


Opportunities From the Digital Revolution: Implications for Researching, Publishing, and Consuming Qualitative Research

SAGE Open
October-December 2016: 1–13
© The Author(s) 2016
DOI: 10.1177/2158244016678912
sagepub.com


Louise Corti¹ and Nigel Fielding²

Abstract

In the 1990s, the term “online” research emerged as a new and vibrant suite of methods, focused on exploitation of sources not collected by traditional social science methods. Today, at least one part of the research life cycle is likely to be carried out “online,” from data collection through to publishing. In this article, we seek to understand emergent modes of doing and reporting qualitative research “online.” With a greater freedom now to term oneself a “researcher,” what opportunities and problems do working with online data sources bring? We explore implications of emerging requirements to submit supporting data for social science journal articles and question whether these demands might disrupt the very nature of and identity of qualitative research. Finally, we examine more recent forms of publishing and communicating research that support outputs where data play an integral role in elucidating context and enhancing the reading experience.

Keywords

secondary analysis, qualitative data, qualitative research, publishing data, online research

Introduction

This article sets out to provide some of the bigger picture issues and context around the Special Issue on “Digital Representations: Opportunities for Re-Using and Publishing Digital Qualitative Data.” As the title of our article suggests, it sets out some of the wonderful opportunities presented to the social scientist for working with digital online sources, yet also considers how to deal with representing, citing and replicating these sources given the fragile world of the Internet. Unknown provenance, lack of insight about sampling, error and bias, and broken links require us to rethink our trust in data and follow new best practice.

In the 1990s, the term “research online” emerged as a new and vibrant field of research methods: the ability to exploit sources that were not collected by established social science methods. As late as 2008, a whole compendium was dedicated to the theme, “The SAGE Handbook of Online Research Methods” (Lee, Fielding, & Blank, 2008). In 2016, at least one part of the research life cycle is likely to be carried out “online,” suggesting that *online methods* can no longer be regarded as distinct. Survey and fieldwork data are collected, uploaded, and stored on the cloud; new forms of data not initially collected as research data are accessed and exploited; online data manipulation and analysis software

and collaboration tools are utilized; and scholarly practice such as publishing findings and data, peer review, communication, and reading manuscripts are likely to be done online. Online opportunities also bring a much greater freedom to term oneself a “researcher,” as demonstrated by the emergence and popularity of citizen science, creating a new space for scientific discovery beyond the domain of academia.

At the same time, this period saw governments and the research base invest in what has been termed “e-infrastructure,” which has turned out to be a great enabler of research. This term denotes the synergy between research and the technological infrastructure required to support it. Furthermore, as appealing evidence starts to mount, we should not underestimate the potential of the “digital research methods” domain—methods that make use of online and digital technologies to collect and analyze research data—as they are utilized by a variety of disciplines. The approach can

¹University of Essex, Colchester, UK

²University of Surrey, Guildford, UK

Corresponding Author:

Louise Corti, U.K. Data Archive, University of Essex,
Colchester CO4 3SQ, UK.
Email: corti@essex.ac.uk



empower users and can increase outputs and their diversity, saving time spent on bespoke preparatory activities like collating and restructuring data that limit time left for analysis. Although it is also easy to hype the significance of online digital media, it is fair to say that the Internet and mobile communications technologies are re-shaping the knowledge discovery process (Nielsen, 2011).

Technology has long spurred development in social research methods at least as much as intellectual trends and information technology (IT) is engendering the rapid pace of developments around the exploitation of new forms of data. The Internet and mobile apps present large opportunities for collecting multifarious and real-time information *from* citizens and *on* citizens, their behavior and their interactions. As of 2015, Eastwood (2015) reports that there were 3 billion Internet users, representing more than 40% of global population. The web has some 600 million websites, Facebook had 1.32 billion active users, and Flickr had more than 87 million users uploading more than 3 million new images every day. The Internet serves as a catalyst of change in policy development and citizen engagement, particularly in the developing world using multi-stakeholder models (Gutterman, 2011).

Contemporary technology can scale up qualitative research to utilize large data sources, and enables innovative analysis techniques required for mining massive amounts of content from social media or open databases. Against a skills deficit in the ability of qualitative researchers to handle large data volumes, larger scale endeavors often bring welcome collaboration and working across disciplinary and geographical boundaries. The qualitative research community involves many specialisms and dispersed locations but can now perform joint work on data over digital networks in a way that previously could only be achieved at periodic professional gatherings.

Beyond academia, opportunities to undertake research have widened. Here, new technologies are not so much providing all-new processes and practices as facilitating what was previously done but in ways limited by available resources. Citizen scientists' reshaping of the discovery process is not confined to the natural sciences, for example, social media-based activities for social and political ends using "crowdsourcing."

In this article, we examine how working with qualitative data "online" requires some differences in approach across the research life cycle. In Part 1, we examine the kinds of sources available, their potential richness, trustworthiness and longevity. In Part 2, we look at the ontological status of doing research online and show opportunities that information on the Internet offers for extending the capacity of scientific knowledge through practices like citizen science. In Part 3, we turn to the matter of the emerging "research transparency" agenda that is starting to play out in some disciplines, where some social science journals are mandating that claims published should be evidenced to data that are fully accessible. We review the role that underlying data sources can play

in helping provide both context and trust in scholarly narrative and what impact these have, as requirements, on publishing qualitative research. Finally in Part 4, we explore how new forms of publishing and communicating research can benefit the reader, where *data* can play an integral role in both elucidating context and creating an enhanced reading experience. Is this a welcome burden for the author and the reader?

Before we commence, we briefly touch on the historical trajectory of qualitative research noting its impact on the term "transparency" used in this piece. In particular, we note the vibrancy of the domain and with it the profusion of competing constructions of what exactly constitutes "qualitative research." In addition to the normal variety of schools of thought in any discipline, there has long been a lack of consensus about what fundamental elements should constitute the discipline. Traditional elements of social scientific inquiry, like the concepts of reliability, replicability, and validity, and their more recent rendering in terms of trustworthiness, veracity, and so on, are subject to question and debate in a way that they are not in the domain of quantitative social science. We return to this issue in Part 3, noting that such critiques of these concepts long pre-date the current period, yet seem to re-emerge with a vigor or a crisis on a regular basis.

Part I: Online Data Sources

Across the world data sources, listed or available via the Internet, have massively increased in number, including new governmental public sector portals, researchers sharing data on repository systems, and journals increasingly publishing data to support reported results.

Availability of Data Online

Open data sources. Governments and organizations have embraced open data in efforts to be more transparent about their activities. By opening up their information for all, the innovation and economic potential of public sector information can be better harnessed. By the end of 2015, almost 300 public data catalogues were registered online by governments and organizations around the world (La Fundación Centro Tecnológico de la Información y la Comunicación, 2015). The U.S. government's Data.gov portal launched in 2009 just 4 months after Obama launched his plans for government transparency (Madrigal, 2009). Equally, the U.K. government's 2012 Open Data White Paper set down standards for timely release of digital open public sector data (data.gov.uk), and by December 2015 its data portal reported more than 22, 000 published datasets.

Other useful public data sources include real-time data feeds, such as current weather reports or stock market share prices. Creating "smart cities" relies on open data; New York City's portal, NYC Open Data (nycopendata.socrata.com)

contains hundreds of datasets provided by agencies and organizations, like information on parking facilities, and electricity consumption by zip code. More recently, organizations like the World Bank and U.K. Meteorological Office have provided programmatic access to open data via application programming interfaces (APIs). The Met Office publishes maps, charts, real-time forecasts, and historical data and runs hackathons, bringing data and data scientists together to develop innovative ideas, which in turn, can lead to new products and services (Met Office, 2015).

Commercial data sources. Search engine and social media site providers have demonstrated how IT can efficiently manage and present massive amounts of data. The large volume of social media data makes it a rich source from which to mine intelligence. Containing both textual and numeric data, commercial brokers now provide search and retrieve platforms for research. The U.K. Collaborative Online Social Media Observatory (COSMOS) helps academic users exploit data such as Twitter feeds, elucidating important methodological, theoretical, and technical dimensions to using social media in research (Sloan, Morgan, Burnap, & Williams, 2015). However, their massive volume and restrictive terms of use and commercial usage clauses obstruct collating an open longer-term data research resource. When the Library of Congress attempted to collect the whole Twitter archive from 2010, the massive scale of storage (half a trillion tweets) and rights issues halted the project (Scola, 2015).

Data like supermarket loyalty programs, mobile phone call records, or swipe cards for security systems are a powerful by-product of transactional systems. These are primarily used for administrative purposes and typically reside in proprietary relational databases, making these data difficult for everyday researchers to access, requiring purchase, or bespoke brokering.

Dedicated academic research data online portals. Archives of digital social research data were established in the United States and Europe in the 1960s to house expensively collected national public opinion and social survey data. As collections grew and the number of archives increased, collaborations like the Data Documentation Initiative (DDI) developed more harmonized approaches to digital data storage, access, and documentation standards. The Council of European Social Service Data Archives (CESSDA) provides an official network of social sciences data services.

As significant resources are invested in quality assessing and documenting their data collections, the quality and integrity of data held by these archives is of a very high standard. They increasingly accept a more diverse range of digital data, from historical databases to qualitative interviews (U.K. Data Service, 2015). Although early data archives long pre-dated the Internet, from the mid-1990s, users increasingly interacted with them online, browsing and downloading data using web visualization systems. Due to the size and

computational power required for working with very large datasets, data will no longer be moved to the researcher, but instead the researcher “moves” to the data. This is an established model in astronomy and climate science, with purpose-built shared data and analysis facilities.

Data sharing policies among research funders have exponentially increased other kinds of data repositories, spanning academic libraries, publisher-related repositories, and dedicated commercial research data storage services. By December 2015, an international registry of research repositories (re3data.org) listed around 1,400 data repositories, 378 specializing in the humanities and social sciences. Academic journals also play a role in ensuring that data underpinning published findings are available for readers and reviewers. Economics, psychology, and, more recently, political science have led the way in social sciences, with journals expecting data to be made available upon request, either submitted as supplemental material or deposited in a suitable public repository. We discuss this relatively new agenda in Part 3.

At this point, we turn to examine the trustworthiness of online sources of information and data for qualitative analysis. Are these sources and outlets different from their more established equivalents? They certainly have greater accessibility and visibility, yet placing trust in them can be a more difficult process.

Trustworthiness of Online Data Sources

Quality of data can refer to its sustainability and integrity. Let us first deal with the *sustainability* of web-based resources, as ironically, in some ways, they can be far more fragile than a box of ageing paper. Unless online data sources have a sustainable location, there is little guarantee of ongoing discoverability and availability, let alone referencing and citation. As a serious data user, it is frustrating to find sources have disappeared, web pages are dead, or new versions of data have replaced older without warning. To stand the test of time it must be possible to revisit sources of data used for research; *persistence* is crucial.

Numerous best practices and protocols for digital data publishing seek to satisfy this need. Data archives have been curating digital data for half a century and have led the way in advocating robust citation of data sources. “FAIR” data publishing principles embrace the principles of **Findability**, **Accessibility**, **Interoperability**, and **Reusability** and focus on the specification of minimally required standard protocols, lightweight interfaces, and formats (The Future of Research Communications and e-Scholarship, 2014). Protocols for “FAIR” data are reaching more eyes as institutions are slowly building their own data curation and publishing infrastructures. “Digital preservation” has become a distinct discipline with dedicated master classes sitting alongside those on curation of medieval manuscripts.

Continued access to a web-based data source relies on the persistence of its citation. Digital Object Identifiers (DOIs)

are used to ensure that online resources remain alive. Represented as a simple string of characters using a Uniform Resource Name (URN), a DOI is a unique global and persistent identifier for an online digital resource, a universal numeric fingerprint. A resolution service for these URNs needs to persist and identify the source data even when the publishing technology or location changes. As a parallel to printed documents, the URN is accompanied by information about author(s), title, and the publication date. Publication style manuals increasingly provide guidance on these sustainable citations for datasets (American Psychological Association, 2013).

Much qualitative data are now published online, especially from the oral history tradition. Notable examples of well-presented audiovisual content include the USC Shoah Foundation Testimonies Archive (vhaonline.usc.edu). Yet many other projects that mount their outputs onto the web rely on short-term funding and may not be sustainable at all. Many websites use poor practice when publishing web content, using “hard-coded” hotlinks, which relies on manual work, and an inflexibility, if web addresses change. There is disappointingly little use of robust metadata schemas that allow structured mark-up of data and context. The U.K. Data Service publishes citable qualitative data to the highest standard and, in the United States, a new Qualitative Data Repository (QDR) seeks to increase its embryonic capacity to do likewise.

Formats also play a significant role in data accessibility. Professional archives have a digital preservation strategy and undertake to migrate formats forward to keep them readable over time and regularly refresh media. Some formats are not suited to easy data extraction, for example, extracting structured information from a pdf. Open formats allow data to be extracted and transferred into a user’s software of choice.

Despite the advantages in accessing data as digital media, Internet pioneers have lately warned of society’s increasing reliance on cloud storage and of the problem of obsolescent technology - systems, software and hardware. Some even advise the public to print their digital photographs because paper may actually be more enduring than contemporary software. A Google executive, Vint Cerf, commented that this could affect not only personal materials but official correspondence residing in emails, court judgments, Twitter traffic, blogs, and videos (Sample, 2015).

Trust and Integrity of Online Data

To assess whether we can trust data, we need detailed information: about the provenance of the data to hand such as the reason for collecting the data, and the sampling and methods used; about the content of the data, such as its shape, format, volume, and topics; and whether it can be used legally and ethically. Meeting such concerns is instructional literature on how to document data so that they are useful for research in

the longer term (Corti, Van den Eynden, Bishop, & Woollard, 2014; Digital Curation Centre, 2015).

Setting out the context in which data have arisen or belongs plays a key role in enabling reuse. This applies not only to primary fieldwork with known collection parameters but to collation of data from unknown sources, like taking dumps of open data from portals or websites. As an example, for a collection of 50 qualitative in-depth interviews available in MS Word format, we could look at whether there exists documentation about the project’s origination and fieldwork approach, a topic guide covering questions asked, and metadata about each interview, such as respondent characteristics and interview settings. Rich description, annotation of classifications in Computer Assisted Qualitative Data Analysis (CAQDAS) packages, and retrospective narrative or interviews are also all useful devices for capturing nuances of study design, fieldwork and data preparation processes, as are notes addressing different levels of study context, e.g., about the fieldwork situation or broader social, cultural or economic context for an interview, or about the data, like annotations for a transcribed interview.

Although we do not wish to rehearse here the debate on the problems of “being there” and “research context” in the role of raw qualitative data as a resource for reanalysis, we note that both of the authors have been fully engaged with this argument in the United Kingdom from the start (Corti & Thompson, 2004; Fielding & Fielding, 2000). The establishment of Qualidata and mandating the sharing of qualitative data from research grants in the 1995 by the Economic and Social Research Council (ESRC) sparked the whole debate. Some qualitative researchers claim that it is too difficult to divorce research data from its context (such as Mauthner, Parry, & Backett-Milburn, 1998). For them, context goes beyond static information about persons, actions, or situations involved in data collection, and considers factors resulting from interaction and interpretive processes. For example, the status of an in-depth interview from a series benefits from situating it within the larger research narrative, or examining a turning point in someone’s life requires information about their biography (Holstein & Gubrium, 2004).

It is true that the further we move from direct engagement in data collection activities the further we step away from context. While original data creators gain benefit from having “been there” as part of the originating context, equally there are situations whereby research investigators who are not themselves directly engaged in fieldwork must rely upon their co-workers or fieldworkers’ documentation of the research process and data being generated. This requires a level of trust and an ability to recreate context. Data archives have made some pragmatic assumptions about the degree to which data documentation enables data to be understood and used independently; and primarily, help reusers in understanding and taking into account the intentions of the primary researchers who collected the data.

Collecting data from online sources, such as social media, presents a significant barrier to context, and trust in the information, and the provider of information becomes more difficult. Onward sharing of digital media enables consumers of research to better validate trust in the data supporting an analysis, and more likely that the method and findings becomes a touchstone in community research endeavor.

Even when context is sufficient, users confronting unfamiliar sources must still take on a detective role, examining and assessing provenance and trust in the data themselves, appreciating its limitations, and setting down assumptions that can help frame the secondary research tasks. Social historians routinely revisit data sources as part of their approach to scholarly practice, encouraging a willingness to embrace the slow and rigorous activities of documentary analysis. The need to evaluate methodically the very sources they are revisiting is second nature. Social scientists are less prone to this approach, yet as the need to justify collecting more data increases, the greater the need to learn from neighboring disciplines (Crow & Edwards, 2012; Kynaston, 2005). Time to evaluate data sources should be built in to any secondary analysis project.

In closing this section, we note that data that are more formally published online bring many advantages with respect to quality and trust. As an example, in 2015, the U.K. Data Service gained “Platinum” level certification for its open data of a large qualitative oral history collection on the Edwardians from the Open Data Institute (ODI, 2015). These certificates require the data publisher to provide evidence, demonstrating transparency in the processes and systems in place to manage and publish data. The evidence focuses on provision of detailed machine-actionable metadata, sustainability, and clarity on ethical and legal matters.

Part 2: Online Research

Although many of the points that one interacts with data—online and offline—are indistinguishable in terms of method, we do suspect that there is greater ontological complexity to be negotiated in working with “online” data. Kitchen’s infamous characterization of “big data” as

huge in volume, consisting of terabytes or petabytes of data; high in velocity, being created in or near real-time; diverse in variety, being structured and unstructured in nature; exhaustive in scope, striving to capture entire populations or systems ($n = \text{all}$); fine-grained in resolution; relational in nature, containing common fields that enable the conjoining of different data sets; flexible, holding the traits of extensionality (can add new fields easily) and scalability (can expand in size rapidly)

neatly sums up the challenges (Kitchin, 2014, p. 262).

Social media and informal writings published on the web are increasingly popular sources of “big data” for sociologists. However, there are numerous challenges facing analysts who wish to use these sources, such as volume, uncertain

provenance, and also methodological traps that can easily be fallen; finding highly, and sometimes amusing, spurious correlations (Vigen, 2015). Stupendous numbers can be assembled to support trivial conclusions. We currently lack accessible, highly specified tools that enable us to extract depth from social media traffic, identify contradictions, or even establish whether the message is facetious. Yet the massive volume of tweets on so many topics allows us to capture phenomena as they arise.

Analysis of blogs as data has developed from novelty to an increasingly mainstream research method gives us insights into how users both produce and consume content while communicating and interacting with each other in an increasingly “confessional culture” in which participants curate and reflect upon their personal lives in the public realm. The material is often appealing because it provides detailed first-person textual accounts of everyday life that are spontaneous and naturalistic. This research agenda is consistent with “rethinking the repertoires of empirical sociology” (Savage & Burrows, 2007, p. 895). Yet the content of blogs is difficult to capture and import into popular qualitative analysis software.

More sophisticated methods tools are needed to handle and exploit complex web content, with functionality that moves beyond semi-automation of coding and retrieval processes. New big data platforms like Hadoop offer data management and cleaning tools and algorithms to cope with massive volumes of text, but do not work by themselves; problems and bias in data must be modeled and scripts run to clean and interrogate the data. These activities are typically beyond the skillset of most social scientists, yet are vital if they wish to exploit the ubiquitous world of data.

Here, we briefly direct attention to the digital devices that create what we are calling the digital environment. Ruppert, Law, and Savage (2013) suggest that digital data, devices, and platforms demand that we rethink our assumptions about social science methods in respect of “transactional actors; heterogeneity; visualization; continuous time; whole populations; granularity; expertise; mobile and mobilizing; and non-coherence” (pp. 22-23).

The trope behind this view is a familiar but appealing one, that of the snake eating its own tail. That is, digital devices are implicated simultaneously in being shaped by our social worlds and acting as agents that shape those worlds. We do not aim to fully engage with this perspective here, as it would move the present discussion beyond exploring new affordances into deeper matters of logic and epistemology. Capturing the sometimes-mediated nature of digital discourse, such as a multi-person email thread, lacks a precise analogy in a pre-digital social world, but that does not mean that pre-digital researchers were unaware that discourse can be mediated. A worthwhile methodological endeavor would be to explore in detail the ways in which the online and the offline worlds of research display both semantically significant technical differences and endless thematic recurrence in terms of our understandings of the status of “data.”

The provision of new data platforms set up specifically to encourage mobilization of the crowd to seek and add to scientific knowledge from Kitchen's "extraordinary" data, means that research is no longer the sole domain of academics; an increasing number of citizens are taking on the role of "researcher." Before considering whether this intrusion into the academic sphere presents a challenge for scholarly practice, we highlight some examples of citizen "social research."

Online Citizen Research

Citizen research is growing around the convergence of several trends, including rising living standards worldwide, the rapid growth of online resources, the increasing number of people with some knowledge of social research and its methods, and the difficulties that modern states have in controlling the online world. The broadening of the research community brings with it new purposes for doing research and new understandings of what it is. In turn, this is likely to radically affect the role of "data archives" and the practice of secondary analysis as explored in this article.

Citizen science is an instance of the fact that social media can assemble and support large numbers of people for collaboration around a shared interest. For example, the Citizen Science Alliance is a collaboration of scientists, software developers, and educators, who collectively develop, manage and utilize Internet-based citizen science projects to advance scientific knowledge and the public understanding of both science and of the scientific process (Citizen Science Alliance, 2015). "Zooniverse" projects range from the classical sciences to climate science, from ecology to planetary science. The Galaxy Zoo Project has around 200,000 members of the public involved in classifying galaxies using images taken by advanced telescopes from the Sloan Digital Sky Survey, dating back to 2007. On its own, the capacity of the astronomy science community is insufficient to directly analyze the vast numbers of images and so citizens act as first "filters." When something interesting is found it is confirmed and further analyzed by professional astronomers. The numbers associated with such initiatives are often dramatic. To their surprise, within 24 hr of launch, almost 70,000 classifications an hour were being received and in the project's first year some 150,000 volunteers helped classify 50 million galaxies, with a Puerto Rican housewife "discovering" two hypervelocity stars.

Another example makes the point that such online citizen science initiatives are natural extensions of classroom teaching, reinforcing pedagogy and providing the excitement of knowledge discovery. The scientific community in Brazil is developing an online catalog of every plant species in the Amazon ecosystem. The "Wikiflora" involves teachers and schoolchildren, as well as other members of the public (Yapp, 2011).

These citizen science examples relating to astronomy and botany both hinge on crowdsourcing, a phenomenon

consistent with Ferdinand Tonnies' conceptualization of modernity's move from "community" to "association" (Tonnies, 1988). For Tonnies, association is a set of social bonds formed around a single shared interest. Technologies like the telegraph and telephone have facilitated such sociality, but online media have accelerated it. Co-presence is no longer required; one can have stronger relationships with people on the other side of the globe than with one's immediate neighbors.

Although citizen research is primarily motivated by the leisure time (or unwaged) interests of the participants, citizen "social science" is often motivated by a socially progressive agenda, or indeed, an oppositional stance toward mainstream institutions, policies, and politics.

A social science example of crowdsourcing is the Democracy Club, which involved politically aware individuals across Britain inputting information about election candidates' views on local constituency matters to a central database (Democracy Club, 2015). Previously, the only way to compare the views across constituencies of a given party's candidates about local issues was to subscribe to hundreds of local newspapers! It became a valuable tool for professional journalists. Another example is that of activists using social media in countries including the United States, the United Kingdom, Spain, and Greece to mobilize support for, and organize, demonstrations, and other interventions, such as free workshops for the unemployed and support for foodbanks. In the aftermath of the 2009 Greek fiscal crisis, researchers in Greece identified online petitions as a freely available source of quantitative and qualitative data regarding community tensions, emergent support for non-mainstream political affiliations, and social cohesion (Briassoulis, 2010). There is now a considerable body of work using social media output as data in analyses of groups like Spain's "Indignados" and the international "Occupy" movement (Gonzalez-Bailon & Borg-Holthoefler, in press).

A further politics example is the use of mobile telephony and social media in Obama's first presidential election campaign, widely seen as superior to that of his opponents. The Obama '08 campaign transformed political participation and civic engagement by creating a nationwide virtual organization that motivated more than 3 million individual contributors and mobilized more than 5 million volunteers (Cogburn & Espinoza-Vasquez, 2011). The feature that made Obama's social media team so effective was their pooling of online information regarding preferences and opinions, which enabled their fundraising calls to potential supporters to be well-informed and attuned to the individual's sensitivities.

Online media for social research. Social researchers are not typically passive observers of phenomena and some use online media to combine a research agenda with empowerment of the objects of study. For instance, in the case of the Athenian anti-austerity protest movement and the Indignados, a practice of "militant ethnography" has emerged (Juris,

2014). Its stance is not only to participate alongside protestors for customary purposes of acquiring data but also to produce analyses that are directed at assisting the protestors, a contemporary form of Action Research.

An example that illustrates empowerment and the new affordances of digital technologies is that of research with First Nation communities in remote parts of Canada. This work includes First Nation people as co-investigators, both directly, in the research team, and in the remote site, where community members become the field team. The research team seeks to advance community interests not only by providing online resources and educational services but also by representing its interests to regional and national government (Beaton, Perley, George, & O'Donnell, in press). The work is underpinned by networked video-teleconferencing applications. This approach can be extended to work in other settings where communities or groups are marginalized, such as inner city Detroit or the banlieues of Paris.

These new research practices and affordances are, of course, not without their problems. Work of this sort with marginalized communities poses new ethical issues. For instance, the researcher may become the pivot point between community members and the government interest, or researchers may be regarded in some forums as biased. Furthermore, community members involved, in common with many citizen scientists, are unlikely to have any social science training. Their efforts may need to be scripted through protocols or closely supervised, or the researcher will need to take time out to provide training. Doubts may be raised about quality control, reliability, and digital rights management.

One of the main problems facing citizen social research is the somewhat intractable matter of physical resource. Researchers without dedicated funding have to negotiate the cost of equipment, software, and access. Many will be using under-specified machines, or have only occasional access, or be reliant on low bandwidth connectivity. The First Nations research example provides a way round the equipment problem by using cloud-based collaboration software.

With these problems in mind, the idea of coproducing research is appealing, especially for organizations without appropriate resource, capacity, or training to gain insight from data. Many civil society organizations collect operational, research and evaluation data but with tight budgets and limited research capabilities few can fully exploit it, or indeed gain insight from potentially useful web-based data sources. Organizations like the charity DataKind UK operate as a “data analysis” broker between these organizations and volunteer data scientists aiming to help them understand, the needs of their beneficiaries, measure impact, plan scenarios, and improve operational effectiveness. For example, DataKind worked with a nongovernmental organization (NGO) that gives money directly to the poorest villages in Kenya and Uganda. The NGO knew that a key indicator of poverty was whether there is a thatched or durable metal roof

on village houses. By using satellite imagery and learning algorithms, DataKind volunteers developed a model to classify roofs so NGO staff on the ground could more effectively identify the poorest communities most in need (DataKind, 2014). Such “analytic philanthropy” harnesses data, analysis, and computational skills to assist in solving social problems.

A Crisis for Methods?

We now consider what the new approaches to research engendered by the digital environment represent in methodological terms. The web enables the compilation of information “mash ups” that mix types of data, and crowdsourcing, where the adequacy of the information compiled is only as good as that of the least motivated participant. But both represent what is effectively a mass uptake of secondary analysis among participants who will mostly be unaware that secondary analysis involves rigor and that, even then, there remains academic debate about its legitimacy (Fowler, Whyatt, Davies, & Ellis, 2013; Wiggin, Newman, Stevenson, & Crowston, 2011). Such concerns little interest participants, who will often have instrumental purposes in mind, like finding good schools for their children.

These observations apply an established social science lens to the new practices. Some argue that such a lens is obsolete. Hardey and Burrows (2008) write of the new developments challenging the “hegemony” of social science. Savage (2013) notes the role of methods in the intellectual differentiation between scientific and humanities expertise and places the turn to make methods themselves an object of inquiry in the context of a “dialectic of transparency” which recasts the relationship between the implicit and explicit. These authors welcome the new developments, and to those who worry about “know-nothing” citizens doing research while lacking a background in the social science canon, our response might be that such a level of public engagement would be the envy of other branches of science.

In fields like physics and biology, special grants have to be given out to promote the public understanding of science. In our case, the general public is already playing our knowledge game, and the need is to help them better understand what makes it a “science” (Bauer, 2009). Compared with the widespread hostility toward GM (genetically modified) crop researchers and nuclear engineers, we should be proud that the core methods of the social sciences are now being used by all sorts of people, facilitated by digital tools that anyone can use. Like any other sea change, the social science community has the choice whether to stand against the inevitable, or to embrace it while seeking to develop new understandings—of “training,” “quality,” and “authority” (Fielding, 2014). If we choose the latter, we will also need to negotiate the obstacles other than educational credentials that constrain citizen social science.

Part 3: Transparency and Replicability in Research

As we saw in Part 1, the ability to make data from a research study available via digital means is not just valuable as future-proofing but also for purposes of “scientific transparency,” accountability, and integrity. For qualitative research, these virtues may be particularly important, because qualitative data analysis constantly remains contested. Indeed, there is an increasingly worrying tendency to rewrite history by asserting that such critical positions began out of the blue in the 1970s (a claim made by Elizabeth St Pierre, 2016, p. 25, among other authors making the same claim), that they then gelled in the work of Denzin, Lincoln, and Guba in the 1980s, and saw their real advance in the 1990s work of Patti Lather (2007).

Those more attentive to the history of social research will be aware that critiques of the fundamental concepts associated with what are essentially positivist (and latterly neo-positivist) criteria for evaluating qualitative inquiry were actually present in the social sciences from the 1920s onward. We should not forget the founding contribution of the Chicago School in developing guiding concepts for qualitative research, as noted by Norman Denzin and Patti Lather themselves at the 2016 *International Congress of Qualitative Inquiry*. Although we acknowledge the importance of such past critical debates, our purpose in the present discussion is to highlight new affordances for qualitative research rather than to directly contribute to the epistemological debate.

Certainly, the field has made real strides in the last 20 years in developing systematic and/or formal approaches to qualitative data analysis. Yet different approaches not only remain but flourish, and rightly so. Enabling different interpretations to be applied to the same dataset offers real promise to enable cumulative knowledge from qualitative research, for example, by promoting independent coding exercises to establish points of convergence and of genuine difference. Opening up data in qualitative research is thus important for demonstrating openness and for enriching and enhancing the context of findings.

The more recent branding of open science, open research, and open data have initiated yet another transparency agenda. In the period since the early 2000s, we have witnessed a surge of support for data sharing in the domain of both research and policy: Internationally, governments are pushing for greater transparency in research and greater reuse of data to maximize the return on science investments; research funders mandate streamlined open access to literature and high quality documented research data; academic publishers demand access to data underpinning findings, for scrutiny or further exploration. Fortunately, human and technology capability and capacity has managed to keep abreast of these drivers, and we now see an increasing number of “transparency organizations” that have been established, including the Research Data Alliance (RDA); the U.S. Center for Open

Science (COS) and the Berkeley Institute for Transparency in the Social Sciences (BITSS); Experiment in governance and politics (egap); and more recently, American Political Science Association (APSA). In the United Kingdom, we have the ODI and the U.K. government’s Public Sector Transparency Board, Open Data User Group, and Research Sector Transparency Board. Some are addressing science as a whole, while others are honing in on supporting how this should and could work for social science research.

Countering these positive drivers that have led to this apparent urgency in this transparency agenda has been the exposure of research fraud—ranging from the faking of experiments and massaging of data to plagiarizing the works of others, and the less dramatic but insidious “desk drawer” problem in which negative or ambiguous findings are less likely to be published. Tracking down such problems offline is laborious but digitization of journal content has enabled bulk validation analysis of articles. Rates of retraction have risen sharply (Times Higher Education, 2014).

The practice of sharing data beyond the original project is now an accepted practice, and some would claim, an art, but only in very few countries is this true for qualitative data. The United Kingdom has been most fortunate in being early adopters in this development and much of the key literature on data sharing, whither and how, has come out of the United Kingdom. The ESRC data policy that was established in the 1970s primarily for the onward sharing of survey data, boldly decided, in 1995, to extend this remit to include qualitative data outputs in 1995. Now of the 3,000 plus collections of primary research data accumulated under this policy, some 900 are from qualitative or mixed methods research, available to download from the U.K. Data Service online access points under fully open or more restrictive access.

With all the national and local resources available to support practical data sharing, creators of qualitative data, and their employing organizations in the United Kingdom, have gained a good understanding of what it means to make data shareable. This well-established data sharing culture is not yet reflected in the United States and, despite National Institutes of Health (NIH) and National Science Foundation (NSF) data sharing policies dating back to 1989, little research data from projects are shared, other than some of the larger social surveys.

To embrace the recent principle of research replicability, the political science community has boldly embraced a set of Data Access and Research Transparency (or DA-RT) principles. In 2012, the powerful APSA took a collective decision to integrate DA-RT principles into its Ethics Guide, adhere to the 2014 Transparency and Openness Promotion (TOP) Guidelines, and adopt a Journal Editors’ Transparency Statement (JETS) (DA-RT, 2015). Although replication procedures are already fairly well embedded for quantitative research-oriented journals, such as the *American Journal of Political Science* (AJPS) or *Research and Politics* which have their own stringent replication requirements, the

position for qualitative research is in an emergent phase. This wave of journal pressure for open access to data may be well placed to stimulate the sparse practice of sharing qualitative data among U.S. social researchers, even if some social science research data cannot be completely open due to ethical and legal restrictions.

The shift in semantics from *open access* to *transparency* in the research space is important but, in some ways, also troublesome. It is important to counter mistrust in research findings and to push for fairness in the even publication of outcomes, for example, in the mission being pursued by the AllTrials campaign that calls for the publishing of all clinical trials results, including negative. Yet the problems in demonstrating transparency in the scientific sense in qualitative research may provoke renewed attacks on qualitative research.

Furthermore, differentiating between whole project data sharing and making available parts to back up specific findings is problematic, and actually, is likely to exacerbate the “context” problem, particularly for qualitative research. Although there is very good guidance and training available on how and where to share “whole project” research data, practical help for authors of qualitative research publications in defining what “supporting data” for an article means is currently lacking. Elman, Kapiszewski, and Lupia have provided some early thinking on transparency for qualitative research as it applies to journals, contributing to DA-RT Guidelines, particularly on providing exemptions for ethical reasons (DA-RT, 2015). DA-RT distinguishes between *production* and *analytic* transparency. *Production* transparency is about revealing the process of data generation and is done effectively by U.K. scholars using approaches discussed in Part 1. *Analytic* transparency, on the other contrary, requires an author to evidence their claims in a paper, suggesting that this is an activity that sits in the “evidence” space between publishing a coherent set of data and including extracts that underpin arguments in a journal piece. One technique proposed for evidencing claims is “Active Citation” (Moravcsik, 2010). This makes use of citations to data, such as an excerpt of a source or a link to the original source, plus an annotation, ideally placed in a “transparency appendix” to the article. We might argue that it is not possible to achieve links to every source referenced for all genres of qualitative research, especially as we move into the interpretative spectrum. Might publishing “chunks” of data out of context threaten the integrity of, for example, an interview narrative? Furthermore, does this approach provide the best “evidence” for a claim and, what if preceding and following paragraphs contradict the claim?

We conclude that a middle ground is likely to be preferable and achievable, where we seek to share as much of the original data as possible in digital format alongside a narrative rationale about claims. This Special Issue highlights the value of inviting readers to view data directly online, so that an interview excerpt can be viewed in its context of the whole interview, and a link to its methods. The U.K.

Qualibank which has published sets of qualitative research in an open online environment, enables this feature, so that a paragraph can be cited and its citation URL resolves back to the original interview transcript, hosted online as part of the full archived data collection (U.K. Data Service, 2014). See, for example, this interview extract as shown (<https://discover.ukdataservice.ac.uk/QualiBank/Document/?cid=q-1dba72b1-d148-40e7-b3dc-a81ae230ca80>; Thompson & Lummis, 2009). Enabling persistence of these links is also a key part of meeting this transparency mission. If links break, the context and “evidence trail” disappear.

Authors who wish to fulfill a transparency mandate from a journal will welcome practical guidance that highlights case studies of particular qualitative research approaches and how their data and context can be most fruitfully presented. However, we hope that, going forward, journals with strict replication requirements will not seek to penalize qualitative research articles through rejection on the basis of insufficient demonstration of analytic transparency.

In some disciplines, notably health care research, we have seen a huge increase (10-fold) in the use of secondary analysis of qualitative data over a 20-year period (Bishop & Kuula-Lummi, 2016, in this Special Issue). This suggests that although the take up of reuse of someone else’s data has been variable, some disciplines have led the way, with others like political science, likely to follow as our stock of available online digital data grows year on year, through active data sharing, and maybe even through encouragement try ones hand at replication. Published raw and processed data offers a resource for both scrutiny and critiquing analysis and results, allowing scholars to rerun models or unearth new insights based on different assumptions of the source data (see Laurence & Elliot, 2016, in this Special Issue; Sutcliffe-Brown, 2016).

As journals seek to enhance the veracity of claims in what research they decide to publish, through greater exposure of data and method, so publishers are seeking to offer real-time interaction with this information. This brings us onto new forms of publishing research.

Part 4: Publishing and Communicating Research Online

As we saw in Part 3, there is a current trend toward data publishing, whether voluntarily or enforced. Publishing research now embraces a much wider portfolio of formats that are increasingly consumed online: blogs, visualizations, video, and interactive media. Although this has been commonplace in journalistic and citizen science projects for a while, academics have tended to lag behind, though popular online outlets such as *The Conversation* has really gained in popularity over the past year or so. Even using basic “hot linking,” modern publishing platforms can offer richer methods for consuming research. The reader can get more involved in an article through exploring an “enhanced publication,” an

appealing term coined by the Dutch organization, SURF (Woutersen-Windhouwer et al., 2009). The idea embraces linking a digital publication with information such as research data, source documents, audio or video segments, graphics, models, and so on. These extras not only provide an important role in the clarification of the context of that publication but also, if well structured and presented by the publisher, allow the reader to navigate a narrative taking time to stop off and enjoy the vistas and viewpoints along the way. This Special Issue has attempted to include articles that have direct links to sustainable online cited data extracts and resources to provide these enrichment features (see Haaker & Morgan-Brett, 2016; Lawrence & Elliot, 2016; Sutcliffe-Brown, 2016, all in this Special Issue).

In a conventional text-based paper, the significance of a supporting reference often goes unrealized until the reader has tracked down the item of interest. In quite frequent cases, it is not even apparent from the citation whether it is a supporting reference, a reference that expands on the point, a reference criticizing or challenging the point, or even a tangential reference. In contrast, the digital dissemination environment enables powerful retrieval and rapid inspection.

The analogy can be drawn with the successful genre of fiction known as Young Adult Literature (YAL) that exploits digital resources to enable author/reader interaction to enhance the experience of already-committed readers. This may involve multiple representational modes, for example, using pop-ups to layer audiovisual media onto text or vice versa with contextual material to supplement the story, engage with social media, and to present alternate storylines (Hundley & Holbrook, 2012). Analogous uses may be attractive in writing up research, but once again there is a likely to be a challenge to established, core practices. For instance, texts can be multiple-authored dynamically over time—which may be an artful way to convey to readers the stages in the development of an analysis, the existence of alternative interpretations, and the message that all analysis is interim, representing only the state of the art at the time of writing. Such work is dynamic and is more the work of a production team than solo author. If the content of a work changes every time it is “used,” it is hard to see how such work could be peer reviewed, or support a case for promotion.

But just because we can enrich the reader’s consumption and experience of our research by getting the reader closer to the evidence, what does this entail for the author, the reader, and the publisher? Some readers may find this conducive as a means of discovery and as opening up a different relationship between author, text, and reader. But, the term “enhanced” suggests more work, which undoubtedly presents an additional burden for all participants.

For the author, sources to be displayed must be available for the longer term which means a reliable website publisher that adopts persistent citation practices. There are choices to make about how to design the route map and viewing places that can be offered to fellow travelers. Data must be prepared

and published in a reliable repository, bringing with it a raft of additional tasks to ensure that accuracy and sufficient context is provided. However, given current emphasis on gaining and identifying “impact” for research, digital publishing affordances are timely. Digital environments are self-documenting, presenting a recoverable trail for digital contact, communication, or transaction and offer a vehicle for serendipitous discovery; what Kane (2004) has tagged “playbour,” denoting an activity of labor that’s feel like play.

For the publisher, an interactive user interface must be slick and easy to use. In the Public Library of Science’s (PLOS) Open Access journal, *PLoS ONE* (plosone.org), articles are enabled with live interactive boxes, such that other researchers can run their own routines on data. Social science journals have been slower on the uptake to embrace such interactivity, although *SAGE Open* provides a basic web-based interface. Interfaces that allow commenting by readers are useful too but have been slow to emerge. Finally, a relatively new form of publication, the “data paper,” goes further by providing a rich description and narrative about the data, its key features, its potential value, and uses. One of the first social science examples is the *Research Data Journal for the Humanities and Social Sciences* (Brill & Data Archiving and Networked Services, 2016).

Finally, *for the reader*, how much time can one invest in absorbing richness and context as opposed to skimming key points?

The significant investment in time spent creating and reading an online interactive article means that there needs to be a clear, identified benefit. Greater transparency can indeed be offered to the reader in how the original author connected the data to their final story. However, on the downside, there are an ever-increasing number of published articles and we might ponder who has time to indulge in this pastime in addition to struggling to keep up with the literature by abstract scanning and conclusion browsing.

We suggest a middle path whereby research based on online data sources uses a minimal set of protocols surrounding the selection and referencing of data that aids sustainability and longevity. Adding linked sources and commenting can transform the once linear article transformed into a dynamic object that presents a richer encounter for readers, yet one must strive to balance author burden versus likely impact.

Conclusion

Our article has presented some of the opportunities and challenges of undertaking research online in the 21st century. The number of online outlets for data is growing exponentially, offering a major digital research resource, yet presenting researchers with deep concerns about trust, sustainability and their own capacity to handle and link so much data. As new and larger data sources come on stream, so methods and

tools need to be adapted to allow us to select, query, and visualize data.

The familiar world we have inhabited as social researchers has for some time now been disjoint with the digital world in which we live as private citizens. The risks and prospects posed by the growth of citizen research, the opening up of data, and the ease of hyper-connectivity suggest that social science is being re-visioned for the digital future.

In that future, “trust” remains a major issue: trust in research integrity, trust in the reliability and validity of data sources, trust in the persistence of online materials, for us to be able to return to them now and in the distant future, trust in intelligent interpretation, trust in the publishing author to narrate a trustworthy story, trust in the content publishers to provide an enjoyable reading experience, and trust in the consumers of research findings to devote some time to exploring relevant evidence so carefully prepared and provided.

We note the connection between epistemological matters and one of the prime affordances that we have presented; that is, around the scope that new digital tools provide for data transparency through robust data publishing. Where the status of qualitative data is contested—and it nearly always is—one large step toward making an analysis accountable to others is to be able to show the relationship between data and interpretation. Neo-positivists and post-positivists may see this possibility as playing to the “replication crisis” that has beset fields as diverse as biochemistry and economics, while postmodernists may see it as an opportunity to build numerous equally plausible interpretations up from the same data, thus demonstrating the impossibility of knowledge.

Whatever one’s perspective, being able to directly examine the data that a researcher adduces in support of their analysis will inevitably change the ground of the epistemological debate. It will also help novices, including students, understand professional norms and standards in a more direct and thorough way than has previously been possible.

To finish, we believe that qualitative researchers need not feel threatened by the research transparency agenda, by new forms of publishing, nor by competing citizen research or by the roller coaster of Big Data methods. Instead, they are advised to embrace the opportunities that rich qualitative data can bring.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research and/or authorship of this article: This article was prepared with support from the Economic and Social Research Council (ESRC) as part of its U.K. data infrastructure funding for the U.K. Data Service.

References

- American Psychological Association. (2013). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: Author.
- Bauer, M. (2009). The evolution of public understanding of science—Discourse and comparative evidence. *Science, Technology and Society, 14*, 221-240. Retrieved from [http://eprints.lse.ac.uk/25640/1/The_evolution_of_public_understanding_of_science_\(LSERO_version\).doc.pdf](http://eprints.lse.ac.uk/25640/1/The_evolution_of_public_understanding_of_science_(LSERO_version).doc.pdf)
- Beaton, B., Perley, D., George, C., & O’Donnell, S. (in press). Conducting appropriate online research with remote indigenous communities. In N. Fielding, et al. (Eds.), *The SAGE handbook of online research methods* (2nd ed.). London, England: Sage.
- Bishop, L., & Kuula-Luumi. (2016). Revisiting qualitative data reuse: A decade on. *SAGE Open*.
- Briassoulis, H. (2010). Online petitions: New tools of secondary analysis? *Qualitative Research, 10*, 715-722.
- Brill & Data Archiving and Networked Services. (2016). *Research data journal for the humanities and social science*. Leiden, NL: Brill. Retrieved from: <http://www.brill.com/products/online-resources/research-data-journal-humanities-and-social-sciences>
- Citizen Science Alliance. (2015). Available from <http://www.citizensciencealliance.org/>
- Cogburn, D., & Espinoza-Vasquez, F. (2011). From networked nominee to networked nation: Examining the impact of web 2.0 and social media on political participation and civic engagement in the 2008 Obama campaign. *Journal of Political Marketing, 10*, 189-213.
- Corti, L., & Thompson, P. (2004). Secondary analysis of archive data. In C. Seale, G. Gobo, J. Gubrium, & D. Silverman (Eds.), *Qualitative research practice* (pp. 327-343). London, England: Sage.
- Corti, L., Van den Eynden, V., Bishop, L., & Woollard, M. (2014). *Managing and sharing research data: a guide to good practice*. London, England: Sage.
- Crow, G., & Edwards, R. (2012). (Eds.). Perspectives on working with archived textual and visual material in social research [Special issue]. *International Journal of Social Research Methods, 15*(4), 259-262. doi:10.1080/13645579.2012.688308
- CTIC (La Fundación Centro Tecnológico de la Información y la Comunicación). (2015). *Public dataset catalogs faceted browser* (Open data @CTIC Sandbox). Asturias, Spain: Fundación CTIC. Retrieved from <http://datos.fundacionctic.org/sandbox/catalog/faceted/>
- Data Access and Research Transparency. (2015). Available from <http://www.dartstatement.org/>
- DataKind. (2014, January 24). *Using satellite images to understand poverty* (Blog post). Author. Retrieved from <http://www.data-kind.org/blog/using-satellite-images-to-understand-poverty/>
- Democracy Club. (2015). *Your next MP*. Author. Available from <https://democracyclub.org.uk/>
- Digital Curation Centre. (2015). *Resources for digital curators*. Author. Retrieved from <http://www.dcc.ac.uk/resources>
- Eastwood, N. (2015, January 13). The Internet can change the developing world. *Huffington Post Tech*. Retrieved from http://www.huffingtonpost.co.uk/nigel-eastwood/the-internet-can-change-t_b_6154540.html

- Fielding, N. (2014). Qualitative research and our digital futures. *Qualitative Inquiry*, 20, 1064-1073.
- Fielding, N., & Fielding, J. (2000). Resistance and adaptation to criminal identity: Using secondary analysis to evaluate classic studies of crime and deviance. *Sociology*, 34, 671-689.
- Fowler, A., Whyatt, D., Davies, G., & Ellis, R. (2013). How reliable are citizen-derived scientific data? Assessing the quality of conrail observations made by the general public. *Transactions in GIS*, 17, 488-506. doi:10.1111/tgis.12034
- The Future of Research Communications and e-Scholarship. (2014). *Guiding principles for findable, accessible, interoperable and re-usable data publishing version b1.0*. Author. Retrieved from <https://www.force11.org/node/6062>
- Gonzalez-Bailon, S., & Borge-Holthoefer. (in press). Advanced methods for social network analysis: Problems and issues. In N. Fielding, et al. (Eds.), *The SAGE handbook of online research methods* (2nd ed.). London, England: Sage.
- Guterman, B. (Ed.). (2011). Internet as a catalyst for change: Access, development, freedoms and innovation. *The Sixth Meeting of the Internet Governance Forum*, Nairobi, Kenya. Retrieved from <https://www.intgovforum.org/cms/documents/publications/176-internet-as-a-catalyst-for-change-access-development-freedoms-and-innovations/file>
- Haaker, M., & Morgan-Brett, B. (2016). Developing research-led teaching: Two case studies of practical data re-use in the classroom. *SAGE Open*.
- Hardey, M., & Burrows, R. (2008). New cartographies of "knowing capitalism" and the changing jurisdictions of empirical sociology. In N. Fielding, G. Blank, & R. M. Lee (Eds.), *The SAGE handbook of online research methods* (pp. 507-518). London, England: Sage.
- Holstein, J. A., & Gubrium, J. F. (2004). Context: Working it up, down and across. In C. Seale, G. Gobo, J. F. Gubrium, & D. Silverman (Eds.), *Qualitative research practice* (pp. 267-281). London, England: Sage.
- Hundley, M., & Holbrook, T. (2012, May 18). Late to the eParty? Academics and the reshaping of authority. *Conference presentation, Eighth Congress of Qualitative Inquiry*. Urbana-Champaign: University of Illinois.
- Juris, J. S. (2014, May 18). Field notes on activism: Beyond established models and pathways. *Cultural Anthropology*. Retrieved from <http://www.culanth.org/fieldsights/532-activism-deviation>
- Kane, P. (2004). *The play ethic*. Basingstoke, UK: Macmillan.
- Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 1-12. doi:10.1177/2053951714528481
- Kynaston, D. (2005). The uses of sociology for real-time history. *Forum Qualitative Social Research Online*, 6(1). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/503>
- Lather, P. (2007). *Getting lost: Feminist efforts towards a double(d) science*. Albany: State University of New York Press.
- Lawrence, J., & Elliot, J. (2016). The emotional economy of unemployment: A re-analysis of testimony from a Sheppey family, 1978-1983. *SAGE Open*.
- Lee, R., Fielding, N., & Blank, G. (2008). *Online research methods*. London, England: Sage.
- Madrigal, A. (2009, September 3). On Science, Obama puts his mouth where his money is. *WIRED*. Retrieved from <http://www.wired.com/2009/03/scintegrity/>
- Mauthner, N., Parry, O., & Backett-Milburn, K. (1998). The data are out there, or are they? Implications for archiving and revisiting qualitative data. *Sociology*, 32, 733-745.
- Met Office. (2015). *Met Office data point*. London, England: Author. Retrieved from <http://www.metoffice.gov.uk/data-point>
- Moravcsik, A. (2010). Active citation: A precondition for replicable qualitative research. *Political Science & Politics*, 43, 29-35.
- Nielsen, M. (2011). *Reinventing discovery: The new era of networked science*. Princeton, NJ: Princeton University Press.
- Open Data Institute. (2015). *ODI certificate for family life and work experience before 1918, 1870-1973*. Author. Retrieved from <https://certificates.theodi.org/en/datasets/1709/certificate>
- Ruppert, E., Law, J., & Savage, M. (2013). Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society*, 30(4), 22-46.
- Sample, I. (2015, February 13). Google boss warns of "forgotten century" with email and photos at risk. *The Guardian*. Retrieved from <http://www.theguardian.com/technology/2015/feb/13/google-boss-warns-forgotten-century-email-photos-vint-cerf>
- Savage, M. (2013). The "social life of methods": A critical introduction. *Theory, Culture & Society*, 30(4), 3-21.
- Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology*, 41, 885-899.
- Scola, M. (2015, July 15). Library of Congress Twitter archive is a huge #FAIL. *POLITICO*. Retrieved from <http://www.politico.com/story/2015/07/library-of-congress-twitter-archive-119698.html#ixzz3hPhSXbQt>
- Sloan, L., Morgan, J., Burnap, P., & Williams, M. (2015). Who tweets? Deriving the demographic characteristics of age, occupation and social class from Twitter user meta-data. *PLoS ONE*. doi:10.1371/journal.pone.0115545. Retrieved from <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0115545>
- St Pierre, E. (2016). "The long reach of logical positivism." In N. Denzin & M. Giardina (Eds.), *Qualitative inquiry through a critical lens* (pp. 1-29). New York, NY: Routledge.
- Sutcliffe-Brown, F. (2016). New perspectives from unstructured interviews: Young women, gender and sexuality on the Isle of Sheppey in 1980. *SAGE Open*.
- Thompson, P., & Lummis, T. (2009). Interview with Mr. Patridge. In *Family life and work experience before 1918, 1870-1973*. U.K. Data Service [distributor], SN: 2000, Para. 619-622. doi:10.5255/UKDA-SN-2000-1. Retrieved from <https://discover.ukdataservice.ac.uk/QualiBank/Document/?cid=q-b2c168e0-70e6-4803-9d1e-7b7f5cbd889c>
- Times Higher Education. (2014, March 20). More seek advice on retractions and peer review. *Times Higher Education Supplement*, p. 7.
- Tonnies, F. (1988). *Community and society*. New Brunswick, NJ: Transaction.
- U.K. Data Service. (2014). *Introducing the UK Data Service Qualibank*. Author. Retrieved from <https://www.ukdataservice.ac.uk/media/428412/qualibankguide.pdf>

- U.K. Data Service. (2015). *Discover data catalogue*. Colchester, UK: University of Essex. Retrieved from [http://discover.ukdataservice.ac.uk/?sf=Data catalogue](http://discover.ukdataservice.ac.uk/?sf=Data+catalogue)
- Vigen, T. (2015). *Spurious correlations* (Spurious Media ILC). Retrieved from <http://www.tylervigen.com/spurious-correlations>
- Wiggin, A., Newman, G., Stevenson, R. D., & Crowston, K. (2011, December). *Mechanisms for data quality and validation in citizen science*. Paper presented at the Computing for Citizen Science Workshop of the IEEE eScience Conference, Stockholm, Sweden. Retrieved from <http://crowston.syr.edu/node/422>
- Woutersen-Windhouver, S., Brandsma, R., Verhaar, P., Hogenaar, A., Hoogerwerf, M., Doorenbosch, P., & Sierman, B. (2009). *Linking publications and research data in digital repositories*. Utrecht, The Netherlands: SURF.
- Yapp, R. (2011, June 20). Brazil to create “botanical Wikipedia” to catalogue the Amazon. *The Telegraph*. Retrieved from <http://www.telegraph.co.uk/news/worldnews/southamerica/brazil/8588013/Brazil-to-create-botanical-Wikipedia-to-catalogue-the-Amazon.html>

Author Biographies

Louise Corti is associate director of the U.K. Data Archive with a background in survey and research methods. She leads the acquisition and research data management teams, and also directs qualitative data activities.

Nigel Fielding is professor of sociology at the University of Surrey and a fellow of the Academy of Social Sciences. His interests are in research methodology, and criminology.