Review of Public Administration and Management

Mayer, Review Pub Administration Manag 2015, 3:1
DOI: 10.4172/2315-7844.1000171

Perspective Open Access

# Expectations and Pitfalls of Big Data in Biomedicine

**Miguel A. Mayer***

*Hospital del Mar Medical Research Institute (IMIM), DCEXS, Universitat Pompeu Fabra (UPF), Barcelona, Spain*

## Perspective

Big Data is one of the trending topics in information management, marketing and also in healthcare. Big Data describes a new generation of technologies and architectures that allow us to extract value from massive data volumes and types by enabling high-velocity capture, discovery and analysis of distributed data. Big Data is characterized by the so-called four V's: *volume* and complexity of data, *velocity* of collecting, storing, processing and analyzing data, *variety* in relation with different types of data (structured, unstructured and semi-structured) and veracity or 'data assurance' about data quality, integrity and credibility [1,2] The International Medical Informatics Association (IMIA) working group on "Data Mining and Big Data Analytics" defined Big Data as data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it [3]. Nevertheless, it is essential to consider that several aspects of the concept and its application can vary by domain, depending on what kinds of software tools are available in each case and what size and types of datasets are more common in a particular field, each type of Big data requires particular analysis methods [4].

The relevance of the new emerging scenarios related to the management of Big Data is so significant that several governments and institutions set up specific and strategic initiatives, to support the development of technologies to manage the huge amount of information accessible in a more efficient manner. In this way, in 2012 the Obama Administration announced the "Big Data Research and Development Initiative" with the aim of advancing, developing and using Big Data technologies applied to several areas such as health, defence, energy and climate change [5]. In 2014 The European Commission outlined a similar strategy on Big Data in order to develop actions in various fields such as health, food security, climate and energy, intelligent transport systems and smart cities [6]. Taking into account that in the past decade, large genomics big datasets have been created and the different 'omics' methodologies such as proteomics, transcriptomics, metabolomics or epigenomics are reaching their maturity, urgent action is needed in order to develop specific biomedical informatics tools to exploit and mine this information [7,8]. On the other hand, the resuse of clinical information for research and the integration of clinical, medical and public health electronic systems is becoming a reality. Through the integration of the diverse sources of clinical information, which at the current time often reside in separated silos, it is possible to analyse big data helping researchers find relationships among variables, not detectable in the past. In addition, new knowledge about the effectiveness of treatments can be generated reshaping biomedical discovery [9-12].

From the technical point of view, there are diverse and critical challenges related to Big Data that need to be solved. Without the appropriate information technologies infrastructures, requirements and analytic tools the potential contributions of Big Data can be misleading or lost [13]. Aspects related to the quality, completeness and reliability of data as well as reproducibility should be tackled carefully in order to better handle the data and the knowledge derived from it [14]. In addition, the use of more sophisticated visualization techniques and algorithms should be developed to facilitate organizations the interpretation and the practical application of the information obtained. On the other hand, the use and development of standards, vocabularies and ontologies, specifically designed for representing and integrating data across multiple sources, is a key component when dealing with Big Data [15]. Finally, there is a need to train a new generation of scientists with a strong background in statistics, information technologies, and computer science and at the same time with a good knowledge of medicine and biology [16].

In addition to the technical aspects, there are some sensitive issues when exploiting health personal data and managing the potential conflicts related to privacy and data protection. Specific regulations and policies to cleanly differentiate public and private information should be established in order to manage the potential risks issues related to Big Data, which should be managed in an appropriate manner. Legal, ethical and policy issues constitute some very critical matters in the context of sharing human-subjects' protection related to Big Data research techniques [17].

To sum up, we are facing a new scenario that offers the possibility of accessing and dealing with health data coming from very different resources such as electronic health records, social media platforms, scientific publications, drug and toxicology databases and all the 'omics' data currently available (genomics, proteomics and metabolomics). By discovering hidden associations and patterns within the data and transforming information into predictive models [18], Big Data analytics has the potential to improve people's health by means of contributing to a more personalized medicine, offering better health services and reducing costs. These changes can help to strengthen real patient-centered care after decades of disease-centered model of care, paving the way for a customization of healthcare and precision medicine. However, projects related to Big Data are still at an early stage, and for the majority of health institutions it is not clear what type of data and analysis are of real relevance and value, and what the strategy should be to cope with the internal changes required, in order to adapt to this new scenario and continuous process, in a sustainable way. For all of these reasons, further investigation into the advantages and limitations of Big Data is imperative as well as a collaborative and interdisciplinary approach.

***Corresponding author:** Mayer MA, Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), DCEXS, Universitat Pompeu Fabra (UPF), C/o Dr. Aiguader, 88, 08003- Barcelona, Spain, Tel: +34 93 316 0648; E-mail: mmayer@imim.es

**Received** August 25, 2015; **Accepted** October 17, 2015; **Published** October 19, 2015

**Citation:** Mayer MA (2015) Expectations and Pitfalls of Big Data in Biomedicine. Review Pub Administration Manag 3: 171. doi:10.4172/2315-7844.1000171

## References

1. Li Y, Chen L (2014) Big biological data: challenges and opportunities. Genomics Proteomics Bioinformatics 12: 187-189.
2. Raghupathi W, Raghupathi V (2014) Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2: 3.
3. Bellazi R (2014) Big data and biomedical informatics: a challenging opportunity. Yearb Med Inform 9: 8-13.
4. Huang T, Lan L, Fang X, An P, Min J, et al. (2015) Promises and challenges of big data computing in health sciences. Big Data Research 2: 2-11.
5. https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release.pdf
6. http://ec.europa.eu/digital-agenda/en/towards-thriving-data-driven-economy
7. Greene CS, Tan J, Ung M, Moore JH, Cheng C (2014) Big data bioinformatics. J Cell Physiol 229:1896-1900.
8. Qin Y, Yalamachili HK, Qin J, Yan B, Wang J (2015) The current status and challenges in computational analysis of genomic big data. Big Data Research 2:12-18.
9. Nash DB (2014) Harnessing the power of big data in healthcare. Am Health Drug Benefits 7: 69-70.
10. Mayer MA, Furlong LI, Torre P, Planas I, Cots F, et al. (2015) Reuse of EHRs to support clinical research in a hospital of reference. Stud Health Technol Inform 210: 224-226.
11. Schneeweiss S (2014) Learning from big health care data. N Engl J Med 370: 2161-2163.
12. Issa NT, Byers SW, Dakshanamurthy S (2014) Big data: the next frontier for innovation in therapeutics and healthcare. Expert Rev Clin Pharmacol 7: 293-298.
13. Roski J, Bo-Linn GW, Andrews TA (2014) Creating Value in health care through big data: opportunities and policy implications. Health Affairs 33: 1115-1122.
14. Cases M, Furlong LI, Albanell J, Altman RB, Bellazi R et al. (2013) Improving data and knowledge management to better integrate health care and research. J Intern Med 274: 321-328.
15. Merelli I, Pérez-Sánchez H, Gesing S, D'Agostino D (2014) Managing, analysing, and integrating big data in medical informatics: open problems and future perspectives. Biomed Res Int 2014: 134023.
16. Martin-Sánchez F, Verspoor K (2014). Big data in medicine is driving big changes. Yearb Med Inform 9: 14-20.
17. Metcalf J (2015) Human-subjects protection and big data: open questions and changing landscapes. Council for Big Data, Ethics, and Society.
18. Costa FF (2014) Big data in biomedicine. Drug Discov Today 19: 443-440.