



Secondary Qualitative Analysis using Online Resources

In: The **SAGE Handbook of Online Research Methods**

By: Patrick Carmichael

Edited by: Nigel G. Fielding, Raymond M. Lee & Grant Blank

Pub. Date: 2017

Access Date: May 21, 2020

Publishing Company: SAGE Publications Ltd

City: 55 City Road

Print ISBN: 9781473918788

Online ISBN: 9781473957992

DOI: <https://dx.doi.org/10.4135/9781473957992>

Print pages: 509-524

© 2017 SAGE Publications Ltd All Rights Reserved.

This PDF has been generated from SAGE Research Methods. Please note that the pagination of the online version will vary from the pagination of the print book.

Secondary Qualitative Analysis using Online Resources

Patrick Carmichael

Introduction

This chapter explores the secondary analysis of qualitative data, the impact that network technologies have had on this, and the research potential for secondary analysis of data that is accessed across both the public Internet and networks of research archives. Secondary analysis allows not only the reassessment of the approaches and arguments of researchers, it also enables individuals and communities not involved in the original research to engage with data in new ways that reflect emerging perspectives or research strategies, some of which may be further supported and enabled by technological developments. Network technologies offer the potential to allow data, analyses, researchers and participants to be distributed in both space and time, but these technologies need to be carefully developed and deployed, and critically evaluated in order to avoid the creation of what have been called 'data tombs' (Fayyad and Uthurusamy, 2002: 32) in which data 'rest in peace' with little opportunity for subsequent access, let alone secondary analysis.

Both the practices of secondary analysis and the technologies to allow them to take place in online environments (or even simply to be undertaken by researchers at a distance) have been dominated by quantitative data and the research practices that generate and consume them. There are long traditions of data sharing and reuse both in the pure and applied sciences (examples being astrophysical, epidemiological and climate data) and in the social sciences, where census data and social and economic indicator data are widely shared. Data archiving and sharing, and the secondary analysis that it might enable, are less well established in qualitative research contexts, with mixed-method and longitudinal studies therefore also often proving difficult to implement.

Since the first edition of this Handbook was published, the technological landscape against which secondary analysis might take place has changed significantly, and alongside this there have been methodological developments; new patterns of participatory and collaborative research have emerged; and there have been changes in the ways in which research is conceptualized, funded, organised and its success and impact measured. The relationships between these developments are complex: the research potential of many emerging technologies (often not designed with research in mind) has only become apparent through experimentation (a good example would be the realisation of the research potential of social media platforms, micro-blogging and 'crowd-sourcing' of data). At the same time, technology developers responding to interest in online research and 'data journalism' have developed more sophisticated search, retrieval and visualisation technologies, many of which feature in other chapters in this Handbook.

However, before considering some of these developments and a number of illustrative examples of how secondary analysis of qualitative sources might be enabled in online research, it is worth exploring some of

the broader debates surrounding the reuse and reanalysis of qualitative sources in particular.

The Nature, Purpose and Challenges of Secondary Analysis of Qualitative Data

Secondary analysis is conventionally defined as involving the use of data collected in prior research to pursue some new direction or answer new research questions, and is thus differentiated from meta-analysis or systematic reviews (Corti and Bishop, 2005; Heaton, 1998, 2004, 2008; Irwin and Winterton, 2011; Szabo and Strang, 1997). Heaton's (2004) review of studies involving secondary analysis of qualitative data revealed that the majority of these involved what are described as 'supplementary analysis', focusing on aspects of data that were not addressed or only partially addressed in the original research. Less common were 'supra-analyses' (which investigated new research questions or issues) and reanalyses of original research questions, the latter involving interrogation or critique of the original researchers' analysis (Heaton, 2004: 39–46). Secondary analysis of data by the original researchers is not precluded: in fact, Heaton (2004: 37) suggests that at least some analysis which is described as 'secondary' may be primary analysis undertaken by the original researchers some time after the original research or may be inspired by a concern that primary analysis was incomplete or inadequate.

Heaton also identifies the importance of secondary analysis as an element in research approaches alongside other qualitative or quantitative data analysis (Heaton, 2004: 47–5). This may take place in the course of comparative or longitudinal studies; where quantitative data such as cohort studies are used to provide a context for qualitative data such as interviews or case studies, or where other research has suggested new lines of enquiry or conceptual frameworks. Irwin and Winterton (2012) offer an interesting reflexive account of their experiences of working with existing data sets gathered in the course of two prior projects as part of the UK Economic and Social Research Council 'Timescapes' programme (Holland, 2011). They carried out secondary analysis of research data from two prior studies ('Work and Family Lives' where data had been collected in Edinburgh in 2007–2010, and 'Men as Fathers', which collected data in East Anglia in 2000–2008) formulating research questions informed by recent empirical and theoretical work on continuity and change in gender, work and care arrangements. These extended and often multi-method approaches contribute to a blurring of the distinction between secondary analysis, meta-analysis and theory-building. This is also the case when analysis takes place in the context of action research in that 'co-interpretation' activities involving participants may involve the revisiting of previously collected data as part of an iterative research design.

Another compelling rationale for undertaking secondary analysis is that it provides access to rich data, particularly those collected in unusual research settings, or which give voice to rarely heard participants – another argument made by Irwin and Winterton (2012) in the previous example. Long-Sutehall *et al.* (2010) describe how secondary analysis of data gathered across multiple previous research projects proved useful when researching sensitive issues with an 'elusive population' (Long-Sutehall *et al.*, 2010: 336) – in this case, family members' experiences of brain-stem death related to organ and tissue donation. It was known that

recruitment to the prior projects had been difficult, and so using what data had been collected was judged to be a more efficient strategy than initiating new data collection. Another example is the body of work built up by Walkerdine (1988), which provides detailed insights into the development of mathematical concepts in pre-school children in their homes, rather than in better-documented school settings. In this example, the sharing of data and its secondary analysis has enabled a rich theoretical debate: records of dialogues between parents and children have been analysed from cognitivist and social perspectives; have contributed to post-structuralist exploration of gender and class; and have informed comparisons between home- and school-based learning. Other arguments for undertaking secondary analysis include time- and cost-effectiveness (although the costs of initial preparation and archiving of data need to be taken into account); reduction of burden on participants; and allowing triangulation and generalisation from findings (Fielding and Fielding, 2000; Hammersley, 1997; Szabo and Strang, 1997).

The diversity of qualitative research and the epistemological underpinnings of different perspectives are reflected in attitudes towards, and approaches to, secondary analysis. Some of the most successful instances of sharing of qualitative data have occurred where contributors and users of data have a domain-specific shared repertoire and a common epistemological basis for their approaches: a good example is the international CHILDES (Child Language Data Exchange System) network, which supports computer-aided conversation analysis and its associated database 'TalkBank'¹ (MacWhinney, 2000a, 2000b). At the same time, there is also a distinctive position on secondary analysis, advanced by Mauthner *et al.* (1998): that not 'having been there' to share the epistemological perspectives of the original researchers is an insurmountable barrier to secondary analysis, and that the primary role of secondary analysis is methodological exploration, rather than substantive engagement with research data. Despite technological developments (for example, video and audio data rather than simply transcripts can now be archived) and metadata schemes that allow original researchers to provide rich contextual information about the settings in which data were collected, this remains the principal objection that some qualitative researchers make to secondary analysis and work with archival data.

An alternative view of secondary analysis stresses its pragmatic and methodologically eclectic character (Heaton, 2004: 116–21), arguing that it may be thought of as 'bricolage' in that it often draws on multiple datasets, sources and methodological approaches. This perspective in general de-emphasizes the importance of personal involvement in original research and challenges the distinction between primary and secondary analysis, pointing to examples from other disciplines where reanalysis is not only the norm but also a distinctive aspect of disciplinary practice (Bishop, 2007, 2009; Fielding, 2004; Thompson, 2000). It also echoes calls for reform of practice within certain social science disciplines: Stenhouse (1978), for example, suggests that educational research might take the lead from historians and calls for a collective effort to develop a 'contemporary history' of education in which quality of analysis and methodological transparency are closely aligned. This would entail research data being explicitly structured and presented so as to encourage reanalysis, with audit trails being provided to allow for critical engagement with the interpretative frameworks and processes of the original researchers.

Although there is a temptation to dismiss 'eclectic' approaches as taking insufficient account of epistemological issues or being post-epistemological, it is important to recognise that such approaches may not simply be opportunistic. Rather, they may represent distinctive methodological and political standpoints informed by commitments to widening participation in research, or to a view of reanalysis and recontextualisation as a continuing knowledge construction activity.

There is room both in what might be described as the 'epistemologically circumscribed' and the 'epistemologically eclectic' perspectives for engagement with 'classic studies', although the identification of a study as a 'classic' makes methodologically and epistemologically neutral reassessment difficult, given the tendency for 'classic' to become 'exemplary' or even to define or circumscribe disciplinary norms (Parry and Mauthner, 2005; Savage, 2005). Savage points out that the normative nature of many texts in social science means that they focus on how researchers should conduct their research, rather than how they actually went about their research; and that even purportedly reflexive accounts are often very selective and reflect highly theorised and subjective stances. This is a strong argument for broadening of the scope of secondary analysis beyond the data collected and archived to include research instruments, fieldnotes and research minutes, which provide insights into the nature of research processes, issues, dilemmas and macro- and micro-political agendas.

Proponents of both perspectives outlined here (the continuing debate is reviewed by Irwin and Winterton (2012) and Medjedović (2011), amongst others) acknowledge that secondary analysis has a role in learning about research methodology in general, as well as about the specific practices and processes employed by researchers. Corti and Bishop (2005) describe a range of specifically pedagogical rationales for engaging students and early-career researchers in secondary analysis, including engagement with rare or exemplary studies; exemplification of enduring issues and questions in research; and allowing students to engage with the complexities of real data in a way that specially constructed datasets developed for teaching purposes do not. Amongst other arguments, Seale (2011), who writes not only about secondary analysis in general, but about his use of online sources in his own research, highlights the value of making available data to students who would not otherwise have the opportunity either to collect them or to engage in the theoretical discussions around their analysis.

Debates about the possibility, nature and purpose of secondary analysis are thrown into high relief in online contexts for a number of reasons. First, there is a set of issues about the actual practices of preparing, describing and archiving data in digital repositories – specifically, how to capture as much as possible of the original researchers' intentions and the nuances of the data collection process. Second, the concerns that critics of secondary analysis have are liable to be exacerbated when secondary analysis is carried out at a distance, by researchers who do not understand the original research context, or who use data selectively. On the other hand, well-documented and structured online resources have the potential to invite and encourage secondary analysis, support longitudinal and comparative research, and support research training and researcher development, opening up new opportunities for students of qualitative methods around the world who would otherwise have little opportunity to engage with classic studies, innovative

methods and 'elusive populations'.

Case study: the Cambridge Conference Archive of Education Evaluations

The issues raised in the previous section – both with enabling reanalysis and supporting research methods training – informed the development of a digital archive of qualitative data collected during a series of influential educational evaluation studies.² This has been developed in association with the members of the 'Cambridge Conference on Evaluation', a network that has existed since 1972 (see MacDonald and Parlett, 1973). Most of the data in this largely qualitative archive were collected in the course of publicly funded evaluations in the UK, US, Australia, and Ireland and carried out by members of the Cambridge Conference between 1975 and 2005. The majority of these are case studies, although they vary widely in their nature and scope. The research outputs, instruments and data from the evaluations are, for the most part, in the public domain, and have been published or made available to the public in some form. However, many date from the period before widespread Internet use became commonplace (and so only a small proportion of the data were 'born digital'), and more widespread awareness and access has, for the most part, been through academic journal articles, limited print-runs of other publications by the evaluation team or commissioning bodies, and informal dissemination.

With the exception of a few examples (such as Stake's *The Art of Case Study Research* (1995), which presented substantive data collected in his case study of Harper School in the US within a highly discursive methodological account), publications have been necessarily selective in the data they have presented, alluding to the scope of the evaluations and discussing methodological issues and research processes rather than documenting them in full.

The initial design of an electronic archive representing these studies was informed by Stenhouse's (1978) vision of 'case data' being made available for secondary analysis across a wider research community, and the potential of emerging multimedia technology to address this (Walker, 2002). This process was assisted by the original participants in the projects, who provided contextual information and guidance as to the roles and interrelationships of specific elements of the data. The evaluations were selected not simply on the grounds of being 'classics' (although some, such as Stake's Harper School study (1995) and the study of bilingual schooling in Boston in the early 1980s – published as *Bread and Dreams* (MacDonald *et al.*, 1982) – arguably have this status amongst education evaluations). Rather, the intention was to represent a broad range of educational evaluations with varied approaches, interpretational frameworks, and patterns of impact and dissemination. Although the archive as a whole can be seen as an historical collection, in that the studies document and reflect the evolution of educational policy, it also illustrates important and enduring issues, questions and dilemmas regarding the role and responsibilities of evaluators; ethical frameworks and informed consent; the purposes and conduct of evaluation; and the relationship of evaluation to the development of policy and practice. The original data are supplemented and placed in context by

interpretative accounts and interviews with original participants in the evaluations, which are presented both as video excerpts and transcripts. Many of the documents in the archive illustrate the point made earlier regarding the conversion of paper archives – they are extensively edited and richly annotated, and as such it proved more appropriate to store these in the electronic archive as image files rather than to attempt optical character recognition or transcription in order to generate text documents.

The development of this particular archive represents a pragmatic response to the issue of subsequent users of the data not having participated in the original study and not having shared the epistemological standpoints of the original participants. The circumstances in which the original evaluations were carried out were frequently complex and, in many cases, innovative research approaches had been developed in order to 'do justice' to these complexities. By working with the original participants it was possible to construct collections of data, research instruments and other 'working documents' in order to represent this as fully as possible. Drafts of interpretative accounts and reports were included in the archive, along with final versions, as were terms of reference, plans and protocols, original fieldnotes, 'found' objects and documents, journals and minutes of meetings at which key decisions had been made. The attribution of significance to specific documents or data was an iterative process. On some occasions, the re-visitation of 'paper' archives of evaluations was a stimulus for the participants to provide reflexive and illuminative commentaries, while on others it was these commentaries and interviews that threw light on the reasons for specific items having been archived in the first place. This led to a shift in emphasis from a phase of archive development that was primarily descriptive (which the project team described as the 'what's in the box?' phase, in recognition of the fact that much of the archive content was in paper form and stored in boxes and box-files), to a phase that was more concerned with why things were 'in the box'.

The question of 'context' or rather, how best to document the various contexts relevant to the data, was gradually elaborated as the archive was developed. Conscious of the concern that the 'context' of data collection is confused with the original interpretative frameworks of the primary researchers (Parry and Mauthner, 2005: 340), the multiple meanings of 'context' were explored in some detail. As well as basic descriptive information, aspects of each evaluation that came to be represented included the political and policy contexts (at local, regional, national and international levels) in which the data were collected; the methodological context and interpretational frameworks; and issues such as different research approaches espoused and employed by members of teams of evaluators. Equally significant was the need to gain an understanding of the contexts in which data had been selected, organised and described as archives (paper and electronic) and were constructed, reconstructed and repurposed. In some cases, original data had been through several phases of reorganisation, secondary analysis and incorporation into pedagogical applications, and so the selection of data and its incorporation into the new electronic archive represented merely the most recent in a sequence of such activities – the process of 'going online' was as context-laden as any other in the history of these project data.

Most of the secondary analysis opportunities for the archive have, at least initially, been predominantly 'pedagogical' and oriented towards providing students and early-career researchers and evaluators with

opportunities to engage with methodological issues. This pattern highlights a final significant aspect of 'context', which we might call the 'context of reuse' and that is itself a product of increased availability of online resource alluded to at the end of the previous section.

In the course of a subsequent major research and development project, 'Ensemble: Semantic Web Technologies for the Enhancement of Case Based Learning', which ran from 2008–2012 (Carmichael, 2012; Martínez-García *et al.*, 2012) the archive was further developed using a range of web tools and approaches informed by 'Semantic Web' approaches (enhanced search facilities; the use of structured taxonomies of terms; and visualisation tools such as maps and timelines (described in detail in Carmichael, 2011)). This new version of the archive was then used as the basis of a number of research capacity-building initiatives aimed at early career researchers in education. As both the archive and the semantic technologies were developing, these were seen as part of a participatory evaluation of their potential – what Boedker and Petersen (2000: 61–2) describe as 'learning-in-use ... understanding and developing use ... once a computer-based artefact has been taken over by users ... [yielding] insight about the developmental aspects of use'. As well as providing useful 'user feedback' on the design of the digital archive itself, and confirming Corti and Bishop's (2005) assertions as to the pedagogical opportunities offered by engaging in secondary analysis, what also emerged from interviews with participants was the increased confidence that they felt they had in using a wider range of data sources, presenting qualitative data as elements of reports, and using what (to them) were unfamiliar and innovative methods. This was particularly evident in cases where they felt constrained by existing reporting procedures, with one evaluator working on a government-funded project stating:

We tried using the pro-forma documents ... [but in our project] we wanted to tell stories and it just didn't seem to fit ... the Buddy project [one of those in the Archive] ... when we saw that, and it is taking on really complicated issues ... but what does the work is *the story that gets told*.

Carmichael, 2011: 333

What is evident in these accounts is the way in which appropriately designed online resources can allow researchers to engage in secondary analysis of richly contextualised qualitative data and make associations between the narratives of the original researchers and their own circumstances.

Electronic Resources for Secondary Analysis

The Cambridge Conference Archive was developed using a combination of established and emerging software, standards and network technologies. In doing so, it drew on practice and advice from a number of major data archives that have developed technologies, archiving practices, ethical and access frameworks oriented towards archiving qualitative data with the intention not only that it is preserved, but available for reuse. These include the UK Data Archive (UKDA),³ the Council of European Social Science Data Archives (CESSDA)⁴ and the Inter-university Consortium for Political and Social Research (ICPSR).⁵ Despite the valuable work carried out by these organisations, national-level infrastructures for qualitative archiving remain

patchy in their coverage and in some cases are limited in their capacity. Medjedović (2011) highlights a continuing lack of national infrastructure and describes how setting up the 'Archive for Life Course Research' (ALLF) at the University of Bremen was in part a response to a lack of alternatives capable of supporting the 'methodological and data-related conditions' required for secondary analysis of qualitative sources.

In terms of the actual technologies required to enable secondary analysis, an enduring concern is how best to adequately describe data and represent the various elements of contextual information which may need to accompany them – and as the account of the development of the Cambridge Conference archive makes clear, this can be a complex task. This involves both provision of metadata to describe documents or other data, participants, the circumstances of the data collection, and the broader research context in which this collection took place, and annotation within texts – even before any coding or memo writing takes place.

Most solutions, both proprietary and non-proprietary, make use of Extensible Markup Language (XML) as a basis for data description and interchange, as well as a means of presenting data consistently across electronic applications and platforms. It allows the representation of the content and structure of data, while at the same time excluding formatting and styling. This makes it ideal for data exchange, and as such makes it an effective basis for both the representation of descriptive metadata and additional annotations within documents, as long as the latter are, themselves, structured as XML content. In the case of the Cambridge Conference archive, most of the archive's contents were scanned and digitised documents, and so the use of XML was limited to metadata records, with subsequent analysis at text level requiring the use of optical character recognition prior to computer-based analysis or a return to pencil-and-paper annotation.

XML is used in combination with metadata and document markup standards, which determine the structure of the documents and the vocabulary that can be used if interoperability and reuse is required. An example of a well-established and widely used XML standard with many applications is the Dublin Core Metadata Set,⁶ which was originally designed to provide a straightforward means of generating human- and machine-readable bibliographical records (Heery, 1996). It can be embedded within the text of web pages and other resources, is used by publishers to disseminate details of publications and forms the basis of the metadata records used in a wide range of databases and electronic repositories. Dublin Core is often used to provide a minimum set of descriptors for documents, although these may be supplemented by descriptive elements drawn from other metadata sets and standards according to the nature of the data concerned. Even then, one of the challenges for data sharing and secondary analysis is that provision of 'lowest common denominator' descriptions can lead to a 'flattening' of complex data and loss of detail. There is a tension between ensuring that data are discoverable and can be used across systems (which tends to involve using widely used standards, vocabularies and formats) and adequate representation of the nuances of discipline-specific, theorised or local research practice.

In the context of development of data for secondary analysis, formatting text as XML does provide a way of representing not only descriptive metadata, but qualitative data itself. Textual data may be represented using minimal markup similar to that used for webpage authoring (XHTML, or Extensible Hypertext Markup

Language), or can be structured using the much richer markup language developed by the Text Encoding Initiative (TEI),⁷ which provides a repertoire of structural elements along with means of representing annotations, deletions and additions to handwritten and typed texts in a consistent form. The TEI is concerned primarily with the electronic representation of written texts, other 'fine-grained' text markup systems allow the representation of variation in tone, pronunciation and conversational practices (such as interjections) and, where transcripts have been generated from video data, the intercalation of data describing movements, gestures and facial expressions.

These codings could also, of course, be related to interpretational frameworks and analytic categories, or could be used to provide extra contextual information to inform secondary analysis – that is, texts could be made available 'part-coded'. This goes much further than the well-established practice of sharing 'code books' alongside analyses, and obviously represents a significant expenditure of time and effort on the part of original researchers, as well as demanding XML-enabled Computer-Aided Qualitative Data Analysis Software (CAQDAS) applications to allow codings and annotations attached 'non-destructively' to texts.⁸ Such an approach would probably be most likely to take place in the context of preparation of electronic resources with clear pedagogical purpose and audience, or of documents of wide interest and with a potentially large audience interested in secondary analysis.

For data other than texts (images, audio and video for example), a wide variety of metadata schemes allow the identification of elements or regions of images; time segments of audio; and multiple descriptive metadata 'streams' describing activity by different participants who are featured. Of these, those most appropriate as the basis of any strategy to promote secondary analysis are those that are non-destructive; the video editing, analysis and coding applications, for example, which typically store metadata separately from the original video data ('tagging' external to the data rather than 'captioning' within it) and thus allows 'cross-coding' and inter-coder reliability checking; collaborative analysis; and, of course, primary and subsequent secondary analysis to proceed independently of each other.

At 'document level' the Dublin Core Metadata Element Set provides a core set of descriptive categories, although, as previously mentioned, these are often supplemented by locally developed and applied descriptors and vocabularies. Fulfilling a comparable role in describing people are metadata schemes such as FOAF or 'Friend of a Friend'.⁹ FOAF allows individuals to be linked to what it describes as 'projects', but these are only loosely defined; more formal are research project description systems such as the Common European Research Interchange Format (CERIF),¹⁰ development of which has been driven by demands for exchange of research information within and beyond the European Union. Cutting across all these levels is the Data Documentation Initiative (DDI),¹¹ which offers a multi-level description vocabulary including not only descriptions of the contents of documents, but also associated projects, methodologies, research instruments used to collect the data they contain and access restrictions applied to them (Blank and Rasmussen, 2004).

If XML is the basis for the representation of structured content, then Resource Definition Format (RDF)

provides the means of representing the relationships between elements of that content, and as such is a key element of the emerging 'Semantic Web'. RDF offers a means of linking data (whether these be whole documents, fragments, annotations or details of authorship or attribution) in arbitrarily complex 'weblike' structures, which may be extended, revised and presented to users in a variety of different formats or viewed through different portals. RDF can, in comparison to other formats, appear verbose and complex, but this belies an underlying simplicity, which allows RDF-enabled resources to be shared, aggregated and incorporated into other resources – even if they originate from different sources.

RDF also allows the prospect of seamless integration of resources located in remote, networked repositories, on local servers, and individual users' 'client-side' data held on their own computers. Characteristically, RDF documents contain only metadata, expressed as 'triples' (for example, 'name-author-document' or 'document-published-date') allowing multiple metadata records to exist (perhaps authored by different individuals) all referring to a single online resource (such as a text) or element within it (such as a paragraph or utterance), without the need for destructive transformation of the original text. The separation of original documents from analyses, which themselves may be analysed, all within a common framework for data exchange, provides the basis for the development of collaborative annotation and analysis tools. Such frameworks would allow multiple researchers to develop parallel coding schemes or analyses of a shared online resource or set of resources. At the same time, this illustrates how the availability of online resources can contribute to blurring of the distinctions among primary analysis, secondary analysis and meta-analysis as data become 'extensible' and reusable beyond the boundaries of original cases.

Alongside these developments in data and document description has come the development of digital repositories, which allow the storage not only of data in multiple formats (texts, images, audio and video) but also the metadata records that accompany them and which may comprise large numbers of RDF statements ('triples') expressing all of the linkages between codes, annotations, memos and data fragments that might emerge in the course of qualitative analysis (which may, of course, involve multiple researchers working collaboratively). These digital repositories may also need to enact data security policies (allowing only selected individuals or groups to access confidential data) while at the same time exposing data and analyses to search engines and human users. Most work has been done in this respect by the developers of the Fedora Digital Repository¹² (Lagoze *et al.*, 2006) and although a full description of the technical aspects of this system is beyond the scope of this chapter, some examples of how it can be configured with a view to data sharing and secondary analysis may prove useful.

The Cambridge Conference Archive described earlier was constructed using Fedora to store project data and metadata records, and was able to draw on these existing metadata records to construct a 'triplestore' – effectively, a single database of every relationship in the entire archive. This enabled the development of much more sophisticated search interfaces allowing users to ask questions of the form 'find me more examples like this' or 'what else has this person commented upon?' It was this that enabled visualisation tools to be incorporated into the archive allowing more intuitive and exploratory approaches by users. In another project, The University of Prince Edward Island's IslandArchives Project¹³ stores a large collection of audio

data relating to the history and cultural heritage of Prince Edward Island in Canada. In this case, the data are exposed via a virtual collaboration environment, which allows university researchers to curate and 'steward' data, build communities of enquiry and invite public participation in further projects.

Martínez-García and Corti (2012) describe another prototype Fedora-based digital archive designed to allow data collected from small-scale student and practitioner projects to be archived, curated, shared and reanalysed. A particular challenge for these types of projects is that they are difficult to aggregate or 'scale up', and it is difficult for subsequent practitioner-researchers to avoid 'reinventing the wheel' when they come to carry out their own enquiries. Martínez-García and Corti (2012: 278–279) describe how they combined a number of the technologies already described:

Data [stored in Fedora] are described using the frameworks of the DDI, to document research instruments; Dublin Core, a popular metadata standard for bibliographical records; and QuDEx¹⁴, a standard published by the UK Data Archive for representing complex qualitative collections (and the relationships between their associated resources) including annotated data ... Finally, it uses the Exhibit Web Application Framework [a well-supported set of web visualisation tools]¹⁵ to allow searching, browsing and display of collection contents.

This allowed the construction of linked collections comprising (typically) a student report or dissertation, research instruments, multimedia content, interview transcripts, survey data and secondary sources such as policy documents on which they had drawn in their projects.

New Developments and Directions

The technical frameworks, developments and examples described above have taken place against a rapidly changing technological, political and methodological backdrop. These developments have meant that secondary analysis of qualitative data now extends far beyond the idea of one researcher carefully reanalysing another's data. The ubiquity of personal technologies capable of gathering image, audio and video data means that the sheer volume of data online is increasing rapidly, and at the same time government drives towards transparency and accountability means that large-scale digitisation and data-sharing initiatives have allowed wider access to hitherto inaccessible research materials. In this concluding section, I will briefly discuss three areas of current interest: the emergence of 'linked', 'open' and 'big' data; opportunities for methodological innovation (and for some established methods to be revisited); and the implications of these developments for qualitative researchers.

Linked, Open and Big Data

The vision of the 'Semantic Web' that was first proposed in the early 2000s stressed seamless and, to the 'end user', invisible, interoperability, with web services and 'agents' (personalised search tools) using machine

reasoning across a web of interconnected data (Berners-Lee *et al.*, 2001). This ambition has been tempered (partly in recognition of the challenges of making data available) and some Semantic Web activity has been reframed as building instead a 'Linked Web of Data' underpinned by use of common data formats, consistent metadata and permanent Internet addresses (Bizer *et al.*, 2009). For example, moves to offer 'open data' from government sources online¹⁶ in order to extend and enhance a linked data cloud have encouraged discussion of opportunities for public engagement and new research possibilities. Although less ambitious than the broader vision of the Semantic Web, the idea of the linked web of data has lowered the bar to participation and to the realization of the benefits of a wide range of Semantic Web technologies. In research settings this has enabled a more pragmatic adoption of Semantic Web technologies with the potential to enhance existing systems and applications including some of those described in this chapter.

The terms 'linked' and 'open' data are often used interchangeably – but it is important to recognise that not all linked data are 'open' (linked data approaches may be used within access controlled data networks) and many 'open' government documents are in proprietary and sometime non-machine readable formats (precluding their easy 'linking'). For the qualitative researcher interested in secondary analysis, the latter can of course prove useful sources, but their use will involve different strategies to those possible with wholly digital data sources, and technological tools may be of limited use. Also sometimes conflated with open and linked data is the notion of 'big' data – the vast volumes of data 'born digital' through Internet-enabled communication, information exchange and surveillance, and in some cases available for public access, again typically through 'open' government data hubs. Gurin (2014) offers an interesting analysis of the relationships between these sometimes technological, but more often rhetorical, claims to openness, linkage and scale, pointing out that much 'big data' is anything but 'open' (and in fact is a fiercely guarded commodity) and that many 'open government' initiatives are highly selective in the data that is actually released. In many cases, it is the disciplinary communities – mainly in the sciences and medicine – that (as already mentioned) have traditions of data sharing and who have made the most rapid progress in using 'big data' approaches in support of their research activities.

For the qualitative researcher, these developments may seem to be of little relevance because Government Data Hubs and other similar online sources seem dominated by large statistical datasets. The UK Data Hub's education collection, for example, seems largely comprised of government 'Statistical Releases' (of enrolments, attendance, exclusions, assessment outcomes, and social and economic indicators, and so on) from national to local level, but it is also a rich source of policy documents, reports and evaluations. Such hubs represent important elements of the 'linked web of data' and exemplify standards and expectations of formats, standards and data quality, as well as giving useful insights into the mindsets of policymakers. 'Big data' initiatives, which at first glance seem very far from the interests and concerns of qualitative researchers, provide not only a context and comparator for qualitative research, but may also provide the technological frameworks and approaches that will allow the primary and secondary analysis of extensive 'big qualitative' data – not only of texts but also of the volumes of multimedia data that are now being produced by researchers and research participants and subjects as part of 'life-streaming', digital ethnography, and oral history and community memory projects. These topics are covered in other chapters in this Handbook, but it

is worth those researchers who plan to use these methods to consider the implications of their methodological choices for subsequent data management, archiving and, potentially, for subsequent secondary analysis.

Methodological Innovation

The second reason that many developments in online technologies are relevant to secondary qualitative analysis is the new patterns of collaboration and novel research relationships that digital networks, virtual collaboration environments and shared digital repositories enable. As research funders increasingly expect to see distributed, international and multidisciplinary research teams bringing together those with distinctive expertise, the likelihood of research data being collaboratively analysed across research teams, either contemporaneously or as part of a sequential research design, increases. In these kinds of research environments, the distinction between primary and secondary analysis that we have already discussed becomes even more blurred.

This means that even within the lifetime of a single research project, data may be generated, analysed, deposited in a shared online environment of some kind (usually accompanied by some descriptive metadata) and then retrieved and analysed again by other members of the research team, who may be located in another location entirely. As described by Laterza *et al.*, (2007) in a study of such research teams as part of a research and development project into virtual research environments, online technologies and the research practices that they enable are co-constituted. Research practice evolves alongside technologies, and a contested notion such as 'secondary analysis' is particularly malleable and subject to changes brought about by the emergence of new technological opportunities.

A good example of this is the resurgence of interest in Grounded Theory, a distinctive approach to qualitative enquiry that characteristically involves close engagement with data, *in vivo* coding and the careful development of codes and categories that are authentic to research participants, rather than being overlain by the researchers' preconceptions and biases. It has proved particularly effective in studies where practice and experiences are poorly understood and undertheorised, for example paramedical staff, care workers and those in emergent and inter-professional settings.

Much contemporary Grounded Theory is informed by the work of Charmaz, who has argued for the use of Grounded Theory approaches within broader research designs informed by social constructivism and reflective practice (Charmaz, 2014), and it is here that the links with archiving and secondary analysis become apparent. From its inception, there has been nothing to prevent Grounded Theory being used by groups of researchers, or for researchers to critically interrogate the conduct of others' enquiries (Corbin and Strauss, 1990: 422). Weiner (2010), drawing on Charmaz's ideas, has described how Grounded Theory can be effectively practiced within 'loose research designs' including those in which teams are separated by time and space, include visitors and where initial analysis by one team member formed the focus for discussion and reanalysis by others. Charmaz (2014) also argues that although the vast majority of Grounded Theory studies have been based on interviews and participant observation, its approaches may be applied to extant sources

and literature, policy documents and the work of other researchers.

There are clearly opportunities for the technologies described here to enable distributed, secondary analysis within Grounded Theory approaches, and it is notable that the work of Long-Sutehall *et al.* (2010) on secondary analysis of 'elusive populations' as part of the 'Timescapes' programme, and which was mentioned earlier in this chapter, does indeed use a collaborative Grounded Theory framework informed by Charmaz's approach. This kind of productive interaction and reworking of research approaches does not have to be restricted to Grounded Theory or to any particular kind of data – there are many opportunities for researchers to explore which different methodological approaches, established or emergent, can be enabled by the use of well-structured original data and metadata, and supported and mediated through well-designed digital environments.

The Changing Role of the Researcher

What is without doubt is that since the first edition of this Handbook, several of the technologies that were at that time emergent, have become much more widely established and the 'technology stack' available to researchers in support of secondary analysis of qualitative data is more extensive and robust. Martínez-García and Corti's (2012) example demonstrates this well when compared with projects from a decade earlier, with their model of student research projects building on underpinning 'layers': QuDEX, RDF, XML and Dublin Core metadata.

It is also clear from even a cursory glance at recent research literature that many scholars who identify themselves as qualitative researchers have been exploring the use of online methods; engaging with data management issues and data archiving; and taking advantage of the range of existing data and other resources that are now online. At the same time, there is a new expectation on the part of the funders and evaluators of research that researchers will have, if not a full range of technological competences, an awareness of how best to increase the reach and impact of their work through careful data preparation, archiving, publication and support for subsequent research – in short, not only to take part in secondary analysis of data, but to enable it as well. In the UK, the expectation from the main research funders is that data emerging from publicly funded research will be made available as a 'public good', and although this might currently involve primarily quantitative data, the requirements do not distinguish between the methods used or the nature of the data generated (Engineering and Physical Sciences Research Council, 2011; Research Councils UK, 2015). Similar expectations are built into the European Union 'Horizon 2020' programme within which an 'Open Data Pilot' (European Commission, 2013) demands that funded research projects have a data management policy that will actively enable data sharing and reuse.

Against this background, the potential for secondary analysis of online qualitative data, supported by methodological innovation, combined with data sharing and open data initiatives and enabled by the ongoing development of digital archive technologies, is considerable. What it depends upon, however, is a continuing dialogue between technologists designing online environments, the authorities responsible for interoperability

standards and researchers willing and able to articulate their practices, concerns and commitments.

Acknowledgements

The author would like to acknowledge the contributions of the members of the Cambridge Conference on Evaluation and the associated ESRC project 'Representing Context in an Archive of Educational Evaluations' and the team of the Economic and Social Research Council–Engineering and Physical Sciences Research Council research project 'Ensemble: Semantic Technologies for the Enhancement of Case Based Learning', particularly Agustina Martínez-García and Louise Corti. I am also grateful to Christina Carmichael for her insights into Grounded Theory.

Notes

1. <http://www.talkbank.org/> (Accessed 1 August 2016).
2. The archive is online and publicly accessible at <http://www.evaluating-education.org.uk> (Accessed 19 October 2016).
3. <http://www.data-archive.ac.uk/> (Accessed 1 August 2016).
4. <http://www.nsd.uib.no/cessda/> (Accessed 1 August 2016).
5. <http://www.icpsr.umich.edu/> (Accessed 1 August 2016).
6. <http://www.dublincore.org/> (Accessed 1 August 2016).
7. <http://www.tei-c.org/> (Accessed 1 August 2016).
8. The idea of non-destructive transformation of texts is important in CAQDAS applications in which multiple codings are attached to a single text – for example, during cross-coder reliability checking. In order to compare patterns of analysis, the 'original' text must remain unchanged by coding or annotation – so the analysis remains external to the text. This contrasts with the editing paradigm of most 'wiki' applications, for example, where multiple authors can alter the text itself (see Silver and Bulloch, this volume).
9. <http://www.foaf-project.org> (Accessed 1 August 2016).
10. <http://www.eurocris.org/cerif/> (Accessed 1 August 2016).
11. <http://www.ddialliance.org/> (Accessed 1 August 2016).
12. <http://fedorarepository.org/> (Accessed 1 August 2016).
13. <http://islandarchives.ca/ivoices> (Accessed 1 August 2016).

14. <http://data-archive.ac.uk/create-manage/projects/qudex> (Accessed 1 August 2016).
 15. <http://www.simile-widgets.org/> (Accessed 1 August 2016).
 16. For example, the US Government data hub at <http://www.data.gov> and the UK data hub at <http://data.gov.uk> (Accessed 1 August 2016).
-

References

- Berners-Lee, T., Hendler, J. and Lassila, O.** (2001) 'The Semantic Web', *Scientific American*, 284(5): 34–43.
- Bishop, L.** (2007) 'A reflexive account of reusing qualitative data: beyond primary-secondary dualism', *Sociological Research Online*, 12(3). Available at: <http://www.socresonline.org.uk/12/3/2.html> (Accessed 1 August 2016).
- Bishop, L.** (2009) 'Ethical sharing and reuse of qualitative data', *Australian Journal of Social Issues*, 44(3): 255–72.
- Bizer, C., Heath, T. and Berners-Lee, T.** (2009) 'Linked data – the story so far', *International Journal of Semantic Web and Information Systems*, 5(3): 1–22.
- Blank, G. and Rasmussen, K.** (2004) 'The Data Documentation Initiative: the value and significance of a worldwide standard', *Social Science Computer Review*, 22(3): 307–18.
- Boedker, S. and Petersen, M.** (2000) 'Design for learning in use', *Scandinavian Journal of Information Systems*, 12: 61–80.
- Carmichael, P.** (2011) 'Research capacity building in education: the role of digital archives', *British Journal of Educational Studies*, 59(3): 323–39.
- Carmichael, P.** (2012) *Ensemble: Semantic Technologies for the Enhancement of Case Based Learning* ESRC End of Award Report RES-139-25-0403-A. Swindon, UK: Economic and Social Research Council.

Charmaz, K. (2014) *Constructing Grounded Theory*. London: Sage Publications.

Corbin, J. and Strauss, A. (1990) 'Grounded theory research: procedures, canons and evaluative criteria', *Zeitschrift fur Soziologie*, 19(6): 418–27.

Corti, L. and Bishop, L. (2005) 'Strategies in teaching secondary analysis of qualitative data', *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 6(1). Available at: <http://nbn-resolving.de/urn:nbn:de:0114-fqs0501470> (Accessed 1 August 2016).

Engineering and Physical Sciences Research Council (EPSRC) (2011) EPSRC Policy framework on Research Data. Available at: <http://www.epsrc.ac.uk/about/standards/researchdata/> (Accessed 1 August 2016).

European Commission (2013) *Guidelines on Data Management in Horizon 2020*. Available at: http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (Accessed 1 August 2016).

Fayyad, U. and Uthurusamy, R. (2002) 'Evolving data into mining solutions for insights', *Communications of the ACM*, 45(8): 28–31.

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

Fielding, N. and Fielding, J.L. (2000) 'Resistance and adaptation to criminal identity: using secondary analysis to evaluate classic studies of crime and deviance', *Sociology*, 34 (4): 671–89.

Gurin, J. (2014) 'Big data and open data: what's what and why does it matter?' *The Guardian*, 15 April. Available at: <http://www.theguardian.com/public-leaders-network/2014/apr/15/big-data-open-data-transform-government> (Accessed 1 August 2016).

Hammersley, M. (1997) 'Qualitative data archiving: some reflections on its prospects and problems', *Sociology*, 31(1): 131–42.

Heaton, J. (1998) 'Social research update 22: secondary analysis of qualitative data'. Technical report, University of Surrey Department of Sociology. Available at: <http://sru.soc.surrey.ac.uk/SRU22.html> (Accessed 1 August 2016).

Heaton, J. (2004) *Reworking Qualitative Data*. London: Sage Publications.

Heaton, J. (2008) 'Secondary analysis of qualitative data: an overview', *Historical Social Research/Historische Sozialforschung*, 33(3): 33–45.

Heery, R.M. (1996) 'Review of metadata formats: iaifa/whois++, marc, text encoding initiative, Dublin core, uniform resource characteristics', *Program*, 30(4): 345–73.

Holland, J. (2011) 'Timescapes: living a qualitative longitudinal study', *Forum Qualitative Sozialforschung/ Forum: Qualitative Social Research*, 12(3). Available at: <http://nbn-resolving.de/urn:nbn:de:0114-fqs110392> (Accessed 1 August 2016).

Irwin, S. and **Winterton, M.** (2011) 'Debates in qualitative secondary analysis: critical reflections. A Timescapes Working Paper Series 4'. Available at: <http://www.timescapes.leeds.ac.uk/assets/files/WP4-March-2011.pdf> (Accessed 1 August 2016).

Irwin, S. and **Winterton, M.** (2012) 'Qualitative secondary analysis and social explanation', *Sociological Research Online*, 17(2): 4. Available at: <http://www.socresonline.org.uk/17/2/4.html> (Accessed 1 August 2016).

Lagoze, C., Payette, S., Shin, E. and **Wilper, C.** (2006) Fedora: an architecture for complex objects and their relationships. *International Journal on Digital Libraries*, 6: 124–38.

Laterza, V., Carmichael, P. and **Procter, R.** (2007) 'The doubtful guest? A virtual research environment for education', *Technology, Pedagogy and Education*, 16(3): 249–67.

Long-Sutehall, T., Sque, M. and **Addington-Hall, J.** (2010) 'Secondary analysis of qualitative data: a valuable method for exploring sensitive issues with an elusive population?', *Journal of Research in Nursing*, 16(4): 335–44.

MacDonald, B. and **Parlett, M.** (1973) 'Rethinking evaluation: notes from the Cambridge Conference', *Cambridge Journal of Education*, 3(2): 74–82.

MacDonald, B., Adelman, C., Kushner, S. and **Walker, R.** (1982) *Bread and Dreams – A Case Study of Bilingual Schooling in the USA*. Norwich, UK: Centre for Applied Research in Education, University of East Anglia.

MacWhinney, B. (2000a) *The CHILDES Project: Tools for Analyzing Talk, Volume 1: Transcriptions Format and Programs*. 3rd edn. Mahwah, NJ: Lawrence Erlbaum.

MacWhinney, B. (2000b) *The CHILDES Project: Tools for Analyzing Talk, Volume 2: The Database*. 3rd edn. Mahwah, NJ: Lawrence Erlbaum.

Martínez-García, A. and **Corti, L.** (2012) 'Supporting student research with semantic technologies and digital archives', *Technology, Pedagogy and Education*, 21(2): 273–88.

Martínez-García, A., Morris, S., Tracy, F., Tscholl, M. and **Carmichael, P.** (2012) 'Case based learning, pedagogical innovation and semantic web technologies', *IEEE Transactions on Learning Technologies*, 5(2): 104–13.

Mauthner, N., Parry, O. and **Milburn, K.** (1998) 'The data are out there, or are they? Implications for

archiving qualitative data', *Sociology*, 32(4): 733–45.

Medjedovic, I. (2011) 'Secondary analysis of qualitative interview data: objections and experiences: results of a German feasibility study', *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 12(3). Available at: <http://nbn-resolving.de/urn:nbn:de:0114-fqs1103104> (Accessed 1 August 2016).

Parry, O. and **Mauthner, N.** (2005) 'Back to basics: who reuses qualitative data and why?', *Sociology*, 39(2): 337–42.

Research Councils UK (RCUK) (2015) RCUK Common Principles on Data Policy. Available at: <http://www.rcuk.ac.uk/research/datapolicy/> (Accessed 1 August 2016).

Savage, M. (2005) 'Revisiting classic qualitative studies', *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 6(1). Available at: <http://nbn-resolving.de/urn:nbn:de:0114-fqs0501312> (Accessed 1 August 2016).

Seale, C. (2011) 'Secondary analysis of qualitative data'. In **D. Silverman** (ed.), *Qualitative Research*. 3rd edn. London: Sage Publications, pp. 347–64.

Stake, R. (1995) *The Art of Case Study Research*. Thousand Oaks, CA: Sage Publications.

Stenhouse, L. (1978) 'Case study and case record: towards a contemporary history of education', *British Education Research Journal*, 4(2): 21–39.

Szabo, V. and **Strang, V.R.** (1997) 'Secondary analysis of qualitative data', *Advances in Nursing Science*, 20(2): 66–74.

Thompson, P. (2000) 'Re-using qualitative research data: a personal account', *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 1(3). Available at: <http://nbn-resolving.de/urn:nbn:de:0114-fqs0003277> (Accessed 1 August 2016).

Walker, R. (2002) 'Case study, case records and multimedia', *Cambridge Journal of Education*, 32(1): 109–27.

Walkerdine, V. (1988) *The Mastery of Reason: Cognitive Development and the Production of Rationality*. London: Routledge.

Weiner, C. (2010) 'Making teams work in grounded theory'. In **A. Bryant** and **K. Charmaz** (eds.), *The SAGE Handbook of Grounded Theory*. London: Sage Publications, pp. 293–310.

<http://dx.doi.org/10.4135/9781473957992.n29>