

Chapter 14

The Challenge of Big Data and Data Mining in Aging Research



Juan Carlos Gómez-Verján and Luis Miguel Gutiérrez-Robledo

“Indeed, we are at the point in history where Big Data should no intimidate, but inspire us”.

-Francis Collins-

Abstract Population growth and fast evolution of several technologies into almost every corner of human activities have become a constant in the modern society nowadays. Consequently, an enormous amount of information is generated every day, from many sources, such as: web information, mobile phones, social media, scientific reports, medical and healthcare information among others, resulting in the so called Big Data. Digitization, storage, collection, and particularly analysis of patterns on these data have led to advances in data sciences (Data Mining) to obtain valuable information from people necessities, demands, hobbies, and activities daily from such databases (Big Knowledge). On the other side, aging has become one of the main challenges of modern societies, since there is an increase on prevalence of chronic pathological conditions and aging associated outcomes. Biomedical and clinical research have experienced several advances in the last years with the implementation of omics technologies and digitization of healthcare, in this context, Big Data and Data Mining become a powerful tool to obtain valuable information that could be used to approach the complexity of the aging process and the clinical implications for older adults. Several companies and governments all over the world have already started to successfully implement these types of technologies from pharmaceutical developments to public health issues. Although future seems promising novel holistic approaches are needed to implement a multidisciplinary agenda

J. C. Gómez-Verján (✉)

Department of Basic Research, National Institute of Geriatrics, Mexico City, Mexico

e-mail: jverjan@inger.gob.mx

L. M. Gutiérrez-Robledo (✉)

National Institute of Geriatrics, Mexico City, Mexico

e-mail: lmgutierrez@inger.gob.mx

that could help us understand the complexities of human aging and develop the appropriate interventions to improve older people's health in healthcare systems.

Keywords Big data · Data mining · Big knowledge · Systems biology · Public health

14.1 Big Data

Along with the population growth worldwide, and the fast evolution of technologies into almost every corner of human activities, an enormous quantity of information is generated every day, from many sources, such as: web information, mobile phone usage, social media activities, consumer preferences, financial systems, climate information, scientific reports, medical and health care information [1], moreover, the software and data business company Domo Inc. on its annual report of the world's data generation called "Data Never Sleeps 5.0" for 2017, indicate that there are almost 3.7 billion users in the global internet population, and there is an output of 2.5 quintillion bytes a day of data created daily [2]. In this context, the digitalization, storage, collection, maintenance and analysis of these data have led to advances in data sciences and into the so called "Big Data Era". This revolution has led to changes in prevalent paradigms, since most of these data cannot be managed by traditional and conventional techniques for data management, and novel infrastructure and approaches for data sciences need to be taken by executives, academics and economist all over the world.

John Mashey a researcher from Pennsylvania University in computer science was responsible for making the term "Big Data" quite popular in the early 1990's decade. Nowadays, Big Data is usually defined as huge amount of data (in the order of Terabytes-10¹² bytes and Petabytes-10¹⁵ bytes) highly complex, that within a reasonable period of time cannot be captured, managed, processed, interpreted and organized as information that could be read by human beings or traditional data processing software [1]. Most of the technologies used for these novel approaches are computational linguistics and machine learning. Table 14.1 indicates some of the main software available for the Big Data processing and management. In this context, accordingly to Davenport, Big Data could be classified in two main groups: machine-generated (data created by machines without human intervention) and human-generated (created with human intervention).

14.1.1 Data Mining and Big Data Analysis

Managing data is one of the most important issues once you start to work with Big Data, several authors have stated different methodologies to start to work with, that in general could be divided by: Data collection, Data measures, Data Analysis and Knowledge discovery [3], accordingly to: Volume (data generation and processing

Table 14.1 Main software available for big data and data mining processing

Software	Characteristics	Type
<i>Apache Hadoop</i>	<i>Is an open-source software for the processing, and storage of big data, using MapReduce programming models on computer clusters, therefore, there is no problem with hardware, it possess the ability to handle limitless concurrent jobs or task</i>	Open source
<i>HPCC systems big data</i>	<i>This is a platform for transforming, querying and data warehousing, implemented on commodity computing clusters to provide high-performance on data parallel processing for big data</i>	Open source
<i>Mongo DB</i>	<i>Is a no-SQL database, written in C++, that helps to store and analyze big data</i>	Open source
<i>R-programming</i>	<i>R programming language, has been designed and implemented as a data mining tool and is widely used for data analysis and machine learning applications, currently there are a series of packages available for big data</i>	Open source
<i>RapidMiner</i>	<i>Is a software platform that provides integrated environment for data preparation, machine learning and predictive analysis</i>	Commercial
<i>Knime</i>	<i>Is a leader in the analysis (machine learning), integration and reports for big data, it possess an open source version available for public in general and a quite remarkable commercial version for specialized data scientist</i>	Open source and commercial
<i>Orange</i>	<i>Is an open-source data visualization, machine learning and data mining highly used tool that possess algorithms for predictive modelling it could be used as a Python library</i>	Open source
<i>Spark</i>	<i>Is an open-source cluster-computing framework that provides an interface programming for big data analysis, maintenance and analysis.</i>	Open source
<i>IBM SPSS modeler</i>	<i>Is a data mining workbench for analyzing and developing predictive models supported by IBM SPSS software</i>	Commercial

numbers), Velocity (data processing in concordance with data generation speediness) and Variety (quality of data i.e. type of data and if it is structured or unstructured).

Data Analysis and Knowledge discovery will make us get novel and useful information from the databases that could be useful for prediction, classification and innovation depending on the type of information we are working with. In this context, the so called Data Mining defined as an analytic process designed to discover or extract patterns, novel information and systematic relationships from variables of large data sets (usually Big Data) [4], involving also the process of storage and processing the data, as well as the complicated process of presenting the results in a way understandable and easily interpretable for everybody. The last goal of Data Mining is the prediction of patterns, the term was first coined by Piatetsky-Shapiro and Frawley in the early 90's when there was a rush for developing novel algorithms for data processing in business by several software and data companies all over the world [5], it was originally called Data Collection by IBM in the 60's. Data mining

techniques are the result of a long process of research in product and business development, that began when the business of data warehousing and collection began to grow.

In order to make Data mining usable as an important tool it involves the complete understanding of the architecture of the data and the analytical methods to be used, since predictive relationships of data may not necessarily represent causes of an action or a behavior [6]. Although a detailed review of the complete Data mining technique is beyond the scope of this chapter, briefly, the process of data mining consists in 3 different stages [7]:

- A) Exploration and Collection of data: At this stage of the analysis we usually focus on the type of data we are going to work with, the type of data, and the nature of the analytical problem we wish to work with. Usually this stage involves the process of data curation [8], i.e. selection of adequate data with complete meta-information for the analysis, transforming data to the correct format for the software analysis, type of variables and predictors to be measured and usually exploratory analysis.
- B) Modelling the data: Building a model involves the knowledge of the kind of answers you wish to apply or the type of prediction you wish to do; since there are several software packages with already known models developed [6]. This stage is critical and involves the highest and most elaborated process in Data mining, in this context, there are a wide variety of techniques, such as: association, normal statistics, classification, Bayesian statistic, and several machine learning techniques [9] such as: clustering, neighborhood, decision trees, neural networks, deep learning, bootstrap aggregating, boosting, etc.
- C) Deployment and validation of the data: Building the model is not the end of the analysis, and once the model has been processed and evaluated, you must validate and interpret the results, their significance, and in case how does the novel information could be classified, remembering that Data mining is not statistics, we do not care how the data is distributed but what is the potential of such data for prediction and, or pattern identification [10]. Moreover, if you gained knowledge from the data you will then need to organize and present such results in a way that anyone in the field could use it. In this stage, there are several examples of informatics tools useful for deployment of the data and the exporting of algorithms, such as: predictive model markup language, portable format for analytics, confusion matrices, SQL, R-algorithms, Python, applications for Java and several programming tools [11].

14.2 Big Data and Health

It is widely known that health involves the production of an enormous quantity of data including diagnostic results and images, medical records, laboratory results, public health registry in the different levels, and data produced for biomedical and

for clinical research [12]. In this context, Ruckenstein et al. [13] proposed the so-called “datafication of health” i.e. the conversion of all aspects of health at its different levels, into quantifiable data, meaning that all techniques that are currently being used for Big Data and Data mining could be applied to analyze health at different levels, the so called “Biomedical Big Data” (BBD). Moreover, the value of health research based on non-traditional data streams services from internet such as: e-mail, online purchasing and video conferences has already been demonstrated [14].

BBD could help us analyze and store people’s information throughout their lives on: diseases, phenotype, genotype, behavior, environmental location, occupation, and clinical data, making therefore, health-predictions easily for individuals and in consequence for populations [12]. In this context, it could provide of the correct tools to governments, public health departments, and decision makers for the implementation of the adequate prevention politics and interventions, in order to improve health on population, for instance, machine learning techniques have allowed several medical disciplines to create more accurate prognostic and diagnostic models based on pattern recognition (computer-aided detection), that could help physicians to improve their diagnosis [15]. Several governments all over the world have started to invest and create departments dedicated only to the analysis of BBD, for instance, the National Institutes of Health (NIH) launched the “Big Data to Knowledge” (BD2K) initiative in 2012 [16], which involves multiple research centers such as the Big Data for Discovery Science Center and the Center for Expanded Data Annotation and Retrieval, as well as a set of focused individual research and training projects, to enable biomedical research and consider novel approaches in data science, to facilitate discovery and support new knowledge, the most important objective of this initiative is to index software to operate these datasets; on the other hand, the e-Health Action plan for 2012–2020 for the European Commission of Health, has allocated 2 billion € under Horizon 2020 program to invest in research and innovation on Big Data [17] as well as, a public-private partnership with a budget of 1.638 billion from the EU commission and 1.425 from other life sciences industries and organizations. Moreover, an Oxford Economics revealed that at least 70% of health-care companies are looking to invest in Big data, Data mining and cloud computing expecting a significant impact in innovation in several topics on healthcare field. In this context, as stated by Vayena et al. [14], it is important to think of health-related big data as an evolving ecosystem.

14.3 Big Data in Epidemiology

Epidemiology and public health, have changed dramatically over the last years, and are now more interconnected with so many other sciences due to the global advances in technology and to the interdisciplinary nature of such sciences. Nowadays, the use of Big Data for epidemiologist is as referred by Salerno J., et al., “the exploration and interpretation of very large and complex datasets derived from pooling

cohorts, from omics projects, electronic stored medical records, and health digital information” [18], since usually a normal epidemiologist turns for primary and secondary data sources to initiate its study, however, in the Big Data era, there is an enormous amount of information and sources of data (medical records, biobanks, geolocalization, shopping habits, genomic data, pharmaceutical prescriptions, social behavior, among others). In this context, novel epidemiologist must be prepared for time-consuming data collection, curation and storage, as well as, novel approaches using Data Mining tools and statistical approaches, as well as novel ethical and legal challenges related to potential harms of Big Data use, including confidentiality and privacy issues and other potential concerns from institutions and public agencies all over the world.

Several examples of epidemiological studies and Big Data have arisen over the last years, among the most important projects is the Nordic Arthroplasty Register Association (NARA) database which includes information concerning implant brands, fixation methods, and implant survival,, since 1995 in Denmark, Finland, Iceland, Norway, and Sweden, generating information from more than 1 million patients from each of the different countries [19] in each of the countries; other interesting example is the Observational Medical Outcomes Partnership (OMOP) which is a public-private partnership formed by several public and private representatives including the Food and Drug Administration (FDA) and members of the Big Pharma industry mainly focused on pharmacoepidemiology using and incentivizing novel healthcare databases through electronic health records allowing therefore, linkage possibilities and the possibility of a lifelong complete follow-up [20]; moreover, South Korea which is a world leader in information technology infrastructure launched in 2011 the so-called Big Data Initiative, which established a pan-governmental big data network and analysis systems, it was so important that in 2014 the ministry of science of such country released the Medical information consulting program to collect medical data and customize treatment and help the national health insurance service become more efficient, providing information to patients such as: duration of illness, cost of treatments, medical services, cases per location, institutions specialized in diseases; and to medical industry of pharmaceutical trends, distribution of drugs, medical equipment and devices most asked for by the population [21]. A very successful case example was the Seoul National University Bundang Hospital, which was the first hospital in the Asia-Pacific region to fully digitalized big data, and doctors and nurses are able to configure systems with precise clinical information, improving time of patient referral from 48 h to 4–6 h, and for example reduce the dosage of antibiotics before surgery [22].

14.4 Big Data Biomedicine Research and Omics

It has been years since the Human Genome Project results were published, and biology experienced a revolution; as well in our understanding of biological mechanisms as well as in access to an enormous amount of biomedical data now freely

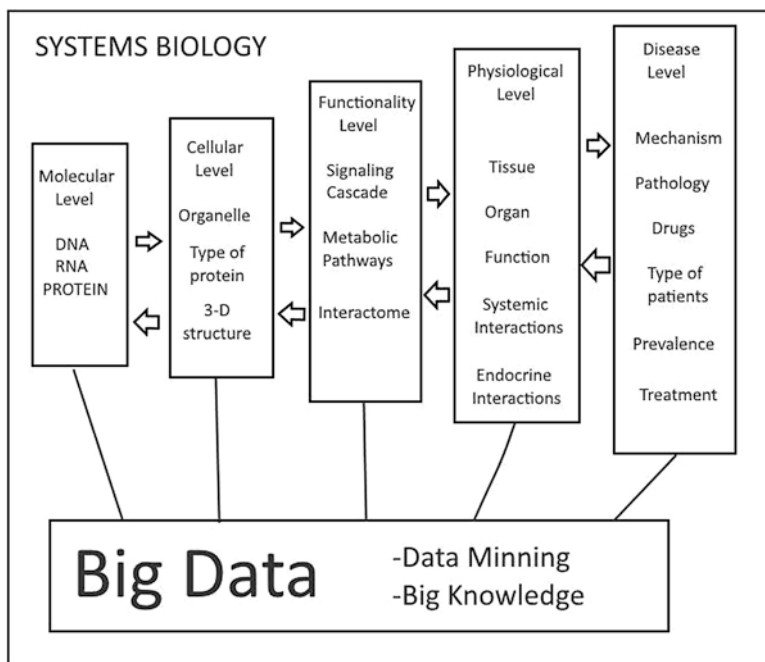


Fig. 14.1 Holistic points of view in systems biology. Analysis coming from different omics technology could only be helpful if we have a holistic approach from different levels this is the objective of systems biology

available. Novel -omics tools are being daily applied to several topics in biomedical research projects in the last few years (genomics, transcriptomics, proteomics, epigenomics, metagenomics, metabolomics, and microbiome, among others) [23]. Moreover, this information now helps us understand relationships between genotype, phenotype, behavior, and many outcomes of human beings through the development of the so-called Systems Biology (Fig. 14.1) [16].

However, information obtained from such technologies increases dramatically, as well as the need of developing public repositories. Besides, each Omic experiment represents challenges, since standard nomenclature and analysis for genes are not the same for proteins or metabolites. Each technology involves the production of different data in different formats (Table 14.2), standards for information and meta-information are needed so that stored data is clear enough for interpretation and re-use from other scientist to compare and reproduce experiments [24].

In the last few years, there has been an international effort from several consortia all-over the world to develop and standardize -omics technologies results to be used and published e.g. for Genomics (Minimum Information about a Genome/Metagenome Sequence from Genomic Standards Consortium), Transcriptomics (Minimum Information about a Microarray Experiment from Functional Genomics Data Society), Proteomics (Minimum Information about a Proteomics experiment

Table 14.2 Main omics used in biomedical research and most common formats

Omics	Technologies used	Formats
Genomics	Genotype microarrays, DNA-seq, Whole genome sequencing, Whole exome sequencing	FASTA, FASTQ, SAM, BAM, VCF, CEL, CSV, GFF, BED, LOH, SNP, IGV, MUT, TSV, JSON
Transcriptomics	RNA-microarrays, RNA-seq, SAGE-seq	FASTA, FASTQ, BAM, SAM, CEL, RES, CSV, GFF, GCT, GFF, BED, IGV, SEG, TSV, JSON
Proteomics	Protein chips, Mass spectrometry, 2-D gel electrophoresis, Mononuclear-NMR	mzML, TraML, mzIdentML, mzXML, mzData, mzQuantML, MSF, tandem, omx, dat, FASTA, PEFF, mzTab, PRIDE XML, MGF, ms2, pkl, TSV, CSV
Epigenomics	ChIP-Seq, MeDIP-Seq, DNase-Seq, Mnase-Seq	FASTA, FASTQ, SAM, BAM, CSV, TSV, GFF, SEG, BED, WIG, txt
Metabolomics	Mono and heteronuclear-NMR, Mass spectrometry, Chromatography and mass spectrometry,	CDF, mzML, mzXML, mzData, mzQuantML, txt, TSV, CSV, CVR, VSN, mzIdentML

from the Human Proteome Organization), Metabolomics (Core information for Metabolomics), Epigenomics/Transcriptomics (Minimum information about a high-throughput Nucleotide sequencing Experiment by Functional Genomics Data Society) [24]. In this burgeoning context, information will soon be reproducible and amenable to knowledge translation into personalized medicine.

14.5 Challenges of Big Data and Aging

It is widely known that life expectancy all over the world has increased in the last few years. According to the World Health Organization report on Aging and Health, in 2016 global life expectancy for females is 73.8 and 69.1 for males [25] as life span grows, the prevalence of chronic conditions and age associated disease increases, and becomes a matter of interest for several public and private stakeholders. Multi-morbidity results from this phenomenon and represents a challenge for health care systems. Diversity increases as we age, and disadvantage plays a significant role in modelling these differences, which vary widely throughout a lifetime. In this very complex and diverse context, recent advances on information, communication and biomedical technologies such as electronic devices, medical digital diagnostics and prescriptions, medical assistive and wearable devices, personalized medicine and genomics, among others, represent an opportunity for an efficient use in order to provide the health-care system with new tools based on anticipation, prevention and an improved continuum of care. Besides, another field of development, as electronic medical records improve and become widespread; could be the use of a Big Data approach to provide decision makers in the public and private sectors with novel information for planning interventions and public health policies;

and to help develop new pharmaceutical and health device resources answering to the needs of an aging population [26, 27].

Modern developments in Big data and Data Mining represent an opportunity to improve multidisciplinary exchanges “Big Knowledge” and accordingly to William Callaghan a novel paradigm in the scientific method, since id. “...comprehensive data coverage unearths causal relationship between phenomena...” [28], which is not far away from most of modern scientific projects for instance, most genomic projects involve first the genome acquisition and consequently the data analysis to develop a hypothesis. However, since Big Data involves a holistic approach multidisciplinary research groups must be created so that results could improve in order to raise the probability of moving towards the right conclusion especially in the field of aging, where a continuous monitoring and correct interventions, as well as prevention are among the main challenges for health systems [26]. However, we must be aware of the fact that already, among the complexity of aging in modern life; the lived reality of big data should be approached with caution because all the data we are shedding every day is too revealing of our intimate selves but may also misrepresent us. Like a fluorescent light in a dark corridor, it can both show too much and not enough [29]. Therefore, we do not really know the true potential or the dangers of big data analysis, and we must be restrained in our approach and look for a balance between enthusiasm and the reality of our limitations. It’s important to mention that any government that adopts any type of Big-Data approach on public health must create policies for the protection of individual data and at the same time promote access and sharing for the use of such data for public benefit [14].

Over the past few years many long-term care programs have been implemented for aging in several countries such as: Japan, Taiwan, USA, Australia, Canada, South Korea, and Denmark, which intended to implement information and communication technologies to improve long-term care systems the so-called aging in place [26]. Even so, Japan’s government made mandatory the long-term care insurance system and implemented the “Guideline to promote the Appropriate Use of Information Systems in Care for the Elderly in Conjunction with At-home Care” to stimulate Big Data networks of information from healthcare providers and licensed in home caregivers, nursing and doctors, as well as information from patients such as residence, medical treatment, nursing, etc. [30]. Another successful example of the implementation of Big Data to healthcare systems in older people, is the Australian e-Health Research Centre which have almost 50 researchers, software engineers and doctorate students developing technologies in partnership with clinicians to improve health care systems. They have developed an innovative platform for cardiac rehabilitation, using wearable devices and mobile phones [31]. Moreover, several start-up companies have moved into the field, such as the New York based Hometeam Inc. which developed its own software to match caregivers with families through mobile technology in a very personalized care planning for older adults. Another successful example is CareZapp Inc. located in United Kingdom; they developed a mobile app that creates an ecosystem platform connecting caregivers, volunteers, family, friend and doctors to improve the health of older patients, by

means of in-home sensors, wearable devices, and mobile phone with notifications 24/7 of patient status [32].

14.6 Globolomics of Aging Research

For decades, researchers in the biology of aging have focused on defining mechanisms that modulate aging by primarily studying a single metric, sometimes described as the “gold standard” lifespan. Increasingly, geroscience research is turning towards defining functional domains of aging such as the cardiovascular system, skeletal integrity, and metabolic health as being a more direct route to understand why tissues decline in function with age (see Chap. 4 on Geroscience). Each model used in aging research has strengths and weaknesses, yet we know surprisingly little about how critical tissues decline in health with increasing age and how the different systems interact. We know very little as well about the interplay between the biological mechanisms of aging and chronic disease.

Over the last few years, research on the molecular foundations of aging and consequently anti -aging therapies, have increased dramatically, just to mention an example, a search for the words: “molecular” and “aging” in PubMed database gets almost 35,563 documents, increasing from 553 documents in 1997 to 3551 in 2017. It becomes clear that several Omic approaches have been critical for progress in the field, performed either with clinical human samples or with animal models (*C. elegans*, *S. cerevisiae*, *M. musculus*, *Rattus norvegicus* etc). In this context, there is an enormous amount of information freely available in the web, consequently, Big Data and Data Mining approaches become relevant in order to systematically extract information and obtain knowledge from such.

As previously mentioned nowadays there are several Omics technologies available, however integration of the data is the main challenge in the field and remains a black box in genomic studies [33], however a new way to improve our understanding of the aging process is acknowledging its complexity and developing new methods capable of embracing it. The use of such technologies and its results, each omic tool, will only give an answer to specific different questions i.e. proteomics will give us information that may be or may be not be related with metabolomics. Therefore globolomics (or “deep phenotyping” leading to molecular or genetic epidemiology, albeit at a finer resolution) approaches must be consider when performing studies in aging so we could have a systems approach of results. Moreover, association with physiological conditions must be consider [34], projects such as the Human Physiome Project become increasingly important, as their analysis could give us a better approach to the proportion of genetic and environmental (nature and nurture) participation in aging and consequently, contribute to the development of a novel aging theory.

14.7 Expectations on Big Data in Aging

It's been quite a few years since the declaration of the Madrid International Plan of Action on Ageing, called for the elimination of the inequalities in health-access and to the development of novel health-care systems policies for older people [35], and although it's still a long and arduous road to travel, novel technologies on informatics, genomics and communication could help us diminish such inequalities and move forth in all the fields of aging research. In this context, there are a lot of international efforts already being successfully developed from private and public sectors, and a number of researchers and governments are already compiling aging and health information in current databases [30, 31, 36]. The more we learn on Big Data and Data Mining the faster we will close this gap.

References

1. Yang C-T, Liu J-C, Chen S-T, Lu H-W (2017) Implementation of a big data accessing and processing platform for medical records in cloud. *J Med Syst* 41:149. <https://doi.org/10.1007/s10916-017-0777-5>
2. Data Never Sleeps 5.0 | Domo [Internet]. [cited 29 Nov 2017]. Available: <https://www.domo.com/learn/data-never-sleeps-5>
3. Vitari C, Raguseo E (2016) Digital data, dynamic capability and financial performance: an empirical investigation in the era of big data. *Systèmes d'Information & Management* 21(3):6392. <https://doi.org/10.3917/sim.163.0063>
4. Han J, Kamber M, Pei J (2012) Data mining trends and research frontiers. *Data Min*:585–631
5. Piatesky-Shapiro G (1994) An overview of knowledge discovery in databases: recent progress and challenges. Workshops in computing pp 1–10
6. Maimon O, Rokach L (2009) Introduction to knowledge discovery and data mining. In: *Data mining and knowledge discovery handbook*, pp 1–15
7. McCue C (2015) Chapter 3 - data mining and predictive analytics. In: McCue C (ed) *Data mining and predictive analysis*, 2nd edn. Butterworth-Heinemann, Boston, pp 31–48
8. Freitas A, Curry E (2016) Big data curation. In: *New horizons for a data-driven economy*, pp 87–118
9. Kononenko I, Kukar M (2007) Chapter 3 - machine learning basics. In: KIK M (ed) *Machine learning and data mining*. Woodhead Publishing, pp 59–105
10. Michael A, AGSL B (2008) Mastering data mining: the art and science of customer relationship management. *Ind Manag Data Syst* 100(5):245–246
11. McCue C (2015) Process models for data mining and predictive analysis. In: *Data mining and predictive analysis*, pp 51–74
12. Vayena E, Blasimme A (2017) Biomedical big data: new models of control over access, use and governance. *J Bioeth Inq* 14(4):501–513. <https://doi.org/10.1007/s11673-017-9809-6>
13. Ruckenstein M, Schüll ND (2017) The datafication of health. *Annu Rev Anthropol* 46:261–278
14. Vayena E, Dzenowagis J, Brownstein JS, Sheikh A (2017) Policy implications of big data in the health sector. *Bull World Health Organ* 96(1):66–68. <https://doi.org/10.2471/BLT.17.197426>
15. Cabitza F, Rasoini R, Gensini GF (2017) Unintended consequences of machine learning in medicine. *JAMA* 318(6):517–518. <https://doi.org/10.1001/jama.2017.7797>
16. Bui AAT, Van Horn JD (2017) NIH BD2K centers consortium. Envisioning the future of “big data” biomedicine. *J Biomed Inform* 69:115–117. <https://doi.org/10.1016/j.jbi.2017.03.017>

17. Kolitsi Z, Thonnet M (2014) New directions in eHealth governance in Europe. In: *Managing eHealth*. Palgrave Macmillan, London, pp 50–60
18. Salerno J, Knoppers BM, Lee LM, Hlaing WM, Goodman KW (2017) Ethics, big data and computing in epidemiology and public health. *Ann Epidemiol* 27(5):297–301. <https://doi.org/10.1016/j.annepidem.2017.05.002>
19. Johanson P-E, Fenstad AM, Furnes O, Garellick G, Havelin LI, Overgaard S et al (2010) Inferior outcome after hip resurfacing arthroplasty than after conventional arthroplasty. Evidence from the nordic arthroplasty register association (NARA) database, 1995 to 2007. *Acta Orthop* 81(5):535–541. <https://doi.org/10.3109/17453674.2010.525193>
20. Stang PE, Ryan PB, Racoosin JA, Marc Overhage J, Hartzema AG, Reich C et al (2010) Advancing the science for active surveillance: rationale and design for the observational medical outcomes partnership. *Ann Intern Med* 153(9):600–606. <https://doi.org/10.7326/0003-4819-153-9-201011020-00010>
21. Jung JJ, Kim P (2017) Big data technologies and applications: 7th international conference, BDTA 2016, Seoul, South Korea, 17–18 Nov, 2016, Proceedings. Springer
22. Yoo S, Hwang H, Jheon S (2016) Hospital information systems: experience at the fully digitized Seoul National University Bundang hospital. *J Thorac Dis* 8(Suppl 8):S637–S641. <https://doi.org/10.21037/jtd.2016.08.44>
23. He KY, Ge D, He MM (2017) Big data analytics for genomic medicine. *Int J Mol Sci* 18(2):1–18. <https://doi.org/10.3390/ijms18020412>
24. Chervitz SA, Deutsch EW, Field D, Parkinson H, Quackenbush J, Rocca-Serra P et al (2011) Data standards for omics data: the basis of data sharing and reuse. *Methods Mol Biol* 719:31–69. https://doi.org/10.1007/978-1-61779-027-0_2
25. OECD (2017) OECD average life expectancy and perceived health, since 2005 [internet]. https://doi.org/10.1787/how_life-2017-graph10-en
26. Song P, Chen Y (2015) Public policy response, aging in place, and big data platforms: creating an effective collaborative system to cope with aging of the population. *Biosci Trends* 9(1):1–6. <https://doi.org/10.5582/bst.2015.01025>
27. Kwon Y, Natori Y, Tanokura M (2017) New approach to generating insights for aging research based on literature mining and knowledge integration. *PLoS One* 12(8):e0183534. <https://doi.org/10.1371/journal.pone.0183534>
28. Callaghan CW (2017) Developing the transdisciplinary aging research agenda: new developments in big data. *Curr Aging Sci* 10. <https://doi.org/10.2174/1874609810666170719100122>
29. Crawford K, Finn M (2014) The limits of crisis data: analytical and ethical challenges of using social and mobile data to understand disasters. *GeoJournal* 80:491–502. <https://doi.org/10.1007/s10708-014-9597-z>
30. Tamiya N, Noguchi H, Nishi A, Reich MR, Ikegami N, Hashimoto H et al (2011) Population ageing and wellbeing: lessons from Japan's long-term care insurance policy. *Lancet* 378(9797):1183–1192. [https://doi.org/10.1016/S0140-6736\(11\)61176-8](https://doi.org/10.1016/S0140-6736(11)61176-8)
31. Hansen DP, Gurney P, Morgan G, Barraclough B (2011) The Australian e-Health research centre: enabling the health care information and communication technology revolution. *Med J Aust* 194(4):S5–S7
32. Sonnega A, Robinson K, Levy H (2016) Home and community-based service and other senior service use: prevalence and characteristics in a national sample. *Home Health Care Serv Q* 36(1):16–28. <https://doi.org/10.1080/01621424.2016.1268552>
33. Lorusso JS, Sviderskiy OA, Labunskyy VM (2017) Emerging omics approaches in aging research. *Antioxid Redox Signal*. <https://doi.org/10.1089/ars.2017.7163>
34. Lund E, Dumeaux V (2008) Systems epidemiology in cancer. *Cancer Epidemiol Biomark Prev* 17(11):2954–2957. <https://doi.org/10.1158/1055-9965.EPI-08-0519>
35. Sidorenko AV, Mikhailova ON (2013) Implementation of the Madrid international plan of action on ageing in the CIS countries: the first 10 years. *Adv Gerontol* 26(4):585–593
36. de Magalhães JP, Stevens M, Thornton D (2017) The business of anti-aging science. *Trends Biotechnol* 35(11):1062–1073. <https://doi.org/10.1016/j.tibtech.2017.07.004>