

Big Data: a non-theory about everything

Description

default watermark

â??Instead of looking through telescopes and microscopes, researchers are increasingly interrogating the world through large-scale, complex instruments and systems that relay observations to large databases to be processed and stored as information and knowledge in computersâ?• (449).

This is how geographers Harvey Miller and Michael Goodchild describe Big Data in their critical article â??Data-driven geography.â?• They add that this data flow amounts to more data than we can analyse. I take this to mean itâ??s more data than any individual or team with a spreadsheet can make use of without the aid of sophisticated statistical analysis, smart algorithms, high bandwidth and powerful processing.

They reference an article by Edd Dumbill of Oâ??Reilly Publishing, who observes that Big data is about: (1) **volume**â??thereâ??s a log of it to be collected and stored; (2) **velocity**â??it needs to be captured and processed at speed, and even in real-time; and (3) **variety**â?? some of it is structured, as in the case of tables and relationships, and some of it unstructured or in diverse formats, e.g. text mixed in with pictures in different formats.

I think these Big Data factors are brought into sharp relief as we think of geographic, urban and other spatial data. Consider a digital map charting a couple of hectares of the countryside at varying levels of detail perhaps down to one centimetre resolution, with height data, material properties at different depths beneath and above the surface, and recorded over past, present and projected time frames, with people, animals and transportation movements across, over and through these surfaces â?? not to mention data about people living in the area.



Now expand that data pool to a whole continent that includes cities, industrial complexes, transportation systems, natural resources, the weather and ecosystems. Thatâ??s Big Data.

We know that itâ??s possible to capture, store and analyse this spatial and attribute data. Following Dumbillâ??s three characteristics of Big Data, itâ??s there in **volume**. It needs to be processed at **velocity**, and itâ??s likely to be in many (**various)** formats.

Is Big Data always good for us?

default watermark

Of course, for all its benefits, Big Data doesnâ??t serve everyone equally. Like digital bandwidth, access to the web, healthcare, and many other social goods there are the haves and have-nots. Any innovation or resource touted as a good has the potential to amplify the difference between those individuals and nations who can afford it and those who canâ??t. Many scholars challenge the utopian speculation that universal global benefit follows the Big Data tidal wave.

Big claims

For an example of Big Data as an ocean not only of data but of magnified claims, see *Big Data: A Revolution That Will Transform How We Live, Work, and Think* by Victor Mayer-Schonberger and Kenneth Cukier. The claim is in the title.

As an example of Big Data benefits, they indicate how â??improving and lowering the cost of healthcare, especially for the worldâ??s poor, will be in large part about automating tasks that currently seem to need human judgment but could be done by computerâ?• (193). Big Data will be able to tell us when we are about to fall ill. They add, â??Soon big data may be able to tell whether weâ??re falling in loveâ?• (192).

Their book follows the common popular-techscience format: (a) exaggerate the revolutionary significance and promise of the new technology; (b) follow with warnings about misuse; (c) exhort readers and politicians to increase their understanding; (d) advocate for judicious use; (e) advocate for further development of the technology.

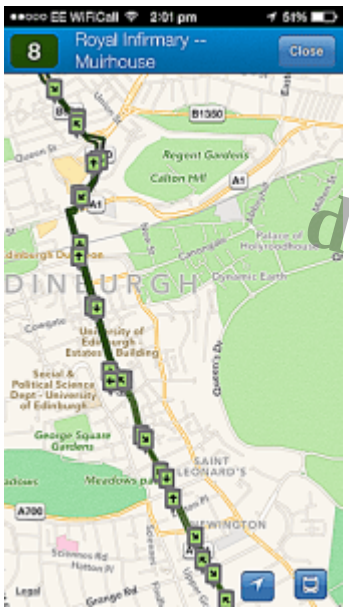
They say, â??There are no foolproof ways to fully prepare for the world of big data; it will require that we establish new principles by which we govern ourselvesâ?• (193).

They acknowledge that Big Data introduces a raft of problems, though they are light on solutions. I think that one solution resides in moderating the claims in the first place, so that less is promised or expected.

Data limits

Critics highlight data challenges, of which I suspect most people working the field are aware.

1. Data bias. Any student of social science knows that any collection of data favours certain interests and diminishes others: e.g. the â??impartialâ?• collection of data on passenger usage on bus routes suggests that routes used less frequently will be discontinued. The very existence of the data skews decision making in a certain direction, in this case in favour of a story about the many (very easy to count) as opposed to those with the greater need (very difficult to measure). Itâ??s hard to argue against statistics.



default watermark

2. Data gathering skews behaviour. Where people are involved, data collection can also skew behaviour. Consider university space audits where pollsters go from room to room counting attendances at lectures. Lecturers encouraged students to attend or theyâ??d lose the space. Knowing that the data is being collected influences peopleâ??s behaviour.

3. Privacy. Where people are involved there are abundant privacy issues. Most of us would be guarded about volunteering the kind of information that gets collected about us at the supermarket, at border controls, or on Facebook â?? if only we knew.

4. The profit motive. Much of the data in Big Data is collected and managed by particular interests: companies with a stake in the data and its uses. According to an online article by Edd Dumbill, â??Itâ??s no coincidence that the lionâ??s share of ideas and tools underpinning big data have emerged from Google, Yahoo, Amazon and Facebook.â?• These are commercial concerns after all. Companies want to maximise revenue, much of it via promotion and advertising. Big Data follows big markets.

Big data fuels the idea of data-rich city planning and management, and the â??smart city.â?• In an article critical of smart city projects, Rob Kitchin identifies a â??neoliberal ethos that prioritises market-led and technological solutions to city governance and developmentâ?• (2). He asserts, â??it is perhaps no surprise that some of the strongest advocates for smart city development are big business (e.g.,

IBM, CISCO, Microsoft, Intel, Siemens, Oracle, SAP)â?• (2).

5. The lure of data. For many designers, artists and others in the arts and humanities, the idea of data (big or small) has little cachÃ©, but such fields donâ??t carry the same influence in political decision making as the worlds of science, management, finance, and governance. In these influential areas data is king.

Already I see research drifting towards projects that have some traction in the world of data. Arguably, Big Data and its emphasis provides further bias away from the arts and humanities and towards the measurable. The supposed digital humanities provide another indication of this drift from the tenets of the interpretive arts, or at least the human art of interpretation, and other forms of evidence-based research. Ethnographic research comes in for a drubbing from various quarters. See Tricia Wangâ??s fight back in an interesting online article, Big Data Needs Thick Data.

These are **practical limits** to data, of which anyone working in fields that rely on data is well aware. I donâ??t think the issues Iâ??ve outlined so far are alien or particularly controversial to anyone working with data.

default watermark

6. A non-theory of everything. But then thereâ??s another category of Big Data enthusiasms that constructs a philosophy around it. In fact some people position Big Data within a pretty old philosophy. That position is best encapsulated in Chris Andersonâ??s short *Wired* article â??The end of theory.â?• Again, the idea is in the title.

Some presume that the data is almost as good as the object being studied. Itâ??s as rich as the physical and social world around us, or at least it can be treated in the same way.

According to Dumbill, thereâ??s an immediacy to big data that obviates the need for theories: â??Having more data beats out having better models: simple bits of math can be unreasonably effective given large amounts of data.â?• Heâ??s mainly thinking of retail, â??If you could run that forecast taking into account 300 factors rather than 6, could you predict demand better?â?•

But does Big Data replace theory? Never mind theory â?? does Big Data replace the thing (phenomenon) from which the data is derived? This reminds me of the quest for perfect digital representations. See post Computer images and realism, and Whatâ??s wrong with parametricism.

Notes

- As another general point of technological critique, there are also those who labour to support the good, and who are invariably denied the same level of access as those who benefit. So iPhones are assembled in factories in China by labourers who are not paid enough to be able to afford them.
- Hereâ??s the full quote from Mayer-Schonberger and Cukier: â??There are no foolproof ways to fully prepare for the world of big data; it will require that we establish new principles by which we govern ourselves. A series of important changes to our practices can help society as it becomes more familiar with big dataâ??s character and shortcomings. We must protect privacy by shifting responsibility away from individuals and toward the data users â?? that is, to accountable use. In a world of predictions, itâ??s vital we ensure that human volition is held sacrosanct and we

preserve not only peopleâ??s capacity for moral choice but individual responsibility for individual acts. And society must design safeguards to allow a new professional class of â??algorithmistsâ?? to assess big-data analytics â?? so that a world which has become less random by dint of big data does not turn into a black box, simply replacing one form of the unknowable with anotherâ?• (193).

References

- Dumbill, Edd. 2012. What is big data? An introduction to the big data landscape. *Oâ??Reilly website*, (https://beta.oreilly.com/ideas/what-is-big-data).
- Kitchin, Rob. 2014. The real-time city? Big data and smart urbanism. *GeoJournal*, (79)1-14.
- Mayer-Schonberger, Victor, and Kenneth Cukier. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston, MA: Eamon Dolan
- Miller, Harvey J., and Michael F. Goodchild. 2015. Data-driven geography. *GeoJournal*, (80)449-461.
- Wang, Tricia. 2013. Big data needs thick data. *Ethnography Matters*, (http://ethnographymatters.net/blog/2013/05/13/big-data-needs-thick-data/).

default watermark

Category

1. Nature

Tags

1. Big Data

Date Created
September 12, 2015
Author
rcoyne99