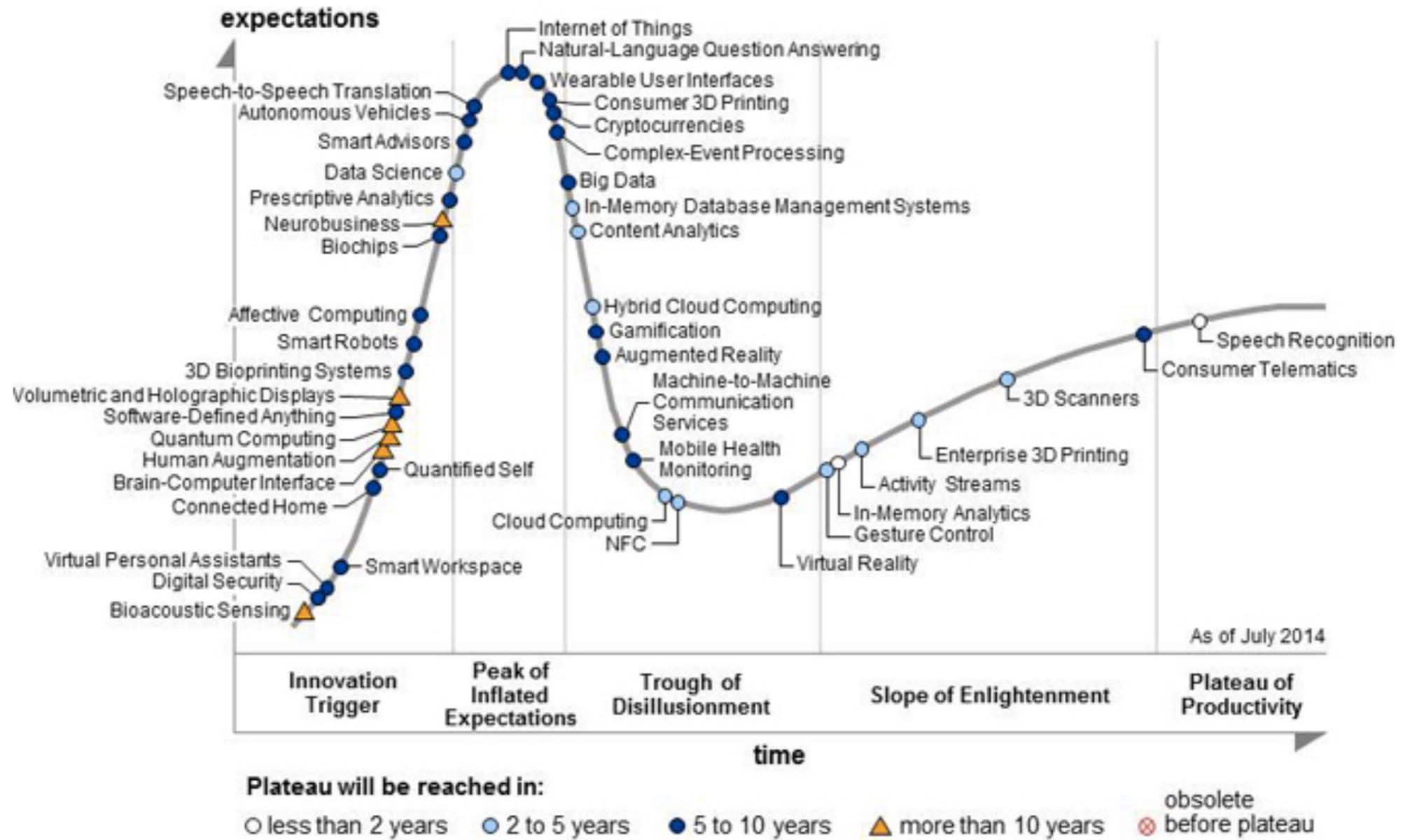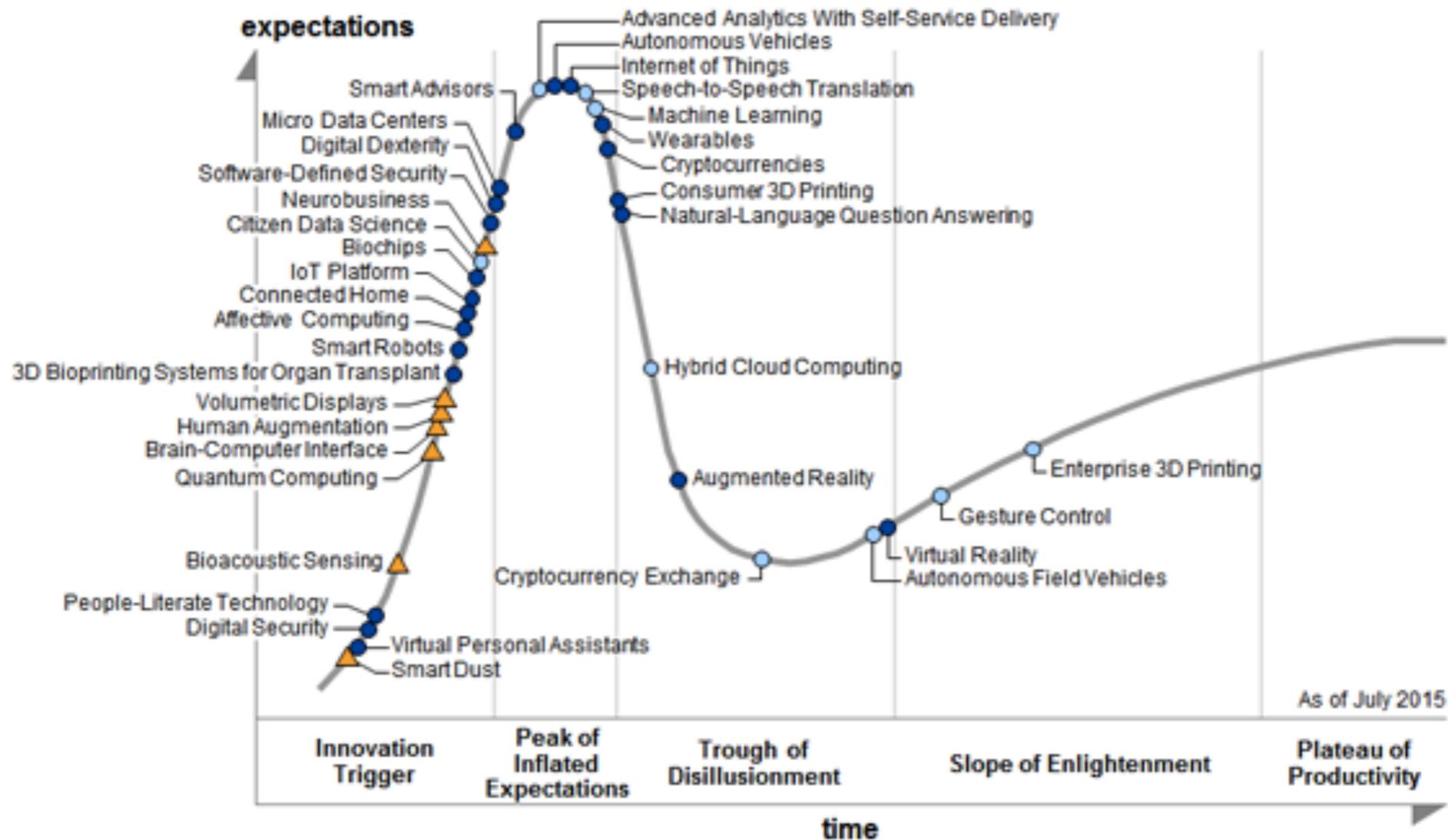# Big (Online) Data Opportunities, Challenges and Ethical Considerations

Simeona Petkova
Assistant Professor
Amsterdam School of International Business, AUAS

(Gartner, 2014)

(Gartner, 2015)

An Overview:

- Big Data Building Blocks

- The Challenges of Online Data

- On Data Algorithms, Software, and Dashboards

- Big Online Data = Ethical Considerations?

What is Big Data?

- "Big data represents the information assets characterised by such a **high volume**, **velocity**, and **variety** to require specific **technology** and **analytical methods** for its transformation into value" (De Mauro et al, 2015: 103).

Additionally:

Big Data is:

- **Exhaustive** (aiming to capture entire population n=all);

- **Rich** (as detailed as possible);

- **Relational** (enabling correlations and lineages of different data sets);

- **Extendable** (new field can be added);

- **Scalable** (can increase in size);
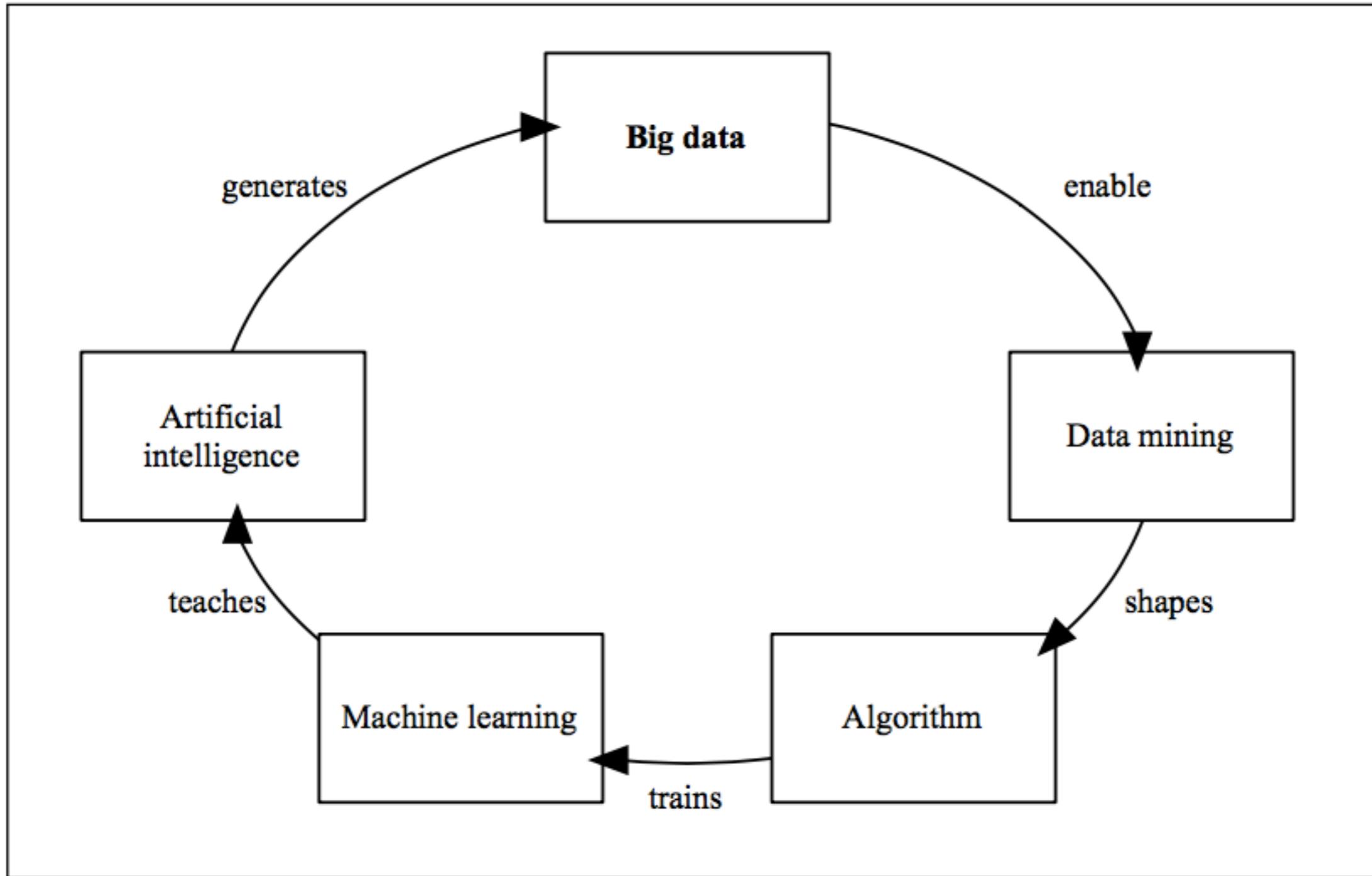
(Kitchin 2014a:68)

The Big Data Promise

- "This is a world where massive amounts of data and applied mathematics replace every other tool that might be bought to bear. Out with every theory of human behaviour, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is that they do it, and we can **track** and **measure** it with unprecedented **fidelity**. With **enough data**, the **numbers speak for themselves**."

(Anderson, 2008)

The Big Data Challenges

- boyd and Crawford (2012) critique gravitates around: the **objectivity**, **accuracy**, and the **size** of big data. The context is very important as well together with **ethical concerns** of how the data was gathered and who has **access** to it. Big data has the potential to change the definition of knowledge.

- Dalton and Thatcher (2014) point out that technology is **never neutral**, warn of the risks of **technological determinism** (the promise of big data to change society), and argue that in some cases big does not equal better data.

Big Data Technology Cycle (Scholz, 2017)

The Challenges of (Online) Data

- "Raw data" is an oxymoron (Gitelman, 2013).

- Big data is (pre-) defined by the process of data gathering, so it has in-built prejudices (Kitchin 2014b).

- Web's data disappoints in terms of quality. It is messy, unstructured and often needs cleaning and correlation with offline data (Rogers, 2013).

The Challenges of Data Mining

- "Big Data means that we can get more precise answers; this is what Bernoulli proved when he showed how the variability in an estimate goes down as the sample size increases. But this apparent precision will delude us if issues such as **selection bias**, **regression to the mean**, **multiple testing**, and **over-interpretation** of associations as causation are not properly taken into account. As data sets get larger, these problems get worse, because the complexity and number of **potential false findings** grow exponentially. Serious statistical skill is required to avoid being misled".

(Spiegelhalter, 2014: 265)

The Challenges of Analytical Software and Algorithms

- Algorithms behind big data are becoming intelligible with inaccuracies (Kleinberg and Mullainathan, 2015).

- Formalisation, algorithmic procedures, data processing can mask partiality (and appear 'objective'). Algorithms can become tools to transform data (Scholz, 2017).

- Challenges with the readability of analytical software (from having access to the source code, to being able to understand it, and assess it's implications (Rieder&Rohle, 2017).

- "By presenting often very complex, messy and often varied data in simplified forms for consumption via a dashboard, sometimes subtle changes take place in how data is understood. Most notably, by introducing a new emphasis on metrics, indicators and measures, it can create a greater focus on operational issues rather than longer-term strategic ones" (Bartlett & Tkacz, 2017).

The Big Online Data Ethical Violations

- "Perhaps the thorniest challenges that exist on the data side, with respect to access and privacy" (Lazer et al., 2009).

- AOL case of making available the search queries of around 600,000 users (for research) in 2006.

- Facebook's "emotional contagion" study on more than 600,000 users in 2012.

- The involvement of Cambridge Analytica in the 2016 USA presidential elections and Brexit.

Big Data research process
Moments of ethical reasoning

Stage 1: Research design — Moment of ethical reasoning

Research interest | Data collection | Data set

Stage 2: Safe data exploration — Moment of ethical reasoning

Data exploration | Generating sample and research questions

Stage 3: Research process — Moment of ethical reasoning

Data analysis | Visualization/anonymization | Publication

Van Schie et al. in van Es & Schäfer eds. (2017)

Thank you!

Bibliography:

Anderson, Chris (2008). The end of theory: The data deluge makes the scientific method obsolete, http://archive.wired.com/science/discoveries/magazine/16-07/ pb_theory, last accessed 10 April 2017.

Bartlett, Jamie & Tkacz, Nathaniel (2017). Governance by dashboards: a policy paper. London: Demos.

boyd, Danah, & Crawford, Kate (2012). Critical questions for big data. Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication & Society 15(5): 662–679.

Dalton, Craig & Thatcher, Jim (2014). What does a critical data studies look like, and why do we care? Seven points for a critical approach to 'big data'. http:// 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, last accessed 10 April 2017.

de Mauro, Andrea, Greco, Marco, & Grimaldi, Michele (2015). What is big data? A consensual definition and a review of key research topics. AIP Conference Proceedings, Madrid, 97–104.

Gitelman, Lisa (Ed.) 2013. "Raw data" is an oxymoron. Cambridge: MIT Press.

Kitchin, Rob (2014a). The data revolution. Big data, open data, data infrastructures & their consequences. Los Angeles: SAGE.

Kitchin, Rob (2014b). The real-time city? Big data and smart urbanism. GeoJournal, 79(1): 1–14.

Kleinberg, Jon & Mullainathan, Sendhil (2015). We built them, but we don't under- stand them. In John Brockman (Ed.). What to think about machines that think (pp. 62–65). New York: Harper Perennial.

Lazer, David, Pentland, Alex, Adamic, Lada, Aral, Sinan, Barabasi, Albert-Laszlo, Brewer, Devon, Christakis, Nicholas, et al. (2009). Computational Social Science. Science 323 (5915):721-723.

Rieder, Bernhard & Rohle, Theo (2017). In Mirko Tobias Schafer & Karin van Es. (Eds.). The Datafied Society: studying culture through data (pp. 109-125). Amsterdam University Press.

Rogers, Richard (2013). Digital Methods. Cambridge: MIT Press.

Spiegelhalter, David J. (2014). The future lies in uncertainty. Science, 345(6194): 264–265.

Scholz, Tobias M. (2017). Big Data in organisations and the role of human resource management: a complex systems theory-based conceptualisation. New York: Peter Lang.