managed (Wutich & Bernard, 2016). While ethical issues related to data sharing in human research are well-known and regulated, documented and monitored by national laws, research ethics committees and scientific integrity bodies (Jones et al., 2018), some issues still raise questions, especially the issue of informed consent. Proposed solutions to this issue include returning to participants to ask for further consent and asking for a blanket informed consent (Chauvette et al., 2019 citing Childs et al., 2014; Heaton, 2008). However, none of these options resolve the problem entirely. Also, it is important to note that it is inaccurate to assume that participants are by default opposed to the sharing of their data. Especially if the research questions are of interest to them, participants are inclined to share their data (Jones et al., 2018), even if it means disclosing their personal identities. It is also wrong to assume that participants' rights take precedence by default over other people's rights and other ethical issues (Bishop, 2009).

A third issue related to data sharing is the tendency to think in black and white. This can be typified by attitudes such as 'every bit of data should be shared or no data at all' or 'data are either completely confidential or not at all confidential'. In reality, datasets are rarely homogenous and often present a range of inherent risks. Furthermore, access to shared data need not be either entirely open or closed: for example, it may be restricted through approval on a case-by-case basis or through data embargoes. Only personal and sensitive data are considered confidential under the GDPR, which is why removing critical information through anonymisation or pseudonymisation techniques is incentivised (Tsai et al., 2016). However, anonymisation and pseudonymisation are not fool-proof processes. On the one hand, they may be fallible, as participants may be re-identified by linking data together and using deductive reasoning (Koot, 2012; Rocher et al., 2019; Tsai et al., 2016). On the other hand, they may alter information to the point where the data become insubstantial and there would be no point in disseminating them (Chauvette et al., 2019; Parry & Mauthner, 2004; Tsai et al., 2016).

It is important to fully understand the difference between anonymisation and other techniques. The anonymisation of personal data can be defined as processing personal data in such a way as to irreversibly prevent anyone from identifying the person to whom it corresponds (Vokinger et al., 2020): 'Anonymisation is the process of creating anonymous information, namely, information which does not relate to an identified or identifiable natural person or to personal data³ rendered anonymous in such a manner that the data subject is not or no longer identifiable' (European Commission, n.d. a). Pseudonymisation is different in the sense that it is not irreversible. According to Article 4, para. 5 of the GDPR,

pseudonymisation 'means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person' (European Union, 2016).

Anonymisation and pseudonymisation have similar outcomes but differ regarding the treatment of the table that matches the pseudonyms to their corresponding entities. With anonymisation, this table is destroyed as soon as possible, while with pseudonymisation, it is saved in a secure place for a determined time. The latter practice is especially widespread in the social sciences, in which longitudinal studies and the triangulation of different sources of data are frequent. Researchers should use all the necessary technical and organisational measures to ensure that 'the pseudonymisation secret', that is, the table matching the entities to their corresponding pseudonyms, is preserved (European Union, 2019). At the operational level, anonymisation and pseudonymisation call for specific techniques. For anonymisation, common practices include generalisation, k-anonymity and randomisation, to name just a few (European Commission, 2014; Simi et al., 2017; Sweeney, 2002). For pseudonymisation, the most common techniques employed are counter, random number generator, cryptographic hash function, message authentication code and encryption (European Union, 2019).

More sophisticated techniques for pseudonymisation and anonymisation, such as data masking or Named Entity Recognition (NER), are currently being developed (Sedkaoui & Simian, 2020). Data masking seeks to obfuscate data so that they are invisible to individuals who have not been granted clearance. This method protects personal data by overwriting signs or other characters on them and comprises various subgenres based on specific techniques and methods (e.g. AI-Multiple, 2020; Tachepun & Thammaboosadee, 2020). NER belongs to the field of Natural Language Processing (NLP), a discipline combining linguistics, artificial intelligence and computer science. It consists of identifying explicitly named entities (often proper names) within a text or a corpus and then classifying them into predefined categories (e.g. 'Name', 'Location' and 'Date') using an algorithm, a piece of software or a system. Its initial purpose is to structure a text; however, recent projects have harnessed this automatic processing task and enhanced it with another layer of automation that allows a recognised and classified entity to be replaced with a label, thereby guaranteeing the entity's anonymity. For example, 'Geneva' becomes 'Place 1', or 'Gedeon Jayden' becomes 'Participant 1'. Only a few solutions currently use NER for pseudonymisation (Kleinberg et al., 2017) because of problems relating to lack of association – for example, treating 'the city at the end of the lake' and 'Geneva' as the same entity (ProjetInterregDecRIPT, 2020).

Anonymisation and pseudonymisation processes are both ontological activities. Researchers must make decisions regarding the portion of text that replaces the original text in order to respect the project's values and the privacy of

participants. For instance, in the case of participants' names, it is considered more user-friendly to read an article with familiar names, such as 'Emily'; however, names are culturally connoted. One possible solution would be to opt for a neutral identifier, like 'Student1'; another, more participatory option may be to ask participants how they would like to be named in the research report. Whatever the selected solution, ORD explicitly considers research data as a full-fledged part of the research process that needs to be opened up. However, current data handling practices reveal a tension between ORD management and privacy regulations (see Figure 1).

Researchers must consider three elements when pseudonymizing or anonymising data: (1) what to pseudonymise and how, taking into account the research question; (2) what parts of the dataset will be made publicly available on FAIR data repositories, balancing the usefulness of the data with the privacy of the research participants (Chauvette et al., 2019) and, finally, (3) which technology should be used to ensure a streamlined process for the researcher.

To summarise, data are deidentified for several reasons: (1) researchers must comply with national and European ethical data-protection laws like the GDPR; (2) public funding institutions strongly recommend that researchers publish their datasets on a data repository and (3) an increasing number of journals require researchers to upload their dataset with their article. This practice invites researchers to make the entire iceberg of their research visible and Open Access, including their dataset, DMP, instruments and article (Tennant et al., 2019) – whereas until recently, only the tip of the iceberg (i.e. the article) was required to be published.

Four recommended pseudonymization practices that are commonly used in qualitative research are (1) keeping the table isolated from the dataset, that is, in separate files and/or folders; (2) deleting the table from any insecure media such as memory storage and systems; (3) enforcing strong access control policies to prevent unauthorised entities to access the table and (4) if the table is stored on a computer, encrypting it and implementing a tight key management and storage policy

for the encryption. Decisions regarding pseudonymisation must always strike a balance between leaving in as much detailed information as possible and not allowing the person to be easily identifiable, which poses a particular challenge for qualitative researchers. It is recommended that, when making these decisions, researchers consider their research question and the requirements it imposes in terms of detailed information, as well as the effort readers would have to make in order to re-identify pseudonymised information.

Method

The methodology developed within this article is based on a Scholarship of Teaching and Learning (SoTL) approach (Boyer, 1990; Haigh & Withell, 2020; Hubball & Clarke, 2010; Miller-Young & Yeo, 2015). It describes a qualitative researcher's progress and reflection on the data deidentification recommendations of the public, democratic and pragmatic schools of Open Science (see Table 1). Table 2 outlines how the outcomes shared within the present article have been produced, using categories from Hubball and Clarke (2010, p. 4).

The study's sample is small, comprising 16 students, 3 tutors, 3 decision makers and 2 participants who held two roles: one as a stakeholder and tutor and the other as a distance learning coordinator and tutor. In terms of dataset size, this represents approximately 150 data chunks.

To develop the tailor-made software, the following constraints were taken into account:

- **Unstructured data:** Two types of data were collected – data generated within training modules hosted on the Moodle learning management system (LMS) during the 2018-19 academic year and interview data collected remotely (due to the public health crisis) during the spring of 2020. Both types of data were unstructured, resulting from essays, open ended templates or personal accounts (e.g. from the training modules).
- **Document format:** Most data chunks were either in docx or in pdf format. In terms of programming and in view of pseudonymising the data, reading and writing in these formats were thus essential.
- **Document structure:** The data contained several images, graphs and tables. It was therefore necessary to keep these intact to preserve the readability and context of the information.
- **Multi-document processing:** With approximately 150 data chunks to pseudonymise, it was important to develop a piece of software that could handle several tasks in succession.
- **Local only:** As the data were sensitive research data, hosting them in the cloud using services like docbyte (<https://www.docbyte.com/services>) was not an option.

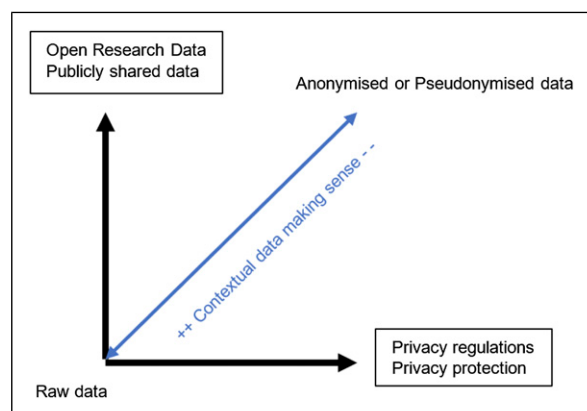
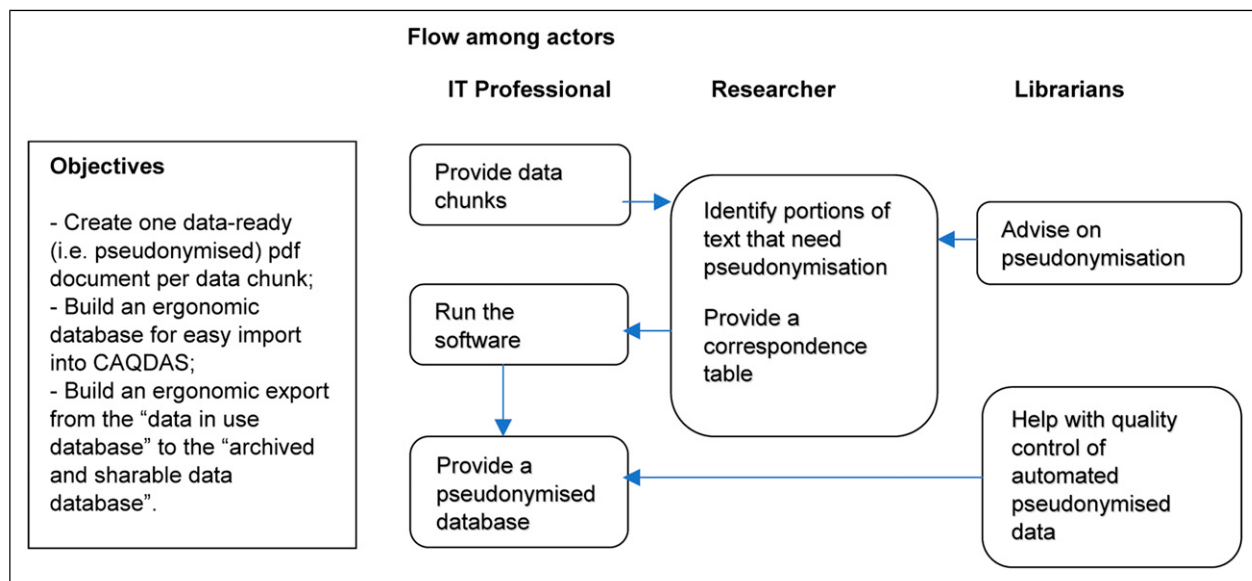


Figure 1. Tension between open research data and privacy regulations in data handling.

The researcher's data-processing objectives were three-fold: (1) to generate a pdf document for each data chunk;⁵

Table 2. SoTL methodology using the categories put forth by [Hubball and Clarke, 2010](#).

SoTL research context	Central SoTL research question	Methodological approach	Data collection methods	General outcomes
Contribute to the discussion on research methods pedagogy, including the handling of qualitative datasets in Open Science perspectives within a Swiss National Science Foundation (SNSF) project. ⁴	Beyond data handling, what could the praxis of qualitative researchers look like in an Open Science perspective?	Action is guided by the DMP and institutional injunctions. Reflection is informed by reflection on institutional positivism (Piron, 2019).	The dataset that informed the researcher's action and reflection is composed of two types of data: 1) real data from the learning environment, that is, stakeholders' productions during learning and teaching and 2) opinion data, that is, interviews with stakeholders.	1) At the level of action, practical recommendations to pseudonymise qualitative data 2) At the level of reflection, epistemological reflection on qualitative research in Open Science perspectives, beyond the issue of data handling.

**Figure 2.** Flow of data handling.

(2) to pseudonymise data in order to optimise data analysis with a computer-assisted qualitative data analysis software (CAQDAS) and (3) to easily generate a dataset that could then be shared on a FAIR data repository.

Figure 2 depicts the flow of data handling among the researcher, the IT professional and the librarians.

Example of the SNSF Project

At a minimum, the researcher needed to pseudonymise the names of the participants, geographical locations, institutions and persons referred to by the participants. The granularity of the pseudonymisation was to be defined by balancing the needs of the research and the 'not easily identifiable' principle. Given that the driving conjecture ([Sandoval, 2014](#)) was '*Teaching research methodology in an open education perspective requires a robust, multifaceted learning environment*', the researcher chose to keep participants' locations at the continental

level (Europe, North Africa and sub-Saharan Africa). Figure 3 shows the different levels of granularity that were considered. Regarding names, cultural connotations were avoided and neutral identifiers, like 'Student1' or 'Stakeholder2', were preferred. Some relevant contextual information was re-inserted in the metadata.⁶

For the interview data, automating the pseudonymisation process entailed modifying the researcher's previously used data-preparation workflow (Figure 4). Specifically, a correspondence table was filled out at step 3 that would be readjusted and used later in step 6. Qualitative researchers typically read and reread their data, so these additions were not felt to be an extra burden. A two-layer reading process was implemented by the researcher at step 3, where the first layer consisted of verifying the accuracy and fidelity of the transcription and the second layer, of identifying entities that needed to be pseudonymised.

When working on the correspondence table for the interview data, the researcher observed that, in terms of data



Figure 3. Example of granularity in the decision process.

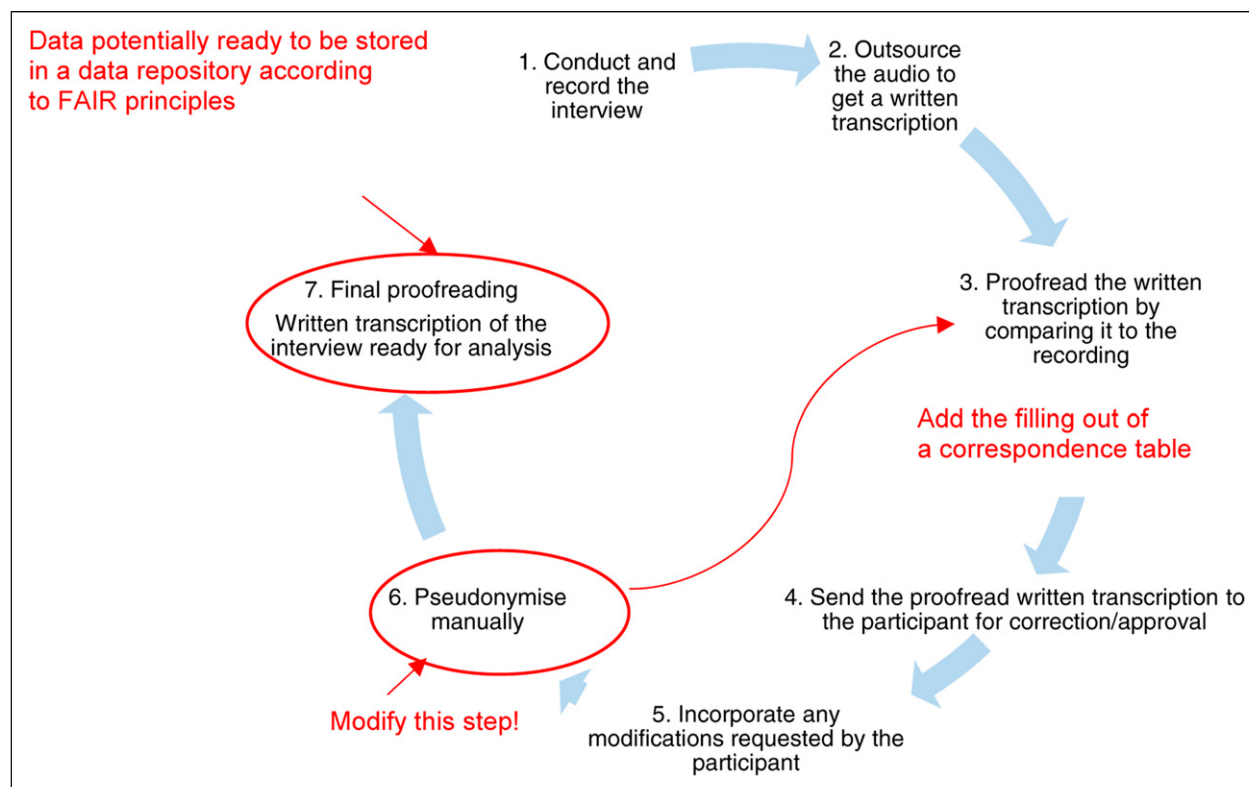


Figure 4. Process to prepare interview data for analysis.

preparation, it was more efficient to create one table per interview document. This resulted in perfectly accurate text replacement, instead of in occasionally broken sentences. However, the process needed to be incremental to ensure consistency, accuracy and efficiency. So, for example, the researcher first processed Interview 2 with the correspondence table created for Interview 1 and then created a new correspondence table just for Interview 2. All terms that had already been replaced for Interview 1 were excluded from this table, which included only new terms found in Interview 2. In parallel, a master table was created which consolidated the data from each individual table into one document. Interview 3 was then pseudonymised with the master table produced from Interview 1 and 2, while a new correspondence table was created to include new terms appearing in Interview 3, and so on.

For the researcher, having this information (i.e. data chunks that have been pseudonymised) for each individual document

could represent a new source of data to consider. Indeed, some interviews contained many names of persons, places, etc., while others had very few. Thus, the process of pseudonymisation could inform new aspects of the research and may be considered as additional data for analysis.

For the data generated by the Moodle LMS, in contrast, the researcher built only one correspondence table upfront. This resulted in some broken sentences when the software was run on this dataset. There were several reasons for this. The first was a lack of appropriate determinants. For instance, ‘At the University of Geneva, PhD training does not take the form of a formal doctoral school’ was pseudonymised as ‘At European University, PhD training does not take the form of a formal doctoral school’ instead of the expected result, ‘Within one European University, PhD training does not take the form of a formal doctoral school’. This problem could only be fixed by human reading at this stage of development of the tailor-made software. The second reason was that the same character chain was used in different contexts

in the dataset. For instance, the name of one participant and the name of one university were identical, so the software replaced it inconsistently. To take another example, the acronym of one participant's place of work (e.g. 'ENA', which was to be replaced by 'a sub-Saharan University') also appeared as a part of other words (e.g. **enable**), so the pseudonymisation resulted in unreadable sentences (e.g. 'they enable the user' was pseudonymised as 'they a sub-Saharan University ble the user'). Again, only human reading could fix these problems. Because of these issues with the automated pseudonymisation of the Moodle LMS data, it was decided to proceed differently with interview data.

Developing the Tailor-Made Software

A review was conducted of existing pseudonymisation software, but no perfect match was found because all programs required structured datasets in order to proceed with the systematic anonymisation/pseudonymisation treatment (e.g. Natural Language Toolkit (NLTK) and National Library of Medicine-Scrubber (NLM-Scrubber)). They also worked by using text extraction, which would have meant losing the information contained in pictures, graphs and tables. However, this review was helpful in that it heightened the researcher's awareness of the requirements and techniques of existing software. Three techniques were found to be particularly interesting due to their simplicity: capital words highlighting (i.e. finding words that start with a capital letter, which often correspond to words that need pseudonymisation), redaction software (i.e. censor information) and search and replace.

Next, the tailor-made software was created using Visual Basic for Applications (VBA), which was selected to keep the file formats (mainly docx, doc and jpg) intact and launch the software through an Excel document. Using the correspondence table filled out by the researcher (Figure 5), the GetFolder(), LoopAllSubFolders() and DocSearch() functions would replace automatically all terms.

To summarise, the pseudonymisation process was divided into tasks completed by the researcher and others by the software:

1. Select the correspondence table and launch the pseudonymisation process (researcher),
2. Select the folder containing the documents that need pseudonymisation (researcher),
3. List all documents, including those in subfolders (software),
4. Identify and open the first docx document (software),
5. Replace all words as specified in the correspondence table (software),
6. Save and close all instances of Word (software) and
7. Process the next document starting at step 4.

Once this process was completed for one document, the process would restart at task 4 for the next document (Figure 6).

In the future, the researcher will seek to make software upgrades in the following five areas: encryption, widening the

A	B	C	D
			Comment
Find	Replace		
Geneva	Europe		
Gedeon Jayden	PhDStudent23		

Figure 5. Find and replace text – Researcher's decision.

string-matching scope, handling document metadata, file types handling and open software. With regard to encryption, the software is limited in that it does not currently encrypt the table. To comply with European recommendations, this is an obligatory feature that must be invested in (European Union, 2019). Technically speaking, it can be achieved by either including a function in the code which would encrypt the selected cells with a password or by using third party tools such as 'obfuscators', which is available on Github.⁷ With regard to the string-matching scope, the software searches for exact matches only, which is problematic in the case of misspellings (e.g. geneve and genèv). This can be fixed by implementing what is called approximate string matching, also known as 'fuzzy search'. Particularly interesting in this regard is the Levenshtein distance⁸ formula, which calculates the difference between two strings. Using this formula would mean programming a simple clause that would require two strings to be considered as the same word if the similarity between them is superior to (for example) 80%.

With regard to handling metadata, systematically erasing the metadata of the original files (such as titles, authors, creation date, etc.) is another feature that needs to be added by using preconceived functions and integrating them into the code. In terms of file types, the software currently only accepts docx. A preliminary workaround for this limitation would consist of converting any proprietary file type into an open file type. For instance, this tailor-made software could be ported into different programming languages in order to process Open Document Format (ODF) files. As this is an open format, there is a large number of text editors created in different programming languages, so there is no preferred porting language. In the same vein, the tailor-made software is shared on Github⁹ with a GNU AGPLv3¹⁰ licence.

More substantial improvements to the software would consist of introducing artificial intelligence such as support vector machine and natural language processing. Concretely, this would consist of porting the software into a more complete programming language such as Python and using libraries such as Natural Language Toolkit (NLTK) to analyse speech patterns. Packaging the software into a user-friendly, intuitive interface that integrates with the qualitative researcher's workbench (e.g. transcription software; CAQDAS and data repositories) would also be essential. More generally speaking, it is important that this software continues to be developed in collaboration with other qualitative researchers,

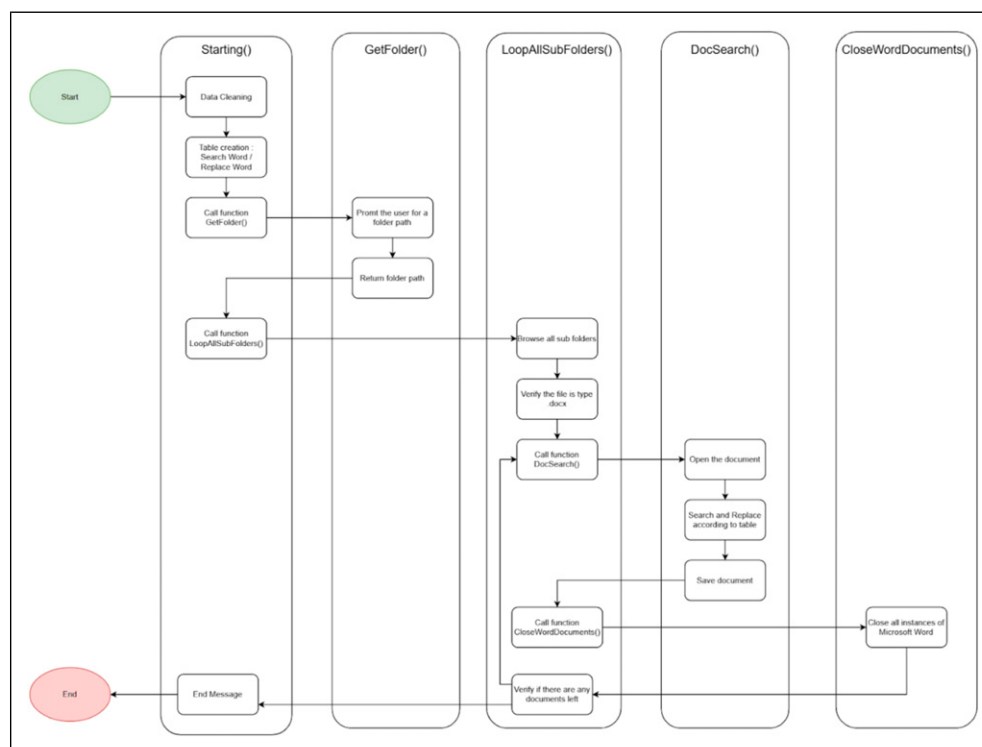


Figure 6. Functioning of the tailor-made software to automate the pseudonymisation process.

taking into consideration the full range of values of qualitative research in line with UNESCO (2020)’s understanding of Open Science.

Epistemological Reflection

Several authors (Brière et al., 2019) have reflected on Open Access in other parts of the world than the Global North (Piron et al., 2017) and have argued against what they call ‘institutional positivism’. Piron (2019) states that, historically, what is called ‘scientific research’ has tried to eliminate any situatedness of individual researchers as manifested in their emotions, personal interests related to their social position, experiences as they relate to their identity or political and social engagements, positing that it is only by controlling those four elements that the human mind is able to access the truth. Of course, this is an extreme positivist position that stands in complete opposition to the constructivist view that scientific knowledge is situated in cultural, political, economic and other contexts. The positivist injunction of neutrality is the outcome of a normative framework with which almost all public academic institutions are linked, namely, the relationship between the ‘knower’ and the ‘other’, where the knower symbolically dominates the other. This normative framework is so deeply anchored in Western scientific thinking that it is inherent to all institutional instruments and orients researchers’ reflection towards

positivism (e.g. ethical forms, funding forms that require a hypothesis, variables and results).

The pseudonymisation efforts described in this article were conducted from within the normative framework, a fact which was not acknowledged or interrogated in the initial stages of the project. Indeed, the researcher’s main concern at these stages was ‘How can pseudonymisation be automated?’ It was only many months later, when conducting interviews with research method teachers, that the researcher began to question this framework, and thus the project’s entire process. The main question subsequently became: ‘Beyond data handling, what could the praxis of qualitative researchers look like in an Open Science perspective?’

To answer this question, an effort was first made to define both qualitative research and Open Science. However, this proved to be a difficult task, as both concepts have been defined in a wide variety of ways. Indeed, Section 2 of this article has shown that emergent interpretations of Open Science can be extremely diverse, and other authors have done the same for qualitative research (e.g. Aspers & Corte, 2019; Brinkmann et al., 2014). The researcher therefore decided to make an active contribution to the conversation by suggesting that both terms be defined in relation to the four foundational constituents of scientific knowledge – axiology, ontology, epistemology and methodology. Building on the encyclopaedia definitions of each of these constituents, the following paragraphs offer a specific understanding of what each one

might mean for qualitative researchers, laying the groundwork for future discussion in this area. The Sage Encyclopaedia of Qualitative Research Methods, in English, was chosen as the initial reference because encyclopaedias are usually the entry point of a literature review and are considered authoritative content, and because English is a dominant language in scientific research.

Axiology, or the philosophy of values, engages with qualitative research in the areas of ethics, research paradigm assumptions and knowledge creation within a scientific inquiry (Hiles, 2008). Axiology seeks to ask questions such as ‘What is the value of knowledge itself? What is the ultimate purpose of human inquiry? How much should inquiry promote human flourishing, individual empowerment, advocacy, activism, relief from oppression and suffering and so on?’ (Hiles, 2008, p. 4). Heron (1996) has emphasised the importance of including research participants as co-researchers throughout the process to avoid their disempowerment, misrepresentation or manipulation. Qualitative researchers are invited to position themselves towards scientific knowledge generation (e.g. awareness of funders’ goals within larger research agendas) and engage in research projects with full awareness of the values being promoted (e.g. the underlying values driving the study). In any research endeavour, researchers should think critically about the values involved, as taking them for granted could mean acting against certain values within which researchers would like to situate their research.

Ontology, or the philosophy of being, is narrowed down to social ontology for the purposes of this article. As such, it seeks to reveal and break up ‘fixed and untested assumptions that illegitimately limit the field of questions and answers that guide empirical inquiry’ (Noonan, 2008, p. 3). Ontology seeks to ask questions such as ‘How did the given social reality come to be constituted as it appears? In what sense is society an object of research? Is social reality distinct from the individuals who make it up?’ (Noonan, 2008, pp. 2–3). Reality is the product of dynamic processes of nature and society and requires critical thinking about the creation of concepts. The activities of classification and categorisation reflect a given structure of human consciousness and carry the risk of locking stakeholders in. Ontology tries to prevent this by questioning assumptions and opening up the field of research without pre-determining answers (Noonan, 2008). Qualitative researchers are invited to adopt a critical perspective towards the research object they are studying and question underlying assumptions of any pre-existing categorisation (e.g. what is the reality in this study and how did it come to exist in its current form?). They are also invited to think critically about any structure that arises from their own research.

Epistemology is the theory of knowledge. The meaning and application of knowledge is diverse and evolving and has long been considered from Western-centred perspectives, with typical worldviews being positivism or constructivism (Stone, 2008). Epistemology seeks to ask questions such as ‘What is

knowing? What is the known? What is knowledge?’ (Stone, 2008, p. 2). A call to epistemic broadening, which encompasses topics such as epistemic justice, is gaining momentum to open up new epistemological perspectives (e.g. Kidd et al., 2017; Rizvi & Caldera, 2021). The goal is to move away from the injunction of neutrality, which requires that researchers remain objective, and move toward building knowledge that is inclusive, heterogeneous, dialogical and aware – both of its roots in culture and language as well as of the connections that exist between its stakeholders (Piron et al., 2016). For qualitative researchers, answering the question ‘What is knowledge’ requires making the effort to go beyond axiology and ontology and potentially position new forms of knowing in the epistemological choices they make. For instance, actor-network theory, epistemology of the link and epistemology of reception all consider the ‘research participant’ as a co-designer of the research journey in an inclusive perspective (Bouilloud, 1997; Fenwick & Edwards, 2018; Nind, 2020; Piron, 2017). Researchers are invited to question the origin of their chosen worldviews and the implications they might have for research design and research findings (e.g. ‘How is knowledge defined within the framework chosen for this study?’).

Methodology is a product of the previous three constituents, and it operationalises and engineers these highly philosophical constituents into an aligned and achievable research design. Methodology, within empirical studies, seeks to ask questions such as ‘What is the guiding paradigm? What is the main research question? What are the conceptual models? Who are the participants and what is the field? What are the procedures for data collection, analysis and interpretation?’ (Schensul, 2008). Answering these questions leads qualitative researchers to take decisions concerning the study they want to conduct.

To summarise, we suggest conceptualising Open Science and qualitative research (in all their interpretations) as being in dialogue with one another and with the four core constituents of scientific inquiry (Figure 7).

At a more practical level, qualitative researchers are invited to consider the full range of interpretations of these concepts

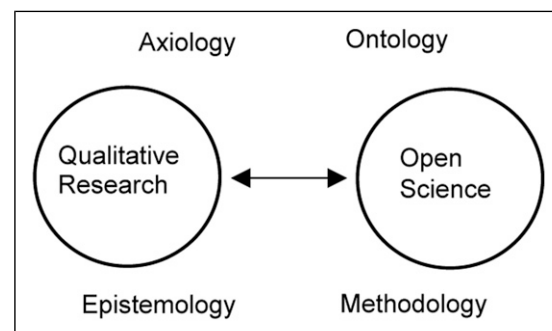


Figure 7. Dialogue between Open Science, qualitative research and the core constituents of scientific knowledge.

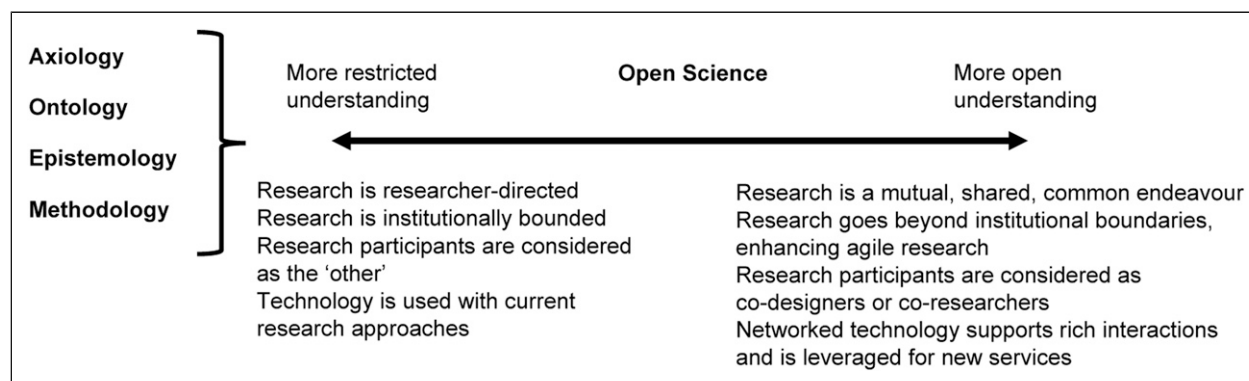


Figure 8. Framework for systematic reflection and positioning a qualitative study in relation to Open Science.

and make an informed decision about where they position themselves in relation to them before undertaking a new research project (Figure 8). They are also invited to think creatively (one of the most important skills in the 21st century) in exploring new possibilities for qualitative research in an Open Science perspective.

Discussion and Conclusion

Open Science presents qualitative researchers with an opportunity to think deeply about what the Open paradigm means for qualitative research at the axiological, ontological, epistemological and methodological levels. This article has shared one qualitative researcher's reflections on changes in praxis introduced to align with Open Science principles, which ultimately led the researcher to suggest a framework for the development of qualitative researchers in their understanding of Open Science (Figures 7 and 8).

Researchers are already in continual professional development and must master multiple competencies. For instance, they must be experts in a specific area of research, including its methodology and certain specialised scientific software, possess advanced project management and engineering skills, be able to successfully navigate funding application procedures and be skilled networkers who pursue and develop interdisciplinary relationships (Djerasimovic & Villani, 2019; Nicholas et al., 2017). Open Science now invites researchers to develop professionally by working together to explore new ways of conducting qualitative research. This article has sought to do this by reporting reflections first from within an 'institutional positivist' framework, focussing quasi-exclusively on Open Research Data and Open Access, and then outside it, as the researcher moved away from these concepts in an effort to think in terms of the deeper, underlying constituents of scientific knowledge production and epistemic opening.

In doing this important work, it may be helpful to take a step back and reconsider critically the history of qualitative research as a source of inspiration. Brinkmann et al. (2014)

have pointed to the six histories of qualitative research – conceptual, internal, marginalising, repressed, social and technological. In their conceptual history, they argue that the 1970s were a pivot decade with, among other things, the emergence of postmodernity. They also give examples of how technology has been used creatively in qualitative research, implicitly calling for more such innovation in the field. However, as researchers explore new possibilities for praxis, they should be careful to keep in mind the intrinsic characteristics of qualitative research – which may not be as clear-cut as one might think (Aspers & Corte, 2019). In any case, qualitative research in an Open Science perspective presents a clear opportunity for deep reflection and methodological reconsideration, in line with Antonio et al. (2020)'s work and with UNESCO (2020)'s draft recommendation for Open Science.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The project is funded by the Swiss National Science Foundation, <http://p3.snf.ch/project-190634>.

ORCID iDs

Barbara Class  <https://orcid.org/0000-0002-5461-2307>

Claire Willemin  <https://orcid.org/0000-0003-2264-5072>

Dimitri Donzé  <https://orcid.org/0000-0002-8742-9971>

Jean-Blaise Claivaz  <https://orcid.org/0000-0001-8782-0922>

Notes

1. To avoid repeating the word 'Open' throughout these definitions, we have deliberately removed it, leaving only the core concept (i.e. 'Open Data' becomes 'Data').
2. 'Accessibility (e.g. open access to publications and research data), transparency (e.g. reproducibility of results and open

- peer review) and inclusivity (e.g. citizen science)' (Beck et al., 2020, p. 3).
3. 'Personal data are any information relating to an identified or identifiable natural person ('Data subject'). An identifiable person is one who can be identified, directly or indirectly. Where an individual is not identifiable, data are said to be anonymous' (EuropeanCommission, no date).
 4. <http://p3.snf.ch/project-190634>.
 5. The broader question of transcribing an oral interview into written text is not considered here. Azevedo et al., (2017).
 6. For instance, 'cities that have been replaced by Europe are cities from Switzerland and France' or '9 students are from North Africa, 5 students are from other Southern francophone countries – for example, sub-Saharan Africa – and 2 students are from Europe'.
 7. To give an example of such a process but at commercial level, the software SpaCy encrypts data, <https://spacy.io/usage/facts-figures#benchmarks>.
 8. https://en.wikipedia.org/wiki/Levenshtein_distance.
 9. <https://github.com/mig2895/VBA-Find-Replace> ; <https://github.com/mig2895/findAndReplace-Python>.
 10. <https://choosealicense.com/licenses/>.
- ## References
- Aguinis, H., Banks, G. C., Rogelberg, S. G., & Cascio, W. F. (2020). Actionable recommendations for narrowing the science-practice gap in open science. *Organizational Behavior and Human Decision Processes*, 158, 27-35. <https://doi.org/10.1016/j.obhdp.2020.02.007>.
- AI-Multiple. (2020). *Data masking: What it is, how it works, types & best practices*. <https://research.aimultiple.com/data-masking/>.
- Antonio, M. G., Schick-Makaroff, K., Doiron, J. M., Shields, L., White, L., & Molzahn, A. (2020). Qualitative data management and analysis within a data repository. *Western Journal of Nursing Research*, 42(8), 640-648. <https://doi.org/10.1177/0193945919881706>.
- Aspers, P., & Corte, U. (2019). What is qualitative in qualitative research. *Qualitative Sociology*, 42(2), 139-160. <https://doi.org/10.1007/s11133-019-9413-7>.
- Ayris, P., Bernal, I., Cavalli, V., Dorch, B., Frey, J., Hallik, M., & Horstmann, W. (2018). *LIBER open science roadmap*. Zenodo. <https://doi.org/10.5281/zenodo.1303002>.
- Azevedo, V., Carvalho, M., Fernandes-Costa, F., Mesquita, S., Soares, J., Teixeira, F., & Maia, Â. (2017). Interview transcription: Conceptual issues, practical guidelines, and challenges. *Revista de Enfermagem Referência*, 4(14), 159-167. <https://doi.org/10.12707/RIV17018>.
- Banks, G. C., Field, J. G., Oswald, F. L., O'Boyle, E. H., Landis, R. S., Rupp, D. E., & Rogelberg, S. G. (2019). Answers to 18 questions about open science practices. *Journal of Business and Psychology*, 34(3), 257-270. <https://doi.org/10.1007/s10869-018-9547-8>.
- Beck, S., Bergenholtz, C., Bogers, M., Brasseur, T.-M., Conradsen, M. L., Di Marco, D., Distel, A. P., Dobusch, L., Dörler, D., Effert, A., Fecher, B., Filiou, D., Frederiksen, L., Gillier, T., Grimpe, C., Gruber, M., Haeussler, C., Heigl, F., Hoisl, K., & Xu, S. M. (2020). The open innovation in science research field: A collaborative conceptualisation approach. *Industry and Innovation*. <https://doi.org/10.1080/13662716.2020.1792274>.
- Bishop, L. (2009). Ethical sharing and reuse of qualitative data. *Australian Journal of Social Issues*, 44(3), 255-272. <https://doi.org/10.1002/j.1839-4655.2009.tb00145.x>.
- Bouilloud, J.-P. (1997). Pour une « épistémologie de la réception ». In *Sociologie et société* (pp. 219-244). Presses Universitaires de France. <https://www.cairn.info/sociologie-et-societe-9782130486312-page-219.htm>.
- Boyer, E. (1990). *Scholarship reconsidered: Priorities of the professoriate*. The Carnegie Foundation for the Advancement of Teaching.
- Brière, L., Lieutenant-Gosselin, M., & Piron, F. (2019). *Et si la recherche scientifique ne pouvait pas être neutre?* Éditions science et bien commun. <https://scienceetbiencommun.pressbooks.pub/neutralite/>.
- Brinkmann, S., Jacobsen, M. H., & Kristiansen, S. (2014). Historical overview of qualitative research in the social sciences. In P. Leavy (Ed.), *The oxford handbook of qualitative research*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199811755.013.017>.
- Broom, A., Cheshire, L., & Emmison, M. (2009). Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sociology-the Journal of the British Sociological Association*, 43(6), 1163-1180. <https://doi.org/10.1177/0038038509345704>.
- Budapest Open Access Initiative. (2002). *Budapest open access initiative* <https://www.budapestopenaccessinitiative.org/>.
- Burgi, P.-Y., Blumer, E., & Makhoul-Shabou, B. (2017). Research data management in Switzerland: National efforts to guarantee the sustainability of research outputs. *IFLA Journal*, 43(1), 5-21. <https://doi.org/10.1177/0340035216678238>.
- CERN. (n.d.). *The birth of the Web*. <https://home.cern/science/computing/birth-web>.
- Chauvette, A., Schick-Makaroff, K., & Molzahn, A. E. (2019). Open data in qualitative research. *International Journal of Qualitative Methods*, 18, 1-6. <https://doi.org/10.1177/1609406918823863>.
- David, P. (2013). The historical origins of 'open science': An essay on patronage, reputation and common agency contracting in the scientific revolution. *Capitalism and Society*, 3(2), 5. <https://ssrn.com/abstract=2209188>.
- de la Fuente, G. (2017-19). *What is open science? Introduction*. FOSTER consortium <https://www.fosteropenscience.eu/index.php/content/what-open-science-introduction>.
- Djerasimovic, S., & Villani, M. (2019). Constructing academic identity in the European higher education space: Experiences of early career educational researchers. *European Educational Research Journal*, 19(1), 147490411986718. <https://doi.org/10.1177/1474904119867186>.

- Dulong de Rosnay, M., & Langlais, P.-C. (2017). Public artworks and the freedom of panorama controversy: A case of Wikimedia influence. *Internet Policy Review*, 6(1), 3-5. <https://doi.org/10.14763/2017.1.447>.
- Eamon, W. (1985). From the secrets of nature to public knowledge: The origins of the concept of openness in science. *Minerva*, 23(3), 321-347. <https://doi.org/10.1007/BF01096442>.
- European Commission. (2014). *Article 29 - Data protection working party*. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf.
- European Commission. (2016). *H2020 programme guidelines on FAIR data management in horizon 2020*. Directorate-General for Research & Innovation. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
- European Commission (n.d. a). *Anonymization*. https://ec.europa.eu/eurostat/cros/content/anonymization_en.
- European Commission. (n.d. b). *Facts and Figures for open research data*. https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en.
- European Union. (2016). *General data protection regulation (GDPR)*. <https://gdprinfo.eu/>.
- European Union. (2019). *Pseudonymisation techniques and best practices. Recommendations on shaping technology according to data protection and privacy provisions*. European Union Agency for Cybersecurity. <https://www.enisa.europa.eu/publications/pseudonymisation-techniques-and-best-practices> <https://www.semanticscholar.org/paper/Qualitative-data-sharing-and-re-use-for-systems-A-Jones-Alexander/b6915b73324f8229f3866a1b135508a51da42154>.
- Fecher, B., & Friesike, S. (2014). Open science: One term, five schools of thought. In S. Bartling & S. Friesike (Eds.), *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing* (pp. 17-47). Springer International Publishing. https://doi.org/10.1007/978-3-319-00026-8_2.
- Feldman, S., & Shaw, L. (2019). The epistemological and ethical challenges of archiving and sharing qualitative data. *American Behavioral Scientist*, 63(6), 699-721. <https://doi.org/10.1177/0002764218796084>.
- Fenwick, T., & Edwards, R. (Eds.). (2018). *Revisiting actor-network theory in education*. Routledge.
- Freire, P. (1994). *Pedagogy of the oppressed*. Continuum.
- García-Peñalvo Francisco, J., García de Figuerola, C., & Merlo José, A. (2010). Open knowledge: Challenges and facts. *Online Information Review*, 34(4), 520-539. <https://doi.org/10.1108/14684521011072963>.
- Haigh, N., & Withell, A. J. (2020). The place of research paradigms in SoTL practice: An inquiry. *Teaching & Learning Inquiry*, 8(2), 17-31. <https://doi.org/10.20343/teachlearninqu.8.2.3>.
- Hammersley, M. (2010). Can we re-use qualitative data via secondary analysis? Notes on some terminological and substantive issues. *Sociological Research Online*, 15(1), 47-53. <https://doi.org/10.5153/sro.2076>.
- Hardy, L. J., Hughes, A., Hulen, E., & Schwartz, A. L. (2016). Implementing qualitative data management plans to ensure ethical standards in multi-partner centers. *Journal of Empirical Research on Human Research Ethics*, 11(2), 191-198. <https://doi.org/10.1177/1556264616636233>.
- Heaton, J. (2008). Secondary analysis of qualitative data: An overview. *Historical Social Research/Historische Sozialforschung*, 33(3 (125)), 33-45. <http://www.jstor.org/stable/20762299>.
- Heron, J. (1996). *Co-operative inquiry: Research into the human condition*. Sage.
- Hiles, D. (2008). Axiology. In L. Given (Ed.), *The SAGE encyclopedia of qualitative research methods*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412963909>.
- Hubball, H., & Clarke, A. (2010). Diverse methodological approaches and considerations for SoTL in higher education. *The Canadian Journal for the Scholarship of Teaching and Learning*, 1(1), 2. <https://doi.org/10.5206/cjsotl-rcacea.2010.1.2>.
- Jones, K., Alexander, S. M., Bennett, N., Bishop, L., Budden, A., Cox, M., Crosas, M., Game, E., Geary, J., Hahn, C., Hardy, D., Johnson, J. D., Karcher, S., LaFevor, M., Motzer, N., Silva, P., Pittman, J., Randell, H., Silva, J., ... Winslow, D. (2018). *Qualitative data sharing and re-use for socio-environmental systems research: A synthesis of opportunities, challenges, resources and approaches*.
- Jones, S., Pergl, R., Hooft, R., Miksa, T., Samors, R., Ungvari, J., Davis, R. I., & Lee, T. (2020). Data management planning: How requirements and solutions are beginning to converge. *Data Intelligence*, 2(1-2), 208-219. https://doi.org/10.1162/dint_a_00043.
- Kidd, I., Medina, J., & Pohlhaus, G. (2017). *The routledge handbook of epistemic injustice*. Routledge.
- Kleinberg, B., Mozes, M., van der Toolen, Y., & Verschuere, B. (2017). NETANOS - Named entity-based text anonymization for open science. *PsyArxiv*. <https://doi.org/10.31219/osf.io/w9nhb>.
- Koot, M. R. (2012). *Measuring and predicting anonymity*. [Ph.D. thesis, University of Amsterdam].
- Kraker, P., Leony, D., Reinhardt, W., & Beham, G. (2011). The case for an open science in technology enhanced learning. *International Journal of Technology Enhanced Learning*, 3(6), 643-654. <https://doi.org/10.1504/ijtel.2011.045454>.
- Langlais, P.-C. (2015). Quand les articles scientifiques ont-ils cessé d'être des communs? *Sciences Communes*. <https://scoms.hypotheses.org/409>.
- Miller-Young, J., & Yeo, M. (2015). Conceptualizing and communicating SoTL: A framework for the field. *Teaching & Learning Inquiry: The ISSOTL Journal*, 3(2), 37-53. <https://doi.org/10.2979/teachlearninqu.3.2.37>.
- Moore, N. (2007). (Re)using qualitative data? *Sociological Research Online*, 12(3), 1-13. <https://doi.org/10.5153/sro.1496>.
- Nicholas, D., Watkinson, A., Boukacem-Zeghmouri, C., Rodríguez-Bravo, B., Xu, J., Abrizah, A., Świgoń, M., & Herman, E. (2017). Early career researchers: Scholarly behaviour and the prospect of change. *Learned Publishing*, 30(2), 157-166. <https://doi.org/10.1002/leap.1098>.

- Nind, M. (2020). *Inclusive research research methods*. Bloomsbury Academic.
- Noonan, J. (2008). Ontology. In L. Given (Ed.), *The SAGE encyclopedia of qualitative research methods*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412963909>.
- OECD. (2015). *Making open science a reality* (Vol. 25). OECD Publishing. <https://doi.org/10.1787/5jrs2f963zs1-en>.
- Parry, O., & Mauthner, N. S. (2004). Whose data are they anyway?: Practical, legal and ethical issues in archiving qualitative research data. *Sociology-the Journal of the British Sociological Association*, 38(1), 139-152. <https://doi.org/10.1177/0038038504039366>.
- Partha, D., & David, P. A. (1994). Toward a new economics of science. *Research Policy*, 23(5), 487-521. [https://doi.org/10.1016/0048-7333\(94\)01002-1](https://doi.org/10.1016/0048-7333(94)01002-1).
- Piron, F. (2017). Méditation haïtienne : répondre à la violence séparatrice de l'épistémologie positiviste par l'épistémologie du lien. *Sociologie et sociétés*, 49(1), 33-60. <https://doi.org/10.7202/1042805ar>.
- Piron, F. (2019). L'amoralité du positivisme institutionnel. L'épistémologie du lien comme résistance In L. Brière, M. Lieutenant-Gosselin, & F. Piron (Eds.), *Et si la recherche scientifique ne pouvait pas être neutre?*. Éditions science et bien commun.
- Piron, F., Diouf, A., Dibounje Madiba, M., Mboa Nkoudou, T., Aubierge Ouangré, Z., Tessy, D., Achaffert, H., Anderson, P., & Lire, Z. (2017). *Le libre accès vu d'Afrique francophone subsaharienne*. Revue française des sciences de l'information et de la communication. <https://doi.org/10.4000/rfsic.3292>.
- Piron, F., Regulus, S., & Djiboune Madiba, M. S. (2016). Introduction : Une autre science est possible. In *Justice cognitive, libre accès et savoirs locaux. Pour une science ouverte juste, au service du développement local durable*. Science et bien commun.
- Pontika, N., Knuth, P., Cancellieri, M., & Pearce, S. (2015). *Fostering open science to research using a taxonomy and an eLearning portal. iKnow: 15th international conference on knowledge technologies and data driven business*, Graz.
- ProjetInterregDecRIPT. (2020). *Résumé du projet INTERREG DECRYPT en vue des entretiens qualitatifs dans le cadre de l'étude de marché*. <http://tesniere.univ-fcomte.fr/projet-decrypt/ressources/INTERREG-DECRYPT-resume-pour-etude-de-marche.pdf>.
- Ramjoué, C. (2015). Towards open science: The vision of the European Commission. *Information Services and Use*, 35(3), 167-170. <https://pdfs.semanticscholar.org/c47a/f7edc0607c6561390bba2a993c89711297c8.pdf>.
- Rizvi, S., & Caldera, C. (2021). Racially-Just Epistemologies and Methodologies that Disrupt Whiteness. *International Journal of Research & Method in Education*. https://think.taylorandfrancis.com/special_issues/epistemologies-methodologies-whiteness/.
- Rocher, L., Hendrickx, J. M., & de Montjoye, Y.-A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, 10(1), 3069. <https://doi.org/10.1038/s41467-019-10933-3>.
- Roulston, K. (2019). *Archiving qualitative data*. QualPage. Examining the world through qualitative inquiry. <https://qualpage.com/2019/05/09/archiving-qualitative-data/comment-page-1/>.
- San-Francisco-Declaration-on-Research-Assessment. (2012). *San Francisco declaration on research assessment. Declaration on research assessment (DORA)*. <https://sfdora.org/read>.
- Sandoval, W. (2014). Conjecture mapping: An approach to systematic educational design research. *Journal of the Learning Sciences*, 23(1), 18-36. <https://doi.org/10.1080/10508406.2013.778204>.
- Schensul, J. (2008). Methodology. In L. Given (Ed.), *The SAGE encyclopedia of qualitative research methods*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412963909>.
- Sedkaoui, S., & Simian, D. (2020). Developed framework based on cognitive computing to support personal data protection under the GDPR. In D. Simian & L. Stoica (Eds.), *Modelling and development of intelligent systems*. Springer. https://doi.org/10.1007/978-3-030-39237-6_8.
- Simi, M., Sankara, K., Sankara Nayaki, S., & Sudheep Elayidom, M. (2017). An extensive study on data anonymization algorithms based on K-anonymity. *IOP Conference Series: Materials Science and Engineering*, 225, 012279.
- Stone, L. (2008). Epistemology. In L. Given (Ed.), *The SAGE encyclopedia of qualitative research methods*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412963909>.
- Sweeney, L. (2002). Achieving k-anonymity privacy protection using generalization and suppression. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10(5), 571-588.
- Tachepun, C., & Thammaboosadee, S. (2020). A Data masking guideline for optimizing insights and privacy under GDPR compliance. In *Proceedings of the 11th international conference on advances in information technology* (pp. 1-9). Association for Computing Machinery.
- Tennant, J., Beamer, J., Bosman, J., Brembs, B., Chung, N., Clement, G., & Turner, A. (2019, January 30). *Foundations for open scholarship strategy development*. [Open Scholarship Strategy](https://open-scholarship-strategy.github.io/site/). <https://open-scholarship-strategy.github.io/site/>.
- Tsai, A. C., Kohrt, B. A., Matthews, L. T., Betancourt, T. S., Lee, J. K., Papachristos, A. V., Weiser, S. D., & Dworkin, S. L. (2016). Promises and pitfalls of data sharing in qualitative research. *Social Science & Medicine*, 169, 191-198. <https://doi.org/10.1016/j.socscimed.2016.08.004>.
- UNESCO. (2020). *First draft of the UNESCO recommendation on open science*. <https://unesdoc.unesco.org/ark:/48223/pf0000374837>.
- Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open science now: A systematic literature review for an integrated definition. *Journal of Business Research*, 88, 428-436. <https://doi.org/10.1016/j.jbusres.2017.12.043>.
- Vokinger, K. N., Stekhoven, D. J., & Krauthammer, M. (2020). Lost in anonymization — A data anonymization reference classification merging legal and technical considerations. *The Journal*

- of Law, Medicine & Ethics*, 48(1), 228-231. <https://doi.org/10.1177/1073110520917025>.
- WikipediaContributors. (2020). *Open science data*. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Open_science_data&oldid=997359416.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship [Comment]. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Wutich, A., & Bernard, H. R. (2016). Sharing qualitative data & analysis. With whom and how widely?: A response to 'Promises and pitfalls of data sharing in qualitative research'. *Social Science & Medicine*, 169, 199-200. <https://doi.org/10.1016/j.socscimed.2016.09.041>.