

Manchester University Press

PROPOSAL GUIDELINES

Please include the following details:

Author / Editor

Name: Hannah Knox

Affiliation: University College London

Work address: Department of Anthropology, 14 Taviton Street,

Work telephone number: +44 (0)20 7679 8837

Work fax number: n/a

E-mail address: h.knox@ucl.ac.uk

Author / Editor 2

Name: Dawn Nafus

Affiliation: Intel

Work address: 2111 NE 25th Ave, JF2-65 Hillsboro, OR 97124

Work telephone number: +1 503 333 8148

Work fax number: n/a

E-mail address: dawn.nafus@intel.com

The Book

Proposed title: Data/Ethnography: Interference, Experiment, Alignment

Brief description of the project's scope and content:

This edited collection aims to reimagine and extend ethnography for a data-saturated world. The book brings together leading scholars in the social sciences who have been interrogating and collaborating with

data scientists working in a range of different settings. The book explores how a repurposed form of ethnography might illuminate the kinds of knowledge that are being produced by data science. It also describes how collaborations between ethnographers and data scientists might lead to new forms of social analysis.

The advent of big data promises to shed new light on social, political and environmental processes, by providing those interested in such relations with comprehensive and often real-time data on the interaction between human and non-human agents. As such, big data opens up new questions for social scientists. These range from questions about how to deal with and understand the scale and patterning of relationships taking place on platforms such as social media; questions about the possibilities of analysing visual big data; and issues surrounding the agency of data itself as a participant in the formation of social worlds. As Burrows and Savage (2007) have already noted for sociology, the advent of data science shifts the grounds upon which our assumptions about the nature of social relations proceeds. Data science puts on the table seemingly new kinds of claims about social relations, potentially refiguring or at least destabilising methods and concepts that have been dominant within the social sciences for many years. As Latour (2009) has pointed out, data science promises a form of total social knowledge that was anticipated over 100 years ago by the social theorist Gabriel Tarde. However, as the social sciences discarded a Tardian approach to populations and social practice in favour of a more community-centred Durkheimian paradigm, the possibilities of a Tardian form of analysis were pushed to the sidelines. Latour and Venturini (2009) suggest, that 'big' transactional forms of data might be seen as an opportunity to re-assert this Tardian paradigm, in ways that radically challenges certain forms of social science thinking. But interestingly they also suggest that the paradigm of big data in fact shares much with ethnographic practice, based on principles of ethnographic holism and inductive research. Similarly data science also works with emergent repositories of 'found' data that are formed without a strong hypothesis or set of analytical parameters.

Whilst a key aim of this book is to think about how ethnographic approaches might complement data science, the chapters collected here nonetheless deviate from the most prominent discussions about the relationship between big data and ethnography that have either used ethnography to simply critique the politics of big data, or taken an 'additive' stance where the depth of ethnographic insight is argued for in light of the reductive or abstractive tendencies of data collection and analysis. In contrast, this collection uses the advent of big data to interrogate the convergences and divergences in the concerns of data scientists and ethnographic researchers.

The contributors to this collection are ethnographers, but they are also trained in, experts on, and interlocutors with practitioners of data science and computing. As ethnographers they are adept at analysing the embeddedness of data in social practice, and the ways in which data is treated as a form of material culture – to be manipulated, fashioned, mined, shared and exchanged. At the same time, understanding the detailed work of crafting big data into meaningful knowledge also requires that ethnographers build up an understanding and literacy of just what data is, where it comes from, how it is organised and configured. This book brings together the unique expertise of data-literate ethnographers, in order to shed new light on just what data science is, what kinds of epistemological promises it hold and how the social sciences might engage critically and productively with the forms of knowledge that big data analytics is able to produce.

Data is understood by all of the authors to be both a powerful participant in contemporary social life, and simultaneously a way of representing it. It is both a material to be crunched, manipulated, and mined, and a form of knowledge production. This dual nature thickens the connections between data and ethnographic practice. This book charts the state-of-the-art in understanding the empirical commitments of both data scientists and contemporary ethnographers by bringing together ethnographic analyses of the social worlds of data science, research on the potential for new forms of data to inform ethnographic

analysis, and experimental studies which are generating interdisciplinary conversations between big data practitioners and ethnographers. Reflecting on the epistemological foundations and methodological commitments of both ethnography and data science, the collection outlines the opportunities and challenges of combining critical ethnographic and data analysis skills in new and productive ways.

Contents list (as well as a list of contributors, if it is an edited collection):

Chapter 1. Introduction (Dawn Nafus and Hannah Knox)

Part 1: What is Data Science? Ethnographic Investigations

Chapter 2: The Making of a Data Science and Scientist (Baki Cakici, Francisca Grommé and Evelyn Ruppert)

Chapter 3: If Everything is Information: instrumenting and archiving the earth (Antonia Walford)

Chapter 4: Baseless Data (Hannah Knox)

Part 2: Doing Ethnography in/of Data

Chapter 5: Intersectional assemblages and analytic affiliations: altered states in code ethnographies (Adrian MacKenzie)

Chapter 6: Data Walking: Ethnographic methods for questioning big data narratives (Alison Powell)

Chapter 7: Engineering Ethnography (Kaiton Williams)

Part 3: Inter-Disciplinarity: Comparison or Convergence?

Chapter 8: Transversal collaboration: an ethnography in/of computational social science (My Madsen, Anders Blok, and Morten Axel Pedersen)

Chapter 9: Participating in Data Through Materials: The Data Sense Project (Dawn Nafus)

Chapter 10: Learning to do Big Data together: from the bottom up and the top down (Farida Vis)

Chapter 11: The Other 90%: Thinking with Data Science (Joseph Dumit)

Chapter 12: Afterthoughts and Trajectories (Hannah Knox and Dawn Nafus)

Synopses of chapters

Chapter 1: Introduction - The Analytics of Data and Ethnography

Hannah Knox and Dawn Nafus

The introduction will outline the theoretical and methodological background to a discussion about the relationship between big data and ethnography and will introduce the chapters of the book.

PART 1: WHAT IS DATA SCIENCE? ETHNOGRAPHIC INVESTIGATIONS

This section will include contributions from ethnographers who have been studying the worlds of data science. It will provide a view on some of the particular epistemological challenges of contemporary data science, and will trace the kinds of techniques and expertise that are being deployed and challenged through the advent of big data in government, corporate and academic settings.

Chapter 2: The making of a data science and scientist

Baki Cakici, Francisca Grommé and Evelyn Ruppert

ARITHMUS, Goldsmiths, University of London

Keywords: Data Science, Capital, Value, Skill, Statisticians, Knowledge

Valuations of Big Data can be extended to the generation of not only economic capital but also the other

forms of capital elaborated by Bourdieu such as cultural and symbolic capital. Struggles over the production, dissemination and exchange of data involve struggles over the valuation of particular skills, education and expertise. How then does the classification of what constitutes the science and scientist of data mobilise specific resources, skills and investments? In this chapter we will approach this question by attending to a particular community of practice within which data science and its professionals are being talked about, conceived and implemented: national statistical institutes. Based on a multi-sited, multi-method and collaborative ethnography of the work of EU national statisticians, we draw on data from reports, discussions, interviews, meetings, and job descriptions to think about how a field and profession come into being in relation to but also as a critique of existing ones such as statistics and statisticians. What is understood to be lacking in the existing field and profession and what are the promises of the new are key questions that we pose to connect these issues to the politics of knowledge more generally.

Chapter 3: If Everything is Information: instrumenting and archiving the earth

Antonia Walford, UCL

Keywords: *Data, Natural Science, Environment, Monitoring, Instrumenting, Mobility*

Although it has not attracted as much attention in anthropology as the Big Data of social media and business, the natural sciences have also over the last two decades experienced a “data deluge”, prompting concerns around the curation, storage and sharing of increasingly enormous data sets. Large scale scientific projects regularly produce peta or exa bytes of data, and data scientists such as James Gray, who worked closely with the natural sciences, have proposed that we understand contemporary data-driven science as functioning under a “fourth paradigm”, to indicate this shift into data. In the earth systems sciences, this is complemented by efforts to increase environmental monitoring by deploying dense wireless sensor networks in some of the most inaccessible and precarious areas of the planet - what has become known as ‘instrumenting the earth’. This shift into real-time extensive data collection, curation and storage has also been accompanied by newly politicised movements around the dissemination of scientific data, collectively known as the Open Data Movement, which call for open sharing of scientific data, so that it circulates as widely as possible beyond the institutions or settings in which it was collected and can be constantly re-used and re-interpreted.

This chapter will explore this drive to collect, store and circulate an ever-increasing quantity of scientific observational data, focusing on how within this almost cybernetic, infrastructural imaginary, “everything is information”, as Geof Bowker puts it, and the effects of this for generating new forms of value and relationality. As Sabina Leonelli has argued, it is data’s capacity to travel between institutions and researchers that is commonly understood to give it its value within a market economy, such that its openness is exactly what contributes to its commodification. But this overarching idiom, critical or otherwise, of openness and sharing eclipses the fact that data can move in lots of different ways, and sometimes not at all. Drawing on fieldwork conducted with a large-scale scientific project in the Brazilian Amazon and on research conducted with the international database the Global Earth Observations System of Systems, this chapter will explore alternative forms of data motility and kinetics that are emerging in the Big Data landscapes of earth systems science; its capacity to pool, leak, overflow, clot, dribble and die. These different kinetic propensities allow for a rich topology of values and social relations to emerge.

Unpacking the ontological alchemy that permits everything to be turned into information also allows for a comparison to be drawn with contemporary anthropological understandings of the ethnographic method, which has at its heart the premise that all details of social life are indeed amenable to ethnographic rendering and subsequent re-use. Tracing the emergence of these two different omni-informational logics in anthropological ethnography and in the data-driven natural sciences provides a way of redefining the relation between the two.

Chapter 4: Baseless Data

Hannah Knox, UCL

Keywords: *Climate data, infrastructure, uncertainty, transformation, baseline, epistemology*

In the context of climate change mitigation, repurposing infrastructure for a changing future is termed 'retrofitting'. In this chapter I build on ethnographic research with data scientists who work with local climate data, to consider one of the challenges of big data: the epistemological challenge of retrofitting new problems to extant data, or what I call 'retro-analysis'. I suggest that one of the challenges of data science, which famously works not only with actively produced data but also with 'found' transactional data, is how to address the problem of how to retro-analyse data, and that, additionally, one of the biggest perceived problems is the difficulty of working with data that appears to have no *baseline*. In an attempt to generate a creative interference between datascience and ethnography, I turn to anthropological discussions of the problem of comparison in ethnography to consider ways in which what I call 'baseless data' might be understood as producing different kinds of meaning. I consider to what extent this ethnographically derived approach to comparative meaning making might provide alternative insights about data to those which are discerned from standard analyses of measurable change over time.

PART 2: DOING ETHNOGRAPHY THROUGH BIG DATA

Having established an understanding of some of the institutional and epistemological commitments of data science, and their parallels with ethnography, in the second section we will turn to the possibilities through which data might transform ethnographic descriptions of social practice. This section will explore ways in which data might become part of ethnography, extending discussions about social science methods from a distinction between quantitative and qualitative methods to an exploration of how big data is disrupting and posing new questions about the processes and practices of representation, visualisation and description through which social scientific knowledge is constructed.

Chapter 5: Intersectional assemblages and analytic affiliations: altered states in code ethnographies

Adrian MacKenzie, Lancaster University

Keywords: *Software Developers, github, coding, relations, recursivity, assemblage*

The chapter concerns an attempt to bring an ethnographic sensibility to the data generated by contemporary software developers. Since launching in late 2007, the code repository Github (Github.com) has become tremendously popular amongst programmers. Github in early 2016 claims to host 29 million code repositories and 6 million coders. Ethnographers first studied coders working on high profile software projects such as Linux in terms of ethics (Coleman 2005), law (Lessig 2004), publics (Kelty 2008) and experiences of freedom (Coleman 2009). Github's growth attests to some substantial transformations in the way coders, coding and code associate with each other. On Github, coding practices have been re-formatted in ways that emulate the traits and tendencies of contemporary social media platforms. 'Sharing,' 'liking,' 'watching' and re-circulation abound. The willingness of coders to affiliate with Github is difficult to characterise. Not only does Github host a wide variety of commercial, industry, government, scientific, educational and civil society software (and non-software) projects, highly dispersed and diverse human and non-human actors congregate there. Github itself thereby generates much data, indeed big data, about the 'social' practices of coding. What is the mode of association then for coders becoming social? Amongst many other changes, coders render accounts of what coders do through analysis of big data generated by other coders. Rather than recursive publics as the mode of association, we see recursive analytics of association. The growth of Github as intersectional assemblage, the re-shaping of coding practices in imitation of social media, and the susceptibility of large-scale public data about coding to analysis by coders alter the scope and focus of ethnographic study. Ethnographic

sensibilities, joined with the right to tinker with the affiliative data, suggest that writing code and writing on Github might also form part of an analytic-ethnographic sensibility.

Chapter 6: Data Walking: Ethnographic methods for questioning big data narratives

Alison Powell, London School of Economics and Political Science

Keywords: *Data workshops, analytics, ethnography, civic assemblage, publics*

Can we really produce ‘big data’ from the bottom up? The assemblages that we consider to be part of the production and positioning of ‘big data’ are themselves large-scale: the computing power required to deal with multiple forms of digital data, the analytics processes required to derive sensible or logical predictions, the institutional meaning-making apparatus required to create frameworks and application spaces for this data are all easier to mobilize top down. In an article of the same name, Nick Couldry and I foreground individual agency and reflexivity as well as the variable ways in which power and participation are constructed and enacted. But the kinds of civic assemblages that we identified as examples of bottom up big data don’t operate in the same way as those from the top down – or they exist conceptually, rather than in practice meaning that critical investigations of big data alternatives have to shift in to a different mode. One way to examine the capacity and opportunity to develop agency and voice is to examine how big data is constructed in the public imagination. This also lets us attend to who is asking the questions about big data, and further, lets us think about how data gets to be ‘big’ in the first place, who asks the questions that make it big (in size as well as importance) and how one might ask different kinds of questions about data.

In the past two years I have held a series of ‘data workshops’ with artists, citizens, town and city planners and researchers in North London and in Cambridge (and soon to be in Exeter) where we observe and document areas of ‘data calm’ ‘data rich’ and ‘body and data interaction’. These events, undertaken with partners from various organizations, produce rich and conflicting reflections about the nature of ‘big data’ and the value of ethnography. But they also provide frameworks for the development of real reflections and openness about the nature of data, generated from the shared confusion and location-based sense-making undertaken by the participants in the workshop. These interventions began as a way to help my students understand the process of gathering and making sense of data and to de-mystify some of the claims about ‘big data’. I wanted them to encounter the experience of both having too much to observe and not knowing which aspects would be most interesting. My first data walk was in 2012. Over the course of several years I used the model in many contexts, including with artist Paula Crutchlow and the Furtherfield Gallery as part of the Museum of Contemporary Commodities project, and with urban planners at a seminar in Cambridge.

This chapter reflects on the data workshop process as a means of surfacing the everyday experiences and reflections that many people have in relation to data, and suggests that ‘top down’ data assemblages need not necessarily be contested with parallel ‘bottom up’ ones but perhaps instead with alternative modes of making sense.

Chapter 7: Engineering Ethnography

Kaiton Williams, Cornell University

Keywords: *Engineering, ethnography, method, expertise, knowledge*

In this chapter, I examine how a combination of approaches from anthropology and data science disciplines has supported my exploration of lives lived at similar intersections. I came to ethnographic work from engineering, where I ran online services that I couldn’t touch, for customers I never saw. Ethnography felt at first an escape—a rejection of the values of that data-driven world. It offered the opportunity to meet the world on different terms, with a richer language to describe

it. However, that data-centred, engineering background has proved central to my ethnographic practice. Freed from earlier constraints, data no longer totalises the discussion. It serves as a complement, beyond evidence, and often as a complete narrative and common tongue.

Building on this background, this chapter describes work I have done at two research sites. One, through self-tracking and the quantified self, is focused internally. The other, with a community of startup developers in Jamaica, is focused on struggles to realise the potential of the global knowledge economy from its margins.

While differing in their geographies and scales, both spaces allow for an interrogation of the potential of combining data science and ethnography: its new methods, modes of inquiry, and modes of expression. For both myself and those I work with, data acts a conduit across borders of nation, history, and flesh, promising new existential and epistemological models, and a means of affecting personal and national transformation. Its analytical lines offer the ability to connect and communicate, to modulate ideas of difference, and to help construct new identities. I discuss the uneven realisation of this potential, and how the attempts at its operationalisation reveal productive complications and reformulations at the convergence of engineering and ethnography.

PART 3: INTER-DISCIPLINARITY: COMPARISON OR CONVERGENCE

In the final section we will hear from those who have been engaged in experimental practice at the interface of big data science and ethnography. What are the challenges of translation, mediation, communication in the intersections between data science and ethnography and what can this tell us about the analytical commitments of these different areas?

Chapter 8: Transversal collaboration: an ethnography in/of computational social science

My Madsen, Anders Blok, and Morten Axel Pedersen, University of Copenhagen

Keywords: *Data collaborations, interdisciplinarity, method, boundaries, ethnography, computer science*

This chapter chronicles and reflects on the experiences of working ethnographically within, alongside, and in collaboration with a large-scale interdisciplinary experiment in so-called computational social science, one of the important transnational frontiers for the mobilization of “big” social data in recent years. Starting in 2013, its three authors have partaken in the Social Fabric/Sensible DTU project ; a large-scale interdisciplinary research program and centre of excellence partly funded by the University of Copenhagen, Denmark. As part of its grand ambition to develop the technical, methodological and theoretical tools needed to transform the “big data” revolution into “deep data” research, Social Fabric/Sensible DTU has sought to make continuous recordings of social interactions at all smart-phone based communication channels (call logs, SMS, Bluetooth, GPS geo-location etc.) among an entire freshman class (N=800) at the Danish Technical University (DTU), while at the same time gathering a variety of other kinds of data via the deployment of more established social science methods ranging from surveys over structured interviews to participant-observation. For three years, a motley 25-plus strong crew of computer scientists, physicists, economists, psychologists, philosophers, sociologists and anthropologists have thus worked to map different components of the ‘social fabric’ of campus life, including via “thick” ethnographic data obtained from ‘embedding’ an anthropology PhD student within the freshmen class for a year. Given her positioning at right angles from both the students under study and the other scientists studying them, the experiences of this anthropologist (the main author of this chapter) offers a privileged access point for addressing a set of pressing questions and challenges within and beyond the social sciences. Here, seemingly disparate data worlds and research practices, associated with computational and ethnographic approaches, come to rub closely off each other.

Based on recounting, from the ethnographer's point of view, a number of 'collaborative moments' at the unusual and awkward intersection of computational data science and ethnographic fieldwork, the chapter explores less the disciplinary commitments involved in this inter-disciplinary encounter and more the nature of that very collaborative relationship itself. Of crucial importance here is the fact that – in ways not too dissimilar from more classical contexts of participant-observation – working as an anthropologists as part of the Social Fabric/Sensible DTU research program has entailed a chronic condition of oscillating between practicing ethnography *in* a (partly) computational social science framework and doing an ethnography *of* the very scientific data practices and infrastructures involved. Rather than thinking about this oscillation either as a distanced gaze akin to that of a science studies (STS) observer, or as a neutral epistemic meeting-point for comparison or convergence, we prefer to think of it as involving what we call 'transversal' collaborations, instantiating forms of non-coherent, intermittent, and yet productively mutual co-shaping among partially connected knowledge practices and practitioners. Such rethinking of the actual and possible relation between computational and ethnographic approaches in practice is crucial, we argue, for understanding new social data 'complementarities' and their epistemological, ethical and political ramifications.

Chapter 9: Participating in Data Through Materials: The Data Sense Project

Dawn Nafus, Intel

Keywords: *Data sense, materiality, visualisation, data production, meaning-making*

There have been many assertions that ethnography is an important supplement to how data scientists generate meaning from large-scale behavioral datasets, particularly those generated by sensors or traces of online activity. This, however, is not the only relationship possible. The Data Sense project explores what might happen when anthropologists also concern themselves with shaping the materials of data production. Data Sense is a software project co-led by an anthropologist and an engineer, designed to support "lay analysis" of personal datasets collected by self-tracking enthusiasts (that is, people who collect activity data, track health symptoms, etc.). We took as our starting point an assertion that it is important that non-data scientists have access to robust tools that can support the ability to make sense of one's own data beyond the narrow framings that consumer devices provide. The design and development of this tool, however, raised difficult questions about what a layperson's analysis might in fact look like, and how that might square with what is and is not technologically feasible. Co-leading such a project required an adjustment in my own assumptions about where ethnographic knowing might take place. I found myself conducting participant observation not just amongst the potential users of the system, but also conducting a participant observation in materials—designing, proposing interactions, involving myself in discussions about database optimization, etc.. Ethnography here meant shaping engineering decisions about how data is aggregated, calculated, and visualized well before any assessment of the meaning of a particular dataset is made. In this way, locating oneself within technologies of production opens up possibilities for three different but related ethnographies to take place—ethnographic knowledge about the likely purposes and uses of the technology built *into* the production process, an ethnography *of* the production, and ethnographies that take place *through* what is produced.

Chapter 10: Learning to do Big Data together: from the bottom up and the top down

Farida Vis, Visual Social Media Lab, University of Sheffield

Keywords: *Images, Visual Culture, Dialogue, Social Media, Viral Media, Scale*

This chapter addresses the issue of learning how to tackle big data from multiple disciplinary perspectives and vantage points, including through ethnographic approaches. Specifically, the chapter seeks to track how this learning has come about, how it has been approached and experienced by a group of interdisciplinary researchers who set up the Visual Social Media Lab in 2014. This research group, interested in studying social media images and online Visual Culture was originally brought together by a

specific funded project (an ESRC Transformative Research grant), and includes researchers from industry and academia: software developers as well as ethnographers and art historians. As a group they all have different ideas and approaches to what a 'bottom up' or 'top down' approach might look like and how -- what on the surface may look like hard to reconcile ontologies and epistemologies -- might productively be brought into dialogue with each other. This chapter therefore seeks to reflect on this process from the earliest point of the *coming together* of this group (when the idea for the grant proposal first emerged) to the first key collective output. This output, a rapid research response report, involved researchers from the lab as well as others and focused on the viral spread of the images of three-year old Syrian refugee Alan Kurdi (originally wrongly identified as 'Aylan'). This report included a collaboration with Google News Lab and opened up the possibility to better understand the spread of a set of images on social media from the scale of 3 billion searches a day, to close to 3 million tweets over a two-week period, and well as via a deep, interpretive, reading of a set of images placing them within the long standing Christian tradition of iconic images of suffering. This initial work has also produced the early stages of a more ethnographic approach into what these images have *done* for different groups of people. The chapter therefore seeks to use this opportunity to reflect on how this work has started to develop promising ways in which different approaches to big data can be brought together productively. Initially this was via a specific focus, a shared drive to better understand the significant global response to the death of a single child.

Chapter 11: The Other 90%: Thinking with Data Science

Joseph Dumit, UC Davis

Keywords: *Data science, translation, epistemology, interrogation, critique*

It is a truism that data science textbooks, online courses, and talks begin by saying that 90% of the time is spent in figuring out the right question to ask, exploring the data to find out if it is right, revising the question, checking and revising the data again, and so on. The other 10% of the time is spent writing algorithms efficiently and running them. Yet more than 90% of most books, courses and talks are spent on the algorithms. Learning how to critically step back and ask: Is this the right question? Do I have the right data? Who matters for these questions? - these are the skills of a social scientist. This paper will analyze the processes of teaching students how to interrogate where data came from ("Data Archaeology"), what is social life of data in the wild and in organizations ("Data Ethnography") and who cares about it at each step of its life cycle or ("Data Interviewing").

Chapter 12: Conclusions: Afterthoughts and Trajectories

Hannah Knox and Dawn Nafus

The final chapter will summarise the opportunities of bringing together big data and ethnography. Drawing on the findings of the ten substantive chapters we will outline a rethinking of big data as a *material practice* rather than representational or descriptive claim. We will suggest that rethinking data as material trace rather than representational index opens up a productive set of methodological opportunities for understanding how big data operates as a practice, a politics and material out of which social and political claims can be made.

Sample chapters (preferably introduction and one other)

We do not yet have any sample chapters available but can produce a draft of the introduction if this is necessary to secure the contract.

Why we should publish your book?

Why *should* we publish? What are the particular strengths?

This work, to our knowledge, will be the first book-length treatment of this topic. Its ethnographic attention to data as embedded in social and material practices —how data pools, clots and leaks, how it sits in physical spaces, and how it lends itself to “hands-on” working and reworking—opens up a set of novel perspectives that allow us to raise substantive and methodological questions simultaneously. While there are critical works on big data such as Kitchin’s *Big Data Revolution* and Gitelman’s *Raw Data is an Oxymoron*, there are none that also delve into the implications for ethnographic practice. Here we explore those issues by addressing new ways of ethnographically knowing social worlds suffused with data, and by addressing how ethnographic practice can in turn reshape data science practices.

Who is it aimed at (think about style, content, theoretical concepts and language etc)?

This book will be of interest to anthropologists, sociologists, material culture, cultural studies, and science and technology studies (STS) scholars and students, and to data scientists interested in the social dynamics of their emerging discipline and the data they deal with day to day. The work offers important new research that contributes to our understanding of data science as a social form that has growing influence in ever more aspects of daily life. It also offers advanced undergraduate students, postgraduate students and researchers, through examples, ideas about how to incorporate data into critically-minded ethnographic practice. It also offers ethnographers who wish to work with and collaborate with data scientists an understanding of the possibilities and the challenges that such collaborations are likely to entail. This book will serve as an important touchstone in the debate that is now only nascent concerning the methodological implications of doing ethnography in data-rich societies.

Who will *buy* the finished publication? Is there a primary and a secondary market?

We anticipate that the book will be bought by a combination of libraries, students registered on courses that are bringing data science into conversation with other social science methods, and researchers interested in embarking on experimental collaborations with data scientists. With an emphasis among research funders on the importance of advanced quantitative methods, there is a strong emerging interest among social researchers from a wide range of different disciplinary fields to understand how big data might become part of their analytical practice. We anticipate that there are many people currently deploying ethnographic methods in anthropology, sociology, cultural geography, organization studies, science and technology studies and material culture studies, who will be interested in finding out more about how they can explore the role of big data produced by everyday activities of shopping, communicating, and browsing online in their research.

Are there any comparable books? Please list the main competitors. Or is there a gap in the currently available literature?

The books that currently address big data which also have somewhat of an ethnographic sensibility (to greater or lesser degrees) are listed below. However, they do not cover the methodological issues we propose to raise.

Gitelman, L. (2013). *Raw data is an oxymoron*. MIT Press.

Boellstorff, T. and B. Maurer (ed). *Data: Now Bigger and Better*. Prickly Pear Press.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

Essential information

What is your proposed submission date for completion of the manuscript?

May 2017

What is your ideal word length? If it is an edited collection, how long is each essay? Please include ALL text (e.g. notes and bibliography).

Introduction – 6000 words including footnotes and references

10 x substantive chapters - 7000 words each including footnotes and references.

Concluding Comments – 4000 words including footnotes and references

Total: 80,000 words

Would you like pictures, graphs or diagrams? How many? Black and white or colour?

NB: the inclusion of illustrations may prevent or delay the publication of your work as an eBook; MUP cannot contribute towards the cost of obtaining images or clearing permissions; colour illustrations require a subvention towards production costs.

Maximum of one illustration per chapter.

Have you contacted other publishers?

No

Has any of the work been published elsewhere (e.g. journal articles, book chapters, university repository)? *If the work is based on a PhD thesis please contact our editors for our guidelines on making the transition from thesis to book project.*

No

Please give the names of two people who you feel would be qualified to comment upon your book proposal.

Nick Seaver

Mike Savage

Please feel free to attach any further information or documents that you feel are relevant to the proposal.