

Is big data governing future memories?

Alessia Ghezzi^a

European Commission, Joint
Research Centre (JRC)
Via Enrico Fermi 2749 Ispra (VA)
210127 Italy
+39 0332789244

alessia.ghezzi@jrc.ec.europa.eu

Estefanía Aguilar–Moreno^a

European Commission, Joint
Research Centre (JRC)
Via Enrico Fermi 2749 Ispra (VA)
210127 Italy
+39 0332789632

estefania.aguilar-
moreno@jrc.ec.europa.eu

Ângela Guimarães Pereira^a

European Commission, Joint
Research Centre (JRC)
Via Enrico Fermi 2749 Ispra (VA)
210127 Italy
+39 0332785340

angela.pereira@jrc.ec.europa.eu

ABSTRACT

In this paper, we will set the basis for a reflection about the meanings of democratisation of digital memories, looking at how, in the big data era, preservation is currently being moved from traditional institutions of memory hands to distributed others.

Categories and Subject Descriptors

K.4 COMPUTERS AND SOCIETY –

K.4.1 Public Policy Issues [**Ethics**]

K.4.2 Social Issues

K.4.m Miscellaneous

General Terms

Management, Documentation, Reliability, Experimentation, Security, Human Factors, Standardization, Theory, Verification.

Keywords

Big Data, Governance, Digital memory, Information management, Democratisation, Citizen Participation.

Knowledge production and its governance are intertwined with memory practices on its various forms and therefore calls for a reflection on ethical dimensions. The development and use of Information and Communication Technologies and the hyper-connectivity momentum have lead to massive content creation, different forms of knowledge and also to humongous amounts of data, that has become known as ‘big data’.

Whereas institutions of memory were dealing with immateriality of contents, trying to find a solution for managing digital memories appropriately, big data is pushing towards a new meaning for memory making and makers. It is estimated that in 2007 only about 7 per cent of data produced was analogue [4], the rest being digital and the phenomenon is progressively and rapidly increasing, in this paper we would like to raise awareness about how collective digital memories could be affected by big data’s governance, in terms of quality of the information and new societal actors, namely algorithms.

The main characteristics of big data are Volume (that need advanced architecture to be managed), Velocity (data generated continuously) and Variety (data created in different formats) [2]. As data became big – as well in the definition, not too technical but catchy – a set of technologies and methodologies with great promises and equally great pitfalls [3] is developing.

When characterising big data some authors also include a fourth V, which stands for Veracity referring to ensuring integrity of data – in terms of formats and structure – that allows their correct management, but not considering their accuracy or exactness. An impressive amount of data is daily generated from very different sources, as well as used and interpreted by actors skilled enough to handle them. Because the cost of storage has fallen so much, it is easier to justify keeping data rather than discarding them. Therefore, the technical capability of preserving huge amounts of data - instead of applying appropriate appraisal procedures - triggers a lower level of attention to quality of contents and hinders a critical approach to governance of future memories, as well as diminishes the importance of the institutions that were traditionally in charge of managing them: institutions of memory, such as archives. Whereas some particular qualities, like provenance, authenticity and accuracy are fundamental to institutional records, these sorts of characteristics seem not to be demanded for big data. In fact, big data predictions rely on a huge amount of inexact data [4], whereas memories, as pieces of evidence, are based on integral and accurate information. The same thing happens to the perceived reliability of the institution (institutions of memory) and the process (memory practices) of preserving records. The truthfulness of records relies on the overarching idea that they are under the control of a trusted (legitimate) authority, which ensures the integrity of the system, its accuracy and reliability. But what are the qualities that make big data trustful? And does it have any kind of “trusted authority” behind data we can rely on?

Big corporations and some governments have the capacity to store massive amount of data for a future use and processing, posing threats to democracy and fundamental rights, such as privacy. This is especially important if we take into account that “*everything about our lives is in the process of becoming data*” [8]. From a concrete institutional perspective, given the exponential rise of contents, preserving everything is an unaffordable task, and appraisal becomes more necessary than ever. So we wonder who is in charge of deciding what data is to be kept or not, and under which criteria? Who is in charge of big

data? Who are the actors using and interpreting big data? Under which processes?

Can we consider big data as records? Records are not data, but “an account officially written and preserved as evidence or testimony” [7] is and this difference should be emphasised in order to avoid considering big data as a main source for memories. Not only institutions of memories but also other institutions dealing with data – like statistical national institutes - have developed a reflection on some critical points in big data management [1]. Can big data interact, overlap or substitute official data provided by institutional sources? Just relying on big data are we able of representing the whole spectrum of society, not just the one that is interacting digitally? Are small or underrepresented voices considered if technologies lead us to deal with trends instead of with stories?

We can also envision the risk in the processes for dealing with big data. Corporations, foundations and other new actors have different interests in the process of generating and preserving data. Such processes are based on algorithms created and performed by big technological companies, with different interests that determine and are determining how and what information is being preserved for the memory to be. Consequently, it is necessary to clarify which is everybody’s part in the digital landscape in order to ensure a consistent and accountable government of memories. In an illiterate algorithmic society, in which just a bunch of actors has the knowledge to deal with such amount of data, these datasets will become black boxes neither accountable, nor traceable. “To prevent this, big data will require monitoring and transparency, which in turn will require new types of expertise and institutions” [3]. So, how should institutions of memory face big data?

In a society collapsed by information, technology and actors behind big data are offering the added value for a concise access to information; yet they also have the capacity to control information as well as society in the long-term. “*The data collected now will have unforeseen uses (and value) in the future*” [6]. Following the same pattern, power has been distributed among these actors in pursuit of a desirable democratisation. In fact, democratised memories carry positive things such as opportunities for validity and enhanced access to information, better quality assurance and accountability, public engagement, and extended peer review and co-production. Nevertheless, some other unclear issues, regarding new memory actors’ goals remained unsolved, there has been a “redistribution of information power from the powerless to the powerful” [5]. Given the fact that big data directly influences decision making, social actors are empowered to collect, analyse and interpret data that may inform decisions. Not knowing *why* but only *what*, with the risk of lost centuries of practices of social understanding.[4] Algorithmisation and data fetishism could lead citizens to take trends, summarization and data visualisation produced by powerful actors as facts, instead of observations organised to respond to specific questions. If big data remains in the hands of a number of companies or governments, it may prevent people to dig into the information, limiting their access under a prisma of infographics and data visualisations, instead of traceable knowledge.

Whereas the opened debate about the bias and obscurantisms of archival procedures triggers attention to a new concept of open and participatory institutions of memory, fewer people seem to be paying attention to the normalisation by big data and technologies

of society, namely through the choice of what becomes visible and invisible, sharable and unsharable. Hence, the questions of ethical nature arise: who governs these data? - data owners, data brokers? Do we have enough constitutional rights against intentional or unintentional disclosure, dual use and abuse of all the data collected by corporations or public institutions? Could data be disclosed to third interested parties, without appropriate consent? For corporations it is manifestly expressed: for commercial use and profit, but could they be also used by public institutions? Are public institutions just interested in demonstrating their accountability and historical purposes or could this situation change in the future?

We have to assume that big data era has just started and it is still under development. We do not know yet the scale or the speed of big data progression, but we have many questions already. No clear rules or principles have been developed for dealing ethically with big data, but four normative ethical values have been formulated: privacy, confidentiality, transparency and identity [6]. Nevertheless if society wants to govern big data, “*who collects, shares and uses data must be made more transparent and accountable*” [6]. We argue that the ethics of memories governance has not been appropriately dealt in this big data context. Our contribution's here is to start a proper societal debate about the future of our memory (and knowledge creation and preservation) and how this is being (co-)constructed in the era of big data.

.^a The opinions of the authors in this paper cannot in any circumstance be taken as official position of the European Commission.

REFERENCES

- [1] Donvito, D. 2013 L’opportunità Big data: sfide e prospettive per la statistica ufficiale. *ISTAT News*. 7 (Feb. 2013), 12. Available at: http://www.istat.it/it/files/2013/02/Big_Data.pdf
- [2] IBM. 2013. *The four V’s of big data*. Available at <http://www.ibmbigdatahub.com/infographic/four-vs-big-data>
- [3] Lohr, S. 2012. How Big data Became so big *NYTimes* (Aug. 2012, 11) Available at http://www.nytimes.com/2012/08/12/business/how-big-data-became-so-big-unboxed.html?_r=0
- [4] Mayer-Schonberger, V., Cukier, K. 2013. *Big data*. John Murray, London.
- [5] Pariser, E. 2012. *The filter bubble: what the internet is hiding from you*. Penguin books, London.
- [6] Richards, N.M.; King, J.H. 2014. Big data ethics. *Wake forest law review*. (May. 2014, 19)
- [7] The American Heritage Dictionary of the English Language, Fourth Edition. 2000. Houghton Mifflin Company.
- [8] Watson, S.M. 2013. You are your data: and you should demand your right to use it. *Future tense*, November 2013. Available at: http://www.slate.com/articles/technology/future_tense/2013/11/quantified_self_self_tracking_data_we_need_a_right_to_use_it.html

