
Article:
Ford, H. (2016) The Person in the (Big) Data: A Selection of Innovative Methods, Strategies and Perspectives for Social Research in the Age of (Big) Data. Working Papers of the Communities & Culture Network+, 8. ISSN: 2052-7268

The Person in the (Big) Data

A Selection of Innovative Methods, Strategies and Perspectives for Social Research in the Age of (Big) Data

April 2016

Edited by Heather Ford, with contributions from Kath Albury, Jean Burgess, Elizabeth Dubois, Stefanie Duguay, R. Stuart Geiger, Ben Light, Giles Moss.

Introduction

In the first few months of 2016, I set about investigating the state of affairs in mixed methods research being proposed by researchers in response to Big Data. Big Data – a term to describe the extremely large datasets that are being computationally analysed to reveal patterns, trends and associations relating to human behavior – has become pervasive in the study of society these days.

In response, some researchers have pushed back on the assumptions and logics underlying Big Data methods for social science research, arguing that such methods do not presuppose the need for smaller, more local analyses, and, in fact, such analyses may even be essential to gaining valuable insights into social behavior and the human experience. The goal of this report is to document the methods of this growing group of researchers who are experimenting with a research practice that makes the most of digital data while retaining the connection to the humans who produce that data.
In my travels, I discovered was a group of researchers from a number of different countries who were experimenting with new methods that made use of data provided by social media platforms or that engaged with the logics with which systems are being built but who, importantly, were trying to understand how that data could be verified or enhanced or 'thickened' using methods that connected very strongly to the people whom that data purports to speak for or to represent.

The thickening of data is a reference to Tricia Wang's (2014) term 'thick data' that was inspired by Clifford Geertz's (1973) writing on 'thick description'. For Geertz, thick description is necessary in order to meaningfully describe the context in which we live, a context that cannot be algorithmically discovered. Geertz wrote that 'culture is not a power, something to which social events, behaviors, institutions, or processes can be causally attributed; it is a context, something within which they can be intelligibly – that is, thickly – described'.

While the majority of tools or instruments for studying 'big' social media data enable us to *observe* patterns of behavior, we need theoretical frameworks to *interpret* that behavior. Theory makes thick description possible, writes Geertz, not by the analyst generalizing across cases but by generalizing within them. Similarly, argues Tricia Wang (2013) 'Big Data needs Thick Data'. While 'Big Data delivers numbers; thick data delivers stories. Big data relies on machine learning; thick data relies on human learning.'

This is not to say that thick data does not rely on data at all. As researchers involved in studying life in an environment suffused by data, we are all (to at least some extent) asking and answering questions about how we employ digital methods in our research practice. First, a word on digital methods. In his groundbreaking work on digital methods, Richard Rogers argued for a move towards natively digital methods. In doing so, Rogers distinguishes between methods that have been digitized (e.g. online surveys)

vs. those that are "born digital" (e.g. recommender systems), arguing that the Internet should not only be seen as an *object* for studying online communities but as a *source* for studying modern life that is now suffused by data. "Digital methods," writes Rogers, "strives to follow the evolving methods of the medium" by the researcher becoming a "native" speaker of online vocabulary and practices.

The risks of going natively digital

There are, however, risks associated with going native. As ethnographers, we recognize the important critical role that we play of bridging different communities and maintaining reflexivity about our research practice at all times and this makes ethnographers great partners in data studies. Going native in this context, in other words, is an appropriate metaphor for both the benefits and risks of digital methods because the risk is not *in using* digital methods but in *focusing too much on data traces*. Having surveyed some of debates about data-centric methodology, I've categorized the risks according to three core themes: 1. accuracy and completeness, 2. access and control, 3. ethical issues.

1. Accuracy and completeness

Focusing only on data traces recorded on platforms rather than data in context can result in problems relating to accuracy and completeness. Critics focus on the multiple meanings of data and the inability of quantified data research to expose the fact that a Facebook 'like', for example, can represent either agreement or sympathy. In my research on Wikipedia editors of the 2011 Egyptian Revolution article, for example, I found through interviews that the most prolific editor of the article had actually shared his user details with two other friends located outside of Egypt. This experience and others led me to recognise that the data traces that we often employ are incomplete at best and inaccurate at worst. Relying only on the data that we can access can be severely problematic as we try to gain stronger insights into experiences that are mediated by digital data.

2. Access and control:

A significant problem with digital methods is that the locus of control over digital data is often heavily weighted towards commercial platforms that can decide when and where to grant access. Rogers describes the ways in which the software that powers platforms can change overnight so that the tools we build as researchers to interface with that software is "scooped" by the objects (p25). Others have argued that, even when we do gain access to data on platforms like Twitter, there is some obscurity over whether we are, in fact, gaining access to the complete set of data and that only chosen researchers will obtain access to preferential access by platforms.

3. **Ethical issues**

   Research that employs digital data introduces a number of ethical questions including issues around the consent of data subjects, questions around the positioning of users as subjects of data and the missing masses of data i.e. people and communities who are not represented by data and are thus made ever more invisible within social systems. In their critical approach to data, Dalton and Thatcher (2015) note that it is important at this stage to ask: “Whose data? On what terms? To what ends?” Others have questioned the positioning of data as somehow “raw” and apolitical, calling for research that exposes the systemic bias of data systems by charting the consequences for “how the world is known, governed and lived-in” (Kitchin and Lauriault, forthcoming).

Solutions

The problem that we as researchers are faced with is that few solutions are offered in order to engage with data in people-centric (as opposed to data-centric) ways. In response to data critiques, many of us have resolutely denied the existence of data, conveniently believing that this is not what ethnographers (or qualitative researchers) *do.*

But there are two problems with this response. The first is that, like it or not, we all do digital methods to some extent, but not all of us use data in considered ways. As my colleague Stuart Geiger argues: "all researchers are mixed-methodological to some extent, and it is dangerous to obscure the more supplementary methods we inevitably find ourselves using." Secondly, when we use digital methods, we're not studying the world apart from the "real" or physical world. Digital methods, as Richard Rogers writes, can be used not only to study the Internet but to use the Internet to study the social world.

This report aims to start to fill this gap, with a series of methods from researchers who are employing innovative strategies for uncovering meaning, motivations and the daily practices that result in data traces. The methods that are being showcased here have two principles in common that make this different from the hang-wringing that often accompanies debates about methods. The first is that data traces are used as a single source among others; the second is a focus on people-centric rather than data-centric models.

1. **Data traces are a single source**

   In order to combat the inevitable incompleteness of digital data traces (both big and small), it is necessary to combine data traces found on digital platforms with other sources of data. Researchers in this edition combine analysis of data traces with interviews, participant observation and critical reflection in order to understand

questions around behavior, motivation and meaning. Giles Moss and others used interviews and focus groups to find out what UK citizens wanted from their politicians during election debates and used these principles to develop an app for evaluating audiences' responses to debates in real time. Stuart Geiger, on the other hand, has learned that the practices and vocabularies unique to platforms like Wikipedia constitute an essential part of what it means to participate in many communities and organizations and that reading through log data can be seen as a form of participation, not just observation in online communities.

Both Moss and Geiger supplement the data traces that they extract from systems with other methods, both in order to enrich the data and to verify the patterns emerging from the data. Recognising the need to employ multiple sources of data is an important principle for rigorous empirical research both because the completeness of data samples is often obscured to researchers and because digital data traces do not stand in for the individual who created it, either actively (in the form of constructed content) or passively (in the form of data exhaust).

2. People-centric vs. data-centric

The researchers participating in this edition also recognize the important role that they play as agents of informed change in society. Instead of just giving a lecture to sex educators about the meanings that young people attribute to selfies, Kath Albury conducted a workshop in which educators learned the basics of media theory and created (and reflected upon) their own selfies. The workshops equipped the educators with an embodied understanding of the media practices that their students were engaging in and a language and vocabulary with which to analyse and understand such practices. Instead of making claims about influential social media users being represented by data through visualisations, Elizabeth Dubois presented those users with visualisations in the interview setting, enabling them to speak back to the data being created about them.

Employing people-centric methods requires refocusing on the person as the object rather than subject of data and asking questions about "Whose data this represents? What was the experience of producing this data? What are the consequences of this data for the person?" In doing so, researchers are able to move from a data-centric to a people-centric model and to focus on human problems and solutions as the subject for research.

Below is a collection of five reflections on mixed methods by researchers that are being used to shine a light on the people behind the massive streams of data that are being produced as a result of online behavior. These research methods use a variety of digital and traditional methods but they share one thing in common: they are aimed at discovering stories. In the methods outlined below, researchers outline how they have

made the most of digital data using innovative methods that uncover the meaning, the context, the stories behind the data. In the end, this is still the critical piece for researchers trying to understand the moment in which we are living. Or, put differently, the ways in which we may want to live but are often prevented from by a system that may reduce the human experience rather than to enable its flourishing.

Five Mixed Methods for Research in the (Big) Data Age

1. Real-time Audience Feedback: The Democratic Reflection Method

Democratic Reflection is a new methodological tool for researching the real-time responses of citizens to televised media content, which was developed by a team of researchers from the University of Leeds and the Open University in the UK as part of a larger research project on TV election debates. The research for the project began by developing an inductive understanding of what people need from TV election debates in order to perform their role as democratic citizens. Drawing on focus groups with a diverse range of participants, the research identified five key demands — or 'democratic entitlements' — that participants felt debates and the political actors involved in them should meet. Participants felt entitled to be: (1) addressed as rational and independent decision makers, (2) given the information needed to make considered political judgements, (3) included in and engaged by the debates, (4) recognised and represented by the political leaders, and (5) provided with meaningful choices that allow them to make a difference politically. In the next phase of the research, the research team developed a new web-based app (accessible via mobile phones, tablets, and laptops), which allows viewers to respond to televised debates in real time and evaluate them using a range of twenty statements based on the five democratic entitlements. An experiment using the Democratic Reflection app was conducted with a panel of 242 participants during the first debate of the 2015 UK General Election, generating a dataset of over 50,000 responses. Analysis of the data provides a valuable new way to understand how viewers respond to election debates: we can explore general patterns of responses, compare different individuals and groups, track changes over time, and examine how specific moments and performances during the debates may relate to particular responses. - *Giles Moss*

More details:
Coleman, Stephen, Simon Buckingham Shum, Anna de Liddo, Giles Moss, Brian Plüss and Paul Wilson. 2014. 'A Novel Method for Capturing Instant, Nuanced Audience Feedback to Televised Election Debates. Election Debate Visualisation. Available at: http://edv-project.net/wp-content/uploads/2014/03/EDV-Briefing2014.04.pdf.

Research projects employing this method:
Coleman, Stephen and Giles Moss. 2016. 'Rethinking Election Debates: What Citizens Are Entitled to Expect', The International Journal of Press/Politics, 21 (1) 3-24.

2. Selfie workshops

The Rethinking media and sexuality education workshops sought to explore the institutional and individual barriers to a 'practice-based' or 'strength-based' approach to young people's digital cultures (and digital literacies) within sex education and/or sexual health promotion settings. The workshops involved a short presentation of three relevant theories of media communications, followed by practical activities adapted from The Selfie Course developed by Kath Albury, Terri Senft and colleagues. These activities invited participants to take (and discuss) their own selfies, as a means of eliciting critical reflection on their understandings of young people's digital practices, and the ways these practices could be linked to other modes of communication, including expressions of sexuality and gender. Participants also reflected the ways young people's digital practices were framed by educators and health professional, and the ways they were supported (or suppressed) by formal and informal workplace cultures and policies. Follow-up surveys assessed the extent to which participants found the workshop's content approach useful, relevant, and applicable to their work. The project generated productive insights into the ways digital practices, platforms and technologies can be employed within qualitative media and cultural studies research, and the role that digital ethnography can play in social engagement. - *Kath Albury*

More details:
'Rethinking media and sexuality education report'
https://www.academia.edu/18618901/Albury_K._and_Byron_P._2015_Rethinking_media_and_sexuality_education_research_report_November_2015_Sydney_University_of_New_South_Wales [open access]

3. Trace ethnography

In my ethnography of Wikipedia, I found it increasingly difficult to explain what I was seeing to other academics without pulling up revision histories and pointing to codes like "rvv" or "{{db-a7}}" in metadata fields. Trace ethnography emerged out of a realization that people in mediated communities and organizations increasingly rely on these kinds of techniques to render their own activities and intentions legible to each other. There are jargons, conventions, and grammars learned as a condition of membership in any group, and people learn how to interact with others by learning these techniques. However, the affordances of mediated platforms are increasingly being used by participants themselves to manage collaboration and context at massive scales and asynchronous latencies. Trace ethnography is based in the realization that these practices around metadata are learned literacies and constitute an essential part of what it means to participate in many communities and organizations. In this understanding, reading through log data can be seen as a form of participation, not just observation -- if this is how members themselves spend their time. However, it is crucial that this approach is distinguished from more passive forms of ethnography ("lurker

ethnography"), as trace ethnography involves an ethnographer's socialization into a group prior to the ability to decode and interpret trace data. - *R. Stuart Geiger*

More details:

"Trace ethnography" was first used in "The Work of Sustaining Order in Wikipedia: The Banning of a Vandal," which I co-authored with my then-advisor David Ribes in the proceedings of the CSCW 2010 conference. We then wrote a followup paper in the proceedings of HICSS 2011 to give a more general introduction to this method, in which we 'inverted' the CSCW 2011 paper, explaining more of the methods we used. We also held a workshop at the 2015 iConference with Amelia Acker and Matt Burton -- the details of that workshop (and the collaborative notes) can be found at http://trace-ethnography.github.io.

Research projects employing this method:

Ford, H. and Geiger, R.S. "Writing up rather than writing down: Becoming Wikipedia literate." Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration. ACM, 2012. http://www.stuartgeiger.com/writing-up-wikisym.pdf

Ribes, D., Jackson, S., Geiger, R.S., Burton, M., & Finholt, T. (2013). Artifacts that organize: Delegation in the distributed organization. Information and Organization, 23(1), 1-14. http://www.stuartgeiger.com/artifacts-that-organize.pdf

Mugar, G., Østerlund, C., Hassman, K. D., Crowston, K., & Jackson, C. B. (2014). Planet hunters and seafloor explorers: legitimate peripheral participation through practice proxies in online citizen science. InProceedings of the 17th ACM conference on Computer supported cooperative work & social computing (pp. 109-119). ACM. http://dl.acm.org/citation.cfm?id=2531721

Howison, J., & Crowston, K. (2014). Collaboration Through Open Superposition: A Theory of the Open Source Way. Mis Quarterly, 38(1), 29-50. http://aisel.aisnet.org/cgi/viewcontent.cgi?article=3156&context=misq

Related methods/projects:

- Documentary ethnography
- Participant-generated ethnography
- Cultural probes

4. Trace interviewing

Trace interviewing is a mixed method approach involving three steps: the visualisation of a user's traces on multiple platforms, the (re)presentation of those visualisations

during the interview process and the analysis of user behavior and interaction after the interview. The researcher and participant work together during the interview to interpret the data in visualizations. Users are encouraged to add missing contextual information, to reflect on the activities that led to the datapoints being created, and to alert researchers to the existence of inaccuracies and/or missing data and information. Trace interviews are useful for enhancing recall, validating trace data-generated results, addressing data joining problems and responding to ethical concerns regarding consent. Potential limitations or challenges include variable levels of visual literacy among interviewees and the need to recognise that, although interviewees are encouraged to co-create the analysis, data being presented to the user is not free of assumptions, and the researcher makes a variety of analytical and editorial decisions when generating useful visualizations. If the challenges of the method are successfully navigated, trace interviewing could allow researchers to respond creatively to new questions about the current, complex political communication environment. - *Elizabeth Dubois and Heather Ford*

More details:
Dubois, E., & Ford, H. (2015). Qualitative Political Communication | Trace Interviews: An Actor-Centered Approach. International Journal Of Communication, 9, 25. Retrieved from http://ijoc.org/index.php/ijoc/article/view/3378

Research projects employing this method:
Ford, H. (2015). Fact Factories: Wikipedia and the Power to Represent. (Doctoral dissertation). University of Oxford.

5. The Walkthrough Method for Mobile Apps

The walkthrough method systematically examines a particular app. It can generate knowledge about its environment of expected use – such as the vision of users put forward by developers, its operating model and its forms of governance. The app itself is also examined exploring registration processes, everyday use and ways of account suspension and closure. Unexpected appropriation may be explored, referring to ways of use not anticipated, and sometimes unsanctioned, by developers. This method draws from science and technology studies and cultural studies while acknowledging user testing practices from human-computer interaction and vernacular walkthroughs, such as those associated with video games. The method can be enacted as an imagined user, where a researcher engages with the app under certain conditions (e.g. by creating a particular kind of user profile). It is also possible to engage the method by sitting alongside people as they engage with an app and its environment of use. - *Jean Burgess, Stefanie Duguay, Ben Light*

More details:
Burgess, J., Light, B., & Duguay, S. (2015). Studying HookUp Apps: A comparative

platform analysis of Tinder, Mixxxer, Squirt and Dattch. Panel presentation at ICA 65th Annual Conference: Communication Across the Life Span. 21-25 May, San Juan, Puerto Rico. Presentation slides: http://www.slideshare.net/doggyb/ica-presentation-final-copy

Research projects employing this method:
Working paper: "Is being #instagay different from an #lgbttakeover? A cross-platform investigation of sexual and gender identity performances" [open access version]

Related methods/projects:
McVeigh-Schultz, J. & Baym, N.K. (2015). Thinking of you: Venacular affordance in the context of the microsocial relationship app, Couple. Social Media +Society, 1(2), 1-13. [open access version]

Light, B., & McGrath, K. (2010). Ethics and social networking sites: A disclosive analysis of Facebook. Information Technology & People, 23(4), 290-311. [open access version]

Conclusion

It is exciting that digital methods have re-energised questions around methods and methodology that have laid in stasis for too long. Research methods are a researcher's craft! When we talk about methods, we're talking about how we work, which tools we choose and how we choose to conduct ourselves in the world. Talking about methods is a wonderful way of recognising that researchers are makers. We make things. In order to make beautiful things, we need to be mindful of the tools we use and the recipes we use to create. In the same way that a potter might deliberate over the choice of paint colors and maintain those paints in neatly labeled jars, or a chef might select the best knives and keep them immaculately clean, the researchers showcased here care very deeply about their craft and the tools they use to make things with. For them, research is not only about what they produce but the ways in which they produce it.

There are many more out there working with data and people in interesting ways – this is just a sampler, an inspiration for others to know that on the edges of Big Data discourse are some really important discussions and experimentation going on about how to more fully engage with the people that such data represents. The hope is that we continue to explore and discover ways to keep research close to the humans it ultimately should aim to serve.

References:

Dalton, C. and Thatcher, J. (2015) 'What does a critical data studies look like, and why do we care? Seven points for a critical approach to 'big data.'' Society and Space Journal. https://societyandspace.com/material/commentaries/craig-dalton-and-jim-thatcher-what-does-a-critical-data-studies-look-like-and-why-do-we-care-seven-points-for-a-critical-approach-to-big-data/

Geertz, C. (1973) 'Thick description: toward an interpretive theory of culture', in: The interpretation of cultures: selected essays. New-York/N.Y./USA etc. 1973: Basic Books, pp. 3-30

Kitchin, R. and Lauriault, T. P. (forthcoming) Towards Critical Data Studies: Charting and Unpacking Data Assemblages and Their Work. Chapter in Eckert, J., Shears, A. and Thatcher, J. (eds) Geoweb and Big Data. University of Nebraska Press. Available at SSRN: http://ssrn.com/abstract=2474112

Rogers, R. (2013) 'Digital Methods'. MIT Press.

Wang, T. (2013) 'Big Data Needs Thick Data'. Ethnography Matters, Ethnomining Edition, 13 May, 2013. Retrieved from http://ethnographymatters.net/blog/2013/05/13/big-data-needs-thick-data/

Acknowledgements

Thanks to the Communities and Culture Network+ for supporting the work that went into producing this report and the Ethnography Matters 'Discovering the Person in the (Big) Data' March and April 2016 edition. You can download all the posts in the edition by visiting http://ethnographymatters.net/editions/the-person-in-the-big-data/.