

# ADVANCED QUALITATIVE RESEARCH

A GUIDE TO USING THEORY

MICHELLE O'REILLY  
& NIKKI KIYIMBA



Los Angeles | London | New Delhi  
Singapore | Washington DC | Boston

# SEVEN

## Primary and secondary data analysis

### CHAPTER CONTENTS

Introduction	129
The difference between data and research data	129
Primary and secondary data analysis	132
Debates about secondary analysis	137
Summary and final thoughts	142
Case study and reflective questions	144

### INTRODUCTION

In this chapter we examine the distinction between primary and secondary data, a distinction that we regard as overly simplistic and misleading. In order to clarify this particular aspect of the debate we differentiate the different kinds of data as they relate to qualitative research. From this platform a critical discussion of the parameters of secondary analysis is given. Within this discussion we consider the potential differences between reuse of one's own data and secondary analysis. This requires us to investigate the fundamental issues of context, quality and ethics. The chapter concludes with consideration of the costs and benefits of archiving qualitative data for the purposes of secondary analysis.

### THE DIFFERENCE BETWEEN DATA AND RESEARCH DATA

In common parlance, the word *datum* is a singular piece of information or fact, and the plural of this is *data* (<http://www.yourdictionary.com/datum>). However,

for the purposes of research this generic definition is insufficient. Generally speaking, collections of information do not become research data until they have been assembled intentionally by the researcher for the purposes of answering a particular research question. However, we acknowledge that the term 'data' is often used by researchers in different ways.

e.g.

A newspaper article may be data in the sense that it is a collection of 'facts', but it only becomes research data when a researcher specifically assembles a number of articles together with the intention of addressing a research question and undertaking analysis. Similarly, a series of Facebook posts are simply an example of social media, but become identifiable as research data when printed as a data corpus for the purposes of research.

A common differentiation between different types of research data has been to categorise them as 'primary data' or 'secondary data'. These have been understood differently within different disciplines. For example, there is a discrepancy between what historians mean by primary versus secondary sources and how social scientists employ the concepts. While the literature on this distinction has been rather sparse, primary data are generally understood to be the initial data specifically collected by the original researcher for their research purposes, whereas secondary data are understood to be collected by someone other than the researcher. In other words, primary data have been considered to be the face-to-face collection (such as interviews and focus groups), whereas secondary data have been understood to be existing available sources. It is important to make a distinction at this point between secondary data and secondary data analysis, which is the reanalysing of data collected by another researcher, a point we return to later in the chapter.

We see this distinction between primary and secondary data as problematic for qualitative research. The reasons for this are that this differentiation fails to account for the arguments related to 'researcher-generated' and 'naturally occurring' data (as discussed in Chapter 4). It also fails to account for the defining parameters that distinguish the particular characteristics of 'research data' from general information which is colloquially referred to as 'data'.

e.g.

A set of recordings of telephone calls to a customer services helpline exists as data in the sense that they are recorded information. However, it becomes a naturally occurring research data set if, having met ethical criteria, it becomes a corpus of research data for the purpose of analysis in order to address a research question.

While we acknowledge that there are several ways in which data might be differentiated, we also propose one particularly useful distinction between privately solicited research data and publicly available data which could form part of a research data corpus. In the example above, pre-existing data were privately

secured by the researcher through negotiations and ethical procedures with the organisation. In this sense they are privately solicited research data. In contrast, some naturally occurring research data are publicly available, such as television documentaries and newspaper and magazine articles. The reason we offer this distinction between privately solicited and publicly available, naturally occurring data is that the former data corpus would solely be available to that individual researcher whereas publicly available data could potentially be simultaneously collected by numerous researchers. This has implications for ways in which secondary data *analysis* may be defined, an argument we address shortly.

e.g.

In the case of analysing YouTube video clips of pedagogical guides to craft activities, these naturally occurring data are publicly available to potentially become research data. However, a researcher may post an advert on the internet requesting permission to record the activities of a specific 'knitting circle'. This would then be naturally occurring, privately solicited research data.

Furthermore, in the case of audio-visual data, some qualitative research traditions in particular argue that the actual recording is the primary data and the transcription is a representation of that recording. The transcript therefore would not in and of itself be considered to be data; rather the original recording would be viewed as actual data. In other words, the transcript itself does not constitute research data; rather it is seen as a tool to enable the practical task of analysis and dissemination. From a pragmatic perspective, however, some researchers argue contrary to this, that the recording and the transcript can be considered and treated as data.

In summary, within the field of qualitative research we argue that the distinction between primary and secondary research data is misleading. We propose instead that more useful ways of differentiating research data are to consider:

- Data and research data – the difference between information and information specifically collected together for the purposes of addressing a research question.
- Naturally occurring and researcher-generated research data – the difference between data that are used for research but would have existed anyway, and data created through interviews or focus groups etc. by a researcher.
- Publicly available research data and privately solicited research data – the difference between data available to anyone which can be used for research purposes, and data which are deliberately collected from a sample for research.
- Research data and representations thereof – the difference between original audio or visual data and written, transcribed representations of those data used for the purposes of convenient analysis and dissemination.

We recognise that these categories are not discrete and are likely to overlap. For example, data may be both privately solicited and naturally occurring, such as video recordings of counselling sessions that are recorded as standard practice

which may be privately collected for research. We have now established what we propose are theoretically and pragmatically more appropriate and sensible ways of differentiating between types of data and research data. This offers us the foundation to now move on to the second issue, which is the importance of distinguishing primary and secondary data *analysis*.

## PRIMARY AND SECONDARY DATA ANALYSIS

There has been a well-established tradition of reanalysing quantitative data, but in qualitative research there has not been a research culture of secondary analysis (Corti and Thompson, 2004), with limited progress of secondary analysis in this field (Fielding, 2004). Traditionally secondary data analysis has been understood as analysis of data gathered for a previous study by another researcher (Heron, 1989), requiring the use of pre-existing qualitative data (Heaton, 2008). Heaton (2008) argued that secondary data analysis in qualitative research can be approached in three modalities as listed in Table 7.1.

**Table 7.1** Modalities of secondary analysis (Heaton, 2008)

<i>Modality</i>	<i>Description</i>
Formal data sharing from publicly available data	The researcher accesses data sets that have been deposited in public archives by other researchers. These are then subjected to secondary analysis. These data were likely to have been well documented for archiving purposes and are likely to have met the necessary ethical and legal requirements for being shared with others.
Informal data sharing between researchers	For this type there are different possibilities. The primary researcher may give their data to another researcher and then have no subsequent involvement in the secondary analysis, or two or more researchers may pool their data sets and work independently in carrying out the secondary analysis.
Reuse of one's own data	The researcher may reuse their own self-collected data as a way of investigating new or additional questions that were not explored in the primary research, or they may use secondary analysis to verify their previous findings. The researcher may also team up with others to combine their respective data sets.

There are a number of proposed benefits for the secondary analysis of qualitative data. Far more than for quantitative data, qualitative research generates significant amounts of material which is rich in detail; this can often mean that much of it remains unanalysed (Äkerström et al., 2004). From an ethical perspective, therefore, it is respectful to participants to make the best use of data that are collected (O'Reilly and Parker, 2014). Furthermore, for potentially 'overresearched' groups or hard-to-reach participant groups, secondary analysis allows for a wider use of the data (Heaton, 1998). In addition, from a very practical point of view,

secondary data analysis is a cost-effective, convenient approach and is arguably a credible method for generating knowledge (Heron, 1989). Thus secondary analysis can be used to investigate additional or new research questions and arguably can be used to verify the findings of previous research (Heaton, 2008). It has also been suggested that secondary analysis allows the researcher an opportunity to view the data with some detachment, which arguably is more challenging for the researcher who originally collected the data (Szabo and Strang, 1997).

e.g. Thorne et al. (2001) conducted a secondary analysis of a large primary data set of interviews with a wide range of participants on the topic of oral health in long-term care. In this study Thorne et al. reanalysed an existing database to address a new research question.

Some of the key benefits and limitations of secondary analysis have been discussed in the literature, and Hammersley (1997, 2010b) has summarised some of these key debates as illustrated in Box 7.1.

### influential voices

#### Box 7.1 Contentions in secondary data analysis

##### Martyn Hammersley

Martyn Hammersley\* is a professor of educational and social research at the Open University, UK. He has written extensively on qualitative methodological issues in relation to theory and practice and has raised some questions about secondary data analysis. Hammersley wrote two seminal articles on this topic in which a number of salient issues were raised. We distil the main points raised by these two articles in order to illustrate the contribution to the debates that have been made.

Hammersley (1997) predicted that data archiving for the purposes of secondary analysis was likely to become much more common in the future. To some extent this prediction has been borne out and there has been a greater drive towards archiving data. However, Hammersley (1997) drew attention to a number of problems for both the archiving process and for secondary analysis, and indeed these have hindered the culture of archiving in contemporary practice. In this article there were several issues raised, and we summarise the three core points here:

- As data are socially constructed through the process of data collection and through the interaction of researchers and their participants, the findings are not *found* in the data but rather are created through those interactions in particular locations and at particular times.

(Continued)

*(Continued)*

- Even where contextual information is available to secondary analysts there is still considerable scope for different views regarding what types of data are required in order to establish the credibility and validity of the claims.
- As secondary analysts are likely to have different concepts and frames of reference from the original researcher, there are likely to be problems of 'fit' causing gaps in information about the data.

Overall, therefore, Hammersley (1997) argued that the archiving of qualitative data had potential for assessing the validity of particular studies and increasing the scope for secondary analysis, but the problems and challenges may limit the contribution that it is able to make.

In a later article, Hammersley (2010b) debated the benefits and limitations of secondary analysis of qualitative data. He noted that there have been conflicting views about the benefits and practical challenges. Within those broad debates Hammersley focused on the challenge of the data not fitting the research questions, that relevant contextual information was mostly absent in secondary analysis; this was a problem which he also argued was found in some primary analysis. While by no means a comprehensive overview of the article, we explicate the key points that he makes in this piece:

- Arguably there is some ambiguity in the meaning of various terms used when talking about secondary analysis, including the word 'data' which is routinely used but rarely defined. Hammersley suggested that we should use these labels in consistent ways in order to avoid confusion.
- It is argued that it is necessary to differentiate between research employing data that have been produced by another researcher and one's own data – in other words, between reanalysing data produced by another researcher and reanalysing one's own data.
- It is important to recognise that what counts as data is determined functionally within the research process. As data cannot exist independently of the research process, they are formed by the process in which they are given meaning. Thus a process is involved within which data are collected and selected in order to provide evidence relevant to candidate answers to the research question.
- The research question is essential to the process and guides the researcher's selection and interpretation of what constitutes appropriate data, and is pertinent to decisions regarding what can be legitimately inferred from that data.
- It is argued that there is a legitimate differentiation between reuse (by a researcher of her or his own data) and secondary analysis of data (produced by others), advocating that both can be of value when employed appropriately.
- In research teams where researchers analyse data collected by another member of the team, this should not be considered either reuse or secondary analysis, but as still constituting primary analysis.

Overall Hammersley (2010b) argued that the effect of the difficulties of secondary qualitative analysis are in reality much less severe if researchers are able to take into account that all data are constituted and reconstituted within the research process, but concludes that there is little that can completely eliminate the difficulties of 'context' and 'fit'.

\*We thank Professor Hammersley for taking the time to read an earlier draft of this chapter.

### Debates about reuse of data as a form of secondary analysis

From reading the contribution of Hammersley, it should be clear that there is some contention regarding definitions and terms. Some researchers consider the reuse of data to be a form of secondary analysis. However, this perspective has been debated with different views represented in the qualitative field. We argue that there is a logical rationale for differentiating between reuse of one's own data at a later stage with the generation of a new research question, and the secondary analysis of data collected by another researcher for different purposes.

Fielding (2004) argued that qualitative researchers need to engage with the debate over secondary analysis by making a clear distinction between what they define as 'reuse of archived data' and what they define as 'secondary analysis'. A difficulty for qualitative research has been in relation to terminology, and specifically to whether the reuse of data from a previous project is sufficiently distinctive to require its own label such as 'secondary analysis' (Hammersley, 2010b).

However, the problem which arises when trying to create a distinction between primary and secondary analysis is that qualitative work by nature tends to be driven by open broad research questions. Several papers may be produced over an extended period of time in attempting to address the general research question. Nevertheless, the mere fact that material has been written up does not mean that it cannot be revisited and reanalysed by the same researcher (Åkerström et al., 2004). This flexibility to return to data sets also compounds the issue that there is a lack of a clear distinction between the terms 'use' and 'reuse' (Hammersley, 2010b). Therefore, distinguishing between primary analysis and the reuse of one's own data for secondary analysis poses challenges in determining whether a new research question is being addressed. We argue that what distinguishes reuse or reanalysis from primary analysis is the generation of a new research question relating to the same data set. We also argue that reuse in a technical sense is not a form of secondary analysis. Furthermore, it is plausible that available data may be simultaneously analysed by different researchers addressing different research questions.



e.g.

Two unconnected researchers may both access the same publicly available data from the internet without being aware that the other is engaging in analysis of those data. In this case both the researchers are engaging in simultaneous primary analysis.

We propose, then, that there are four forms of analysis: primary analysis, reuse/reanalysis, secondary analysis and simultaneous analysis (as illustrated in Table 7.2). The differentiation between these four types of analysis is an important clarification as the simple polemic of either primary or secondary analysis does not truly reflect the subtleties of actual research practice. Such a definition of terms is important, as previously indicated when we argued for a more sophisticated understanding of the issues related to types of data.

**Table 7.2** Modes of analysis

<i>Mode of analysis</i>	<i>Description</i>
Primary analysis	The initial analysis of the data collected by the researcher to address the main research question.
Reuse or reanalysis	Subsequent analysis of the initial data set by the same researcher to address a new research question.
Secondary analysis	The analysis of data collected by a different researcher for the purposes of answering a different research question.
Simultaneous analysis	Analysis of the same data set by different researchers concurrently, either in the same research team or independently. This type of analysis can be either simultaneous primary analysis or simultaneous secondary analysis.

We now turn our attention to consideration of the general criticisms aimed at secondary analysis. There are three main arguments that call into question the credibility of what has been broadly referred to as ‘secondary analysis’ which has historically also included reuse of one’s own data. We argue that the arguments put forward against secondary analysis represent legitimate concerns in relation to our now more specifically defined understanding of secondary analysis, but are of less concern (although still potentially relevant) when applied to the reuse or reanalysis of data or to primary data analysis. This is therefore an important reason for making the distinction that we do between reuse of one’s own data and secondary analysis of another researcher’s data.

### Reflective space

Consider and discuss the distinctions between data and research data.

At this point we invite the reader to review the distinctive types of analysis presented in Table 7.2 and think about why these distinctions in terminology are important.

## DEBATES ABOUT SECONDARY ANALYSIS

There are three main areas causing contention in relation to secondary analysis. First, there is the problem that a researcher analysing data they did not collect is at the disadvantage of not having the full details of context relating to how and where the data were collected. Second, there is the problem of quality in relation to qualitative research, which is mainly focused on issues of transparency, reflexivity and trustworthiness. Third, there is the problem of ethics, in the sense that participants may not have been fully informed about the secondary analysis of data at a later date and may not have consented to have their responses used in that way. Each of these issues is considered in turn.

### Context

We acknowledge that the notion of context encompasses a broad range of issues and that the term is used in different ways. In relation to the criticisms put forward regarding the deficiencies of secondary analysis, there are three key arguments. First, questions are raised in terms of the difficulties of having limited access to information about the situated nature of the data collection such as the historical, political or cultural setting. Second, there is concern that data generated for the purposes of answering a particular research question may not be a good 'fit' with the purpose of the secondary analysis. Third, related to fit, the usual research trajectory begins with an epistemology which informs methodological choices. The secondary analyst may not have access to complete information about the original researcher's epistemological position from which the data were collected. Therefore, secondary analysts should exercise caution and question their assumptions about the assumed congruence between their own epistemological positions and that from which the data were collected.

Depending upon where the data were accessed, the researcher who is performing the secondary analysis has a greater or lesser degree of information relating to the specific contextual detail of the original research project for which those data were collected (Hammersley, 2010b). Access to this information is important as the situated nature of the data collection, its cultural, institutional and interactional context, have a significant bearing on the way the data are interpreted (Bishop, 2007). Thus there is a risk that the influence of the particular features of the original data set may not be obvious to the secondary analyst (Thorne, 1994). Secondary analysis therefore raises risks regarding potential misrepresentation of data due to selective interpretation (Corti and Thompson, 2004).

Within the broad spectrum of qualitative research there is significant variability in depth and breadth of topics addressed and relative flexibility in the nature of the research design (Heaton, 2008). The problem of limited access to the original context in which the data were collected means there may be a lack of 'fit' between the data available and the requirements of secondary analysis, which can

lead to potential difficulties in interpretation and unintentional errors (Gillies and Edwards, 2005). Thus the compatibility of data depends on the closeness of 'fit' between the purpose of the secondary analysis and the nature of the original data (Thorne, 1994). The purpose of any data collected will be tailored to address a particular research question and therefore it is likely that there will be a gap between the projects (Hammersley, 2010b). The secondary analyst will need to account for the relationship between the original and secondary research questions but they will also have to think about the implications of explicit and implicit methodological variations between the original pieces of research (Thorne, 1994).

As there is such diversity in relation to epistemology and methodology within the qualitative approach, the methods employed for data collection are varied and ultimately informed by the theoretical stance of the researcher.

e.g. A researcher employing phenomenology is likely to conduct interviews which seek to gather detailed information on the lived experience of participants. This shapes the trajectory of the interview and influences the style and content of questioning.

It is important for any would-be secondary analyst to be aware of this framework within which the data were collected as the secondary analyst may be attempting to employ a form of analysis that is incongruent with the way in which data were collected. We argue, therefore, that data collected are not necessarily interchangeable across methodologies (refer back to Chapter 5).

e.g. In grounded theory there is a successive process of collection, coding and analysis which is an organic process led by the interest in developing a theory, and cannot be replicated in secondary analysis (Szabo and Strang, 1997).

Mauthner et al. (1998) caution researchers engaged in secondary analysis to make their epistemological position explicit and to openly address the difficulties related to any potential mismatch between their position and that of the original researcher. Some researchers argue that these epistemological concerns need not negate the pragmatic application of secondary analysis (Fielding, 2004). We would contend, however, that while taking a pragmatist position is in many ways attractive, the quality of the secondary analysis may be compromised. In order to defend against this possible criticism we propose that the secondary analyst is transparent in their reflexive consideration of matters of methodological fit.

## Quality issues

Another of the key concerns about secondary analysis in qualitative research relates to issues of quality. It is important that secondary analysts attend to issues

of quality as a mechanism for ensuring methodological rigour (Szabo and Strang, 1997). Notwithstanding debates about universal markers of quality, these concerns relate primarily to the main quality criteria of transparency, reflexivity, transferability, ethicality and integrity, as outlined in Chapter 2. In relation to the overall credibility of the research project, the broader issue of trustworthiness in secondary analysis is brought into question due to limitations in the way that transparency can be managed. With regard to transparency the secondary analyst faces a twofold challenge. First, they must report openly and clearly about issues and events in which they were directly involved. Second, they have the difficulty of being limited in illustrating details relating to the collection of data in which they were not privileged to be involved in the first instance. Similarly, secondary analysts face obstacles in trying to communicate reflexivity in the dissemination of their findings because they were not present during the data collection phase of the research.

The trustworthiness and credibility of qualitative secondary data analysis are brought into question when limited information about the process of data collection is available. While it is possible to assess the credibility of new secondary findings by recourse to archived information (Hammersley, 1997), it is more challenging and relies on considerable knowledge about the original project. Problematically, the secondary analyst has a lack of control regarding how the original data set were conceived, generated and recorded which may restrict the quality of the analysis (Szabo and Strang, 1997). The reason for this is that the data were not originally collected for the specific purposes of the secondary analyst's research question.

e.g.

Regarding the issue of transcription (which is an active process; see Chapter 6), the secondary analyst is limited in the information they may be able to access about the rationale behind decisions related to transcription, such as who transcribed the data and how (Bishop, 2007).

These are important factors which should be reported by original researchers in the interest of transparency in a way that may facilitate the possibility of secondary analysis later. Additionally, it is necessary for the secondary analyst to recognise the complexity of the work and account for any limitations of access which may impact upon their interpretations in regard to the design, methods and issues (Heaton, 1998).

An essential aspect of qualitative work is that it requires the monitoring of the effects of the research process through reflexivity. A fundamental methodological issue in qualitative research is the intersubjective relationship between the researcher and the participants, and there is concern that there will be a significant deterioration in the quality of reflexive insight in secondary analysis (Heaton, 1998). Arguably, however, because data are in effect reflexively constructed, some researchers believe that the loss of original context is less of a problem (Moore, 2007).

Some researchers position this issue of quality as a practical matter which is argued to be more easily overcome (Fielding, 2004). However, our stance is that issues of quality in qualitative research are inherently tied to the epistemological position of the researcher. While there are some overarching quality criteria which arguably can be usefully applied to any methodology, the specificity of particular qualitative approaches dictates that additional measures of quality within that discipline are attended to.

### Argument of ethics

The argument of ethics is twofold. On the one hand, it is ethically advantageous to perform secondary analysis as it negates the need to recruit additional participants and expose them to potential risks and burdens (Heaton, 1998). On the other hand, however, it is often the case that original researchers do not recruit participants with future secondary analysis in mind and therefore fail to weave this into the informed consent procedures. Nonetheless, if the original researcher has made provision for the possibility of later secondary analysis then it resolves the ethical issue in a straightforward way (Szabo and Strang, 1997). An alternative option available in some cases is for the researcher to seek additional consent from participants for secondary analysis, but this may not be feasible and is generally not typically done (Hinds et al., 1997).

In cases where explicit written consent for secondary analysis cannot be or has not been obtained (whereby obtaining consent is preferred), judgements about the ethicality of reuse or secondary analysis of data for a different purpose need to be grounded in assessments of risk to participants (Bishop, 2007). However, in qualitative research the use of sensitive data is common (Heaton, 1998). This raises three issues. First, the nature of the sensitive data is more likely to raise concerns from the participants' perspective in terms of how the data are to be stored and used. Second, with sensitive data there are likely to be tighter parameters imposed by ethics committees with regard to a definitive date for data destruction. Third, there is a greater risk of harm to participants in relation to breaches of confidentiality, failures of anonymity and the potential for distress.

e.g.

In our own work on family therapy there was no intention of later use of data for secondary analysis by new researchers. This was due to the sensitive nature of the video recordings of naturally occurring family therapy sessions. The topic and setting were sensitive by definition. Had we asked the families for consent to store the data for other research teams to view, this probably would have resulted in a low or zero response rate as we could not have given assurances as to how the data would be treated. It is unlikely that the ethics committee would have allowed an indefinite time-frame for the storage of such sensitive material. Furthermore, if the data had been stored and accessed by new

research teams, there would have been a greater risk of breach of anonymity and confidentiality, with the potential to distress the families involved. Thus neither the data nor the transcripts were ever made available for secondary analysis.

## Archiving

Despite the issues that we have discussed so far in this chapter, it is worth noting that in contemporary research practice there has been a drive towards archiving data for secondary analysis. Much of this has been driven by funding organisations. For example, in the UK, the Economic and Social Research Council (ESRC) commissioned a survey which revealed that 90 per cent of data were lost or at risk and that the 10 per cent archived failed to meet the requirements of an archive such as public access or security (Thompson, 1991). Thus, in 1994 the Qualidata qualitative archiving project was launched with the support of the ESRC; this was not an archive in itself but an action unit with a remit to locate and evaluate research data sets, organise their transfer to suitable archives and catalogue them (Corti and Thompson, 2004). However, there are many countries that have not yet put such infrastructures in place.

The drive towards archiving qualitative data for later secondary analysis has been motivated by consideration of the benefits of both archiving and secondary analysis itself. Fielding (2004) argued that there are three core benefits:

- The process of archiving a data set in itself ensures that the data corpus is kept in an organised and accessible way.
- The requirements of the archiving process dictate that data need to meet minimum standards whereby recordings have to be audible, documents have to be legible and there are no significant gaps in the material.
- A necessary aspect of archiving is that original researchers are constrained to ensure that appropriate informed consent from participants is obtained and that legal standards are met.

One of the benefits of a data archive is that researchers are able to conduct comparative analysis of historical and contemporary materials, events and issues (Corti and Thompson, 2004). This enables the exploitation of older data which can be a rich source of material regarding historical attributes and behaviours of individuals, groups, societies and organisations (see <http://bit.ly/1tQwvgE>). It has been argued that an additional benefit of using archived data for secondary analysis is that original research might be verified through the process, in a similar way to quantitative research (Heaton, 2008). However, the use of data for the purpose of verification through secondary analysis is controversial (Heaton, 2004), and there is little evidence to suggest that archived data are being used in this way (Corti and Thompson, 2004).

There have been a number of further criticisms put forward in relation to the shortfalls inherent in archiving. First, researchers may be reluctant to archive their data until their own analysis has been exhausted (Mauthner et al., 1998), which may be difficult to judge. Second, researchers may feel vulnerable in terms of others viewing their data and being made accountable and open to criticism (Corti and Thompson, 2004). Third, there are issues of ownership of the data, and complex legal and copyright matters (Mauthner et al., 1998). Finally, there are questions regarding the cost-effectiveness of archiving material as it is a time-consuming process and researchers typically have pressing administrative and teaching responsibilities placing demands on their time available for such matters (Mauthner et al., 1998).

## SUMMARY AND FINAL THOUGHTS

In this chapter we have explained how the typical discourses of primary and secondary data have created an unnecessary dichotomy. In reality, what has been regarded as 'secondary data' is actually the same data as primary data and therefore the distinction is misleading. Rather we have clarified the distinctions between different types of data in a more sensible way. By recognising the broader qualitative literature on researcher-generated and naturally occurring data, solicited and unsolicited mechanisms for collecting data and the possibility of simultaneous data collection, we have provided clarity as to why the notion of secondary data is a misleading concept.

While negating the logic of primary and secondary data concepts, we acknowledge and recognise the necessity of differentiating primary and secondary data analysis. Within the broad topic of secondary analysis we have argued that there is a distinct and qualitative difference between reuse of one's own data and secondary analysis of data collected by another. However, we recognise that qualitative analysis, whether of one's own data or data collected by someone else, should always be rigorous and of a high standard and we recommend engagement with textbooks which explicate the process (see, for example, Bryman, 2008a). We argue that clearly defining what constitutes reuse of one's own data and distinguishing it from secondary analysis of another's data facilitates the attendance to issues of quality.

In relation to issues of quality assurance in secondary analysis, we have discussed the three main arguments that have caused contention for its credibility. These were problems of context and data 'fit', transparency and reflexivity, and ethics. Each of these was addressed in turn and illustrated some of the complexities and difficulties of conducting secondary analysis while maintaining the quality criteria for qualitative research. Expounding the rationale for these quality arguments demonstrates the distinctive differences between reuse and secondary analysis.

As a consequence of the arguments we have presented in this chapter, we propose that the issue is more complex than may have previously been presumed. Particular issues related to qualitative research, which are problematic in a way that does not apply to the analysis of quantitative data, relate to two broad issues: the fact that qualitative research is an iterative and reflexive process, and that epistemology needs to be taken into account during all phases of the process.

With regard to the non-linear process of qualitative research, the use of data for secondary analysis can restrict the iterative process and confines the availability of the reflexive aspect of the original researcher. However, the benefits of making data available for secondary analysis by others are that there is potential for a different kind of transparency which embraces the possibility of new insights from researchers or research teams. The promotion of such archiving of data has the potential to foster a culture among qualitative researchers which is more open, collaborative and supportive. Ideally, original researchers could make themselves available to secondary analysts for consultation to contextualise the material where needed (Heaton, 1998). Within this environment students are likely to benefit from this culture in a more tangible way, making them part of a broader research community with opportunities for managing their research time constraints.

As archived data sets continue to grow there will be a greater requirement for systematic cataloguing which is meaningful to would-be secondary analysts. In order to facilitate the methodological, epistemological and quality issues that are pertinent in qualitative research we propose a number of suggestions. First, in relation to the potential paucity of contextual information, we advocate that the original researchers supply as much detail as possible about context alongside their data set. Supply of original documentation such as field notes (where they exist) may help to recover a degree of context and may also enhance the secondary analysis (Corti and Thompson, 2004). Second, with regard to ethical considerations, we suggest that where practical, consent from participants is obtained for later use of the data. Third, bearing in mind the epistemological variance within qualitative research, we propose that it would be useful to catalogue data sets in accordance with their epistemological positions, or have a cross-referencing system in place. While this places a greater onus on the original researcher, we argue that the benefits for knowledge and research community culture are worth the effort. These benefits are threefold. First, the original researcher is more likely to take greater care with these important issues in order to meet criteria for archiving, resulting in better-quality information for reuse or secondary analysis. Second, the stipulations would ensure that if sensitive data were being collected, there would be more stringent procedures in place to manage and protect the data. Third, if the necessity for all researchers to approach a primary source is mitigated, there is potential to protect vulnerable and overly researched groups from further exposure to risk.



---

## Case study and reflective questions

---

### *Group discussion activity*

Stephanie is a criminology graduate and is interested in hoax calls requesting assistance from police services. She has been employed as an independent researcher by the police commission as the number of hoax calls has increased, causing significant costs to already stretched budgets. Stephanie has identified five research papers spanning 20 years which have used small-scale data sets of such calls. She intends to look at general themes and trends in the calls, as well as the demographics of hoax callers. In order to do so she plans to contact each of the five researchers to request access to their data sets for secondary analysis.

**Q:** What will the benefits be of using existing data for secondary analysis?

**Q:** What difficulties is Stephanie likely to encounter?

---

Suggestions for answering these questions are available at the end of the book.

## FURTHER READING

Corti, L. and Thompson, P. (2004) Secondary analysis of archived data, in C. Seale, G. Gobo, J. Gubrium and D. Silverman (eds), *Qualitative Research Practice*. London: Sage. pp. 327–343.

In this book chapter the authors provide some context and history for archiving qualitative data and illustrate some of the modern methods for utilising archived data. They discuss some of the challenges that analysts face when undertaking secondary analysis of qualitative data and some of the issues related to ownership.

Hammersley, M. (2010) Can we re-use qualitative data via secondary analysis? Notes on some terminological and substantive issues. *Sociological Research Online*, 15(1): 5. doi: 10.5153/sro.2076.

In this article the author considers some of the benefits and challenges of undertaking secondary analysis in qualitative research and considers some of the key arguments in the literature. More specifically, he examines the problems of 'fit' and 'context' and the argument that there are difficulties in attempting to resolve such challenges.

Heaton, J. (2008) Secondary analysis of qualitative data: An overview. *Historical Social Research*, 33(3): 33–45.

This is a useful article as it provides an overview of secondary analysis in social research. The author provides definitions of secondary analysis and looks at the history of this in the qualitative field. She provides examples of how secondary analysis has been used and looks at some of the challenges that researchers face.

Thorne, S. (1994) Secondary analysis in qualitative research: Issues and implications, in J. Morse (ed.), *Critical Issues in Qualitative Research Methods*. Thousand Oaks, CA: Sage. pp. 263–279.

Although over 20 years old, this book chapter still provides some useful information about the issues inherent in secondary analysis. The author provides a critical discussion of the possibilities and implications of secondary analysis and considers some of the opportunities that this provides for the academic community.

