

# Beyond Replication: Secondary Qualitative Data Analysis in Political Science

Comparative Political Studies  
2023, Vol. 0(0) 1–33  
© The Author(s) 2023



Article reuse guidelines:  
[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/00104140221139388  
[journals.sagepub.com/home/cps](https://journals.sagepub.com/home/cps)



Florian G Kern<sup>1</sup>  and Katariina Mustasilta<sup>2</sup> 

## Abstract

Shared qualitative data – such as interview or focus group transcripts – can be used for secondary qualitative data analysis (SQDA). Yet, much archived qualitative data remains unused after primary analysis. Applications and guidance on how to employ SQDA are rare. We use an example application of SQDA studying informal institutions and resilience in Sub-Saharan Africa to show: First, SQDA depends on how primary researchers share ‘raw’ qualitative data and additional documentation to understand primary context. Second, deductive and inductive uses of SQDA require varying engagement with primary data. Third, current practices of participant consent often do not consider potential SQDA. Fourth, SQDA is not less time-consuming than primary data research but offers different benefits, such as expanding the comparative sample of cases or avoiding research fatigue of studied communities. Going forward, SQDA requires greater consensus on the instruments (e.g. transcripts and participant consent forms) used by researchers and further applications of hypothesis-testing and hypothesis-generating designs.

## Keywords

secondary qualitative data analysis, qualitative research, transparency, resilience, Sub-Saharan Africa

---

<sup>1</sup>University of Essex, Colchester, UK

<sup>2</sup>Finnish Institute of International Affairs, Helsinki, Finland

## Corresponding Author:

Florian G Kern, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK.

Email: [fkern@essex.ac.uk](mailto:fkern@essex.ac.uk)

## Introduction

In recent years, the social sciences have experienced calls for greater transparency and more replicable and reliable research. In political science, new principles of Data Access and Research Transparency (DA-RT) have been incorporated in the ethics guides of the American Political Science Association (APSA, 2012, 2022). Sharing data for replication is an established convention in quantitative political science (King, 1995). Yet, for qualitative approaches, the debate on whether and how to practice transparency and data sharing is on-going (see Bütthe et al., 2015; Elman & Kapiszewski, 2014b; Kapiszewski & Karcher, 2021; Lupia and Elman, 2014), best illustrated by the Qualitative Transparency Deliberations (QTD) sponsored by the APSA section for Qualitative and Multi-Method Research (Jacobs et al., 2021).

Beyond increasing transparency and enabling replication, making data available has additional potential: rather than attempting to replicate an original finding, the shared primary data (e.g. interview and focus group transcripts, field notes, archival texts and audio recordings) can be used by fellow scholars to answer different research questions through secondary qualitative data analysis (SQDA) (Bütthe & Jacobs, 2015, 56; Elman et al., 2010, 2013; Lupia & Elman, 2014, 22). This is different from the standard definition of replication, that is, ‘that sufficient information exists with which to understand, evaluate, and a build upon a prior work if a third party could could replicate the results without any additional information from the author’ (King, 1995, 44). The latter aims to either reproduce a result using the same data and same question (a standard in observational quantitative social science) or to replicate to see if a result holds with the same design but new data (an increasingly common practice in experimental social science). SQDA, instead, is about asking new questions with existing, archived qualitative data. Yet, for a variety of reasons, many scholars remain hesitant about sharing their qualitative data.<sup>1</sup>

Even when archived, the majority of qualitative data remains unused for SQDA. There is little guidance and applications in political science showcasing how to prepare qualitative data for secondary analysis and how to run a re-analysis (with some notable exceptions in more interdisciplinary fields, especially Hughes & Tarrant, 2019; Irwin et al., 2012; Watkins, 2022). This is unfortunate, as, in principle, SQDA may have much potential to deepen our understanding of various social issues by using readily available data. For instance, assume one researcher has archived interview transcripts from a study on electoral candidates’ intrinsic motivation of running for office in Uganda. A second researcher could study candidates’ assessment of voter preferences by basing their investigation (partially or fully) on a secondary analysis of the shared primary interview transcripts. However, if the only option of the secondary analyst in the above example is to collect new, primary

data, she possibly faces the problem of research fatigue in the study population. This is particularly pertinent when research focuses on vulnerable populations for which scholars are mindful to not overstretch participants' time spent providing information. SQDA can also enable researchers with scarce resources (e.g. in the Global South), who face severe hurdles when trying to conduct original data collection, to analyse transcripts or field notes relevant to their research problem. Moreover, SQDA could provide avenues of research where original data collection is difficult (as is the case for many research environments during the COVID-19 pandemic). Ultimately, SQDA is much akin to what qualitative research has always relied on: the interpretation and analysis of primary sources (say, presidential speeches or archival texts) – but in SQDA these sources have been recorded or compiled by another scholar. While SQDA is not a replacement of primary data collection in political science, both approaches could be used in conjunction.<sup>2</sup>

In this paper, we ask: How can political scientists employ archived qualitative data for secondary data analysis? What information is needed from the original studies for this purpose beyond the raw qualitative materials? We show, first, that SQDA largely depends on how primary researchers share their 'raw' qualitative data, that is, interview and focus group transcripts, as well as additional documentation to understand primary context; second, that the deductive and inductive uses of SQDA require a different engagement of the secondary analyst with the primary data; third, that current ethics practices of participant consent have to be re-considered to incorporate the consent for further analysis of archived data; fourth, that SQDA is not less time-consuming than primary data collection and analysis, but offers different benefits, such as, among other things, expanding the comparative sample of cases for the analysis or avoiding research fatigue of repeatedly studied communities.

To illustrate SQDA, and given our own substantive research expertise in informal institutions (Goist & Kern, 2018; Holzinger et al., 2016, 2020; Mustasilta, 2019, 2021), we use an exemplary research project with the research question: how do informal institutions affect community resilience in Sub-Saharan Africa? We select data from studies which have previously shared their transcripts from qualitative interviews and focus group discussions on the UK Data Archive (UKDA), overall consisting of 992 interview transcripts and 151 focus group transcripts.<sup>3</sup> In order to demonstrate challenges and opportunities when working on a new research problem compared to the primary study, we have formulated our research question prior to probing for data availability and then chose studies and data that do not explicitly focus on our selected research question.<sup>4</sup> This broad focus allows us to better avoid framing our research question solely on data availability and to open our sample of data collections up to those from various disciplines, given that resilience is an interdisciplinary topic of study. In principle, secondary

analysis is not limited to qualitative inference, but also applied quantitative data and mixed methods (Watkins, 2022). The re-analysis of qualitative data sources can be attempted with methods of qualitative inference (e.g. process tracing and discourse analysis) or computational techniques of text analysis. However, we focus our attention solely on qualitative approaches to SQDA.

## Understanding Secondary Qualitative Data Analysis

### What is SQDA?

SQDA refers to the ‘re-use of pre-existing qualitative data derived from previous research studies’ (Heaton, 2008, 34). The following process constitutes the precondition of SQDA: a study yields original data in a shareable format (e.g. paper, digital or audio), which a researcher archives accompanied by documentation (e.g. manuscripts, transcripts, code books and field notes). A different scholar can use the shared data for an analysis separate from the primary study. While an established practice in quantitative data analysis, SQDA remains rare in political science. Ethical questions concerning the sharing of sensitive data as well as doubts regarding the transferability across research projects of highly contextual qualitative material remain challenges in the practice of SQDA (Bishop, 2007, 2014; Coltart et al., 2013; Heaton, 2008; Long-Sutehall et al., 2011). However, the development of data archives (see e.g. the UKDA and its Qualibank, the Harvard Dataverse or the Qualitative Data Repository at Syracuse University) that store and enable access to qualitative primary data has sparked increasing interest in the reuse of qualitative data.

SQDA should be differentiated from systematic reviews of qualitative research that aggregate findings across state-of-the-art research. In SQDA, the focus is on using primary data of previous research to answer substantive research questions (Bishop, 2007; Coltart et al., 2013; Heaton, 2008). SQDA can aim at verifying or testing the findings of the primary analysis, usually referred to as replication. This form of SQDA remains controversial as the interpretative nature of qualitative inference is considered to be at odds with verification (Heaton, 2008), though Becker (2020) has recently presented a thoughtful teaching simulation using qualitative replication. The focus of our research here is to use SQDA in order to study *new* and *distinct* research questions (Lupia & Elman, 2014, 22). For example, in her study on convenience food and choice, Bishop (2007) used two datasets from her own earlier research to study a research question that had not been under investigation when the data was first collected. Our study expands Bishop’s objectives. Our aim is to explore the use of pre-existing qualitative data from a range of previous research projects we have not been involved in, to study a research question different from the original studies.

## Challenges of SQDA and Contribution

Qualitative research has the advantage of allowing researchers to examine the multitude of meanings, perceptions and preferences attached to different social processes in an iterative process in which the scholar updates findings based on new information. Qualitative research requires particular attention to the cultural, political and local context in which the qualitative information is collected (Coltart et al., 2013; Long-Sutehall et al., 2011). A related challenge is the positionality of the primary researcher in that context: What was the relationship between the primary analysts and human subjects? What information is not included in the primary data, and why? These questions of re-contextualisation of the primary analysis relate to the problem of 'not having been there' (Heaton, 2008). Some scholars remain sceptical of the idea of reuse of qualitative data produced by primary researchers even if the secondary analysis is carried out by the same researchers that originally collected the primary data (Coltart et al., 2013; Mauthner et al., 1998), though if qualitative data is really only comprehensible to the primary researcher, then even co-authors of a primary study who do not enter the field may be unable to fully co-produce joint analyses.

However, others view SQDA as a promising methodological approach as long as it is conscious of the above challenges (Bishop, 2014; Bishop & Kuula-Luumi, 2017; Fielding, 2004; Heaton, 2008; Long-Sutehall et al., 2011). Scholars emphasise the similarities between primary and secondary analysis of qualitative data. Fielding (2004, 99) notes:

'Qualitative researchers have always been in the position of having to weigh the evidence, and often have to deal with incomplete information or speculate about what may have happened if a researcher had not been there. The difficulty is not, therefore, epistemological but practical'.

Scholars agree that the more detailed information can be derived with regard to the context of the primary data collection the better the quality of the secondary analysis will be. How primary researchers store the original data and information on the context can significantly facilitate or complicate SQDA.

Ethical issues abound when using qualitative data collected on the basis of participant consent in the primary research project (Bishop, 2014; Heaton, 2008). Consenting to participate in the original research project does not automatically mean that the data generated can be used in a secondary analysis. Reflecting these challenges, Bishop (2007) recognises five inter-related steps in carrying out SQDA: understanding the context, defining the subject area, finding data and sampling, later sampling and topic refinement and handling of transcripts. At its best, SQDA provides insights into patterns

in the data originally not considered by primary analysts (Mason, 2007). In addition, our application suggests that using primary data for SQDA can at times allow secondary analysts to access relatively unbiased data in regard to their separate research question.

Our project contributes to the on-going methodological discussion on transparency (Anderson, 2013; Elman & Kapiszewski, 2014a, 2013; Gelman, 2013; Humphreys et al., 2013; Lupia & Elman, 2014; Miguel et al., 2014; Moravcsik et al., 2013) by exploring SQDA and its purpose in political science, where it has rarely been applied. Most of SQDA work has been carried out in sociology, focusing on topics such as health, nursing and family, often using data from only one or two previous projects (Bishop & Kuula-Luumi, 2017). Moreover, our substantive focus on a topic of comparative political science reveals important additional challenges and opportunities of SQDA so far overlooked. The size and the nature of our data sample differs from previous applications of SQDA. We use data from nine research projects that have collected interview and focus group transcripts from various countries, including 992 interview and 151 focus group transcripts that touch upon the themes of informal institutions and/or community resilience but have not been collected with the aim of exploring these topics comparatively in the first place. We study how the heterogeneity of the data influences the application of SQDA. We see the comparative focus both as a challenge (how does one compare data that have been collected in multiple, different contexts) and as an opportunity (SQDA allows using in-depth data from different contexts for comparison).

## Applying SQDA

To investigate how SQDA could be applied in political science, we provide an illustrative application in the comparative study of local politics: how informal institutions shape community resilience in Sub-Saharan Africa. We choose this research problem based on our own research expertise, but also to capture interdisciplinary data across the social sciences. We do not attempt to offer a comprehensive analysis of this research problem, given that our objective is a methodological contribution, and we thus cannot do justice to the in-depth analysis necessary to answer this research question. Focusing on a specific research question and formulating testable hypotheses allows us to exemplify the various challenges and opportunities of SQDA in comparative political science. In the following, we provide a brief theoretical discussion of the example project, before outlining the key methodological motivations and steps behind the application of SQDA.

### *Secondary Question: Informal Institutions and Community Resilience*

*Research Motivation and Research Question.* Resilience – ‘the ability of a system, community or society exposed to hazards to resist, absorb,

accommodate, adapt to, transform and recover from the effects of a hazard in a timely and efficient manner, including through the preservation and restoration of its essential basic structures and functions through risk management.’ (United Nations Office for Disaster Risk Reduction definition) – is pivotal for development. Scholars have frequently focused on how formal institutions, such as governments, administrations and laws, influence resilience. Yet, as Jones (2011, 1) notes, many analytical frameworks that focus on formal institutions have little applicability at the community level where ‘the majority of adaptation action will inevitably occur’. Community resilience is also a function of informal institutions, that is, ‘social norms; customary laws and codes of conduct’ that do not derive from the state’s codified system (Mazzucato & Niemeijer, 2002, 172). These informal institutions can be instrumental for community resilience as they mediate access to resources and shape local coping and adaptive strategies (Jones, 2011; Jordan, 2015). We know surprisingly little about how different types of informal institutions influence resilience in varying contexts. Therefore, below, our example research project asks: how do informal institutions shape community resilience? Regionally, our project focuses on Sub-Saharan Africa, as resilience is particularly vulnerable there (Bell & Keys, 2016).<sup>5</sup>

*Definitions and Theoretical Expectations.* Existing research identifies two dimensions of resilience (Adger et al., 2007; Jordan, 2015). First, *reactive* resilience refers to a community’s capacity to cope with and withstand the current hazardous situations it faces. Reactive resilience captures the strategies and practices, such as food sharing during droughts, that communities employ to continue functioning despite a hazardous situation. *Proactive* resilience, in contrast, refers to a community’s capacity to adapt and change its vulnerability to future shocks by developing practices and structures (e.g. diversifying livelihoods) that decrease the negative effects of future hazards.

We distinguish two types of informal institutions: first, *ordered* informal institutions, which include customary authority and governance structures (e.g. chiefs or councils of elders), non-codified customary law and norms, as well as kinship links and family ties. Ordered informal institutions are characterised by systematic structures with clearly defined roles and relations of individuals, for example, the chief and the subjects, or the father and the children. Second, *unordered* informal institutions, such as women’s and youth networks or savings groups, are more loosely organised with less hierarchical or structured roles and relations between individuals. Based on these definitions, we formulate two propositions to explore in our application.

**Proposition 1.** Ordered informal institutions are more effective in strengthening reactive resilience than proactive resilience.

Informal institutions that are more structured, such as customary governance institutions, create accountability mechanisms and social cohesion. They channel interactions during crises and carry institutional memory about previous emergency relief, contributing to communities' reactive resilience. However, rather than also contributing to proactive resilience, ordered informal institutions may hinder the flexibility necessary to develop proactive resilience when threats are evolving rapidly. Proactive capacity can be restrained by cultural practices limiting available strategies to use in adaptation.

**Proposition 2.** Unordered informal institutions are more effective in strengthening proactive resilience than reactive resilience.

In contrast, unordered informal institutions, such as women's networks or local land management groups, can be more effective and flexible in contributing to proactive, adaptive resilience and in decreasing future vulnerability. Proactive resilience requires more active changes in the norms and practices in place. Informal institutions with less rigid ordering principles can help individuals alter norms or habits that hinder their adaptive capacities. However, unordered informal institutions may lack the resources and accountability mechanisms to promote reactive resilience.

*Research Design.* To test the two hypotheses, our illustrative research project requires an empirical strategy able to capture and analyse two key concepts – informal institutions and resilience – varying along the values of ordered versus unordered institutions and proactive and reactive resilience.

Our theoretical and comparative approach is methodologically challenging as it requires rich local-level (e.g. village and community) data with considerable variation. One option to proceed would be to focus on a qualitative within-case comparison of ordered and unordered institutions in a location with enough over time variation to observe how different informal institutions – say chieftaincy structures and communal resource sharing practices – form and connect to reactive and proactive sides of resilience.

Considering a quantitative approach is challenging, as informal institutions escape large-n and cross-sectional quantified measurement almost by definition, as their presence is highly context-specific and rarely codified in systematic and formal ways. Whilst some cross-sectional data projects include indicators of the most common ordered informal institutions (e.g. traditional authorities), no such data sources exist for unordered informal institutions. Finding cross-sectional databases capturing how communities withstand crisis situations and prepare for them in advance is equally challenging. The nature of the research problem and the theoretical propositions therefore push us to identify and analyse nuanced qualitative data that is able to identify and distinguish locally specific informal institutions and the processes linking



these to community resilience. Yet, focusing solely on one such context (even with over time variation) would necessarily restrict the breadth (e.g. external validity) of the study, precisely because informal institutions are highly context-specific. Hence, ideally, while focusing on collecting and analysing context-specific qualitative data, the project would cover more than one community or case. The case-selection process would also need to consider the implications of variance not only in the main concepts of interest but also regarding the various other contextual factors influencing a community's resilience and the institutions on the ground.

*Data.* The conventional approach would be to collect new primary data 'from the field'. A structured focused comparison of two most-similar cases, for example, could be conducted based on data collected through semi-structured interviews and/or focus group discussions run by the analyst team (or their local partners) in one or two Sub-Saharan African countries and selected localities. A plausible case-selection strategy in this scenario would be to focus on typical cases of local communities and governance conditions, given the little comparative research available and the desire to start from representative cases. However, this approach would (1) require significant financial resources and time; (2) risk a 'research fatigue' of the communities and individuals being studied, as the safety and accessibility factors regarding primary data collection tend to make certain research sites more often visited than others; and (3) given the primary data collection expenses, the study would most likely need to settle for a small number of case studies, thus lowering the breadth of the project.

Alternatively, the analysts could consider whether suitable primary data already exists, for example, in-depth interviews and other qualitative materials, that include information on the two key concepts and their relationship. Whilst there is little cross-sectional data focusing comparatively on informal institutions and resilience, topics falling under the phenomena, such as food security, land use, climate and environment or governance provisions, have gained considerable research attention also at a local level, particularly in qualitative research. Compiling data on such studies from multiple countries and contexts could offer both wider empirical breadth and a cost-effective data collection strategy. Given the challenges related to primary data collection around this research question and the understudied nature of the topic, a secondary analysis of existing data from multiple sources could provide a credible first step in understanding the relationship under scrutiny.

Therefore, rather than collecting original data on the link between informal institutions and community resilience, the project's methodological approach employs SQDA as a first choice and analyses archived qualitative data from various available studies. Our main source of data for the secondary analysis is the UKDA which stores data of quantitative and qualitative studies.<sup>6</sup> We

restrict our focus on transcripts of qualitative interviews and focus group discussions as they are the most frequently available shared qualitative types of data in the UKDA. These data are also particularly suited to illustrate re-coding and re-analysis in SQDA given their text format.

As of August 2018, the catalogue stores 7348 data collections for all data types. To discern studies having shared interview and focus group transcripts of relevance to the question of how informal institutions shape community resilience, we need to visit the catalogue's search tools and engage in a form of purposeful sampling (Palinkas et al., 2015). Suspecting that studies focusing explicitly on the role of informal institutions may be rare, we decide to cast a wide net for data identification. We refine our catalogue search as follows: (1) only displaying {Data type: Qualitative and mixed methods data} (1074 data collections); (2) only displaying data collections for which the assigned {Country} is in Sub-Saharan Africa or for which the pre-defined UKDA {Country} descriptor contains a set of countries such as {Africa}. For some countries, such as Chad or the Central African Republic, no descriptor exists, which suggests no data for these countries is available in the UKDA. For others, more than one data collection is available, for example, Kenya (11) or South Africa (19). Refining the data search with these two steps leaves 83 data collections.<sup>7</sup> As we discuss further in the next section, we also try out other keywords and search techniques. However, we decide to keep the original search terms substantively vague so as to avoid choosing our data on pre-defined understanding of the nature of the dependent variable, resilience.

We now delve into each of the 83 data collections with a view to (a) whether their descriptions broadly apply to the topic of informal institutions and community resilience; (b) whether the data provided are transcripts of qualitative interviews and focus group discussions and (c) whether the transcripts are actually available on the UK Data Archive, as for some studies the links are not working or data access requires further authentication – which we assume disqualifies the data for reuse. Regarding the first criterion, we include studies that have collected data on a village and/or communal level on topics related to public goods and basic services or local livelihoods and resources. Using this time intensive process, we identify nine pertinent studies (see Table 1). All of these studies' archived data provide information on the link between the latent concepts of unordered/ordered informal institutions and reactive/proactive community resilience. We decide to include all these data projects in our analysis.

As a result, we have a dataset of 992 interview transcripts and 151 focus group transcripts, altogether from 12 Sub-Saharan African countries. The data covers West Africa, the Horn of Africa, Central Africa and Southern Africa, therefore giving us a heterogeneous data sample including each sub-region. However, this sample still represents a minority of Sub-Saharan African countries and some countries, such as Kenya and Malawi that are relatively

**Table 1.** Available Data Sources (All Information According to UK Data Archive).

Data creator	Project title	Type and extent of data
Ansell, 2010	Averting 'New Variant Famine' in Southern Africa: building food-secure livelihoods with AIDS-affected young people	142 interview transcripts
Baker, 2008a	Evaluating of the effectiveness of police local partnership boards in Sierra Leone	11 interview transcripts, 4 focus group transcripts
Baker, 2008b	Multi-Choice Policing Resources for Post-Conflict Situations: Rwanda and Liberia, 2006-2007	355 interview transcripts, 25 focus group transcripts
Copestake, 2016	Impact assessment in complex contexts of rural livelihood transformations in Africa. Part 2 - Interview data	136 interview transcripts, 32 focus group transcripts
Edwards-Jones, 2010	Comparative Assessment of Environmental Community and Nutritional Impact of Consuming Fruit and Vegetables Produced Locally and Overseas	64 interview transcripts, 7 focus group transcripts
Hall, 2017	Land and agricultural commercialisation in Africa	110 interview transcripts
Hossain et al., 2015	Food riots and food rights: The moral and political economy of accountability for hunger	43 focus group transcripts
Heald, 2015	Making rural law in Kenya	11 interview transcripts
Urban, 2017	China goes global dams project	163 interview transcripts, 40 focus group transcripts
Overall		992 interview transcripts, 151 focus group transcripts

peaceful and stable, are over-represented, which could bias our analysis down the line if very vulnerable, conflict-affected countries (Chad) are not included. Most importantly, whilst not a random sample, we have not actively produced this sample ourselves either. Rather, the primary analysts' specific research priorities and their access to specific research sites have facilitated this particular data sample for us. This is both a challenge and an opportunity, as discussed below. Naturally, purposefully selected data (based on the used keywords and data availability in the UKDA) from 12 countries and specific subnational research areas in these countries does not constitute a representative sample of the whole continent and caution must be practiced when

generalising the findings. Yet, relative to other plausible methodologies, SQDA provides key advantages. Firstly, accessing data on 12 countries and hundreds of interview transcripts on a variety of topics related to informal institutions and resilience provides significantly more comparative breadth than a within-case or small-n case study. Second, SQDA enables maintenance of depth and internal credibility of the data better than reliance on existing cross-sectional datasets that do not capture the topic of interest here. Using SQDA provides a first step in accumulating comparative knowledge.

Two caveats about data source selection merit mentioning here. First, we select only from those studies and data collections for which information has been archived. Given that the norm of data sharing is not established yet in qualitative data, there must be a majority of studies relevant to the topic of informal institutions and resilience for which we cannot access the data, which may mean that important insights escape our secondary analysis. However, we would argue that it is unlikely that there is a substantive bias as to why some studies are archived and others are not, beyond those mentioned above, that is, the most fragile countries are generally understudied. Some of the 'archive bias' is likely due to funding requirements in some locations to store data. Moreover, there could be a methodological bias, in that it is easier to share transcripts than to share, for example, field notes. Second, the primary researchers have selected their study populations and there may be an inherent, unobserved bias in the selected studies, meaning unobserved variation or information relevant to our new research question. This is a severe problem and difficult to solve. The secondary analyst should make sure to control for how subjects and study sites were selected in each study and see whether imbalances appear to bias the overall sample. Yet, conscious of these biases, we still believe that the sheer number of transcripts we have access to across contexts allows for learning about our research interest.

*Operationalisation and Measurement.* A conceptual framework based on the two theoretical propositions outlined above and the distinctions between (1) ordered and unordered informal institutions and (2) reactive and proactive community resilience forms the basis for our analysis. Informal institutions are operationalised as micro-level perceptions and descriptions stemming from the interview and focus group data concerning communally shared rules, practices and decision-making structures and goods that are not primarily imposed by the state authorities. To discern between ordered and unordered institutions, we focus on whether the institution in question builds around centralised decision-making structure and top-down imposing of rules and practices. We operationalise resilience as micro-level perceptions of the community's ability to minimise negative implications arising from empirically perceived shocks (reactive resilience) and perceptions of the ability to decrease future threats to the community's well-being. As the primary projects

focus on a range of topics, for the re-analysis and measurement of our variables and their connection, we assume that the available transcript data will be re-coded and re-analysed using computer assisted qualitative data analysis software (CAQDAS, e.g. NVivo). There is a need for re-coding as all data projects differ and crucially are not specifically designed for our study.<sup>8</sup>

## Challenges

*Hypothesis-Testing and Hypothesis-Generating.* Before deciding to conduct a SQDA, it is important to clarify whether answering the research question is even feasible using SQDA. It is inevitable to opt for the collection of new data if no or only insufficient secondary data is available. Furthermore, the researcher conducting the SQDA should determine whether she is following a more exploratory, hypothesis-generating approach or pursuing a hypothesis-testing research design. Both are feasible using SQDA, but each requires different theoretical preparations and iterative procedures, and challenges of SQDA will differ across qualitative approaches. To illustrate, with a hypothesis-generating approach using a small number of secondary data sources that are rich in their content may be sufficient to formulate theoretical propositions. When the aim is hypothesis-testing, relying on only one or two data collections can prove difficult unless they speak directly to the specific topic. Even then, the external validity of the study may remain low if the SQDA relies entirely on data from a specific data collection covering a carefully defined set of cases. In the case of our example study, the two theoretical propositions make it clear that we opted for an approach akin to the hypothesis-testing design and a deductive strategy to SQDA, and thus, challenges and opportunities described here focus on those related to hypothesis-testing. The context-specific and latent nature of the main variables meant that more than one data source would be needed to test our comparative hypotheses and enable external validity.

SQDA can differ significantly depending on the type of data to be re-analysed. For instance, field notes and transcripts that paraphrase the responses of participants provide information that is more edited than verbatim transcripts of open-ended and semi-structured interviews. The latter may be a better fit for deductive SQDA, while those data with more direct authorship of the primary researcher may be fruitful sources for inductive research. In a project such as our illustrative example in which the secondary analysts aim at capturing highly context-specific and latent phenomena, heavily edited data (such as field notes) present particular challenges as they already constitute interpretations (that of the primary analyst) of the phenomena to be (re-)coded. Conversely, verbatim transcripts of interviews and/or focus group discussions from multiple primary projects can provide a rich dataset for secondary analysts' hypothesis testing.

*Data availability and Data Structure.* To run a SQDA, shared data, such as in our case verbatim transcripts of interviews and focus group discussions touching upon informal institutions and community resilience, must be available (Elman & Kapiszewski, 2014b). Yet, to date, no broad consensus exists on whether and how these types of data should be shared. Should transcripts be stored in their raw version, that is, simply containing the verbatim text? If transcripts have been coded using computer assisted qualitative data analysis software, should the coded transcript be archived? What type of documentation and metadata should be provided? The challenge here goes beyond the mere availability of data. The data also have to be informative for the specific research question to be studied.

As clarified in the research design section, in order to establish the availability of pertinent data, we consulted the catalogue of the UKDA where researchers have stored their qualitative data of research projects with relevance to informal institutions and community resilience. While we were able to find a number of studies speaking to our research puzzle, several challenges pertained to the data identification and collection process.

First of all, the identification of data for our research project was complicated by the fact that the descriptors of the UKDA are not consistently applied across archived data projects. For instance, refining the search by including only {Data format: Text} reduces the number of data collections from 83 to 55. However, manually going through the excluded cases reveals that this filtering in fact makes the researcher miss some studies that have stored interview and focus group transcripts (but that have not used the descriptors accordingly). Some of the descriptors are assigned based on the individual study itself and not as general categories. For instance, data by Ansell (2010) has the country descriptor {Malawi and Lesotho}, but refining results by selecting the individual descriptors {Malawi} or {Lesotho} will not include the data collection in the search results. In our example case, we decide against refining further by searching for the word {resilience} in the 83 collections (which is possible), as we would like to include studies not explicitly concerned with resilience. Such casting of a wide net when searching through the existing data archives may be recommendable and necessary also more generally due to the unsystematic use of the descriptors. Hence, the secondary analyst will likely need to spend time in manually going through the potential pool of data projects, as in our case.

Broadly speaking, if possible in the archive they use, secondary analysts should discriminate their search by data type and format, geographic region and sub-regional units (if applicable) and possibly employ search terms that are conceptually reflecting the independent variables of the hypothesis testing design. In our case, this would mean to further revising the search by unpacking the concepts of ordered and unordered institutions. However, the analyst may also risk to reduce the sample of data collection this way,

excluding studies that would speak to the explanations, but do not specifically appear as relevant to the identified explanations.

The lack of agreed-upon principles for transcription becomes apparent in Figure 1. Some data are stored in Word document format – some of which are colour coded (Hossain et al., 2015) while others are not (Edwards-Jones, 2010); some transcripts are saved in Excel files (Copestake, 2016), others in NVivo format (Ansell, 2010, not displayed). Demonstrating these differences is not a criticism directed at the authors, but rather highlights the challenging variance in data structure that scholars conducting SQDA need to engage with. Data collections also differ significantly in the number and length of transcripts shared.

**Lack of Contextual and Topical Knowledge.** A particular strength of qualitative approaches is their sensitivity to context and nuance reflected in the data. Qualitative research produces detailed insights into the particularities of the subject studied. This is a further challenge for SQDA, as researchers may have to develop at least minimal contextual knowledge and topical expertise for a variety of original studies (Arjona et al., 2019; Shesterinina et al., 2019). Table 1 highlights this dilemma: while the example project focuses on the relationship between informal institutions and community resilience, none of

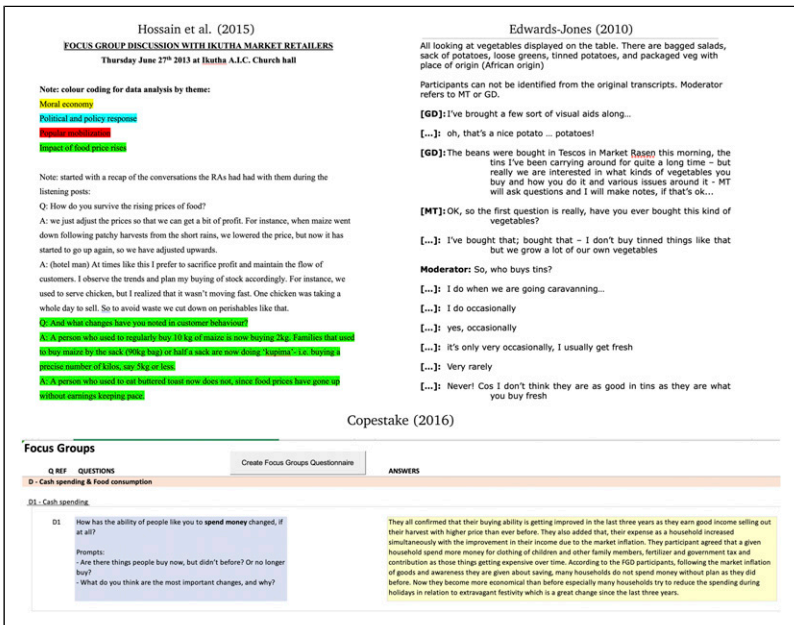


Figure 1. Variance in focus group discussion transcripts.

the identified studies primarily focuses on this topic. Topics range from food security, to environmental consequences of dams or policing in countries as diverse as Ghana, Kenya, Ethiopia, Malawi, Lesotho, Liberia, Rwanda or Sierra Leone. The researcher conducting the SQDA is unlikely to be an expert in all of these issues.

To understand the context in each study, the researcher will read through the studies produced with the data. Aside from familiarising the SQDA researcher with the contexts of the primary data, this step builds an understanding on how the primary analysts have conceptualised and operationalised their research objects and what type of substantive insights their project has focused on. This generates necessary metadata for the secondary analyst to be able to use and possibly re-code the data at hand. In principle, secondary analysts can also try to contact authors of original studies to further understand context. Yet, if these authors are unavailable (or deceased), secondary analysts face similar challenges as to those scholars working with archival data in terms of reconstructing context. That said, many of the stored data collections also include additional documentation. For instance, [Baker \(2008a\)](#) provides a study guide with context descriptions on policing in Sierra Leone. [Ansell \(2010\)](#) includes further documentation accompanying the study such as stakeholder leaflets handed out during the research containing contextual information. Using this information, the secondary researcher can familiarise herself with some of the context and topical knowledge necessary to understand each study's data before engaging in re-coding and re-analysis.

Beyond the necessity of familiarising oneself carefully with the original data sources, the secondary analyst will need to engage in standard background research to understand each context of the original data. As with any research project and method, using background sources to understand the relevant political, socio-economic and, for example, environmental context is vital to contextualise the raw data. The secondary researcher analysing data from different time points in addition to the data stemming from different contexts needs to be particularly attentive to time-variant aspects in the data. In general, the researcher's background and level of expertise in the studied area and context influences the extensiveness of the background research needed.

*Researcher Positionality and Bias.* Assuming that the scholar pursuing SQDA is not re-analysing her own data, she was not present in the field site. Thus, she lacks the knowledge of the relationship between the researcher and the subjects and cannot fully reproduce the background and positionality of the scholars conducting the original study ([Shesterinina et al., 2019, 4](#)). Possible biases of the scholars running the original study, which skew the latter's analysis, are difficult to track for the purposes of SQDA. Before the re-analysis the secondary analyst must clarify a variety of aspects that may lead to bias, for example, the gender of the scholar and subjects, differences in nationality or



racial identities of researchers, subjects and translators or funding sources of the original study.

Regarding the nine data collections identified above, it is quite challenging to clarify positionality and bias for the entire sample. There is much variance in how much information is provided on the background of researchers and participants, possibly depending on the sensitivity of each study's topic and context. Exemplary applications are offered, for example, by Baker (2008a,b) who provides a list of interviews containing information on the country and location of the interview, the date of interview, the gender of the participant(s) for interviews and focus groups or the organisation the interviewee works for.<sup>9</sup> This information offers a first glimpse into positionality and bias. The scholar conducting the SQDA may spend considerable time delving into each identified project's documentation, data and publications and even seek out one of the original study's principal investigators to discuss these issues for greater insight.

The severity of the challenge regarding the positionality and biases of the original data collectors at least partially depends on the secondary researcher's questions and aims. If the secondary analyst's research question closely approximates that of the primary researcher, there may also be a higher risk of skewed secondary analysis based on the positionality of the original researcher (if this is not appropriately uncovered and dealt with). On the other hand, asking different research questions and therefore focusing on different aspects of the data than the primary analyst somewhat decreases the risk of systematic biases. Similarly, SQDA might be particularly challenging to use for study projects that aim at interpreting subjective emotions or reactions based on the raw data, particularly if there is limited metadata on the data generating processes and situations.

In our example case, we ask a research question that is different from any of the original studies and deals with understanding of meso-level institutional dynamics that we can assume do not change rapidly or depend on the data collection moment. We find this to lower the risk that the contextualities, positionalities and biases of the original data would be reproduced systematically in our research. Moreover, we ask a question that does not lead us to study aspects in the data that would be very sensitive to the positionality of the primary researcher and the original data collection context, such as subjective emotions.

Notably, the way the data are originally transcribed and coded reflects a primary analyst's positionality and bias. The data from Copestake (2016) illustrate this. In the same dataset, some of the interviews are transcribed using the interviewee's voice (e.g. 'I used to do this but now I have changed into...') while other interviews have been archived using a narrative voice from the interviewer (e.g. 'He said he used to do this but has now changed into...').

From a secondary analyst's perspective, the former appears less influenced by the interviewer than the latter.

*Ethical Concerns.* Human subject research requires informed consent from participants. In the identification of the data projects for our project, we included only those sharing information about consent. Some of the projects included even provide further information, such as [Hossain et al. \(2015\)](#), who include a methodological guide discussing ethical concerns related to the study and how these have been addressed. Ethical questions regarding the confidentiality and anonymity of the responses are discussed clearly in seven out of nine projects. Yet, not all consent forms in the identified sample of primary data specify clearly the possibility of re-analysis of the anonymised transcripts. [Ansell \(2010\)](#), for example, states that 'We might wish to use quotations from the interview in materials (including presentations and publications) produced from the research. Please tell us if you would prefer us not to do so. We will not, however, use your name'. This could refer to data sharing, but not necessarily includes the possibility of secondary analysis. [Copestake \(2016\)](#) includes the following statements in the consent form:

All information gathered is confidential and *will be used only for the research*. Your name will not be shared with anyone and not revealed in any reports we produce as a result of this study. [...] But *we do intend to feedback some of our findings to you and to others* working to promote farming and food security here and in other places like it. (Emphasis added)

Hence, the participants consent to their data being used solely for the purposes of the original study. Yet, participants also consent to the anonymised transcripts being shared for further consultation. Finally, the consent form of [Edwards-Jones \(2010\)](#) states clearly that data will be shared: 'The information I supply will normally be preserved in the UK Data Archive at the University of Essex and will be kept confidential unless I give permission for my name to be used'. Out of the nine projects, this is the only one that unequivocally requests consent for archiving of the data in a public archive where it can be used for research purposes.

The ethical question looms on whether consenting to the production of anonymised data and its archiving also means that data can be reused for different purposes. If consent forms are not explicit about the possibility of re-analysis, one may wonder if and why transcripts of interviews and focus groups should be shared at all. While there is no consensus on how consent forms should incorporate SQDA, the Qualitative Data Repository team has put forward templates to adhere to standards of informed consent and archiving of data.<sup>10</sup>

## Opportunities

*Secondary Value.* SQDA constitutes a cost-effective complement or alternative to primary data collection. As [Bishop and Kuula-Luumi \(2017\)](#) note, it is desirable to reuse data already collected in order to maximise the returns of the committed investments. Applying SQDA helps to maximise the use of carefully collected primary qualitative data beyond the original analysis. Collecting primary data is both time-consuming and expensive, from identifying participants to organising the researchers' data collection process and scheduling this in line with the participants, possible fieldwork travels and finding qualified assistants. SQDA allows for scholars with a lack of resources (e.g. in the Global South) to investigate research questions based on primary data without the related costs. In turn, scholars in the Global South can share their own qualitative data for SQDA, providing these marginalised researchers a way to gain recognition and reputation as primary producers of data.

In addition to SQDA lowering direct costs for researchers, primary collection of qualitative information through fieldwork and direct interaction with human subjects often presents real risks for the participants (e.g. privacy or even safety issues, see [Arjona et al., 2019](#), 12) especially when studying vulnerable populations. Moreover, high opportunity costs and research fatigue of participants can be an issue if researchers repeatedly query the same population (e.g. MPs). Using SQDA can reduce these costs for participants.

Our substantive research interest exemplifies these opportunities related to SQDA. The datasets that our analysis relies on draw empirical evidence from multiple African countries and from various subnational regions, societal levels and political and cultural contexts. Given the substantive topic and comparative objective, collecting our own primary data would be very expensive. Moreover, primary data collection raises a set of crucial ethical and practical questions. How would another data collection project intervening in the lives of often vulnerable individuals and groups (given our topic) contribute to knowledge production and the lives of the individuals participating in the research? How would we choose the empirical focus countries and contexts in such a manner that would enable comparative research? How would we identify and organise translators in these multiple contexts and guarantee an in-depth understanding of the contexts across countries and localities? With these questions in mind one may find SQDA a well-suited approach for our substantive interest.

Within the framework of SQDA, we have been able to access a sample of datasets focusing on Sub-Saharan Africa, which, while not specifically concerned with community resilience, all touch upon institutions that influence coping and adaptive strategies of communities vulnerable to environmental and political crises. For instance, [Hall \(2017\)](#) collected both data on changing land management practices in Kenya, Zambia and Ghana. The

archived data consist of 55 transcripts from Kenya and additional transcripts from interview and focus group discussions in Ghana and Zambia. Overall, the nine projects consist of 1143 transcripts. The heterogeneity of these data can be essential and useful for our aim of conducting comparative research. Moreover, these data collections also provide in-depth, micro-level and contextually valid data that is crucial in order to understand the topic we are interested in. For example, the data from [Hossain et al. \(2015\)](#) contain information on the practice of burning and selling charcoal in order to cope with food crises in Kenya. The practice appears to constitute an unordered institution among the locals that supports coping resilience even though it is recognised to be harmful in the long-term. Through SQDA, we reuse data that have been carefully collected and archived by scholars with an in-depth understanding of that specific context. Maximising the use of these rich data sources allows us to approach a substantive topic from a perspective that otherwise would have been difficult to conduct without sufficient resources.

*More Substantive Insights.* Besides maximising the use of primary qualitative research and data, we find SQDA particularly effective to produce in-depth, context-valid comparative research to generate new substantial insights. Our hypothetical research project is located in the field of African politics, governance and development that is studied with various methodologies. Comparative, large-n research has examined macro-level institutions and variables to explain variation in outcomes such as economic development and state stability. When disaggregated to subnational levels, the bulk of this research has focused on quantifiable measures, such as local state capacity, demographic variables and natural resources to explain local variation of societal outcomes ([Raleigh & Hegre, 2009](#); [Tollefsen et al., 2012](#)). Yet, there is an increasing awareness of the shortcomings of such observational studies. Focusing on externally valid indicators, many crucial topics, political agents, groups and areas have been overlooked (due to difficulties in accessing quantifiable data). One response to this shortcoming is provided by the growth of experimental studies that have explored local interactions and political dynamics in specific micro-level contexts (e.g. [Fearon et al., 2015](#); [Humphreys & Weinstein, 2009](#); [Goist & Kern, 2018](#)). Qualitative studies focusing on specific countries or areas have long contributed to a more nuanced understanding of the complexity of political processes, development and stability. However, as the focus of these studies has usually been limited to a small number of cases, extrapolating more general inferences is difficult.

SQDA can bring added-value in comparing political dynamics across contexts while still retaining a more nuanced understanding of the local specificities. Indeed, similarly to the benefits of SQDA in contexts where information is hard to collect because of its sensitivity, our example study on informal institutions and community resilience in Africa demonstrates that

SQDA can enlarge the comparative scope of a study without losing context-specific information of each individual case.

Overall, as secondary analysts with background in researching local institutions in Sub-Saharan Africa, going through the nine primary data collections offered significant new substantive insights on informal institutions that we had not previously studied ourselves, related to communal policing, food security and livelihoods changes. Notably, despite the stated general challenges regarding data availability and documentation, our background in the broader field and reading through the primary studies, transcripts and metadata provided us with insights to understand the data and their context to some extent. In such a case, the potential for new substantive insights is noteworthy. Consider the example of two separate data sources: [Hossain et al. \(2015\)](#) examine local-level responses to food crises in the contexts of Mozambique and Kenya. [Copestake \(2016\)](#) focuses on assessing the livelihood transformation in Malawi and Ethiopia. Both studies collect rich qualitative interviews with constituents in vulnerable contexts and generate data that would in practice be difficult to gather from distance or in such (quantitative) form that could be readily compared. Despite the different contexts, both data collections include ample transcripts that discuss the themes of adverse changes in the physical and/or political environment and the local responses to these changes. An initial look at the data suggest intriguing similarities between different types of informal institutions (ordered vs. unordered) and how these appear to constitute reactive and proactive resilience in these different contexts.

[Table A1](#) (see [Appendix](#)) illustrates how SQDA can be used to analyse a general substantive question using data from very different original research projects. The table contains extracts from the transcript data of the different studies used in our example study, re-coded along the theoretical framework. Though a simple exercise based on a surface review of the data, [Table A1](#) already suggests similarities across the different data in terms of what type of resilience is more frequently referred to (i.e. reactive resilience) and what the effects of each type of informal institutions are. Expanding the table to a full-fledged analysis using SQDA of the nine selected data collections – and further unpacking each statement – could generate substantive new comparative insights.

*Secondary Analyst Independence.* A third opportunity to explore with regard to the application of SQDA has to do with the secondary analysts' relatively independent position from the original data generating process. We previously discussed the positionality of the primary researcher as one of the challenges in SQDA. However, the review of the data that we have identified for our example study suggests that the independence between the primary research space and the secondary analysis can constitute an opportunity to use data that

has relative little bias in regard to the topic of interest. Going back to our examples of [Hossain et al. \(2015\)](#) and [Copestake \(2016\)](#), neither of the research projects originally focused on informal institutions and community resilience. However, the interview and focus groups transcripts include ample material of local informal institutions influencing capacities to cope or adapt to a changing environment. These topics have been brought up rather ad hoc by the participants in the focus groups discussions and interviews. The data of interest to us can thus be regarded as relatively unprompted and unbiased by the primary analysts with respect to the secondary research question.

Moreover, a secondary analyst with less involvement in one research context, but broader knowledge of data across cases, is maybe less prone to emphasise certain data points based on contextual expectations. The distance between the immediate data collection context and the secondary researcher can allow the secondary analyst to approach the data with different priors and expectations.

## **Observations: What to Expect from SQDA**

Having outlined challenges and opportunities of SQDA using qualitative tools of analysis, we present four concluding observations: First, given the described biases intrinsic to primary data collection and its original purpose, coding and analysis, we find that SQDA is most effective if the primary researchers deposit raw, anonymised qualitative data in the voice of the human subjects on a data repository. This would mean storing verbatim, typed transcripts of qualitative interviews and focus group discussions redacted for sensitive or identifiable information. Encouraging the sharing of raw data constitutes the most straightforward way to enable SQDA, instead of, for example, trying to standardise coding rules. The transcripts should ideally reflect (a) which participant made which statement and (b) to which item in interview questionnaire responses refer. In case a more formal questionnaire did not exist, a separate protocol of questions of interest could be provided. Access to the raw data in the voice of human participants enables SQDA researchers to re-code and re-analyse while minimising the ‘noise’ of the original study.

Second, as we have argued above, SQDA can follow inductive or deductive scientific reasoning. Each approach has different consequences for the conduct of SQDA. For instance, a deductive study such as the illustrated project may require more data in order to generate valid inference. Primary transcript or data collections will offer varying amounts of information on the subject of the secondary analysis. In fact, it may be the triangulation of different data sources that allows for new insights for the purpose of the new question. Thus, deductive SQDA may require the researcher to collect various sources, which may be additionally impeded by current norms of qualitative

data sharing. An inductive SQDA may be based on fewer sources since the content of these sources feeds more directly into the theories being formulated. Here, the secondary scholar can move iteratively between the data sources and the secondary research project, re-formulating the questions and propositions based on the data sources. Researchers should of course be aware of the dangers of basing their new theories on a small amount of primary data. Then again, even a small number of transcripts can be helpful to identify an interesting research puzzle.<sup>11</sup> For the inductive approach, the secondary researcher may also want to acquaint herself with the background of the primary scholar and project in order to identify biases stemming from the previous researchers' inference and the study's context.

Third, participants' consent remains central to SQDA, and this requires actions from both the primary researcher and the secondary analyst. As is clear from our application, consent forms are not consistently designed to allow SQDA. In our view, simply sharing anonymised data for which consent has been given is insufficient to conduct large-scale, valid SQDA. A primary researcher wanting to archive their data *and* make it available for SQDA should explicitly state on the consent form that a non-identifiable version of the information provided by participants will be stored in a data repository and possibly used by other researchers for the analysis of a related, but different research problem. We are aware that this addition to a consent form may increase the likelihood of more vulnerable participants to reject participation. Yet, this is the purpose of consent, that is, to allow subjects to consider the risks of participation. One solution to this problem may be to make it optional for subjects to allow for their anonymised data to be used for SQDA. Political scientists could also consider ways to incorporate dynamic consent (as discussed in the natural sciences, see [Kaye et al., 2015](#)). Bar greater consensus on how the possibility of SQDA is incorporated in consent forms, a secondary researcher should check the provided consent forms of original projects carefully when identifying primary data. Even once this consent is provided, secondary researchers should make sure that the data they use is non-identifiable. There may in fact be the need for more discussion in the social sciences on the ethics of consent for secondary analysis. In our view, the current state of affairs is unsatisfactory in this respect.

Fourth, while SQDA may be less expensive or generate greater value of existing data, it is not less time-consuming to pursue than primary research. Secondary analysts should not expect quick returns. The secondary researcher needs to identify available data sources and develop an understanding of the primary data, including the content of the data itself and how it was collected. For the example project on informal institutions and community resilience using the list of projects in [Table 1](#), this would take a long research-intensive period. After this, developing an overarching analytical framework based on which to subsequently re-code and re-analyse the primary data may take

considerable time. Following SQDA, the secondary analyst may also archive the data used for secondary analysis together with the secondary documentation. The latter can range from information on where to find the original data, new datasets compiled for the secondary analysis, or code books and further information detailing how the new, secondary inference was generated. This would allow further SQDA of the now aggregated, raw data. Additional materials can in principle allow for replication of the SQDA in the future. In turn, for the primary researcher, producing, for example, complete transcripts of a series of qualitative interviews come with significant costs in time and money.<sup>12</sup> The possible payoff, that is, further attention to the original study and additional citations, will present itself with delay. Yet, SQDA may offer benefits that are different from original data collection, for example, the greater value generated from primary data, the comparative potential of accessing sources from a larger number of different cases and the time saved for research subjects and participants in volatile contexts.

## **Conclusion**

We discuss the challenges and opportunities related to secondary qualitative data analysis in political science. We do so by using an applied illustration of an example study on how informal institutions shape community resilience in Sub-Saharan Africa. Based on the application, we show, first, how SQDA largely depends on how primary researchers share their ‘raw’ qualitative data, that is, interview and focus group transcripts in our example study, as well as the additional documentation to understand primary context; second, deductive and inductive uses of SQDA require a different engagement of the secondary analyst with the primary’s study data – while more data is needed for deductive SQDA, in-depth contextual knowledge may be particularly important for inductive SQDA; third, for SQDA to be more widely used, current ethics practices of participant consent have to be re-considered to incorporate the consent for further analysis of archived data; fourth, SQDA is not less time-consuming than primary data collection and analysis, but offers different benefits, such as expanding the comparative sample of cases for the analysis or avoiding research fatigue of repeatedly studied communities.

We argue that many of the challenges of SQDA are due to a lack of consensus regarding the design of instruments used in qualitative research, rather than the inference produced. For SQDA, especially when following deductive reasoning such as in the illustrated example, the format of transcripts or participants’ consent forms is at least as important as the inference presented by primary researchers. Moreover, as we have only focused on the use of interview and focus group transcripts here, it remains to be seen how amenable other types of qualitative data (audio recording, field notes and diaries) are to SQDA. The variety of qualitative approaches makes it likely



that opportunities and challenges differ significantly across approaches, and SQDA may even be irrelevant or impossible for the most interpretative approaches. A scaled pilot implementing SQDA across methods would be desirable.

The challenges remaining for SQDA to be more widely adopted are not specific to qualitative inference. Quantitative data sharing and re-analysis can face similar obstacles, with regard to standards of data sharing, researcher bias or lack of expertise of the analyst. Yet, qualitative data sharing for SQDA cannot follow the same rules as that of numeric data. To reap the opportunities of SQDA, further comprehensive applications are required. These could also tackle related methodological research problems, such as the use of computational text analysis for SQDA.

## Appendix

**Table AI.** Data re-coding examples based on theoretical propositions.

	Unordered institutions	Ordered institutions
Proactive resilience	<p>'He said there is a social self-help organization named "dabo". It is declining these days because many people are opting for working with money (as daily labourers) than helping each other'. (Copestake, 2016, Ethiopia)</p> <p>'Q: When Scheme Idea started &amp; reason? A: The scheme idea started in 2005. Droughts became recurrent so survival became difficult and we came up with the plan of sugar scheme. We realised the Kafue river nearby would be a lifeblood through irrigation' (Hall, 2017, Zambia)</p> <p>'We are prepared to move (change settlement pattern) to give way for cane growing because we have seen the benefits from those who are currently doing it. We also know that when they shift you and destroy your house to pave way for cane, they will compensate you'. (Hall, 2017, Zambia)</p>	<p>'We have lost interest in the resettlement issue because we have not been told anything. We don't know what is going on. We want our Emir and our village heads to be fully involved during the compensation exercise. We need to be fairly compensated so that our situation will not worsen'. (Urban, 2017, Nigeria)</p> <p>'A: [another male youth leader] The community department has several youth groups [wings of the PLPB, not clubs]. We are using them to catch thieves. Q: How do you communicate with the community? A: [Chairman of] the [zonal] committees send two representatives to the [main] Board. They and the chiefs pass on the information'. (Baker, 2008a, Sierra Leone)</p>

(continued)

**Table A1.** (continued)

	Unordered institutions	Ordered institutions
Reactive resilience	<p>'Q: So with the rising cost of food what changes do you make to your business/work to enable you to cope? A: You then try to burn more, depending on the size of the need. A: If you need to burn a lot in a short time, you call a mwethya (work party) and you cook for them while they help you to cut down the trees. You go to the shop and take things on credit, e.g. sugar'. (Hossain et al., 2015, Kenya)</p> <p>'They (the community) have their own way to help each other through a system called iddir since earlier times. In the last three years some rich farmers started to help the poor farmers by buying them fertilizer and seed so that they will pay back after harvesting. Even though there are attempts made by government to strengthen the social tie through structures called "I-5 grouping", it is not successful because people fear to order each other so that there is a power relationship within these structures he said. Due to this, the community prefers its own cultural self-help organizations like "iddir" (Copestake, 2016, Ethiopia)</p>	<p>'Q: Can you not simply apply for a licence to make the charcoal (one of the RAs refers to another group elsewhere that registered and got licenced)? A: We are told you need a letter from the chief before you can apply for the licence; but the chief will want to eat a lot. If I give the chief Shs.1000, yet the two sacks I cut in one day are only worth shs.400'. (Hossain et al., 2015, Kenya)</p> <p>'They also deal with conflicts through a cultural conflict management process of community elders. They prefer to deal conflicting cases through community elders in order to save time and cost which is the case in the formal justice system due to its bureaucratic' (Copestake, 2016, Ethiopia)</p>

## Acknowledgments

This research was generously supported with funding from the PVC Research Strategic Fund as well as the Centre on the Politics of Representation in Crisis at the University of Essex. We are grateful to comments on the argument presented here by Royce Carroll, Ismene Gizelis, Alexandra Hennessy, Shane Martin, Jonathan Slapin, and Martin Steinwand as well as the graduate student participants of the Grant Writing Workshop at the University of Essex (June 2018). We also thank Alan Jacobs for his feedback, as well as participants of the panel "Methodology in African Studies: Using Evidence" at the conference of the African Studies Association of the UK (Birmingham, September 2018), especially Kakia Chatsiou, Nic Cheeseman, Susan Dodsworth, Mareike Schomerus and Gina Yanitell Reinhardt. Rebecca Cordell and Dragana Vidovic provided excellent research assistance. We are grateful to helpful

comments received at the SISP (2022), PolMeth Europe (2021) and at POLIS, University of Warwick (2021).

### **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: This work was supported by the PVC Research Strategic Fund.

### **ORCID iDs**

Florian G Kern  <https://orcid.org/0000-0003-4222-4478>

Katariina Mustasilta  <https://orcid.org/0000-0002-5273-4490>

### **Notes**

1. The reports emanating from the QTD concentrate on discussing the consequences of transparency for various approaches and contexts, but two reports mention secondary data analysis briefly with reference to replication and risks for researchers and participants (Arjona et al., 2019, 12, and Shesterinina et al., 2019, 3/4).
2. We do not argue that SQDA be adopted by all qualitative scholars indiscriminately. Opportunities and challenges of SQDA will vary across traditions of qualitative research, which includes interpretive diverse methods such ethnographies, or qualitative comparative analysis (QCA). For some approaches, SQDA may even be impossible to implement. That said, SQDA can benefit political scientists under specific circumstances, interested in comparative, qualitative empirical evidence, and it is worth exploring what these opportunities – and related challenges – are.
3. For the exemplary application, we restrict our data search by relying only on the UKDA, which has long allowed depositing qualitative data. Sources from other archives, such as the Harvard Dataverse or the Syracuse Qualitative Data Repository, could be consulted as well. We discuss these where applicable below.
4. We decide to only rely on qualitative interview and focus group transcripts as data sources, given that these are currently the most frequently archived types of qualitative data. In fact, conducting SQDA may be very different, or even infeasible, for other types of qualitative information, for example, ethnographic field notes.
5. We decided to formulate this question prior to engaging in purposeful search for archived data, in order to not let data availability drive our research interest. A more inductive approach than the one employed here may allow for an iterative

- engagement and reformulation of the research problem when discovering new insights while reviewing primary archived data.
6. See UK Data Service catalogue at <https://discover.ukdataservice.ac.uk>
  7. By comparison, following similar steps in the advanced search of the Harvard Dataverse by restricting the 'Subject' to {Social Sciences} and the 'Kind of Data' to {qualitative} yields 22 deposited datasets – of overall over 90,000 datasets. While the Harvard Dataverse does not allow to directly discriminate by focus country, a review of the 22 identified studies reveals that only six of those focus on Sub-Saharan Africa.
  8. Researchers are currently trying to develop a cross-platform standard for exchanging data across CAQDAS, see [www.qdasoftware.org](http://www.qdasoftware.org).
  9. These interview appendices have also been recommended elsewhere when using interview-based research (see Bleich & Pekkanen, 2015).
  10. See templates here: [https://qdr.syr.edu/guidance/templates#informed\\_consent](https://qdr.syr.edu/guidance/templates#informed_consent).
  11. Many research projects follow the sequence of, for example, collecting qualitative evidence first to inform subsequent quantitative analysis or to scope contexts using qualitative interviews to follow up with more in-depth focus groups. However, here, we refer specifically to consulting archived, primary qualitative data from studies other than your own to induce new theories.
  12. The additional costs of transcription should be taken into account by organisations providing research funding, as the National Science Foundation does under 'data curation costs' in their 2021 Proposal and Award Policies and Procedures Guide.

## References

- Adger, N., Agrawala, S., Monirul, M., Mirza, Q., Conde, C., O'Brien, K., Juan, P., Pulwarty, R., Smit, B., & Takahashi, K. (2007). Assessment of adaptation practices, options, constraints and capacity. In M. Parry, & O. Canziani (Eds.), *Climate change 2007: Impacts, adaptation and vulnerability. Contribution of working group II to the fourth assessment report of the intergovernmental panel on climate change* (pp. 717–743). Cambridge University Press.
- Anderson, R. G. (2013). Registration and replication: A comment. *Political Analysis*, 21(1), 38–39. <https://doi.org/10.1093/pan/mps034>
- Ansell, N. (2010). *Averting 'new variant famine' in southern Africa: Building food-secure livelihoods with AIDS-affected young people*. Data collection. UK Data Service. <https://doi.org/10.5255/UKDA-SN-850364>
- APSA (2012). *A guide to professional ethics in political science*. American Political Science Association.
- APSA (2022). *A guide to professional ethics in political science*. American Political Science Association.
- Arjona, A. M., Mampilly, Z. C., & Pearlman, W. (2019). Research in violent or post-conflict political settings. *American Political Science Association Organized*

- Section for Qualitative and Multi-Method Research, Qualitative Transparency Deliberations, Working Group Final Reports, Report IV.2 (2018).*
- Baker, B. (2008a). *Evaluation of the effectiveness of police local partnership boards in Sierra Leone, 2006* (p. 5751). Data Collection. UK Data Service. <https://doi.org/10.5255/UKDA-SN-5751-1>
- Baker, B. (2008b). *Multi-choice policing resources for post-conflict situations: Rwanda and Liberia, 2006-2007*. Data collection. UK Data Service. <https://doi.org/10.5255/UKDA-SN-6039-1>
- Becker, M. (2020). Qualitative replication as a pedagogical approach to teaching research methods. *PS: Political Science & Politics*, 53(4), 802–806. <https://doi.org/10.1017/S1049096520000864>
- Bell, C., & Keys, P. W. (2016). Conditional relationships between drought and civil conflict in sub-saharan Africa. *Foreign Policy Analysis*, 14(1), 1–23. <https://doi.org/10.1093/fpa/orw002>
- Bishop, L. (2007). A reflexive account of reusing qualitative data: Beyond primary/secondary Dualism. *Sociological Research Online*, 12(3). <https://doi.org/10.5153/sro.155>
- Bishop, L. (2014). Re-Using qualitative data: A little evidence, on-going issues and modest reflections. *Studia Socjologiczne*, 3(214), 167–176. ISSN 0039–3371
- Bishop, L., & Kuula-Luumi, A. (2017). Revisiting qualitative data reuse: A decade on. *SAGE Open*, 7(1). <https://doi.org/10.1177/2158244016685136>
- Bleich, E., & Pekkanen, R. J. (2015). Data access, research transparency, and interview: The interview methods appendix. *Newsletter of the American Political Association Organized Section for Qualitative and Multi-Method Research*, 13(1, Spring).
- Büthe, T., & Jacobs, A. M. (2015). Conclusion: Research transparency for a diverse discipline. *Qualitative and Multi-Method Research*, 13(1), 52–64. <https://doi.org/10.5281/zenodo.892983>
- Büthe, T., Jacobs, A. M., Bleich, E., Pekkanen, R., Trachtenberg, M., Cramer, K., Shih, V., Parkinson, S., Wood, E. J., & Pachirat, T. (2015). Symposium: Transparency in multi-method research. *Qualitative and Multi-Method Research*, 13(1).
- Coltart, C., Henwood, K., & Shirani, F. (2013). Qualitative secondary analysis in austere times: Ethical, professional and methodological considerations. *Historical Social Research/Historische Sozialforschung*, 14(1), 271–292. <https://doi.org/10.17169/fqs-14.1.1885>
- Copestake, J. (2016). *Impact assessment in complex contexts of rural livelihood transformations in Africa. Part 2 - interview data*. Data collection. UK Data Service. <https://doi.org/10.5255/UKDA-SN-852065>
- Edwards-Jones, G. (2010). *Comparative assessment of environmental, community and nutritional impacts of consuming fruit and vegetables produced locally and overseas, 2004-2008* (p. 6573). Data collection. UK Data Service. <https://doi.org/10.5255/UKDA-SN-6573-1>

- Elman, C., & Kapiszewski, D. (2013). A guide to sharing qualitative data. *Qualitative data repository (QDR), center for qualitative and Multi method inquiry (CQMI)*. Syracuse University. Version 1.3.
- Elman, C., & Kapiszewski, D. (2014a). Data access and research transparency in the qualitative tradition. *PS: Political Science & Politics*, 47(1), 43–47. <https://doi.org/10.1017/S1049096513001777>
- Elman, C., & Kapiszewski, D. (2014b). Sharing qualitative data: The launch of the qualitative data repository. *The Washington Post: Monkey Cage*. [https://wapo.st/1hKiYm?tid=ss\\_tw&utm\\_term=.394f37714736](https://wapo.st/1hKiYm?tid=ss_tw&utm_term=.394f37714736)
- Elman, C., Kapiszewski, D., & Vinuela, L. (2010). Qualitative data archiving: Rewards and challenges. *PS: Political Science & Politics*, 43(1), 23–27. <https://doi.org/10.1017/S104909651099077X>
- Fearon, J. D., Humphreys, M.A., & Weinstein, J. M. (2015). How does development assistance affect collective action capacity? Results from a field experiment in post-conflict Liberia. *American Political Science Review*, 109(3), 450–469. <https://doi.org/10.1017/S0003055415000283>
- Fielding, N. (2004). Getting the most from archived qualitative data: Epistemological, practical and professional obstacles. *International Journal of Social Research Methodology*, 7(1), 97–104. <https://doi.org/10.1080/13645570310001640699>
- Gelman, A. (2013). Preregistration of studies and mock reports. *Political Analysis*, 21(1), 40–41. <https://doi.org/10.1093/pan/mps032>
- Goist, M., & Kern, Florian G. (2018). Traditional institutions and social cooperation: Experimental evidence from buganda kingdom. *Research & Politics*, 5(1), 1–9. <https://doi.org/10.1177/2053168017753925>
- Hall, R. (2017). *Land and agricultural commercialisation in Africa*. Data collection. UK Data Service. <https://doi.org/10.5255/UKDA-SN-852392>
- Heald, S. (2015). *Making rural law in Kenya*. Data collection. UK Data Service. <https://doi.org/10.5255/UKDA-SN-851834>
- Heaton, J. (2008). Secondary analysis of qualitative data: An overview. *Historical Social Research*, 33(3), 33–45. <https://doi.org/10.2307/20762299>
- Holzinger, K., Kern, F. G., & Kromrey, D. (2016). The dualism of contemporary traditional governance and the State: Institutional setups and political consequences. *Political Research Quarterly*, 69(3), 469–481. <https://doi.org/10.1177/1065912916648013>
- Holzinger, K., Kern, F. G., & Kromrey, D. (2020). Explaining the constitutional integration and resurgence of traditional political institutions in sub-saharan Africa. *Political Studies*, 68(4), 973–995. <https://doi.org/10.1177/0032321719884>
- Hossain, N., Scott-Villiers, P., Shankland, A., Joshi, A., Jahan, F., Musembi, C., Britto, L., Sinha, D., Patnaik, B., Kalita, D., & Benequista, N. (2015). *Food riots and food rights: The moral and political economy of accountability for hunger*. Data collection. UK Data Service. <https://doi.org/10.5255/UKDA-SN-851951>
- Hughes, K., & Tarrant, A. (Eds.). (2019). *Qualitative secondary analysis*. Sage.

- Humphreys, M., Sanchez de la Sierra, R., & van der Windt, P. (2013). Fishing, commitment, and communication: A proposal for comprehensive nonbinding research registration. *Political Analysis*, 21(1), 1–20. <https://doi.org/10.1093/pan/mps021>
- Humphreys, M., & Weinstein, J. M. (2009). Field experiments and the political economy of development. *Annual Review of Political Science*, 12, 367–378. <https://doi.org/10.1146/annurev.polisci.12.060107.155922>
- Irwin, S., Bornat, J., & Winterton, M. (2012). Timescapes secondary analysis: Comparison, context and working across data sets. *Qualitative Research*, 12(1), 66–80. <https://doi.org/10.1177/1468794111426234>
- Jacobs, A. M., Tim Büthe, Arjona, A., Leonardo, R. A., Bellin, E., Bennett, A., Björkman, L., Bleich, E., Elkins, Z., & Fairfield, T. (2021). The qualitative transparency deliberations: Insights and implications. *Perspectives on Politics*, 19(1), 171–208. <https://doi.org/10.1017/S1537592720001164>
- Jones, L. 2011. Towards a holistic conceptualisation of adaptive capacity at the local level: Insights from the local adaptive capacity framework (LAC). *Overseas Development Institute background paper*. <https://odi.org/en/publications/towards-a-holistic-conceptualisation-of-adaptive-capacity-at-the-local-level-insights-from-the-local-adaptive-capacity-framework-lac/>
- Jordan, J. C. (2015). Swimming alone? The role of social capital in enhancing local resilience to climate stress: A case study from Bangladesh. *Climate and Development*, 7(2), 110–123. <https://doi.org/10.1080/17565529.2014.934771>
- Kapiszewski, D., & Karcher, S. (2021). Transparency in practice in qualitative research. *PS: Political Science & Politics*, 54(2), 285–291. <https://doi.org/10.1017/S1049096520000955>
- Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H., & Melham, K. (2015). Dynamic consent: A patient interface for twenty-first century research networks. *European Journal of Human Genetics*, 23(2), 141. <https://doi.org/10.1038/ejhg.2014.71>
- King, G. (1995). Replication, replication. *PS: Political Science & Politics*, 28(3), 444–452. <https://doi.org/10.2307/420301>
- Long-Suthehall, T., Sque, M., & Addington-Hall, J. (2011). Secondary analysis of qualitative data: A valuable method for exploring sensitive issues with an elusive population? *Journal of Research in Nursing*, 16(4), 335–344. <https://doi.org/10.1177/1744987110381553>
- Lupia, A., & Elman, C. (2014). Openness in political science: Data access and research transparency. *PS: Political Science & Politics*, 47(1), 19–42. <https://doi.org/10.1017/S1049096513001716>
- Mason, J. (2007). ‘Re-using’ qualitative data: On the merits of an investigative epistemology. *Sociological Research Online*, 12(3), 4–7. <https://doi.org/10.5153/sro.1507>

- Mauthner, N. S., Parry, O., & Backett-Milburn, K. (1998). The data are out there, or are they? Implications for archiving and revisiting qualitative data. *Sociology*, 32(4), 733–745. <https://doi.org/10.1177/0038038598032004006>
- Mazzucato, V., & Niemeijer, D. (2002). Population growth and the environment in Africa: Local informal institutions, the missing link. *Economic Geography*, 78(2), 171–193. <https://doi.org/10.2307/4140786>
- Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K. M., Gerber, A., Glennerster, R., Green, D. P., Humphreys, M., & Guido, I. (2014). Promoting transparency in social science research. *Science*, 343(6166), 30–31. <https://doi.org/10.1126/science.1245317>
- Moravcsik, A., Elman, C., & Kapiszewski, D. (2013). A guide to active citation. *Qualitative data repository (QDR), Center for qualitative and Multi method inquiry (CQMI)*. Syracuse University. Version 1.8.
- Mustasilta, K. (2019). Including chiefs, maintaining peace? Examining the effects of state–traditional governance interaction on civil peace in sub-Saharan Africa. *Journal of Peace Research*, 56(2), 203–219. <https://doi.org/10.1177/0022343318790780>
- Mustasilta, K. (2021). The implications of traditional authority contest for local-level stability – Evidence from South Africa. *Conflict Management and Peace Science*, 38(4), 457–480. <https://doi.org/10.1177/0738894220959655>
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Jennifer, P. W., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533–544. <https://doi.org/10.1007/s10488-013-0528-y>
- Raleigh, C., & Hegre, H. (2009). Population size, concentration, and civil war. A geographically disaggregated analysis. *Political Geography*, 28(4), 224–238. <https://doi.org/10.1016/j.polgeo.2009.05.007>
- Shesterinina, A., Pollack, M. A., & Arriola, L. R. (2019). Evidence from Researcher Interactions with Human Participants (February 12, 2019). American Political Science Association Organized Section for Qualitative and Multi-Method Research, Qualitative Transparency Deliberations, Working Group Final Reports, Report II.2 (December 2018). Available at SSRN. <https://ssrn.com/abstract=3333392>
- Tollefsen, A. F., Håvard Strand, & Buhaug, H. (2012). PRIO-GRID: A unified spatial data structure. *Journal of Peace Research*, 49(2), 363–374. <https://doi.org/10.1177/0022343311431287>
- Urban, F. (2017). *China goes global dams project*. Data collection. UK Data Service. <https://doi.org/10.5255/UKDA-SN-852499>
- Watkins, D. C. (2022). *Secondary data in mixed methods research*. SAGE Publications.



### **Author Biographies**

**Florian G. Kern** is a Senior Lecturer (Associate Professor) at the Department of Government, University of Essex.

**Katariina Mustasilta** is a Senior Research Fellow at the Finnish Institute of International Affairs.