

# Rethinking the Geoweb and Big Data

## Future Research Directions

Mark Graham

This is a pre-publication version of the following chapter: Graham, M. 2018. Rethinking the Geoweb and Big Data: Future Research Directions. In *Thinking Big Data in Geography: New Regimes, New Research*. Thatcher, J., Eckert, J., and Shears, A. (eds). University of Nebraska Press. Lincoln. 231-236.

This short chapter is a reflection on future directions that research on the geoweb and big data could take. It is derived from a reflection that the editors of this volume asked me to provide to a session on the geoweb and big data at the 2014 meeting of the Association of American Geographers. Panelists were asked to summarize some of the day's themes and to debate how they speak to future directions in the discipline. My reflections are organized into seven themes.

## Defining the Object of Our Inquiry

It is important to point out that different researchers are referring to very different things with the terms “big data” and “geoweb.” This is not necessarily a problem, but I hope that this book has helped to clarify what exactly the geoweb is. Words, as signifiers, are able to not just represent but also reproduce and enact change in the social world.<sup>1</sup>

As such, it is crucial to unpack what we mean and what we are trying to convey when we refer to the geoweb. Where are the boundaries between the web and the geoweb? (I am not sure that I clearly see them.) Where are the boundaries between the geoweb and what we might think of as the underlying/offline/material geo that seems to underpin, augment, or inform it? I'm also not sure I clearly see those boundaries, in part because of the ways that place is always

transduced: constantly remade and reenacted.<sup>2</sup> The point here is not that we need to necessarily agree on any definitions but rather that we should avoid taking for granted some of the assumptions wrapped into these very powerful terms.

## Mixed Methods and Brandolini's Law

Much effort has gone into arguing for the need for more mixed-methods research, for less reliance on big data, and for the nuance and context that only qualitative research can provide.<sup>3</sup> These are sentiments that I fully endorse, but here the so-called Brandolini's Law (i.e., "the amount of energy necessary to refute bullshit is an order of magnitude bigger than the amount required to produce it") can potentially provide some useful guidance.

In other words our time is not necessarily best spent when focused on caricatures that we choose to argue against. How many people are actually saying that big data can answer all facets of all societal questions? The answer is, surprisingly few. Let's then move away from arguing for the need for small data and mixed methods and instead begin channeling some of that energy into creating, carrying out, and enacting those hybrid approaches.

## Disciplinary Cross-Pollination

This book illustrates some of the potentials for cross-pollination between geography, computer science, information studies, Internet studies, and other social sciences. As many other disciplines are having their own "spatial turns," there is much that we can (and should) contribute as geographers.

But geographers also need to make sure that we are not reinventing the wheel. For instance, there is a growing amount of work in geography that focuses on crowdsourcing and volunteered geographic information.<sup>4</sup> But much of this work being carried out by geographers

unfortunately omits research being done in information studies, psychology, and Internet studies that also is trying to understand motivations for crowdsourcing. More could therefore be done to make our disciplinary boundaries a little more porous and allow some of that work to cross over to geography and geoweb research (and then feed back into it).

## Human and Machine Labor

Conversations about big data often seem to neglect the truly massive amount of paid human labor that goes into the filtering, sorting, cleaning, manipulating, and managing of it. Big data is often spoken about as something that pings around between sensors, data sets, machines, and algorithms.

But it is ultimately humans, through digital labor, who are creating much of the content that makes up the geoweb.<sup>5</sup> It is thus crucial to keep focusing on, and asking critical questions about, the digital sweatshops, the micro workers, the click workers, the gold farmers—those laborers in the background who are keeping our networks chugging along.<sup>6</sup> And I hope that we will begin to see more of this work and to remember that automation is often an illusion. What should we be asking about those millions of workers in the shadows, doing unorganized, low-paid, alienated work and making many of our big data ecosystems function?

## Privacy and Transparency

Privacy is one of the most pressing issues of our time, and as geographers we are unwisely ceding much of the debate on it to computer scientists (who tend to be informed on the topic) and to politicians (who don't). Because many privacy implications center on location—its collection, its data structures, its precision, and its accuracy—we can ask what we should be doing and saying and researching as geographers, to draw on our expertise and the strengths of our

discipline to make a difference in this new world of always-on tracking and monitoring and the datafication of everything.

At the same time, how do we also make sure that privacy isn't used as an excuse for the wholesale locking away of social data by large companies, meaning that we can't use those data to address the social and human questions that really matter. There have been many projects that have pointed to the value of open geographic data (e.g., Follow the Things, Fair Tracing, and Wikichains).<sup>7</sup> However, it may be that some of these goals stand in direct competition with desires for increased privacy. Where, therefore, do we stand on the transparency/privacy spectrum? And what should we be doing about it?

## Digital Exclusion

One of the most important questions that we can ask as researchers of digital information and networks is, What things, what persons, and what places are being left out? Castells points to this issue when he says that “the cost of exclusion from networks increases faster than do the benefits of inclusion in those same networks.”<sup>8</sup> This is an area of work that we tend to do well as geographers (and it is a set of questions that people in other disciplines often seem to miss), but it is only a first step. How can we move beyond it? What can or should we do about digital inequalities and exclusions? If we establish that the digital layers that augment place are inherently uneven, unrepresentative, and imbalanced, what can we do with that knowledge? What *should* we do with that knowledge?

We should also think about the other side of this issue. While there has been a lot of focus on sparse data, or contexts where data might not be able to capture the complexities of any given situation, what about contexts where we have too much data? Some of the chapters in this book guide us through methods for dealing with big data, and we need more of this sort of work.

Let's have conversations about cluster computing, graph databases, agent-based models, and other methods for grappling with unmanageable volumes of data. Yes, we always need to remember what those data leave out, but unless we want to abandon the whole big data project we should also be—critically—trying to figure out what those data sets *do* tell us about society and how they help us to answer the big questions that we need to ask.

## The Big Questions

Finally, let's keep our eyes on the prize. Let make sure that we are asking the questions that matter and not being too driven by just what data are available. Let's make sure our research continues to focus on questions about things like inequality, power, voice, control, and human welfare. And I say *continue* because much of this work is already happening (as is evidenced by many of the chapters in this book).

We also can make sure that we are shaping not just the questions being asked but also the data being collected. This may mean doing things such as always being explicit that there is never any such thing as raw data. Data are always socially and humanly constructed and can therefore participate in the constructing and shaping of data.

Platforms, algorithms, and the people and organizations that control them are also playing an ever more central role in the lives of any Internet user. This is because they are making editorial decisions that shape our lives.<sup>9</sup> Organizations such as Google and Apple and platforms like Wikipedia and Facebook are arbiters of not just what we see and read but of what we know about our world and how we navigate through our world.

Therefore, as the geoweb becomes ever more integral to our lives, it might be instructive to turn to some words spoken by the late Tony Benn, a British Labour Party politician. He

famously had a set of five questions<sup>1</sup> he said that we should always ask any powerful person:

“What power have you got? Where did you get it from? In whose interests do you exercise it? To whom are you accountable? And how can we get rid of you?”

Many of the platforms and mediators that we rely on for geographic information are for-profit entities that do not necessarily have issues of justice, equality, human rights, and peace at the top of their list of concerns. We could therefore use moments of reflection to dream up and practice radical and democratic alternatives in the geoweb. At the very least we should use Tony Benn’s provocations to hold the data intermediaries, systems, platforms, and algorithms accountable for the power that they wield.

Let’s continue to ask, “What power have you got? Where did you get it from? In whose interests do you exercise it? To whom are you accountable? And how can we get rid of you?”

## Notes

1. I. Parker, *Discourse Dynamics*, 1; Howarth, *Discourse*, 9–12.
  2. M. Graham, Zook, and Boulton, “Augmented Reality.”
  3. Ruppert, “Rethinking Empirical Social Sciences”; Gorman, “Danger of a Big Data Episteme”; M. Graham and Shelton, “Geography and the Future of Big Data.”
  4. Sui, Elwood, and Goodchild, *Crowdsourcing Geographic Knowledge*.
  5. Scholz, *Digital Labor*; Fuchs, *Digital Labour and Karl Marx*; Graham, Hjorth, and Lehdonvirta, “Digital Labour and Development.” *Transfer*. <https://doi.org/10.1177/1024258916687250>.
  6. There are large and nascent networks of researchers already tackling these issues (see, for instance, the groups formed around the Institute for Distributed Creativity mailing list or the **ERC COST** network on the dynamics of virtual work). My own work also increasingly focuses on questions about who benefits and who doesn’t from low-skilled forms of digital labor (see the Geonet project at the Oxford Internet Institute [OII], [geonet.oii.ox.ac.uk](http://geonet.oii.ox.ac.uk), and the Connectivity, Inclusion, and Inequality Group of the OII, [cii.oii.ox.ac.uk](http://cii.oii.ox.ac.uk)).
  7. Kleine, “Negotiating Partnerships”; M. Graham and Haarstad, “Transparency and Development.”
  8. Castells, “Network Theory of Power,” 774.
  9. Kitchin and Dodge, *Code/Space*; M. Graham, Zook, and Boulton, “Augmented Reality.”
- 
- 1 We use these questions to structure a critical discussion about the role of Google in everyday urban life in the following publication: Shaw, J. and Graham, M. 2017. An Informational Right to the City? Code, Content, Control, and the Urbanization of Information. *Antipode*. 10.1111/anti.12312