Open Access and Sensitive Social Sciences Data in Different Legislative Contexts: The Case of Strategic Selection “Elite” Interviewing in Sweden

E. C. H. Keskitalo

Abstract
Open access has moved from the periphery to the mainstream in recent years (Suber, 2012). Originating in broader movements towards open access (e.g., Costa and Leite, 2016), in Europe the open access movement has recently been driven by a focus within the European Commission on making “research data that result from publicly funded research publicly accessible, useable and reusable” (2012/417/EU). The aim is to make research more accessible, transparent, and trustworthy, and of greater benefit to society. The underpinning argument is that the provision of open access to research data will allow anyone to examine whether “the right conclusions are drawn” given the research data used for a particular study.

While the basic logic of this argument is hardly disputable, strategic selection organisation-focused or “elite” interviewing, which is potentially fundamental to possibilities for research on ongoing policy and legislative implementation, could potentially be made impossible in some cases as a result of implementing open access to research data. At the same time, these types of data may be crucial for “whistleblowing” and researching structural or organisational disorders, and such research cannot be undertaken unless interviewees are accurately protected.

It would thereby be of concern if open access policies, in their implementation in a specific legal context, would themselves result in limiting research in areas that themselves foundationally aim to improve transparency and support a democratic society. While this may not be the case in all legislative contexts, nor for all types of data, this article discusses a specific case in a specific legal context where open access taken together with the legislative situation at large may compromise research. The focus is specifically placed on elite interviewing and the conditions for it in Sweden. Sweden is hereby not a typical case, but rather stands out at as a case

Introduction
Open access has moved from the periphery to the mainstream in recent years (Suber, 2012). Originating in broader movements towards open access (e.g., Costa and Leite, 2016), in Europe the open access movement has recently been driven by a focus within the European Commission on making “research data that result from publicly funded research publicly accessible, useable and reusable” (2012/417/EU). The aim is to make research more accessible, transparent, and trustworthy, and of greater benefit to society. The underpinning argument is that the provision of open access to research data will allow anyone to examine whether “the right conclusions are drawn” given the research data used for a particular study.

While the basic logic of this argument is hardly disputable, strategic selection organisation-focused or “elite” interviewing, which is potentially fundamental to possibilities for research on ongoing policy and legislative implementation, could potentially be made impossible in some cases as a result of implementing open access to research data. At the same time, these types of data may be crucial for “whistleblowing” and researching structural or organisational disorders, and such research cannot be undertaken unless interviewees are accurately protected.

It would thereby be of concern if open access policies, in their implementation in a specific legal context, would themselves result in limiting research in areas that themselves foundationally aim to improve transparency and support a democratic society. While this may not be the case in all legislative contexts, nor for all types of data, this article discusses a specific case in a specific legal context where open access taken together with the legislative situation at large may compromise research. The focus is specifically placed on elite interviewing and the conditions for it in Sweden. Sweden is hereby not a typical case, but rather stands out at as a case

1Department of Geography, Umeå University, Umeå, Sweden

Corresponding Author:
E. C. H. Keskitalo, Umeå universitetet Institutionen för geografi, Umeå Campus Area, Umeå 901 87, Sweden.
Email: Carina.Keskitalo@umu.se
where a strong focus in legislative frameworks is already placed on openness and transparency (cf. Griffin & Leibetseder, 2019). The sections below discuss, firstly, the nature of the specific qualitative social science data in focus here, and then the present issues with sufficiently protecting research subjects so as to allow for the research to be ethically conducted, within different national frameworks, as a basis for the discussion of the Swedish case. The final sections discuss the implications and requirements for implementing open access in relation to sensitive qualitative social sciences research data.

The paper builds on a review and summary of the issues seen as impacting the integration of open access to research data in Sweden. These include both the existing legislative situation with a strong focus on transparency, as well as the current implementation of open access to research data in which catologuing of data is coming to be required. As this is a developing area of interpretation, following upon governmental requirements implemented from 2016 onwards, the selection of material for review and summary with regard to the Swedish case has been informed by public meetings on the issue organised by universities, the Swedish National Data Service, the Swedish Research Council, and the Swedish Association of University Teachers and Researchers, contact with university legal services, and documents developed at university services including university instructions to scholars. The material thus includes both legislation and cases referred to, policy documents such as governmental or agency policies, reports and the like, and university documents attempting to clarify the legal situation and its interpretation. The methodology with regard to material for the article is thus akin to tracing a development – in this case, an issue and the implications of it – through different sources (cf. George & Bennett, 2005; Walgrave & Varone, 2008). As the area is still in development, the documentation drawn upon here constitutes the best information available to the author at the time of writing.

The Nature of Qualitative Social Science Data, in Relation to the Specific Focus of Organisational and Institutional Studies

The open access initiative has developed over a long time, with the purpose of making research publicly available (e.g., Costa & Leite, 2016; Bartling & Friesike, 2014 eds.). It has often been seen as a positive development. For instance, Suber (2012) describes open access as being “easy, fast, inexpensive, legal, and beneficial”, and there are plenty of statements in the literature on the potential benefits of open access to research data and secondary analysis of data (e.g., Tsai et al., 2016; Kuula, 2011; Sherif, 2018). The development of open access across Europe in recent years can be seen as linked to the development of regulation at European Commission level. A 2012 Recommendation, for instance, calls on member states to implement the general aim of, for instance, making research data “accessible, useable and reusable”. However, it also opens for variations in application as it is very generally phrased, does not discuss different types of data, and notes among other things the need to take into account intellectual property rights (2012/417/EU, see also European Commission, 2021).

Despite a relatively positive assessment of open data in literature (e.g., Suber, 2012; Tsai et al., 2016; Sherif, 2018), the literature in this field from a social science perspective varies. Some of it targets the potential benefits of open access to research data, and notes that acceptance of open access is mainly a cultural question (e.g., Kuula, 2011). However, it is also social science literature that has been perhaps the most critical of it and most clearly states the prerequisites required for, particularly, qualitative social science data to be made open access (e.g., Kirilova & Karcher, 2017; Chauvette et al., 2019; DuBois et al., 2018). For instance, Elman et al. note that “[q]ualitative data sharing presents a range of challenges that differ both in kind and degree from those involved in quantitative data archiving” (Elman et al., 2010: 24). It is also generally noted that research open access development has been grounded in and mainly targeted other data than qualitative social science data. This potentially also rests on a relatively strong historical emphasis on natural science as “science”, in relation to which a lesser focus has been placed on social science and in particular qualitative studies (see e.g., Snow, 1961; Henderson, 1993; Hollis & Smith, 1991; Worster, 1996 for a discussion of the “two cultures” in research). In addition, the social science literature generally mentions the context dependence, interpretative role, and sometimes also sensitive role of qualitative social sciences data – in short, that it also matters who collected the data and under what circumstances, factors that can be hard to describe in meta code for any anonymised dataset (e.g., Kirilova & Karcher, 2017). Contrary to a focus on the openness of data, it has also been noted that “[d]ata archives worldwide stipulate that personal information concerning research participants is to be kept confidential and that guarantees of confidentiality and anonymity are to be honoured” (Parry & Mauthner, 2004: 143).

In the social sciences, one common method for collecting research data in focus here is the use of in-depth, for instance semi-structured, interviews. Such interviews are typically conducted about a topic that is highly sensitive — or that may even be so current and contextually dependent that there exists no clear written or “objective” data on it. In some cases, it would not even be possible to develop clear written or “objective” data on a topic. Crucially, social science concerns not only “hard facts” or neutral data (as in fact no data is ever neutral) but often also explicitly opinions, conflicts, barriers to development, crisis management, systemic disorders, issues related to power relations, social structures, and how people think and act in relation to these social structures, barriers, and conflicts (e.g., Schön & Rein, 1994; Macnaghten & Urry, 1998). Such research
can be done in multiple areas. For instance, interviewees may be employees at companies or in the government, and may describe issues not in line with company or government policy, such as the organisation in fact not following regulations. Making such statements could be interpreted as the interviewee being disloyal to their employer and could thereby be grounds for their dismissal, for instance (Fransson, 2013). At the same time, it is data that is crucially important to conduct research on, for instance environmental norms or why environmental regulation is not implemented (cf. Eriksen et al., 2015). Accordingly, the data that is sensitive in these ways is just the type of data that many social scientists seek to collect, and can also be crucially important data for research (cf. Wellstead et al., 2013; Wellstead et al., 2014). (See Appendix 1 for a background on how qualitative social sciences data may be collected and treated.)

**Elite Interviewing as a Sensitive Case**

These types of issues, which have started being described above, fall into several specific branches of research: among others, those involving implementation or organisational working and compliance, which may be undertaken to a great extent in organisational and institutional research, such as business or policy studies. Such studies often, but not always, focus on specific organisations and positions in them. Some may even focus on “elite” interviewing, which selects highly specific groups and individuals who may be seen as placed in a vulnerable position because of the power and position they hold. The case of elite interviewing will here be used as one example of a sensitive case, to illustrate the complexity that may exist even beyond for instance cases that may more commonly be seen as sensitive, such as cases in criminological or health research (e.g., Israel, 2004; Gibson et al., 2013).

With regard to elite interviewing, it has been noted that even within social science, “[t]he particularities of conducting policy research in a highly politicised domain, and the vulnerabilities of participants usually conceptualized as ‘elite’ within these spheres, have been underexplored in the qualitative methods literature” (Lancaster, 2017: 101, cf. Littig, 2009 on differences and similarities between expert and elite interviews). While elite research participants are often selected because they are actors with certain power, this power in fact makes them sensitive to power plays by others, should their sensitive information get out. As Lancaster notes, “[w]hile policy processes research may not fit the usual categories of ‘sensitivity’ to which ethics committees are attuned (such as traumatic events, violence, death or drug use), the vulnerabilities of participants … were at times significant due to the personal, professional and political issues in play” (Lancaster, 2017: 101). The privileged and specific nature of participants, e.g., in specific high-level panels or organisations, which cannot be anonymised per se as their nature is a crucial focus of the research, means that such interviews or other types of research data (such as observations) require considerable consideration. Knowledge that sensitive information exist if, for instance, meta data on it is made available may even lead to this information being particularly sought out and requests made that data be shared. Thus, Lancaster notes that “‘sensitivity’ is not an ‘unproblematic or commonsensical’ concept but rather relational and emergent” (Lancaster, 2017: 101).

What is more, in the case of elite interviewing in which the selected groups or panels are not large enough to allow for the successful anonymisation/pseudonymisation of data, it could also be expected that the sensitivity of the data may remain over time. Even if confidentiality is applied broadly to certain material, legal confidentiality may in some cases be reassessed or removed after a certain time (such as 10 years). This is a matter that may so far have been comparatively little discussed in qualitative method, as the focus has often been placed more on best practice (e.g., Bos, 2020), and the variations in legal requirement on data may differ highly between countries (e.g., Corti et al., 2000; Riksarkivet, 2022; OFS, 2009). However, in the increasingly formalised setting of research, also these legal frameworks may come to increasingly play a role. With regard to elite interviewing, people on, for instance, high-level panels or boards may at a later point in their careers be found in even more sensitive positions, acting for instance as national or EU politicians or high-level civil servants or on company boards. Elite interviews, for example in the case of international high-level boards, may thereby constitute highly sensitive material for long periods into the future, with considerable ethical implications for researchers in conducting research, and possibly paramount implications for research participants.

This type of elite interviewing is by nature not necessarily a large research area (and the direct contribution of qualitative research to policy-making is sometimes even discussed, cf. Veltri et al., 2014). It is also quite possible to undertake studies in these areas that do not have the problems highlighted here (for instance where the focus is on lower-level organisations that can be de-personified/pseudonymised by removing areal identifiers, such as local councils). However, despite the potentially limited and numerically relatively small role of these specific types of studies on elite type participants, who cannot be easily anonymised or pseudonymised, they can be of great importance indeed, and even play a role in democratic functioning. Research into for instance the development and implementation (or non-implementation or areas of avoidance) in relation to regulation and policy is crucial for understanding the extent to which, for instance, law is applied in specific cases, and what may hinder this. Research on compliance is crucial for understanding, for example, how different companies, authorities, and courts manage – or do not manage – something like environmental demands. It is thereby not only broader society that may support a “whistleblower” function (cf. e.g., Lewis et al., 2014; Hollings, 2011) but also research (Martin, 2014).

In this way, while journalism has often been seen as a means of transmission (sometimes struggling to be recognised as research or working within academic frameworks, e.g.,...
Following from the previous section, in sensitive elite interviewing – at national, EU, or international levels, but sometimes also at regional levels where it becomes apparent where the case is situated – data may thus both be a particularly high-level and crucial type of research data, but also constitute high-risk data for participants. At the same time and for the same reasons it might be difficult to share this type of research data openly while properly protecting the interviewee.

With regard to protecting data, a set of this type of personal data has come to be increasingly protected in the EU under GDPR legislation and various types of formal causes for confidentiality that exist in many countries. In GDPR regulation as well as in many countries, data on political, ethnic, or sexual belonging, or similar topics, is by nature regarded as falling under specific ethical grounds for consideration, so as to remove requirements for open access from this specific data (in Sweden PUL 1998:204). However, these grounds may not be applicable in all the cases we discuss here, such as in relation to implementation, organisational working, or compliance. Data may also be made confidential due to the existence of sensitive personal data beyond these categories. However, as will be discussed below, these are not sufficient in all cases to entirely remove interview data from open access. To manage data under the requirements both for data sharing and for protecting sensitive personal data, the focus in much open access literature and supporting documents in different countries has come to lie on the means for de-personifying data through the use of pseudonyms, whereby specific direct identifiers are replaced with more general statements that convey the approximate meaning of a term (for instance, stating “city” instead of the city name). (https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation/qualitative.aspx). Risks of indirect identification remain even under this method, as full pseudonymisation would need to remove not only formal identity markers (name, locations, direct identifiers) but also details like dialectal expressions, typical statements, organisational references, references to social structures, and information that in summary throughout a full interview may in some way identify the interviewee or the social context to which he/she belongs (e.g., organisation, company, etc.). Removing these details does thereby not necessarily prevent as much as make “indirect identification” through the summary information provided in an interview more difficult – particularly if the selection of interviewees already targets a smaller identified group. The issue is compounded by that what is formally regarded as “research data” may include not only transcripts from interviews or recordings, but also sound files or any other recordings which may be considerably more difficult to fully de-personify.

How far the protection of interviewees under these circumstances can go, then, seems to differ according to country. Examples in the literature from Canada highlight the possibility to select only some interviews for open access in order to limit risks of identification: for instance in a study including interviews in several areas, only those that were conducted in one area were made open access (after de-personification/pseudonymisation in the interview as well) (cf. Chauvette et al., 2019). As another workaround, research agencies in different countries, such as the UK, advise that one consider such concerns already when designing the study: Perhaps sensitive issues that may pinpoint an interviewee can be divided into several interviews, or delimited in topic (https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation/qualitative.aspx). However, this type of procedure may result in issues that may bias an interview, or that may result in the interviewer not being able to freely choose the topic of study to best match concerns. While qualitative interviewing typically strives for rapport and for the interviewee to speak freely (e.g., Obelené, 2009), dividing a larger relevant area into sub-topics, with separate interviews held for each one, might itself impact the openness of discussion or make the interviews less relevant to the larger topic.

Another possibility is that the research, due to its sensitive nature and potential implications for interviewees, may in its entirety be classed as confidential, in which case data will then on principle not be shared. However, it may differ between countries to which extent the researchers themselves are able to themselves make the decision on what data to share or not. The conditions under which confidentiality may be broken may also vary to a large extent between countries and specific cases (e.g., Israel, 2004).

The Swedish Case

Sweden may stand out in this respect, as one of the countries with the strongest requirements for open access to information at large, as all data that is received at an authority, as a rule, is considered public (cf. Prop. 1994/95:19, p. 461). Under Swedish law, the digital record of, for instance, an interview received at the university as a state body falls under public transparency requirements. These requirements are the same
at a public research body as at any other public body (e.g., Borglund & Engvall, 2014), and would include any sound or text file that is received as part of an employee’s (in this case, the researcher’s) work. It is thus not possible to ensure participants that only anonymised data will be shared, or to at all guarantee that specific ethical requirements (such as not sharing specific data) will be maintained. This is because confidentiality is not only assessed based on the research itself, but also on who is requesting to see or use it, and for what purpose (e.g., OSL 24 ch. 8 § third para). This means that it will be very difficult for the individual researcher at a public body – a large part of research – to assess how well any conducted research and the research subjects therein are protected, as well as what protection they can be guaranteed.4

The general requirements for transparency are thereby so far-reaching that interviewing in sensitive cases may mainly have been possible due to the fact that interview data has seldom (if ever) been centrally stored, labelled, or made available for easy identification and access. These practical and informal delimitations to the far-reaching formal transparency – and researchers’ limited knowledge of these formal requirements – may thereby have meant that sensitive interviewing could only be undertaken as long as the practical delimitations meant that it was seldom easy to identify or request interview data. As open access is presently implemented as part of a government bill (prop. 2016/17: 50) and related assignments to Swedish authorities and institutions, concerns for research may in particular result from the interaction between existing Swedish regulation and broader open access requirements.

The Swedish policy framework that impacts open access is thus not only open access-specific but rather include general principles that are gaining increasing impact through a general implementation of open access. Thus, these types of requirements have long been present in Sweden with regard to public institutions, but as research data previously has commonly not been made subject to procedures for public cataloguing or available to the open access movement and its institutionalisation through EU requirements, potential issues regarding interviewee protection and transparency have perhaps most often not been recognised, making this type of research possible to undertake nevertheless (but see Swedish National Agency for Higher Education, 2005). Restrictions to public availability in the Swedish case mainly center on the specific grounds for confidentiality outlined above (such as political and ethnic belonging), general confidentiality requirements (e.g., OSL 24 ch. 8 § third para), and now also taking into account GDPR grounds for confidentiality. However, it has also been noted that GDPR in Sweden does not necessarily prevent data from being made public, or even allow the person from which the data originated to retract it (Swedish legislation decision 2018-10-02 dnr 327/18, decision 2019-02-13 in court case 5437-18). As Swedish legislation, together with for instance GDPR requirements, constitutes the legal situation against which assessments are made, the “right to be forgotten” in GDPR is not always applicable, to the extent that data cannot be fully retracted (Swedish legislation decision 2018-10-02 dnr 327/18, decision 2019-02-13 in court case 5437-18). This will have effects not least on the right of interviewees to retract statements, for instance if an interviewee would wish to do so late in the process after realising any potential impact of their statements. Thus, as with all these issues, there are significant impacts on the rights of interviewees – the actual original “data owners”.

This situation also implies that the role of restrictions to data that are stated by researchers at a public institution in any data management plan they develop is delimited. While a data management plan is to include statements by the researcher on the sensitivity of the data and could, for instance, include a decision on the researcher’s part to not make the data open, upon any request data availability is in Sweden assessed first under the law rather than in relation to researcher statements. Assessment in this situation is made not necessarily by the researcher but by those assigned by the university to assess data availability and public access; that is, persons who are typically not active in the particular field of the research and may thereby not necessarily be able to assess what may be sensitive data in the particular situation and case (e.g., with regard to level of de-personification required to hinder interviewee identification). The transparency requirements in Sweden have already led to legal assessments that sensitive data must be shared, contrary to the researcher’s ethical assessment of it (cf. Swedish National Agency for Higher Education, 2005). In one sensitive court case in which open access was requested for interviews, the researcher even destroyed the data rather than share it (e.g., Swedish National Agency for Higher Education, 2005, for an overview of this development and concerns related to it).

Thus, it has broadly been concluded that a researcher cannot guarantee confidentiality of data in Sweden (e.g., Swedish Research Agency, 2017).

If a request for data is received, the university will assess whether it falls under the legal requirements for confidentiality (general or specific) and is sufficiently protected personal data. Depending on assessment, the university may make data or specific parts of data available, which in the case of sensitive interviewing in limited groups could allow for the indirect identification of interviewees. Even if data is placed under confidentiality requirements, such requirements may be lifted after specific time periods (OFS, 2009:400). For elite interviews, however, as has been noted above, data may not necessarily be less sensitive for interview participants far in the future (as they may potentially at a later time hold even higher positions, thus being more politically vulnerable). This may imply potential risks to interviewees that an ethically minded researcher working on sensitive areas (for instance, implementation or legislation issues relevant to elite interviewing) cannot expose research subjects to. It also means that it is not possible for the researcher to know beforehand exactly
how the data use may be delimited, given that some assessments are only made at the point at which data is requested.

As a result, the general discussion upon which international literature on responsible qualitative data sharing has focused (for instance responsible anonymisation, e.g., Chauvette et al., 2019, https://www.ukdataservice.ac.uk/manage-data/legal-ethical/anonymisation/qualitative.aspx) is not sufficient for data at public institutions in Sweden, as it is not possible to know which data – including original non-anonymised or non-depersonified files – may be judged as legally obliged to share with whom, and when. This also means that in Sweden you cannot know exactly how data will be treated in the future. At the very least, you cannot promise your research participants confidentiality.

The present requirements for making information available about what data exists as well as for depositing data (archiving) in common university-based repositories thus fundamentally mean that the researcher cannot limit the access to sensitive data. As a result, some research of the type we have discussed here may be both ethically and practically (obtaining agreement from participants) impossible to undertake, as no guarantees can be made to participants.

Discussion and Conclusion

In relation to the specific nature of the data discussed here, the question remains as to whether it is at all ethically possible to undertake specific research dependent on sensitive data under the legal and open access requirements in some contexts. In the Swedish case, the answer for some data may be “no”: however, the implementation of open access is still under development and a crucial question may be to what extent the requirements or social sciences sensitive data is taken into account in this implementation.

Crucial research in qualitative social science include studies that may come the closest to revealing why we are not taking action on environmental, organisational, or other change in institutions, and how this could be incentivised. However, if the integrity of the research subject cannot be protected, social science researchers will likely not be able to collect this kind of research data in the first place, and accordingly social science research will be limited to research topics that do not take up issues like power relations, conflicts, organisational structures, relations, and structural and systemic inequalities in elite bodies; i.e., those that may be the most important to policy research and implementation. Ultimately, this raises the question of whether initiative towards transparency and open access, despite all good intentions, is in this case in fact limiting rather than extending possibilities for research. In the Swedish case, open access implementation without consideration to pre-existing legal requirements may risk limiting crucial research in qualitative social science and thereby even make the research it intends to make publicly available impossible to conduct in the first place.

Thus, in order to successfully implement open access while avoiding inadvertently limiting qualitative social science, there are a set of issues that need to be resolved. While the discussion here focuses on the Swedish case, it is, as literature has illustrated, also crucial in other countries and contexts to assure that “ethical considerations are of paramount importance in the archiving of qualitative research data” and that research can be undertaken even in cases where “some qualitative research material may be intrinsically impossible to archive” (Corti et al., 2000, sections 7 and 5, respectively). The Swedish case regarding research at public institutions can thereby illustrate – potentially in an atypical or even extreme way – the concerns around qualitative data that have been highlighted in literature (e.g., Wessels et al., 2014; Elman et al., 2010; Kirilova & Karcher, 2017), and the need to consider the nature of qualitative social sciences research data and the specific requirements of this data in relation to open access.

This manuscript has noted the ways in which qualitative data has perhaps not been the primary target of regulations regarding open access to research data, and that considerations specifically involving qualitative data, particularly in the types of cases discussed here, must be given more attention (e.g., Chauvette et al., 2019; Wessels et al., 2014). The quality of research that is enabled by a specific research design must be primary to the considerations that open access places on issues: such as revising the study design, dividing up interviews, or even not undertaking certain studies. This means that the researcher must be the one who determines open access to sensitive data, and hold intellectual property rights to the data so that they can limit its use or availability in different cases.5 This means that the researcher, not the university or other body where the research is undertaken, must be able to act so as to, on an ethical basis, limit access or use in cases in which this is clearly motivated. Thus, the importance of according a role to the researcher – and in the Swedish case, potentially treating research differently than other documents created or maintained at public bodies – may be paramount. It does not challenge open access to research data as a general principle – its benefit has been noted widely – but recognises the types of data that have so far been peripheral to discussions of open access in relation to research regulation and policy (cf. Class et al., 2021).

Following this, legislation and policy need to correctly enable and take account of research data generated in the social sciences. This might require review of existing legislation in consideration of sensitive social science data. In Sweden, this is in line with the requirement that researchers in Sweden, in accordance with the Swedish Higher Education Act, should be able to independently choose their problem and method (Högskolelag (1992:1434) t.o.m. SFS 2021:1282 1 kap. 6 §). This might mean, for instance, heeding the calls that have already been made at university level to review issues in national context (Swedish National Agency for Higher Education, 2005). In Sweden, one way to attend to the...
issue could relate to exempting data gathered at Swedish public institutions for the purposes of research from general transparency requirements, in line with what seems to be the case for data in several other countries (e.g., Corti et al., 2000). In addition, the EU should clarify potential considerations in relation to its existing open access regulation and develop a discussion of different types of research data, taking care to monitor implementation so that EU regulation does not in fact create more problems than it solves – making data less rather than more possible to use or even collect.

Appendix 1

A Background on Qualitative Social Science Data Gathering Methods and Data Treatment, with a Focus on Strategic Selection in Organisational Studies

In the social sciences, qualitative data from in-depth semi-structured interviews is typically made up of a core recording, typically a voice recording – often of about an hour, with the researcher asking questions that are often relatively open in that they allow for a broad range of interviewee responses and do not bias or predetermine what they might say. A desirable condition for the interview is rapport; that is, that the interviewee feels it is possible to speak freely on the topic of the interview (Obelené, 2009). In organisational or institutional studies interviewees are often selected strategically; for instance selecting those who work with a specific topic in a range of organisations and are thereby able to speak on this topic. This may even identify specific positions in different companies or agencies, meaning that a more limited selection of people is available already from the start, and that the risk of identification of interviewees is greater. The focus in strategic selection is thus on interviewees being able to express their understanding of, for instance, how they have worked with something and what has impacted the work, based on their specific position in the organisation. The researcher often asks questions based on predetermined themes within this purview, sometimes pre-phrased so that the researcher will not inadvertently provide a value statement that interviewees simply respond to (rather than expressing their own opinion and understanding) (e.g., Kvale, 1996).

This type of approach, particularly if the interviewee feels able to speak freely, may result in interview statements that cannot be foreseen. Statements may include for instance complaints about bosses, the organisation, implementation, or other statements that can be interpreted as disloyal to the employer. These are statements which if taken out of context may be seen as defamatory (criticism of people at another agency, in the government, or the like), or, if the interviewee is identified, could even result in them losing their job (Fransson, 2013). Examples of this type of sensitivity can be seen in the following statements about the process by qualitative researchers:

“Through … [a] narrative about a particular case, general criteria of decisionmaking in recruitment become as clear as it becomes apparent how criteria of decision-making in an organisation, in connection with gender-related structures of prejudice, bring about a practice which is typical and not unfamiliar. While narrating his story the interviewee becomes aware that he is being recorded and by and by is tangled up in a narrative the unfolding and outcome of which he could not clearly foresee from the start. In doing so he reveals more about the relevance structures underlying his behaviour than he would actually do – and would be able to do – if he were asked directly. It can be presumed that the knowledge coming to light is located at the border of practical and discursive consciousness” (Meuser & Nagel, 2009: 33)

Another example of the often inherent sensitivity of qualitative interviewing is provided by Obelené:

“I collected plenty ‘off-the-record’ or ‘not for press’ episodes during the interview. However, there was also plenty of apparently sensitive information without such markers involved. It is obvious that when at first the researcher struggles to recruit respondents amid the risk they may decline the sensitive topic, now the researcher faces decisions how to go about the sensitive data collected, possibly by wrongly assessing which issues are sensitive. The guidelines the experts give may be vague. What should, for example, the phrase ‘I trust you’ pronounced recurrently throughout the interview entail for what to do with data?” (Obelené, 2009: 195)

These examples illustrate how qualitative interviews can both lead to unexpected statements – which may then be crucial not to spread, as the interviewee might not have intended them, for instance when signing consent. This situation may also require a position of trust between interviewer and interviewee in order to at all discuss sensitive issues, meaning that not all the material can be shared. This results in the notion that “consent should weigh heavily on the researcher, who must take exceptional care in analysis, editing, production, and distribution” (Banister, 2007: 4).

Typically, these considerations have been managed through processes for dealing with detailed, sometimes unforeseen and multi-interpretable, data. Data has generally not been conceived of under an edict of data sharing but rather under one in which the assumption has been that qualitative interviews will not be shared. Thus, the focus has been placed on developing an interview situation which supports portraying as much detail as possible, in order to then be able to accurately portray the data in relation to the research topics in focus (while the detailed data itself, including potential sensitivities, is generally not made available). For these reasons, the recording of material is typically transcribed verbatim, so that the researcher can check that they did not inadvertently say something that may have biased the interviewee and thus impacted the results (sections, for instance whole topic areas, that may in that case then be omitted from use).
As the intent is for different interviewees in different organisations to be able to highlight, for instance, what has impacted their implementation of the same legislation or policy, the factors interviewees highlight and the point in an interview at which they do so may thus differ substantially. Their individual perspectives and the way they express them may also be so specific as to pinpoint an individual to others in the field, should data be shared.

However, if the interviewees are of a wide enough selection (organisations of different types or other different prerequisites) or even part of a full study (all organisations working with a specific topic in a region or country), the factors they discuss will together provide a clear understanding of the factors that impact work on the topic, such as the implementation – or non-implementation – of an organisational change, legislation, or policy (for instance, environmental policy or the like) (cf. e.g., Flyvbjerg, 2006; Ruddin, 2006). This will make it possible to generalise the research to theory, for instance regarding what factors impact implementation. This is a type of generalisation that differs from that in quantitative studies, but is paramount in qualitative studies.

In order to identify these different factors and the conditions under which they are important in the results from research and in research publications, the researcher thereby typically codes the interview transcript: identifying themes, for instance inductively based on interviewees’ statements or deductively in relation to theoretical assumptions in the area of research (e.g., policy implementation, environmental action), ascertaining that no content is lost in the process. Result descriptions and analyses thereby tend to include (anonymised or de-personified) quotations typical of a theme, so that the reader can check that the analysis and understanding conveyed by the researcher, e.g., of a theme, seem to reproduce those directly expressed by the interviewees, as visible in specific reproduced segments (where sensitive information and potential individual identifiers have been removed). As a rule, the researcher thus takes care to omit any sensitive statements of the type discussed earlier – instead highlighting the factors of action or organisation that resulted in interviewee criticism of the organisation, employer, or the like – and ascertains that no information is provided in their result description that can identify the interviewee, even indirectly (e.g., Lune & Berg, 2017).

Thus, strategic selection in organisational or institutional studies with a focus on, for instance, implementation results in both recording and transcription data that may provide high detail and that could harm the interviewee if that data would be shared and s/he is identified. Given the detail and context available in the data, this type of research may also run a larger risk of its subjects being identified than other studies do.

For these reasons, it also may not be possible to entirely de-personify data without the loss of the role and context of the research (Mauthner et al., 1998). Researcher knowledge may be crucial in contextualising data and assuring that statements in it are not overinterpreted or understood in an incorrect way. As Mauthner et al. note:

“Our experiences of revisiting our own data, discussed in this paper, have highlighted the central part played by the researcher not only in the interpretation and theorising of data, but also in their construction. Because of this, we contend that the conditions under which data are produced are inescapable, thus rendering their re-interpretation at some later date problematic” (Mauthner et al., 1998: 733).7

This type of researcher knowledge may also include knowledge of what questions might for instance have biased an interviewee, resulting in that material following these needing to be omitted from use.

In general, qualitative studies with this type of design may therefore need to be acknowledged as specific cases, rather than as falling under general – natural science-developed – open access foci.

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) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

E. C. H. Keskitalo https://orcid.org/0000-0003-3244-7272

Notes

1. Similarly, Wessels et al. note that the “perceived benefits rest on an approach that sees data in very broad terms. This approach does not fully consider the specificity of different types of data and research practice in relation to making data openly available” (Wessels et al., 2014: 52).

2. As noted above, for cases other than those discussed here, in which the smaller group identification is a result of the selection of a specific locality where the research is done, it may be possible to de-identify the location, making it potentially any village or city rather than a specific one. This type of data, by nature of the possibility to de-identify it through the removal of direct personal data (as not only permitted but required under GDPR), may be far less sensitive.

3. Regarding, for instance, the implementation of specific legislation or policy, with concerns about this implementation possibly themselves being broad-ranging and relevant to cover in one interview rather than attempting to divide them into several smaller blocks or parts of studies.

4. In addition, if interviews are undertaken at EU level but by a Swedish researcher at a public institution (for instance a university) and are thereby seen as received there, they are presumably subject to the same regulation that any Swedish data gathering falls under.
This would imply that also interviews undertaken outside Sweden by researchers at Swedish public institutions can result in risks to interviewing in sensitive areas in this respect.

5. While copyright is not applied to data that does not include the types of design and co-construction that qualitative data does, it can be argued that qualitative data of the type discussed here should fall under intellectual property protection as well. In a review of legislation, Guibault and Wiebe note: “To be eligible for copyright protection, collections of data, tables and compilations must therefore show a sufficient degree of originality in their selection and/or arrangement. Whether collections of scientific research data meet the criterion of originality is a question to be determined on a case-by-case basis” (Guibault et al., 2013: 21). In addition to acknowledging copyright, it is also crucial to acknowledge the moral rights that go with intellectual property protection: As Parry and Mauthner note, copyright expires while moral rights (for instance, the right to prohibit the distortion of one’s work) remain. As they note, these moral rights reside with the interviewees and researchers rather than, for instance, the institution where the research is undertaken (Parry & Mauthner, 2004, cf. Mauthner et al., 1998).

6. In addition, issues of respondent consent are “problematic because there are multiple and unknown uses to which an archived data set may be put. As such, it is beyond the power of the primary researchers to guarantee that future outputs from the data will be acceptable to the respondents” (Parry & Mauthner, 2004: 146–147).

7. As Mauthner et al. continue: “Thus, we argue that, while ‘reflexivity’ is now recognised as a key feature of qualitative research across social science disciplines, the implications of this for the archiving and reuse of qualitative data have yet to be adequately addressed. Failure to acknowledge this epistemological issue leads researchers unwittingly to adopt a ‘naïve realist’ position” (Mauthner et al., 1998: 733).

References

2012/417/EU European Commission (2012) Commission recommendation of 17 July 2012 on access to and preservation of scientific information (2012/417/EU) official Journal of the European union L 194/39 21.7.2012 (Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32012H0417&rid=1).

Bacon, W. (2006). Journalism as research? Australian Journalism Review, 28(2), 147–157. https://doi.org/10.3316/ielapa.20070367

Banister, S. (2007). Ethical issues and qualitative methods in the 21st century: How can digital technologies be embraced in the research community? Journal of Ethnographic & Qualitative Research, 1(1), 1–10.

Bartling, S., & Friesike, S. (2014, eds.). Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing. Springer.

Borglund, E., & Engvall, T. (2014). Open data? Data, information, document or record? Records Management Journal, 1(3), 163–180. https://doi.org/10.1108/RMJ-01-2014-0012

Bos, J. (2020). Research ethics for students in the social sciences. Springer.

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

Class, B., de Bruyne, M., Wuillemin, C., Donzé, D., & Claivaz, J. B. (2021). Towards open science for the qualitative researcher: From a positivist to an open interpretation. International Journal of Qualitative Methods 20: 1–15. https://doi.org/10.1177/16094069211034641

Corti, L., Day, A., & Backhouse, G. (2000). Confidentiality and informed consent: Issues for consideration in the preservation of and provision of access to qualitative data archives. Forum Qualitative Sozialforschung/Forum: Qualitative Social Research 1(3): Art. 7.

Costa, M. P. D., & Leite, F. C. L. (2016). Open access in the world and Latin America: A review since the budapest open access initiative. TransInformação, 28(1), 33–46. https://doi.org/10.1590/2318-08892016002800003

DuBois, J. M., Strait, M., & Walsh, H. (2018). Is it time to share qualitative research data? Qualitative Psychology, 5(3): 380. https://doi.org/10.1037/cqp0000076

Elman, C., Kapiszewski, D., & Vinuela, L. (2010). Qualitative data archiving: Rewards and challenges. PS: Political Science and Politics, 43(1), 23–27. https://doi.org/10.1017/S1049096509990771

Eriksen, S. H., Nightingale, A. J., & Eakin, H. (2015) Reframing adaptation: The political nature of climate change adaptation. Global Environmental Change 35: 523–533. https://doi.org/10.1016/j.gloenvcha.2015.09.014

European Commission (2021) Open access. Available at: https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-access_en (Accessed March 3, 2021).

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. Qualitative inquiry, 12(2), 219–245. https://doi.org/10.1177/1077800405284363

Fransson, S. (2013) Tittandefrihet och whistleblowing - Om gränserna för anställdas kritikrätt. Premiss förlag.

George, A., & Bennett, A. (2005) Case studies and theory development in the social sciences. MIT Press.

Gibson, S., Benson, O., & Brand, S. L. (2013). Talking about suicide: Confidentiality and anonymity in qualitative research. Nursing Ethics, 20(1), 18–29. https://doi.org/10.1177/0969733012452684

Griffin, G., & Leibetseder, D. (2019). “Only applies to research conducted in Sweden ...”: Dilemmas in gaining ethics approval in transnational qualitative research. International Journal of Qualitative Methods 18: 1–10. https://doi.org/10.1177/1609406919869444

Guibault, L., Wiebe, A., Dietrich, N., Guibault, L., Margoni, T., Siewicz, K., Spindler, G., & Wieb, A.) (2013). Safe to be open. Study on the protection of research data and recommendations for access and usage. Universitätsverlag Göttingen.

Henderson, D. K. (1993). Interpretation and explanation in the social sciences. SUNY Press.

Högskolelag (1992). Högskolelag 1992: 1434 t.o.m. SFS, 2021, 1282. https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-författningssamling/hogskolelag-19921434_sfs-1992-1434.
