Organized Communities as a Hybrid Form of Data Sharing: Experiences from the Global STEP Project

Isabell Stamm

Abstract: With this article, I explore a new way of how social scientists can share primary qualitative data with each other. More specifically, I examine organized research communities, which are small membership groups of scholars. This hybrid form of data sharing is positioned between informal sharing through collaboration and institutionalized sharing through accessing research archives. Using the global "Successful Transgenerational Entrepreneurship Practices" (STEP) project as an example, I draw attention to the pragmatic practices of data sharing in such communities. Through ongoing negotiations, organized communities can, at least temporarily, put forward sharing policies and create a culture of data sharing that elevates the re-use of qualitative data while being mindful of the data's intersubjective and processual character.

Table of Contents

1. Introduction

Information sharing is critical to scientific progress, as it allows cumulative research and increases efficiency. Robert MERTON (1973) defined the unconditional sharing of knowledge as one of the essential features of academic life. While social scientists share their findings publicly through research articles, working papers, blogs, presentations, etc., a more complex issue is what HAEUSSLER, JIAN, THURSBY and THURSBY (2014) called the "specific sharing" of raw or preprocessed data, either through repositories or direct correspondence. Examples of this are data sets, transcripts, or ethnographic field notes that have directly emerged from interaction with research subjects, which I will refer to as primary data. Youngseek KIM and Melissa ADLER (2015) suggest that sharing primary qualitative data among social scientists is far from being a common practice. Researchers, who have invested time and effort into collecting rich, and often, sensitive data are reluctant to share for reasons both individual (e.g., fear of not getting published, emotional attachment) and institutional (e.g., confidentiality requirements, lacking incentives from universities). [1]

Among qualitative social scientists, the debate about data sharing is particularly critical (MASON, 2007; MAUTHNER, PARRY & BACKETT-MILBURN, 1998). From a constructivist perspective, qualitative methods such as open interviews, focus groups or participant observation represent intersubjective and interpretative processes. Researcher and participant are intimately linked in the creation of knowledge, which creates a dialectical tension between individual ownership of data and the ethics of protectionism (BROOM, CHESHIRE & EMMISON, 2009, p.1167). These methodological underpinnings suggest that primary data is highly context dependent, i.e., the situation of data collection shapes and biases the collected data (such as texts and field notes). In the process of data analysis, a careful reflection upon these contextual factors (e.g., demographics of the interviewer, research agenda, interview setting, and timing) and their potential influence on the produced data is thus an important step in order to arrive at valid and reliable findings (CRESWELL, 2018; FLICK, 2015). In fact, some scholars suggest qualitative data analysis to be an 'insider activity' (MAUTHNER et al., 1998) and doubt data analysis could ever be untangled from the original (research) context (e.g., FINK, 2000; HAMMERSLEY, 1997). [2]

More recently, however, the discourse on qualitative data sharing has shifted from a doubtful critique towards a more pragmatic discussion of how data sharing can be organized (e.g., SMIOSKI, 2013). We can observe a growing awareness among qualitative scholars of the value of preserving and sharing primary data. These benefits include an increase of transparency in the interpretation process, facilitating analysis from multiple perspectives, reduced costs, or contribution to the education of students (KIM & ADLER, 2015). This shift has significantly accelerated the emergence of secondary analysis as a research method, which re-uses primary data to glean new social scientific and/or methodological understanding (IRWIN & WINTERTON, 2011). [3]

Inextricably linked to this more pragmatic discourse are immediate changes in the institutional environment of academic research. Major funding agencies in North America (e.g., National Science Foundation and the National Institutes of Health) and Europe (e.g., Economic and Social Research Council and the Deutsche Forschungsgemeinschaft) require data management plans (KVALHEIM & KVAMME, 2014), in which researchers have to justify that they have only collected data when no suitable archival data is available, and also that wherever possible they made newly acquired primary data accessible for future use by others (IRWIN & WINTERTON, 2011). A growing number of archives have been quick to notice these trends and offer data curation services, and scholars have put forward guides for good practice in preparing primary data for sharing (CLIGGETT, 2013; SAUNDERS, KITZINGER & KITZINGER, 2015). These institutional changes occur at a time, when qualitative research is increasingly expected to be team based and publicly accountable (BROOM et al., 2009, p.1175); social media and big data have altered the perception of data ownership and the tolerance for private data to be publicly exposed. [4]

Before these recent developments, data was rarely shared, and if shared done informally between research collaborators. Now a division of labor is emerging between primary researchers conducting and analyzing primary data, and secondary researchers accessing preprocessed and stored qualitative data through archives and repositories. Through this dualism qualitative data sharing is starting to professionalize, allowing secondary analysis to mature as a research method. What remains, however, is the challenge that an increasing distance between primary and secondary researchers imposes on the intersubjective nature of qualitative research, a challenge that sparks the need to revisit and revise the methodological underpinnings of qualitative research. [5]

Data sharing in organized research communities—small, preselected groups of scholars jointly investigating a shared research question—offers an alternative approach to this challenge. Such communities resemble informal collaborations, as researchers personally know each other and are familiar with the research context of the primary data. They also resemble professional repositories, as they put forward codified data sharing policies. In this hybrid form of data sharing, primary researchers and secondary analysts simultaneously use and re-use qualitative data thereby conserving the intersubjective nature of the research process. [6]

Despite the innovative potential of organized research communities, hybrid forms of data sharing have thus far received scant attention in the literature on data sharing (BISHOP, 2007; COLTART, HENWOOD & SHIRANI, 2013). With this article, I examine the opportunities and drawbacks of such communities, drawing on a single case study from the "Successful Transgenerational Entrepreneurship Practices" (STEP) project. Its long duration, global reach, and large number of scholars make the STEP project a particularly informative case when it comes to examining data sharing policies and practices. This organized research community has found innovative answers to the most pressing questions of data sharing, including confidentiality, quality, and timing. This article contributes to the

broader discussion on data sharing by suggesting that organized research
communities—at least temporarily—create an institutionalized setting suitable for
sharing qualitative data. In the following, I will discuss sharing within organized
research communities in contrast to the ideals of informal and formal sharing
(Section 2). A brief description of my methodological design (Section 3) and an
introduction to the case (Section 4) form the basis for the presentation of my
findings on the potentials and limits of organized research communities as hybrid
form of sharing qualitative data (Section 5). [7]

2. What are Organized Research Communities and how do They Share Data?

In traditional curation, the preservation of texts, photos, or audio data as
documented history is an end in itself. Data sharing, on the other hand, is a
practice among researchers that explicitly facilitates the re-use of research data
for teaching or research purposes (CORTI & THOMPSON, 2008). As such, data
sharing strongly caters to the needs of the re-user, while having to respect and
capture the intimate and often confidential relationship primary researchers have
established with their data. This specific character makes secondary analysis
distinct from document analysis, as practiced for example in history or sociology,
which draws from material produced in "every-day" or non-research settings
(THORNE, 1994). Secondary analysis as method depends upon the availability of
primary material for re-use. The following section presents informal and formal
sharing as ideal typical ways of sharing primary materials and then situates
organized research communities as hybrid form. These forms differ in the way
primary and secondary researcher encounter each other and negotiate about the
terms of data sharing. [8]

2.1 Informal sharing of primary data materials

A common way of sharing qualitative data, especially prior to it being available in
archives and repositories, was to "simply" contact a fellow social scientist to
request re-use of their material (KIM & ADLER, 2015). In the following, I refer to
this form of personal, dyadic exchange of primary material among researchers as
informal data sharing. [9]

One of the major challenges in informal data sharing poses the issue of allocating
adequate primary data for secondary analysis. As long as qualitative scholars
store their primary materials individually and for their own use, the question of
who has what kind of data remains opaque to fellow social scientists. By
identifying adequate primary data, researchers are dependent on network
contacts and studying method sections in research publications. [10]

Once they have tracked down the desired data, secondary researchers need to
convince data proprietors to make their material available. KIM and ADLER
(2015) identify individual motives, such as privacy issues, concerns about misuse,
loss of publication opportunities, and institutional factors, such as weak data
sharing requirements of academic journals obstructing data sharing practices.

Qualitative researchers may have used confidentiality agreements or are bound by specifications of an approval process (such as a university's Internal Review Board [IRB] in the US context), which may not have included the option of data sharing in the first place. In addition, Barry BOZEMAN and Monica GAUGHAN (2011) argue that during their careers researchers have chosen different collaborating strategies that assess the usefulness of a collaboration for one's own career progress, experience or mentoring skills. Informal data sharing is a reciprocal process rather than an altruistic act, with co-authorship being an explicit product of a scientific collaboration (LI, LIAO & YEN, 2013). [11]

Once implemented, informal sharing of data allows primary researchers to retain direct control over the re-used data, while secondary researchers can profit from rich context knowledge, an emotional connectedness to research participants, and possible access to previous electronic coding (HEATON, 2008). The secondary researcher can draw on the primary researcher's experience, the implicit understanding and memories of the data collection process, which may play an important role in making sense of the data (HAMMERSLEY, 2009). At the same time, informal data sharing entails the risk that the data may be lacking consent for re-use or documentation; or that the actual data content is misaligned with the secondary researcher's project aims (HEATON, 2008). [12]

2.2 Formal sharing of primary data materials

In the sense of traditional data curation, archives and repositories increasingly acquire, prepare, and enhance research data and facilitate its re-use for research purposes. I refer to the depositing of primary research material in a professional archive or repository for the purpose of re-use as formal data sharing. Such archives and repositories can take various forms ranging from national archives that store data of all sorts and topics centrally (e.g., Qualidata in the UK, Finish Social Science Data Archive [FSD] or the Wiener Institute for Social Science Data Documentation and Methods [WISDOM]) to more disperse local, topic-centered or data type specific collections (CORTI, 2011). [13]

In general, storing qualitative data in archives and repositories enables a wider audience to participate in data sharing. Searchable meta-documentation fosters transparency in the allocation of primary material and allows secondary researchers to assess the content and quality of that material (IRWIN & WINTERTON, 2011). Driven by such potential benefits, the development of qualitative archives has moved forward (for an overview see BISHOP & NEALE, 2011; MEDJEDOVIĆ & WITZEL, 2010; SLAVNIC, 2013). [14]

The rise of qualitative data archiving has sparked an intense discussion about the modalities of archiving. What once began as an imitation of archiving quantitative data sets has turned into a debate about how archiving can accommodate the epistemological and ethical characteristics of qualitative research data. In particular, the context dependency of qualitative interviews or observations makes the archiving and re-use of qualitative data methodologically more difficult (CARUSI & JIROTKA, 2009). Proponents of institutionalized data sharing argue

that "the idea that only those involved in initial data generation can understand the context enough to interpret the data is not only anti-historical but it puts enormous epistemological weight onto the notion of 'successful reflexivity'" (MASON, 2007, §3.3). In an attempt to acknowledge the context dependence of qualitative data (e.g., FIELDING, 2004), information on the original research process, for example, a sampling plan or the original research question, is stored as supplemental material if possible. [15]

A number of journal articles, case studies (e.g., McNEILL, 2016; SMIOSKI, 2011), and process descriptions offered by archives and repositories provide guidelines for data curation and elaborate upon data sharing policies. Achievements have been made in terms of access protocols, ethical guidelines, intellectual property, metadata standards, and preparing data for deposition (BROOM et al., 2009). Interestingly, formal data sharing does not provide a single solution, but offers a range of possibilities for archiving qualitative data. [16]

A major concern in the discourse on formal sharing of primary data, and a main reason for the reluctance of qualitative scholars to actually archive their datasets, is the maintenance of ethical standards and the protection of participants' rights (e.g., CORTI, DAY & BACKHOUSE, 2000; GEBEL et al., 2015; LEH, 2000; PARRY & MAUTHNER, 2004). Primary researchers feel that they potentially lose control over who will access the data and thus violate their duty to protect their research participants. Annamaria CARUSI and Marina JIROTKA (2009) point out, that these ethical risks compound with digital data because it is inherently susceptible to being copied, manipulated, and de- or re-contextualized. One measure typically taken is the anonymization of the stored data material; a work-intensive task often imposed on the depositing researcher, which constraints their willingness to engage in data curation. Large repositories thus offer anonymization services or primary researchers can integrate this task in their applications for research funding. Nevertheless, not all anonymization efforts are beneficial to secondary analysis. Alex BROOM et al. (2009) remind us that "steps taken to unsure participant confidentiality extend well beyond changing or deleting names, and may require removing contextual material that is vital to understanding the research setting in its entirety" (p.1166). Removing background information or textual data can thus create serious validity issues for secondary analysis or even disqualify the data material for re-use (CARUSI & JIROTKA, 2009). [17]

If an anonymization of the data material is not possible or conducive, access restrictions may solve a potential ethical conflict. Access to primary data in repositories can vary greatly ranging from fully anonymized and publicly available to clear-name material, only accessible for accredited individuals on site, to embargoed data that is locked up from use for a defined number of years (see for example the access classification at WISDOM, SMIOSKI, 2013). This variation in access produces variation in the actual sharing relationship, which allow for more or less direct exchange between primary and secondary researchers. [18]

In sum, formal data sharing practices cover a wide variety of how primary data is actually prepared and made available for re-use, which addresses different data sensitivities and disciplinary practices (CARUSI & JIROTKA, 2009). In any case, formal sharing starts from the assumption that a primary researcher has ended the research study and the repository stores that data or, at least, manages data access (with the exception of living archives, a form that we will return to later). The terms and conditions of formal sharing are subject to negotiations that predominantly occurs between primary researcher (on behalf of the researched) and the data repository (p.290). The repository thus acts as mediator between primary and secondary researchers in establishing a data sharing relationship. In this position, archives and repositories contribute to more clarity in the sharing relationship and escalate the professionalism of data sharing. At the same time, the mediator position increases the distance between secondary analysis and the original research context. [19]

The standards and good practices developed in formal sharing contribute to a reevaluation of the epistemological understanding of qualitative research itself, which lets secondary analysis raise as valid methodology. In countries with an advanced infrastructure for formal sharing—such as the UK or Finland—re-use has grown into the mainstream of qualitative research (BISHOP & KUULA-LUUMI, 2017), whereas in other countries (and specific research cultures) sharing of qualitative data has yet to arrive at a point where it is a taken-for-granted practice (CORTI, 2011). [20]

2.3 The community mode of sharing data

Besides personal collaboration (informal sharing) and access through data curation (formal sharing), Lizzie RICHARDSON (2015) identifies communities as a third form of sharing practice. In general, communities create a peer-to-peer experience within institutional structures that offer routines, policies, and a sense of security. In business, typical examples are the internationally known platforms *AirBnB* (private and commercial offers of short-term housing) or *Uber* (private and commercial offers of car rides) that create communities sharing accommodation or transportation. These communities emerge within a specific organizational setting and use platforms for their interactions. Yet can we picture a community mode of sharing as a viable hybrid in the context of sharing qualitative data? [21]

To be clear, a community mode of sharing is distinct from what is generally known as "research community" in terms of a limited, yet unspecific number of individual researchers sharing a common research interest and institutionalized rules of communication. A community mode of sharing is further distinct from the open access movement, which strives to implement open and immediate knowledge sharing. A community mode of sharing data requires an organizational setting, defined membership, and data sharing policies creating a formal framework for the peer-to-peer experience of sharing data. I thus refer to these as organized research communities. [22]

Three striking examples are the Global Social Media Impact study, a multi-partner global research collaboration simultaneously conducting and sharing ethnographic data on social media use; the Timescapes project, a multi-affiliation project that followed the lives of over 300 individuals in seven empirical projects explicitly sharing and depositing their data; and the STEP project, a multi-affiliation project conducting and sharing over 125 case studies on global transgenerational entrepreneurship. Each of these organized research communities has restricted their membership either through predefining their members during the grant-writing process or through installing a membership application procedure. They further provide their members with an infrastructure for sharing data with fellow primary researchers. [23]

A key feature of organized research communities is the simultaneous engagement in conducting and re-using qualitative data, which bridges the distance and divide in labor between primary and secondary researchers. The co-involvement of primary and secondary researchers means that, compared to re-using archived data, "there may be greater awareness of the context of primary work, and sensitivity to the feelings of the researched" (HEATON, 2008, p.515). Although a member may not conduct primary material personally, the researcher can still relate to the setting from which the material emerged. This setting promises to be particularly beneficial for sharing qualitative data, which is context dependent and intersubjective in its character. [24]

As a hybrid form, organized research communities bring primary research and archiving "into a closer and more productive alignment through ongoing communication and negotiation" (NEALE, HENWOOD & HOLLAND, 2012, p.11). As the data collection is still ongoing, organized research communities embrace elements of living archives. In addition, through their common scope of interest, organized research communities function as a domain repository and can improve the storage and discovery of data in a particular realm (McNEILL, 2016). In difference to a classic repository, however, organized research communities only moderate the negotiated terms and conditions of data sharing between primary and secondary researcher. [25]

As such, data sharing in organized research communities shakes up and reconfigures the performance of sharing qualitative data. It conserves the positive aspects of collaborative data sharing (informal sharing) while creating searchable and comparable data by negotiating a formal sharing protocol. As these organized research communities are specifically set up to share data, one can expect that questions of data privacy are addressed beforehand, and that researchers have a higher chance of receiving qualitative data suitable for secondary analysis. Thus, organized research communities promise to be a viable form of data sharing particularly suitable for qualitative data. [26]

Yet we know very little about actual sharing practices and the fulfillment of these promises. With this study I examine abstract and pragmatic issues that arise when qualitative data is shared within organized research communities. Of particular interest is the way organized research communities solve common

issues in data sharing, such as protecting the research subject and handling the context dependence of qualitative data. [27]

3. Method

To study the outlined questions, I chose a single case study design (CRESWELL, 2018; YIN, 2013). This approach offers the necessary depth and openness needed for a full exploration of the specifics of an organized research community. I selected the global STEP project as a case illustrative of an organized research community sharing qualitative data. A short synopsis of this case reads: In 2005, under the lead of Timothy HABBERSHON, six European universities and business schools joined forces to explore entrepreneurial mindsets and capabilities that enable the creation of value through recurring entrepreneurship across many generations. They agreed to interview business families, i.e., families that own and/or run one or multiple businesses, and researchers shared their data. In return for field access, the researchers reflected on their gained knowledge and reported back to participating business families in annual summits. At the summits researchers also met the families that participated in the cases conducted by other researchers. The conducted case studies further informed teaching at the affiliated institutions, thereby educating the next generation of family business leaders. A membership model allowed researchers to jointly engage in a common research question and to accumulate knowledge. Today, the STEP project counts 37 affiliated institutions and approximately 175 researchers globally. As of January 2017, STEP has hosted 18 summits with business families around the world, 56 academic meetings in four regions, has authored 5 books and yielded more than 17 academic articles published using STEP data. Over the past decade 129 case studies and recently a global quantitative survey have been conducted. Over time, the STEP project institutionalized a sophisticated data sharing protocol. The long tenure of this community and lively discussions of data sharing, make the STEP project an ideal case for examining the potential and limitations of sharing qualitative data in a community mode. [28]

My initial engagement with the STEP project was driven by a strong interest in this large data pool that I was hoping to use for a secondary analysis on life course dynamics in business families. In 2013, I reached out to the administrative team at Babson College in Boston, MA, but as an individual, I was unable to secure the necessary funds and accomplish the required workload in order to receive access. Hence, in 2014 I joined the team of Stetson University Florida, one of the new North American affiliates, while working at the University of California Berkeley. A co-authored publication emerged from the collaboration on our first case study. Through my work on the Stetson team, I gained data access to the North American data pool. In addition, I worked closely with the affiliated research teams to identify original data materials for my secondary analysis. [29]

My active member status granted me access to the STEP community, which enabled me to conduct a participant observation of data sharing (KNOBLAUCH, 2014). This method promised unique insights into sharing practices and

considerations of various actors involved in the sharing-process over time. Early on I proposed my research interest to the STEP global board, which allowed me to sit in on their quarterly teleconferences and to approach members for interviews. Overall, this study builds on field work spanning two-years and includes field observations, interviews and document analysis (DAVIES, 2010). From 2014-2016, I took field notes at STEP board meetings, regional academic meetings and conferences (EMERSON, FRETZ & SHAW, 2011). In addition, I conducted eleven semi-structured interviews with current and former STEP members (HELFERICH, 2014). The interview plan included: 1. informants that could speak about the history of the STEP project from its beginning until today, 2. representatives of all governance units of the STEP project (including members of the Babson administrative team, the global board, and members of all four regional councils, and 3. informants that have practiced data sharing within the STEP project (one interviewee can fulfill multiple functions within STEP and in the interview plan). This interview plan expanded the examined time span from two years to the full trajectory of the STEP project and systematically includes viewpoints of various actor groups. With the verbal consent of each interview participant, I recorded the interviews. These data sources were further triangulated by document analysis (FLICK, 2015) of newsletters, the complete collection of STEP board meeting minutes, the STEP academic information package for new members, and presentations about the STEP project. Such data triangulation, i.e., the use of multiple data sources, helped to cross-check information and to ensure validity (JOHNSON, 1997). [30]

Following a case study methodology (YIN, 2013), the analysis occurred in an iterative process going back and forth between coding various data sources and consulting previous research literature. The analytical approach entailed three phases: First, I uncovered common arguments that STEP members voiced when reflecting about data sharing case studies vis-a-vis survey data in the framework of the current data sharing protocol. Second, I shifted my focus to the special character of qualitative data and potential ways to share these. I clustered emerging themes around four areas that resembled common problems of sharing qualitative data in the research literature. These are: 1. selection of sharing collaborators, 2. privacy issues, and 3. quality assurance. Emerging from the material, I added 4. timing of data sharing. For each of these areas I coded the arguments feeding into the data sharing protocol and the sharing mechanisms that the protocol enables. Third, I contextualized these codes with regard to the evolution of the data sharing protocol. This phase required assembly of a detailed timeline of leadership decisions that pertained to the data sharing protocol and captured its continuous adaption to the needs of the community. [31]

Throughout the analysis, I constantly compared the emerging argumentative patterns and sharing mechanisms with the typical modes of informal and formal data sharing as described in the literature. During this process, I actively engaged in critical self-reflection about my own potential bias, intensively discussed my thoughts and findings with trusted colleagues, and asked interview participants to read early drafts of this article (JOHNSON, 1997). I thus arrived at the distinct character of organized research communities for sharing qualitative data. This

character includes four key features: moderated reciprocity in the selection of collaborators, the community as confidentiality steward, peer pressure as mechanism of quality assurance, and a mid-term time range for data sharing. [32]

In the following section, I report my empirical material and present my observations on the emergence and practice of data sharing policies at the STEP project. On these grounds, I discuss my findings in each of the four analytical categories and evaluate the potential and limits of sharing qualitative data in organized research communities. To ensure privacy, I use pseudonyms and a high-level descriptor of their location in any reference I make to study participants. This article has undergone an approval process with the STEP board. [33]

4. Data Sharing Policies and Practices at the STEP Project

This section provides systematic information about the case. It gives an impression of the richness of the collected materials and their interpretation, thereby adhering to replicability as norm for presenting qualitative research (LaROSSA, 2012). Table 1 depicts the development of the STEP project and its data sharing policies from its inception until today. The following three sections elaborate on the practice of sharing data within this organized community during its formation phase, its institutionalization and how data sharing is envisioned in the future.

Table 1: History of the STEP project. Click here to download the PDF file. [34]

4.1 Early days: Accumulating cases and codifying sharing policies

Timothy HABERSHON envisioned a substantiation of his model of transgenerational entrepreneurship in a collaborative effort both between researchers and with business families, an approach that would allow accumulating knowledge on a research subject that is typically hard to access. In 2005, the STEP project was launched officially at Bocconi University in Milan, Italy, as a collaboration between Babson College and six European founding institutions. The project was strategically located at Babson College due to its high international reputation for its research on entrepreneurship. The "Global Entrepreneurship Monitor" also located at Babson College, provided a role model for assembling national survey data about entrepreneurial efforts from around the globe. [35]

From its beginning, the STEP project was built as a self-sustaining group fully funded by annual membership fees and independent from a funding agency. The STEP project's strategy explicitly included a systematic organization of a communal research collaboration. In a presentation to interested universities and new members in 2006, the vision for the STEP project read:

> "Babson is building a global partnership network of leading academic institutions in four regions of the world—Europe, Latin America, Pacific Rim, and North America.

> Our goal is to have a total of 40 institutions worldwide who are committed to entrepreneurship and family business research" (Internal Document: STEP recruitment, 2006). [36]

The affiliated institutions shall work together so that each partner can leverage their institutional relationships with business families in their region. Organizing this research community follows two objectives: First, the initiators of STEP wanted to create a "shared learning format that allows the findings to have an immediate application" (ibid.). Second, the initiators of STEP aimed at scaling up research on transgenerational entrepreneurship through accumulating and sharing data. While each member contributes a "minimal amount of research per institution, the combined data creates a large global data base that can be used for comparative case and quantitative stakeholder analysis" (ibid.). [37]

Researchers active during the formative years of the STEP project reported that they

> "had a pretty collegial atmosphere. People really liked hanging out with each other. We had a lot of face-to-face meetings, every one getting together multiple times a year. It built trust, we cared about and enjoyed working with each other" (Scott, North America). [38]

Together they laid out a research design, presented their respective empirical findings and substantiated the theoretical model. Although the methodological design includes both qualitative and quantitative methods, the researchers first focused on conducting cases. The participating scholars specified an interview guideline, a selection scheme for cases and a sample outline for a case report. [39]

Shortly after the Milan launch in 2006, STEP initiators released a methodological note that summarized the consensus found among STEP members about what kind of data they intended to share with each other:

> "All interviews should be taped and transcribed. The verbatim interview transcription does not have to be translated into English in the first phase, but should be saved in a Word document (one for each interview). This will be our raw data and will be compiled in a shared data base" (Internal Document: STEP methodological note, 2006). [40]

In addition, each research team compiled a rich and thick case description structured around the themes of the interview guide including many direct quotes from the interviews. In a separate section researchers protocol emerging analytical ideas, possible interpretations and other reflections that spring to mind when working with the data. These case description documents are written (translated) in English and enter a shared database using a qualitative research software package. This work method should facilitate "coding and later comparisons between cases and countries, even if the first aim should be to understand the uniqueness and story of each individual case" (ibid.) These case

reports are thus a first step in preparing the raw data for analysis and not fully developed case studies. [41]

STEP members provided hands-on-guidance to each other in conducting interviews and writing up reports. Although the desired form of the data had been explicitly described, not all members adhered to the agreed upon sharing policies. As Sven, Europe, recalled:

> "Some teams were reluctant to share transcripts simply because they did not exist. They had written up their cases without transcribing them verbatim, which was the agreed upon policy. The other rationale was the team realized that it was sensitive to share the transcripts. It was quite difficult to get permission from the respondents to share the detailed transcripts in such a global project." [42]

Given the dominance of the strategic goal to grow the project and organize a vivid research community, the STEP project codified looser criteria on shared data. They proceeded with storing and sharing only the case reports and provided meta-data about each case. Transcripts and audio files remained with the original researcher, who could make these materials available upon request. [43]

With more and more members joining the STEP project, a discussion emerged about how new members would be selected and when they would be granted access to the data stock. Established regions in the organized community voiced concerns about "free riding" (René, Europe), i.e., unearned data access. Eventually, the STEP project introduced an application procedure, which considered new members based on regional coverage and the applicant's potential to fulfill project requirements. This discussion also yielded a tiered-access to the accumulated data: after submitting the first case study, new members receive access to their regional database of case reports; after the second case, they have full access to the global database of case reports. [44]

During these formative years, the STEP community not only laid the groundwork for the project's methodological design, but also formulated key rules for who can be part of the project and under what conditions. The agreed upon policies imprint important principles for what is being shared and how it can be accessed for the years to come. [45]

4.2 Institutionalization: Reducing the workload and enforcing sharing policies

In 2011, the STEP project reached the desired 40 affiliated teams as a critical mass of North American institutions joined and formed their own regional cluster. The growing size of the research community required formalization of the organically grown governance structure. In 2010, the global board and regional councils became the official government units of the STEP project. The global board, which was constituted by a global director, a global board chair and a representative of each region, was responsible for the strategic direction of the STEP project. The regional councils organize activities in each region, which

operate independently of one another within the defined STEP protocols. The global faculty director, who was assisted by an administrative team located at Babson College, served as liaison between these institutions and mediates between interests. With the new structure in place, discussions and decision making about sharing policies became more formal. [46]

A major undertaking for the global board was to complete STEP's methodological design by drafting and implementing the STEP global survey. In a long and detailed process, a committee of global board members lead the development of the questionnaire, the protocol, and the search for previously used measures. Aside from creating a rigorous research tool that finds consensus across the regions, this process included an intensive discussion about data sharing. The solution found adheres to the imprinted idea of a tiered approach. In addition, research contributors were granted a head start of using the data for analysis purposes. In contrast to sharing qualitative data, these discussions showed little concern about confidentiality issues due to the coded character of the shared material. [47]

STEP members perceived the global survey as an excellent opportunity to generate unique data that would not only move the project forward, but also enhance the field of family business research. At the same time, they now had to recruit survey participants on top of conducting case studies, which required additional resources and time. The increasing demands of the STEP project seemed to occur at a point in time, when most members already struggled to meet the qualitative requirements. The global board engaged a task force comprised of board members and researchers in different regions to revise the STEP good standing status. As a result, the global board increased the timeframe during which the required six cases studies had to be conducted from three to five years. In hindsight and from an observer's perspective, these changes turned out to be too moderate. In 2011 and 2015, the global board lowered the good standing criteria once more (after investigations of respective task forces), first to three cases within three years plus 20 surveys, and then to a choice between engagement in the qualitative or quantitative workloads. These gradual reductions to the STEP work package granted STEP members more time to meet the criteria to remain in good standing. At the same time, these reductions slowed the growth of the cumulative database and as such the kind and amount of data available for sharing. [48]

These multiple reductions of the workload were largely intertwined with a discussion about data sharing protocols. For example, in parallel to a newly revision of the good standing status in 2013, the global board formed a subcommittee that was given the task of reviewing current documents related to data access and usage. Over the years, researchers voiced a desire to get quicker access to the data and a concern about a responsible use of data conducted by others. Eventually, the global board did not lower the threshold for case access—the tiered approach remained in place. Instead, they pointed to co-authorship with existing partners as an option to gain immediate data access. Revisions of the data usage and access policies mainly touch upon formal

permission to data sharing and crediting primary researchers. For example, in 2013 the STEP protocol specified the following procedure for data sharing:

> "When engaging in case comparative work, STEP Protocol further requires that the member gains written permission from the case author and considers co-authorship with the case author. If doing case comparative work on a meta-level (i.e. using no names or case details) permission of the global board vs. individual permissions of case authors is required. Any publication using STEP data must clearly state that the data was gathered as part of the Global STEP Project and acknowledge the author(s) and institution(s) of the case studies used" (Internal Document: STEP protocol, 2013). [49]

With this and similar text passages, STEP policies once more underline the overall goal of creating research collaborations within an organized community. They further specify the relationship between primary and secondary researchers, thereby offering orientation to a community in flux. This community no longer exists of researchers who personally know each other, but is made up of founding members, established scholars in the field and those interested in learning the method and contributing to the field. [50]

The refined good standing criteria is accompanied by increasingly strict sanctioning policies in case these criteria are violated. For example, in 2014, the global board added a passage to the STEP policies explicating that members who were no longer in good standing lose their data access privileges. These rules also put the governance units of the STEP project in a position of ensuring adherence to these defined standards of collaborating, a role that the global board fulfills effectively. For example, as the global board learned that a STEP member used the case of another STEP member without permission and without anonymization for a conference proposal, they immediately contacted the submitting member. The issue was resolved promptly and cooperatively. The member offered to withdraw the submitted abstract. [51]

During these years of institutionalization, the STEP community completed their methodological design, reached the desired size of the community and implemented formal mechanisms for the discussion, codification, and enforcement of data sharing policies. The STEP project kept adjusting the object of data sharing and its conditions to the needs of the growing and changing community. Remarkably, these changes did not affect the very principle of data sharing as a corner stone of the STEP project nor the kind of data being shared (case reports and survey data). Instead the changes were focused on offering guidance in the design of collaborations between primary and secondary researchers. [52]

4.3 STEP 2.0: Rethinking data sharing for future use

A revision of the STEP membership agreement in April 2015 symbolizes the starting point of an emerging new self-definition. This membership agreement more strongly than ever emphasizes that members should engage in research dissemination efforts (e.g., book chapters, journal articles, and teaching cases). This shift in self-understanding occurs in an academic environment that becomes more and more competitive as entrepreneurship and family business research have developed into established research fields (DE MASSIS, SHARMA & CHUA, 2012; MELIN, NORDQVIST & SHARMA, 2013). Articles actually unlocking the case comparative potentials of the accumulated STEP data are few; out of the seventeen academic journal articles based on STEP data, only four draw on multiple case studies conducted by a single institution (IRAVA & MOORES, 2010; KANSIKAS, LAAKKONEN & VALTONEN, 2011; KANSIKAS, LAAKKONEN, SARPO & KONTINEN, 2012; ZELLWEGER & SIEGER, 2012). Only one article presents case comparative work based on data sharing across institutions (SIEGER, ZELLWEGER, NASON & CLINTON, 2011). Similarly, the contributions covered in the five STEP books mainly pertain to single case studies and conceptual advancements. [53]

STEP members largely agree to increase the number of publications based on that data, which involves both more effectively using the accumulated data stock and STEP's collaborative framework to conduct new data collection. During a board meeting in August 2015, a board member summarized these new trends as follows:

> "STEP is at a turning point. There is a change in methodology. In the beginning, we were very much focused on case studies and on the actual interaction between researchers. Now we have more developed regions. Scholars have evolved in their careers. We are able to do multisite research projects—harnessing our data" (Observational note: STEP board meeting, August 2015). [54]

This spirit sparks a lively discussion about the future of the STEP project. On the table are not only adjustments to the good standing requirements or data access policies, but rather the whole set up of the project. In reaction to these discussions, the global board passed a strategy paper, fondly called STEP 2.0, which reinforces the idea of being an organized research community with adaptation to new research requirements. The strategy paper defines seven priorities. These include increased collaboration within and between regions, rigorous research in published outlets, and new methods of collecting and sharing research data. The full STEP project for family enterprising strategy is depicted in the Appendix. [55]

Following the release of the strategy paper in the STEP board meeting minutes in November 2015, the global board initiated a number of changes. Two immediate responses were the commissioning of a new web-portal and the organization of an academic conference. The web-portal is expected to function as platform for researchers looking for collaborators on specific questions, as a repository for the

collected data, and as new tool to collect data. This platform is also expected to help present academic research to business families in a digestible form. The STEP academic conference, which is exclusive to STEP members and invited external scholars was held for the first time in October 2016, to foster new research collaborations and help advance publication efforts. In addition, the global board introduced an annual call for members, and offers the status of a "collaborator" to affiliations that are not full-members themselves, but work closely with members on the STEP data. STEP Europe has hired a "research champion" to assist affiliates in realizing case comparative work. Most recently, the STEP project has held its first webinar on how to build a successful research collaboration within STEP. [56]

The ongoing redefinition of the STEP project still values the idea of data-sharing. As Inga, Europe, voiced during an interview: "After all, data is what has brought us together." Yet, new ideas emerge of how this data sharing can be organized in the future. The vision departs from sharing data openly within the whole community that became increasingly dispersed and less personal. In fact, the idea to store clear transcripts centrally is seen very critically due to confidentiality reasons; there has even been an initiative to anonymize case reports, which however has not crystallized. Instead, the favored model includes the formation of smaller research teams on themes of their interest, in which each researcher conducts a small amount of data and contributes to the whole research team. This model reproduces the original idea of blending primary and secondary research on a smaller scale, more tightly coupled to specific research questions and limited in their duration of inquiry. What remains unanswered, at least at the moment, is how the gathered data treasure can be lifted and with what kind of methods researchers can go deeper into the accumulated cases. [57]

5. Potentials and Limits of Organized Research Communities

In this section, I reflect on the peculiarities of sharing data in the STEP project in contrast to informal and formal modes of sharing. First, I address potential reasons for the reluctance of STEP members to share primary data and to engage in re-use. Then I explore the limits and potentials of organized research communities in the dimensions of access, confidentiality, quality, and time horizon. [58]

5.1 Reluctance to share

Looking at the case as presented in the previous section a question arises: If the STEP members only make such little use of actual data sharing, is STEP and, more generally, are organized research communities an effective mode of sharing qualitative data? In this section, I examine the reluctance of STEP members to share and argue that a low degree of actual use is not indicative of the potential this form of sharing holds. Neither informal nor formal sharing are largely used practices despite the benefits reuse of data holds. [59]

Based on my empirical observations, I detect at least four components that may explain the reluctance to share among STEP members. The importance of each of these components varied tremendously over time, but they all led to the same result: a disinterest in harnessing the collected qualitative data. First, not all STEP members actually have a strong research interest. Especially in the South American and Asian regions, teaching is the dominant focus of STEP scholars and the main reason for joining the project. STEP offers a large collection of cases on successful family businesses, which serve as teaching examples. In this sense data sharing within STEP functions very well. Second, a large number of researchers engaged in STEP are actually more interested in participating in the quantitative survey and harnessing its results. Again, in this sense data sharing within STEP functions very well. Third, during my study I noticed a low degree of knowledge about secondary analysis and the discourse on archiving qualitative data among STEP scholars. As in the social sciences in general, this research practice has not yet reached the mainstream. Fourth, the STEP project conducts data around a very specific topic, which allows only for a small scope of alternative questions for secondary analysis. [60]

Nevertheless, STEP is still a compelling example as a hybrid mode of sharing qualitative data. This organized research community is not only constituted around the very idea of sharing data, but has preserved this intention over more than a decade. The STEP project succeeds in committing primary researchers to make their primary materials searchable and available within a predefined membership group. The attractiveness of the STEP project largely feeds from the ability to engage in networking and form collaborations, as the members of the global board agree. Such collaborations allow individual scholars to be part of a broader team beyond their home department. Each involved member has knowledge about the sampling plan, the interview guidelines and the theoretical underpinnings of the research framework. As such each STEP member has a basic, albeit limited, understanding of the context in which interviews are conducted, a feature detrimental to reanalyze qualitative data (McNEILL, 2016). STEP as organized research community expands the character of a co-production from an interviewer-interviewee situation to an interviewer-interviewee-community relationship. This is different from both other forms of sharing. This case study will thus look at the ways organized research communities structure the availability of qualitative data for sharing, its limitations and potentials in shaping an alternative approach to formal and informal sharing. [61]

5.2 Confidentiality: Data stewardship in research communities

In terms of protecting the research participants' confidentiality in organized research communities, the organizing entity acts as additional data steward. As the STEP case illustrates, the responsibility to safeguard sensitive information from misuse no longer rests upon the primary researcher alone (as in informal sharing). Similar to archives or repositories, the organizing entity acts as additional data steward with a reputation to ensure an ethically acceptable handling of the data, albeit the organized research community is unable to fully

take over this role as in some variations of formal sharing, i.e. when the deposited data are turned over into an archive's proprietorship. [62]

Confidentiality issues are a major concern in the scholarly debate about data sharing and within the STEP project. While the first data sharing policy of STEP encouraged open sharing or transcripts and audios within the community, this ideal was never actually practiced. Situated at the Babson College, the STEP project has followed their IRB requirements with regard to protecting research subjects. The issue of a confidential treatment has been a reoccurring theme in regional and global meetings. The most recent debate about potentially masking case reports illustrates the topicality of this issue. What is at stake here? [63]

During the semi-structured interviews, participants tell their story "beyond just the business aspect, it also relates to the family aspect" (Tiffany, North America). Interview participants reveal hints about intimate relations, give detailed accounts of family dramas, or narrate personal anecdotes. After an interview, business families are concerned that "information may come out that they do not want others to see in terms of their personal information about the family or something along those lines" (David, North America). As all qualitative researchers, STEP scholars have the ethical responsibility to handle sensitive data material confidentially. As Louise CORTI and others emphasize, it is usually the primary researchers, the ones the field, who provide guarantees as to how the data will be used and assure anonymity and representation (e.g., CORTI & THOMPSON, 2008; CORTI et al., 2000; PARRY & MAUTHNER, 2004). In addition, Scott, a colleague from North America, remarks that a misuse of data may not only harm the privacy of the research subjects, but also negatively affect the relationships between primary researcher and research subject. He says "it's their connections to these quite powerful families, many of which are sponsors of their school that they draw on a lot for their institutions." Aside from the duty to protect research participants, STEP members may thus feel an increased pressure to maintain the positive relationship they have established. [64]

The STEP data policies reflect these circumstances. According to the STEP Confidentiality Agreement the primary researchers will "serve as guardian of the data on behalf of the interview participants" (Internal Document: Version 2007). The decision to share interviews and transcripts remains with the primary researcher. As Jacob stated, "in my mind, whether I share the transcripts or not would always depend on whether the family was ok with it." This requires that primary researchers contact their interview subjects once more in order to obtain permission for re-use, or that the primary researcher assesses the sensitive nature of the transcript. In the latter case, sharing of primary data becomes dependent on an assumed willingness of the family to share (which in some countries even violates data protection laws). Having said this, the ultimate decision about the desired level of confidentiality depends upon the consent of the participating individual and any attempt to act upon their behalf can have paternalistic or patronizing effects resulting either in under- or overprotection of research participants that also constraint a climate for sharing with in the community (CARUSI & JIROTKA, 2009). [65]

Among STEP scholars, there is further a high awareness that case reports may be used by other scholars, which may cause researchers to withhold or downplay some of the sensitive information they have encountered with the best interest of their research subjects in mind. This bias of case reports certainly affects their use in case comparisons. Hence, for a valid and reliable data set, researchers always need to strike a sharing relationship and secure access to primary data materials. [66]

Yet, it is not only the primary researcher who is visible to the research participant as business families agree upon participating in the STEP project, a community of researchers jointly conducting and sharing data. The STEP project as organization appears on the website, consent forms, and business families meet fellow STEP researchers at local and global summits. In other words, from the very beginning the organizing entity of STEP lends additional legitimization and conveys an impression of professionalism to the research subject. Aware of this position, the STEP project has defined a number of standards that should ensure the privacy protection of participating families at all times and secure this unique data pool. These measures include signed consent forms of all participating families, non-disclosure agreements of all STEP researchers and collaborators, policies requiring written permission by the global board in order to use masked STEP data in research publications and written permission by primary researchers to use unmasked materials. The STEP board and administrative team take compliance to these policies seriously. In case of violation (e.g., using materials without written permission), they implement quick and consequential actions (e.g., warnings, downgrading the tired-access, expelling from the community). These measures taken to protect the privacy of research participants not only define use and misuse of data within the community, but release primary researchers to some extent from their duties as data stewards. For researchers, the negotiated protocols offer guidance in the process of sharing qualitative data; beyond that these policies imply an obligation to do so. [67]

The organizing entity as additional data steward and the primary researcher as last resort forms a construction that comes with a number of advantages with regard to confidentiality issues. It may be easier to convince interview participants as the organizing entity provides additional legitimization, reliable mechanisms release primary researchers from their duties of data protection. At the same time, we can learn from STEP that the organizing entity does not fully substitute the primary researcher as gatekeeper of the conducted materials in organized research communities, which creates a trade-off situation between exchanging sensitive data and protecting the research subject. The struggle for confidentiality in organized research communities becomes highly dependable on the level of trust generated within the community. [68]

5.3 Access: Trust and reciprocity within research communities

In terms of accessing primary data, a community mode of sharing structures the dialog between primary and secondary researcher. As the STEP case illustrates, sharing policies moderate the negotiations between both parties about the terms and conditions of sharing. [69]

From the perspective of a primary researcher, the STEP project demands trust in fellow STEP members to handle primary data appropriately. We learn from the STEP project that defining clear membership criteria and tasks contributes to building a trust within the community. These membership criteria are, however, in need of ongoing adjustment in order to maintain the level of trust among community members. Within STEP, new member and good standing criteria have been subject to constant discussion, especially on the board level, which have resulted in numerous changes over the past decade. Adjustments to membership status shall ensure trust in a professional and confidential use of data by fellow researchers even under changing conditions such as an increase in membership size, changing regional disparity, or dynamic research requirements. Heterogeneity among community members seems to go along with a heightened investment in creating trust within the community. Recent desires to share data in smaller, topic-centered teams indicate that the community may have reached a size where trust among all members is increasingly challenging to guarantee. [70]

STEP's tiered access to data symbolizes efforts to grant primary researchers more control over who gains access to their materials. The policy emerged as a reaction to the communities' growth and increased regional scope. As Scott, a STEP member from North America, voiced: "Uncertainty about new people coming in and other regions created the need to have some kind of tiered structure for it." The tiered access warrants familiarity with the methodological design and an understanding for the data's sensitivity prior to accessing other researcher's cases. In addition, the tiered approach implements an element of earning trust by proving research merit and dedication to the project. This data-policy can at least partially compensate for lacking personal knowledge of members accessing one's data. [71]

Finally, the STEP project only demands from primary researchers that they share case reports and case meta-data. Hence, the primary researcher remains in sole possession of full access to the primary data material and can decide upon individual requests, if she or he is willing to share. As such, the STEP project imitates a number of measures widely applied in formal sharing in order to increase the willingness of primary researchers to share their data. They define access limitations both through catering only to a specific interest group (STEP members) and through a tired-access model, in addition they only share classified materials (case-reports vs. transcripts). [72]

From the perspective of a secondary researcher, a topic-centered stock of data eases the task of locating reusable and relevant primary data, as long as they share an interest in the overall research topic of the community. In addition, the

STEP project provides guidelines and data policies that ensure a common standard in conducting and preparing the data material, which increases comparability. These features largely resemble domain repositories. But unlike most variations of formal sharing, secondary researchers need to engage actively with the primary researcher in order to gain full access to the data materials. This allows secondary researchers to draw from knowledge about the interview setting —or even to collaborate with primary researchers in the sampling of additional cases. STEP as organized research community realizes a preselection of potential collaborators through their membership application process. [73]

Carolin HAEUSSLER et al. (2014) point out that the likelihood of researchers' sharing of primary data increases when the organized research community is perceived to follow the norm of communalism. The experiences of the STEP project confirm this relationship. Although the primary researcher remains the locus of control over the primary materials, the underlying norm is that these materials are common goods. Therefore, when secondary researchers approach primary researcher to share data the question no longer is why a primary researcher should give another community member access to the data, but rather why they should not. Rene, Europe, speaks of a "reversal of the burden of proof" that suggests that openly sharing data follows the spirit of the STEP project unless, in individual situations, there are good reasons that speak against it. This norm of communalism eases the task of a secondary researcher to convince a primary researcher to share in comparison to informal sharing. [74]

Asking a data proprietor to share opens an exchange relationship similar to informal sharing. Previous studies describe successful informal data sharing as a practice of exchange for mutual benefit—primary and secondary researchers create immediate reciprocity. In this process, data proprietors consider perceived career benefits (such as co-authorship or citations), perceived efforts (such as the hours needed to anonymize transcripts or shared workload), and perceived risks (such as the loss of publication opportunities or criticism by other scientists) (KIM & ADLER, 2015; TENOPIR et al., 2011). Steven, North America, explains that for a collaboration "to be successful' it needs 'to meet the institutions' and individual scholars' expectations." He refers to tenure or improved teaching as examples of such expectations. Hence, sharing of primary data among STEP members still includes strategic considerations as in informal sharing. After all, some community members may be competing for positions, journal publications, grants, etc. in an increasingly competitive field—a situation that at least implicitly creates ambivalence to data sharing. [75]

STEP data policies moderate these negotiations and strategic considerations among STEP members and as such reduce the complexity of this task in comparison to informal sharing. For example, the use of case reports for members in good standing requires a citation of the data source and an acknowledgment of the primary researcher and their institution as common practice in formal sharing. The use of case reports with new or not yet affiliated members requires co-authorship with affiliated members as in informal sharing. The regulations in terms of citation or co-authorship in case of reusing transcripts

or audios, however, is not regulated and hence topic of negotiations among primary and secondary researcher. In addition, a tiered data access guarantees an ongoing inflow of new cases in return for access to the data stock. At the same time, the organized research community requires and allows for more influence of the secondary researcher in designing the exchange relationship as in formal data sharing arrangements. [76]

These reflections yield to the discovery of moderated reciprocity as particularity of accessing data within organized research communities. In sum, this mode of data sharing offers a realm in which primary researchers are approached by preselected members, follow an ideology of data sharing, and are offered guidelines in striking up collaborations including potential incentives (i.e., co-authorship). These measures moderate the exchange for mutual benefit. Organized research communities can thus affect how the reciprocal relationships of sharing primary data is designed, without acting as a mediating instance. [77]

5.4 Quality: Peer pressure in research communities

Accessing primary data allows an assessment of the quality, rigor, and adequacy of the conducted interview and the quality of data processing. Consequently, scholars who are convinced of the high quality of their interviews, transcripts and case reports, or are willing to be exposed to critique, may be more likely to share their primary materials. Inconsistency in the data material is an often unspoken hurdle in data sharing. [78]

The STEP project provides a number of documents supporting STEP scholars in the process of conducting a case study. Sampling rules, an interview guide, a sample consent form, and an exemplary case report set minimum requirements and ensure comparability across cases. During the write up of our first case report, we consulted multiple other case studies to get a sense of the good practices that go beyond these minimum standards. In conversations with other STEP scholars, I noticed that most of them were well informed of each other's case reports in terms of content, rigor, and style. Knowing that one's case report will be open to all other qualified STEP scholars creates peer pressure within the community. The sharing of case reports creates a constant orientation around the performance of others, continuously increasing best practices and setting trends in the case reporting style. [79]

The same type of peer pressure, however, appears to constrain sharing primary data within the STEP project. For example, when I contacted a STEP scholar hoping to receive access to interview transcripts, my request was denied, with the explanation that the transcripts had only been done by a student and not well and that I could not possibly use them for a secondary analysis. When I mentioned this anecdote to other STEP scholars, they explained that the STEP project does not offer explicit standards of good practice when it comes to interviewing, transcribing, and the steps of analysis—as they do for case reports. Further, as interviews are not shared openly or stored centrally, there is no institutional pressure to actually transcribe well or at all. Hence, the quality of interviews and

transcripts varies dramatically. During my interviews, Rebecca, a STEP member from Latin America, explained that in some regions scholars are under high demand for teaching with limited training in interviewing and time capacity to engage in research. She admitted to being unsure how many interviews have actually been conducted, "but I am almost sure that it is not professional." On the other hand, Charlotte, a STEP member from Europe, spoke elaborately about the measures they have taken to ensure the highest standards of interviewing and about the money they have invested for professional transcription services. In other words, the quality of conducted interviews and of data processing is strongly contingent upon the performance and resources of affiliated institutions. [80]

As these findings suggest, in organized research communities the obligation to expose data (i.e., case reports) openly has the potential to function as a powerful mechanism to professionalize qualitative research. Although Martyn HAMMERSLEY (1997) doubts the impact of data re-use as a mode of auditing in an organized research community, data sharing induces peer control as an efficient and flexible means of increasing research standards. We can learn from STEP that for this positive effect to unfold, it is not sufficient to develop standards of good interviewing, data processing or management. The need to expose oneself in a community of colleagues and collaborators, who create the peer pressure, calls primary researchers to go beyond these standards. This feature is unique to organized research communities and to some forms of formal sharing, where the circle of participating scholars is limited, part of the same field and competing for reputation, jobs, and research funding. In informal sharing, however, peer pressure can prohibit the installation of sharing relationship and in open, formal sharing the pressure created by an anonymous crowd may not be as forceful. [81]

5.5 Time horizon: The finitude of community sharing

The last area addressed in this article touches upon the timing of data sharing, an issue that emerged as relevant for a community mode of sharing during the analysis. With timing, I refer to the ideal timeframe in which data sharing is fruitful. [82]

In informal sharing, the ideal time to access data material is immediately after primary researchers have finished their study. The maximum timeframe for data sharing is limited to the life times of the researchers, but the advantage of drawing from their rich context knowledge decreases rapidly over time. Hence, the time horizon to share data in an informal mode is short. In contrast, data access in formal sharing occurs even later then in informal sharing as primary researchers have to prepare their materials for data curation. Secondary analysis of curated data is dependent on the context documentation of the primary researcher and requires a treatment of these data as historic documents, but allows for an extended re-use period. [83]

Organized research communities can shorten the time lag between primary and secondary use of data materials substantially, which also increases the

advantage for drawing from rich context knowledge. In the STEP project, data material becomes available immediately after it has been conducted (if transcripts and audios are requested directly) or during the analysis as soon as the STEP member has submitted the case report—usually within one year. The opportunity to shape and direct the re-use of data in synch with primary research is unique to research collaborations as fostered in organized research communities and has been described as valuable form of secondary analysis (BISHOP, 2007; HINDS, VOGEL & CLARKE-STEFFEN, 1997). Quick re-use and close collaboration with primary researchers allows harnessing of qualitative data to the fullest. [84]

In organizing a research community, interactive data sharing creates momentum in comparison to informal sharing. In the STEP research community, data sharing seems to work best when multiple researchers are conducting and analyzing case studies simultaneously. For example, during the early stages of the STEP project researchers within the European region had already conducted several case studies. To investigate their issue of interest further, they requested re-use of case reports of several other newly joined members who had just finished their case reports. Sven remembers: "This was a hot, hot phase. They [the new STEP members] looked at us as champions squeezing the best out of their cases and afraid that nothing would be left for them. Eventually, they understood that we are acting in the spirit of the STEP project and do what we are supposed to— compare our cases." The research team invested a lot of time and effort into establishing sharing relationships with other members always pronouncing that their cases are open for re-use as well. Other successful comparative case studies have been conducted by groups of doctoral students that met within the STEP community and joined forces in investigating their topics of interest. A shared midrange time-horizon of the engagement in doing case studies can be communicated well to the participating families, and appears to provide researchers with a perspective for joint publication, and ultimately to achieve their individual academic goals (such as tenure). The vision to create topic centered teams within the STEP project that jointly work on a specific question for a set time, maybe able to leverage the positive effects of a community mode of sharing qualitative data. [85]

Once the time range of simultaneous research has passed, the interest in the case reports preserved in the STEP repository seems to decrease. When prompting STEP scholars with this observation, they voice concerns about the comparability of the case studies, some of which are over a decade old. They explain that over time new theoretical nuances of research impose themselves on the focus of the interviews and their analysis, which lessens the comparability between interviews and between cases. In addition, primary researchers may have left the affiliated institution or the STEP project and thus their primary material is no longer available to the community. [86]

After this period of simultaneous engagement, the primary data appears to turn into historic documents that require different forms of treatment even within organized research communities (BROOM et al., 2009). There is a need to adapt the way these "aging" primary materials are stored (for example in a multi-media

archive as in the Timescapes project) and to provide more detailed meta-data on changes in theoretical approaches and selected focus topics (BIRKE & MAYER-AHUJA, 2017; HOLLAND, 2011; IRWIN & WINTERTON, 2011). Hence, after some time, data sharing in organized research communities increasingly resembles elements of formal data sharing. A community mode of sharing is thus a midrange option. [87]

6. Conclusion

This article contributes to the discourse on sharing qualitative data as it suggests organized research communities as a distinct mode to practice sharing. Drawing on the example of the STEP project, I have argued that a community can establish sharing norms, implement rewards, and provide resources that may be beneficial to the sharing of qualitative primary data. This, however, involves ongoing effort, encouragement, and diplomacy to negotiate and nourish sharing policies that accommodate differing priorities among scholars (NEALE et al., 2012). [88]

A key feature of organized research communities is the simultaneous engagement in the use and re-use of qualitative data. This allows scholars to harness data jointly for research output. This also implies an intimate familiarity of all scholars—primary or secondary—with the research context, increasing reflexivity upon the intersubjective and processual character captured in qualitative data (BISHOP, 2007; HINDS et al., 1997). [89]

The above discussion highlights the potential of organized research communities to upscale qualitative data sharing. Clearly negotiated sharing policies provide guidance in the implementation of data sharing and increases the transparency of who will get access to one's data. These sharing policies also encapsulate previously negotiated ways of achieving reciprocity among sharing participants, which reduces complexity in striking up a sharing relationship—an opportunity, which I have termed moderated reciprocity. Within organized research communities, peer pressure can function as an effective measure to raise quality standards and rigor of data conduction and data processing. Pragmatic challenges arise in detailing and adapting data sharing policies. Specifically, with respect to privacy issues, this article has shown that organized research communities can function as additional data stewards, but do not resolve the issue of dependency on the primary researcher to decide ultimately on the protection of a research subject. [90]

The elaboration upon the particularities of STEP as an example of an organized research community has evidenced a number of parallels to formal data sharing. I thus suggest that organized research communities should look at repositories as role models for designing their own sharing-policies. Best practices in formal sharing; for example, with regard to informed consent forms, levels of access, or standards in anonymization, can function as valuable guidelines. In addition, organized research communities can learn from formal sharing about the importance to facilitate reuse and to teach about secondary analysis. In turn,

organized research communities may function as cultural agents, as in this realm scholars practice the sharing of data and learn about the usefulness of using material that one has not conducted. [91]

The strength of this study is its in-depth investigation of a single case, which at the same time forms its largest limitation. The findings presented here are subject to research bias, although a number of measures (reflexivity, data triangulation, and peer-review) were used to ensure validity of data and replicability of their interpretation. Future research that explores other current or historic examples of organized research communities will help to further our understanding of this type of data sharing. In addition, I have touched upon a number of pragmatic aspects that are worth exploring in future research. Specifically, an analysis of teamwork and group dynamics in organized communities would deepen our understanding of how communities negotiate, uphold and discard data sharing policies. The effects of member exit on research communities and data sharing, as well as variance in the treatment of different kinds of research data (ethnographic data or interview data) present further avenues for future research. [92]

Overall, I have argued that organized research communities are a hybrid form of data sharing particularly adequate for the context depend and intersubjective character of qualitative data. They offer a sense of collaborative, sustainable, and interoperable way of dealing with qualitative data (CARLIN & VAUGHAN, 2016). Organized research communities uniquely facilitate an intertwinement of primary and secondary analysis, and thus embrace the very idea of what SMIOSKI (2013) calls a living archive. In organized research communities, the depositing and sharing of data is not an end state, but is considered early in the research process, in which primary and secondary researchers are equally and actively involved (ibid.). [93]

This form of organizing data sharing, however, is a mid-range solution, as simultaneous engagement in use and re-use of qualitative data is of a temporary nature. At a yet undefined point in time, the collected data material turns into historic data tilting the community mode of sharing data towards data curation. This implies that to share qualitative data it could be useful to consider two stages of data sharing. Each of these stages go along with different sets of secondary analysis methods that can be applied to the data material. Each of these stages could be explicitly communicated to interview participants thus giving them a choice in terms of how long and under what conditions they would like to make their data available for academic research. [94]

Acknowledgments

Appendix: Reprint of the STEP Project for Family Enterprising Strategy (2015-2017)

Click here to download the PDF file.

References

Birke, Peter & Mayer-Ahuja, Nicole (2017). Sekundäranalyse qualitativer Organisationsdaten. In Stefan Liebig, Wenzel Matiaske, & Sophie Rosenbohm (Eds.), *Handbuch Empirische Organisationsforschung* (pp.105-126). Wiesbaden: Springer Gabler.

Bishop, Libby (2007). A reflexive account of reusing qualitative data: Beyond primary /secondary dualism. *Sociological Research Online*, *12*(3), http://www.socresonline.org.uk/12/3/2.html [Accessed: January 2, 2018].

Bishop, Libby & Kuula-Luumi, Arja (2017). Revisiting qualitative data reuse: A decade on. *SAGE Open*, *7*(1), http://journals.sagepub.com/doi/10.1177/2158244016685136 [Accessed: January 2, 2018].

Bishop, Libby & Neale, Bren (2011). Sharing qualitative and qualitative longitudinal data in the UK: Archiving strategies and development. *IASSIST Quaterly*, *34-35*, 23-29, http://repository.essex.ac.uk/2448/ [Accessed: January 2, 2018].

Bozeman, Barry & Gaughan, Monica (2011). How do men and women differ in research collaborations? An analysis of the collaborative motives and strategies of academic researchers. *Research Policy*, *40*(10), 1393-1402.

Broom, Alex; Cheshire, Lynda & Emmison, Michael (2009). Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sociology*, *43*(6), 1163-1181.

Carlin, David & Vaughan, Laurene (Eds.) (2016). *Digital research in the arts and humanities*. *Performing digital: Multiple perspectives on a living archive*. London: Routledge.

Carusi, Annamaria & Jirotka, Marina (2009). From data archive to ethical labyrinth. *Qualitative Research*, *9*(3), 285-298.

Cliggett, Lisa (2013). Qualitative data archiving in the digital age: Strategies for data preservation and sharing. *Anthropology Faculty Publications*, *1*, https://uknowledge.uky.edu/anthro_facpub/1 [Accessed: January 13, 2018].

Coltart, Carrie; Henwood, Caren & Shirani, Fiona (2013). Qualitative secondary analysis in austere times: Ethical, professional and methodological considerations. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *14*(1), Art. 18, http://dx.doi.org/10.17169/fqs-14.1.1885 [Accessed: January 2, 2018].

Corti, Louise (2011). The European landscape of qualitative social research archives: methodological and practical issues. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *12*(3), Art. 11, http://dx.doi.org/10.17169/fqs-12.3.1746 [Accessed: January 2, 2018].

Corti, Louise & Thompson, Paul (2008). Secondary analysis of archived data. In Clive Seale (Ed.), *Qualitative research practice* (pp.327-343). London: Sage.

Corti, Louise; Day, Annette & Backhouse, Gill (2000). Confidentiality and informed consent: Issues for consideration in the preservation of and provision of access to qualitative data archives. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *1*(3), Art. 7, http://dx.doi.org/10.17169/fqs-1.3.1024 [Accessed: January 2, 2018].

Creswell, John W. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Los Angeles, CA: Sage.

Davies, Charlotte Aull (Ed.) (2010). *Reflexive ethnography: A guide to researching selves and others* (2. ed.). London: Routledge.

De Massis, Alfredo; Sharma, Pramodita & Chua, Jess H. (2012). State of the art of family business research. In Alfredo de Massis, Pramodita Sharma & Jess H. Chua (Eds.), *Family business studies. An annotated bibliography* (pp.10-46). Cheltenham: Edward Elgar Publishing.

Emerson, Robert M.; Fretz, Rachel I. & Shaw, Linda L. (2011). *Writing ethnographic fieldnotes* (2. ed.). Chicago, IL: The University of Chicago Press.

Fielding, Nigel (2004). Getting the most from archived qualitative data: epistemological, practical and professional obstacles. *International Journal of Social Research Methodology*, 7(1), 97-104.

Fink, Anne S. (2000). The role of the researcher in the qualitative research process. A potential barrier to archiving qualitative data. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *1*(3), Art. 4, http://dx.doi.org/10.17169/fqs-1.3.1021 [Accessed: June January 2, 2018].

Flick, Uwe (2015). Mapping the field. In Uwe Flick (Ed.), *The Sage handbook of qualitative data analysis* (pp.3-18). London: Sage.

Gebel, Tobias; Grenzer, Matthis; Kreusch, Julia; Liebig, Stefan; Schuster, Heidi; Tscherwinka, Ralf; Watteler, Oliver & Witzel, Andreas (2015). Verboten ist, was nicht ausdrücklich erlaubt ist: Datenschutz in qualitativen Interviews. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *16*(2), Art. 27, http://dx.doi.org/10.17169/fqs-16.2.2266 [Accessed: January 2, 2018].

Haeussler, Carolin; Jian, Lin; Thursby, Jerry & Thursby, Marie (2014). Specific and general information sharing among competing academic researchers. *Research Policy*, *43*(3), 465-475.

Hammersley, Martyn (1997). Qualitative data archiving: Some reflections on its prospects and problems. *Sociology*, *31*(1), 131-142.

Hammersley, Martyn (2009). Can we re-use qualitative data via secondary analysis? Notes on some terminological and substantive issues. *Sociological Research Online*, *15*(1), http://www.socresonline.org.uk/15/1/5.html [Accessed: January 2, 2018].

Heaton, Janet (2008). Secondary analysis of qualitative data. In Pertti Alasuutari, Leonard Bickman, & Julia Brannen (Eds.), *The Sage handbook of social research methods* (pp.506-519). London: Sage.

Helferich, Cornelia (2014). Leitfaden- und Experteninterviews. In Nina Baur & Jörg Blasius (Eds.), *Handbuch Methoden der Empirischen Sozialforschung* (pp. 559-574). Wiesbaden; Springer VS.

Hinds, Pamela S.; Vogel, Ralph J. & Clarke-Steffen, Laura (1997). The possibilities and pitfalls of doing a secondary analysis of a qualitative data set. *Qualitative Health Research*, *7*(3), 408-424.

Holland, Janet (2011). Timescapes: Living a qualitative longitudinal study. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *12*(3), Art. 9, http://dx.doi.org/10.17169/fqs-12.3.1729 [Accessed: January 2, 2018]

Irava, J. Wayne & Moores, Ken (2010). Clarifying the strategic advantage of familiness: Unbundling its dimensions and highlighting its paradoxes. *Journal of Family Business Strategy*, *1*(3), 131-144.

Irwin, Sarah & Winterton, Mandy (2011). Debates in qualitative secondary analysis. *Timescapes Working Paper Series*. *4*, http://www.timescapes.leeds.ac.uk/assets/files/WP4-March-2011.pdf [Accessed: January 2, 2018].

Johnson, Burke R. (1997). Examining the validity structure of qualitative research. *Education*, *118*(2), 282-292.

Kansikas, Juha; Laakkonen, Anne & Valtonen, Heli (2011). In search of family business continuity: The case of transgenerational family entrepreneurship. *International Journal of Entrepreneurship and Small Business*, *13*(2), 193-207.

Kansikas, Juha; Laakkonen, Anne; Sarpo, Ville & Kontinen, Tanja (2012). Entrepreneurial leadership and familiness as resources for strategic entrepreneurship. *International Journal of Entrepreneurial Behaviour & Research*, *18*(2), 141-158.

Kim, Youngseek & Adler, Melissa (2015). Social scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. *International Journal of Information Management*, *35*(4), 408-418.

Knoblauch, Hubert (2014). Ethnographie. In Nina Baur & Jörg Blasius (Eds.), *Handbuch Methoden der Empirischen Sozialforschung* (pp. 521-528). Wiesbaden: Springer VS.

Kvalheim, Vigdis & Kvamme, Trond (2014). Policies for sharing research data in social sciences and humanities: A survey about research funders' data policies. *International Federation of Data Organization for Social Sciences*, Norway. http://ifdo.org/wordpress/wp-content/uploads/2015/07/ifdo_survey_report.pdf [Accessed: January 2, 2018].

LaRossa, Ralph (2012). Writing and reviewing manuscripts in the multidimensional world of qualitative research. *Journal of Marriage and Family*, *74*(4), 643-659.

Leh, Almut (2000). Problems of archiving oral history interviews. The example of the Archive "German Memory". *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *1*(3), Art. 8, http://dx.doi.org/10.17169/fqs-1.3.1025 [Accessed: January 2, 2018].

Li, Eldon Y.; Liao, Chien H. & Yen, Hsiuju R. (2013). Co-authorship networks and research impact: A social capital perspective. *Research Policy*, *42*(9), 1515-1530.

Mason, Jennifer (2007). "Re-using" qualitative data: On the merits of an investigative epistemology. *Sociological Research Online*, *12*(3), http://www.socresonline.org.uk/12/3/3.html [Accessed: January 2, 2018].

Mauthner, Natasha S.; Parry, Odette & Backett-Milburn, Kathryn (1998). The data are out there, or are they? Implications for archiving and revisiting qualitative data. *Sociology*, *32*(4), 733-745.

McNeill, Katherine (2016). Repository options for research data. In Burton B. Callicott, David Scherer & Andrew Wesolek (Eds.), *Charleston insights in library, archival and information sciences. Making institutional repositories work* (pp.15-30). West Lafayette, IN: Purdue University Press.

Medjedović, Irena & Witzel, Andreas (2010). *Wiederverwendung qualitativer Daten: Archivierung und Sekundärnutzung qualitativer Interviewtranskripte*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Melin, Leif; Nordqvist, Matthias & Sharma, Pramodita (Eds.) (2013). *The Sage handbook of family business*. London: Sage.

Merton, Robert K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago, IL: University of Chicago Press.

Neale, Bren; Henwood, Karen & Holland, Janet (2012). Researching lives through time: An introduction to the Timescapes approach. *Qualitative Research*, *12*(1), 4-15.

Parry, Odette & Mauthner, Natasha S. (2004). Whose data are they anyway? Practical, legal and ethical issues in archiving qualitative research data. *Sociology*, *38*(1), 139-152.

Richardson, Lizzie (2015). Performing the sharing economy. *Geoforum*, *67*, 121-129.

Saunders, Benjamin; Kitzinger, Jenny & Kitzinger, Celia (2015). Anonymising interview data: Challenges and compromise in practice. *Qualitative Research*, *15*(5), 616-632.

Sieger, Philip; Zellweger, Thomas; Nason, Robert S. & Clinton, Eric (2011). Portfolio entrepreneurship in family firms: A resource-based perspective. *Strategic Management Journal*, *5*(4), 327-351.

Slavnic, Zoran (2013). Towards qualitative data preservation and re-use—Policy trends and academic controversies in UK and Sweden. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *14*(2), Art. 10, http://dx.doi.org/10.17169/fqs-14.2.1872 [Accessed: January 2, 2018].

Smioski (formerly Jesser), Andrea (2011). Infrastructure, acquisition, documentation, distribution. Experiences from WISDOM, the Austrian Data Archive. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *12*(3), Art. 18, http://dx.doi.org/10.17169/fqs-12.3.1734 [Accessed: January 2, 2018].

Smioski, Andrea (2013). Archiving strategies for qualitative data. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, *14*(3), Art. 5, http://dx.doi.org/10.17169/fqs-14.3.1958 [Accessed: January 2, 2018].

Tenopir, Carol; Allard, Suzie; Douglass, Kimberley; Aydinoglu, Arsev Umur; Wu, Lei; Read, Eleanor; Manoff, Maribeth & Frame, Mike (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE*, *6*(6), e21101, http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0021101 [Accessed: January 2, 2018].

Thorne, Sally (1994). Secondary analysis in qualitative research: Issues and implications. In Janice M. Morse (Ed.), *Critical issues in qualitative research methods* (pp.263-279). Thousand Oaks, CA: Sage.

Yin, Robert K. (2013). *Case study research: Design and methods* (5. *ed.).* Los Angeles, CA: Sage.

Zellweger, Thomas & Sieger, Philip (2012). Entrepreneurial orientation in long-lived family firms. *Small Business Economics*, *38*(1), 67-84.

Author

Dr. *Isabell STAMM* is head of the research group "Entrepreneurial Group Dynamics," founded by Volkswagen Foundation and located at the Department for Sociology at the Technical University Berlin. Beside her engagement in sociology of entrepreneurship, she is interested in cooperative aspects of social research. In the past she has been a fellow at the Berkeley Institute of Data Science, a member of the D-LAB Qualitative working group and has taught qualitative methods on all levels. She has engaged in data sharing at the STEP project and is currently combining online teaching and crowdsourcing to generate data about the trajectories of entrepreneurial groups.

Contact:

Dr. Isabell Stamm

Technical University Berlin
Department for Sociology
Fraunhofer Str. 33-36
10578 Berlin, Germany

E-Mail: Isabell.stamm@tu-berlin.de
URL: http://www.isabellstamm.de/,
http://www.entrepreneurialgroups.org/

Citation