

# Big Data: Are Biomedical and Health Informatics Training Programs Ready?

## Contribution of the IMIA Working Group for Health and Medical Informatics Education

P. Otero<sup>1</sup>, W. Hersh<sup>2</sup>, A. U. Jai Ganesh<sup>3</sup>

<sup>1</sup> IMIA WGEEd Chair, Department of Health Informatics, Hospital, Italiano de Buenos Aires, Buenos Aires, Argentina

<sup>2</sup> IMIA WGEEd Past-Chair, Department of Medical Informatics & Clinical Epidemiology, Oregon Health & Science University, Portland, OR, USA

<sup>3</sup> IMIA WGEEd Co-Chair, Sri Sathya Sai Central Trust, Prasanthi Nilayam, Puttaparthi, Andhra Pradesh, India

### Summary

**Objective:** The growing volume and diversity of health and biomedical data indicate that the era of Big Data has arrived for healthcare. This has many implications for informatics, not only in terms of implementing and evaluating information systems, but also for the work and training of informatics researchers and professionals. This article addresses the question: What do biomedical and health informaticians working in analytics and Big Data need to know?

**Methods:** We hypothesize a set of skills that we hope will be discussed among academic and other informaticians.

**Results:** The set of skills includes: Programming - especially with data-oriented tools, such as SQL and statistical programming languages; Statistics - working knowledge to apply tools and techniques; Domain knowledge - depending on one's area of work, bioscience or health care; and Communication - being able to understand needs of people and organizations, and articulate results back to them.

**Conclusion:** Biomedical and health informatics educational programs must introduce concepts of analytics, Big Data, and the underlying skills to use and apply them into their curricula. The development of new coursework should focus on those who will become experts, with training aiming to provide skills in "deep analytical talent" as well as those who need knowledge to support such individuals.

### Keywords

Data interpretation, statistical; medical informatics, education; databases as topic; individualized medicine; patient participation

Yearb Med Inform 2014;177-81

<http://dx.doi.org/10.15265/IY-2014-0007>

Published online August 15, 2014

Healthcare data increasingly come from many sources. In the past, most data were obtained from patient records, public health surveillance reports, or research results. These data could be in a structured, unstructured, or semi-structured form when digitally recorded. With the advancement of health information technology (HIT) in the last decade, more data can be used for analyzing information and the creation of new knowledge. The implementation and use of HIT systems that integrate the electronic health record (EHR) has increased the amount of data that are digitally gathered at every encounter. When a patient seeks care in a healthcare facility with a better integration of clinical care, it may improve the care delivered to this patient, but also to others if the information can be re-used for research.

In addition to the sources described above, "new" kinds of health data created and managed by patients have emerged from fitness and personal health data capture devices, social media, and genomics and related sources. These data are presently not usually available at the point of care because they are difficult to capture, organize, and integrate for proper analysis [1].

### Healthcare Challenges

There is a growing concern over waste and inefficiency in health care, in part due to poor use and integration of data. The US Institute

of Medicine (IOM) has estimated annual excess costs of care in the US of around \$750 billion (out of \$2.5 trillion expended) and also resulting in approximately 75,000 annual premature deaths [2]. The IOM has categorized the causes of waste and harm as unnecessary services provided, services inefficiently delivered, prices too high relative to costs, excess administrative costs, missed opportunities for prevention, and fraud. While some of these problems may be unique to the US, it is likely that they exist in most other countries.

The convergence of all data related to healthcare from these disparate sources can help in developing what the IOM has termed the learning health system (LHS), which aims to lead to a situation where "each patient-care experience naturally reflects the best available evidence, and, in turn, adds seamlessly to learning what works best in different circumstances" in order to provide information to improve healthcare decisions, encourage patient empowerment by means of education on health issues and self-management, help in defining public health strategies, and provide the necessary support research and the development of new knowledge that can provide feedback to the LHS as a virtuous cycle. The data that is obtained from the integration of various related, but at the same time, disseminated sources can create a continuous learning environment facilitated by HIT. It is important to consider that in order to have an LHS, there is a need

to build an infrastructure that can collect data from the health activity carried out, no matter the domain (clinical or research) or the situation where it is generated by taking into account that all data should be captured once, and use for all the purposes where it can be helpful or needed [3].

A recent IOM report provided a big-picture view of the LHS and asserted that implementing standard practices from those of other industries could result in [2]:

- Records immediately updated and available for use by patients,
- Care delivered proven to be “reliable at the core and tailored at the margins”,
- Patient and family needs and preferences being a central part of the decision process,
- All healthcare team members fully informed about each other’s activities in real time,
- Prices and total costs fully transparent to all participants in the care process,
- Incentives for payment structured to “reward outcomes and value, not volume”,
- Errors promptly identified and corrected,
- Outcomes routinely captured and used for continuous improvement.

The IOM report also identifies four “characteristics of a continuously learning healthcare system” [2]. These include:

1. Science and informatics - real-time access to knowledge and digital capture of the entire care experience
2. Patient-clinician partnerships - engaged, empowered patients
3. Incentives - aligned for value with full transparency
4. Culture - instilled by leadership and with supportive system competencies

## Big Data Opportunities

The growing array of data and its importance to health is not limited to the healthcare system. Indeed, when we consider advances in areas such as genomics, imaging, quantified personal data, and more, the sum and synergy of this new digital information is often referred to as Big Data. Big Data has a large potential impact for healthcare since it can contribute to primary and secondary disease

prevention, especially in the current burden of chronic diseases at a population level, monitor the safety of healthcare systems and also help in the implementation of appropriate treatment paths for patients, and support clinical improvement by research [4].

Big Data has different attributes that can be divided into four dimensions (Volume, Variety, Velocity and Veracity) where each one poses a challenge [5]. The first dimension is **Volume**, the amount of data is a challenge since all these sources (healthcare related applications, patients and devices) that need to be integrated generate an amount of data that ranges from terabytes, to petabytes, exabytes, and beyond. In healthcare most of this data is stored in “silos” that do not interoperate with other sources outside the walls of a single institution. The second dimension is **Variety**, since the data come from different resources that are likely not standardized. Such sources may be structured (in the best scenario), but usually they are unstructured or semi-structured with metadata that can help define the meaning of the data. The third dimension is **Velocity**, in which data is generated at a substantial rate and is constantly changing. The final dimension is **Veracity**, or trustability, which requires measures to ensure that the data is reliable, but also that it securely protects the identity (privacy and confidentiality) of the person from where it came [4, 6].

Although many definitions of Big Data emphasize its size, few of them actually quantify what that size is [7]. An alternative definition comes from the new Associate Director for Data Science for the US National Institutes of Health, who notes that Big Data is more about clinical, research, and other health-related organizations making maximal use of all of their data assets [8].

There are many potential benefits in the use of Big Data to improve healthcare. The use of Big Data needs to involve patients as active participants and providers of information to improve not only a given patient’s health but also the health of the population as a whole. It also can transform the way healthcare is delivered since patients may be able to access related information from all the sources (electronic health records, monitoring devices, and self-reported data) integrated as one. This can empower pa-

tients by allowing them to make informed decisions for their well-being. Research can be difficult to carry out, especially in developing economies or concerning orphan diseases or regional endemic diseases. The integration of information can help the provision of observational evidence that can answer clinical, epidemiological and public health questions [9]. There are also, however, challenges to make use of data in operational clinical systems. Data captured during the provision of care may be incorrect, incomplete, of uncertain provenance, and of uneven granularity [10]. Such data is, by definition, operational, and may not be controlled for confounders as might be in a randomized controlled trial [11]. A classic example of this was the Women’s Health Initiative, where a large-scale RCT overturned previous evidence from observational studies on the use of postmenopausal hormone replacement therapy [12]. Nonetheless, there is value for the sheer quantity of data in clinical systems, and informaticians are poised to leverage it [13]. Informatics, defined as the field devoted to the use of data, information, and knowledge to improve health, healthcare, public health, and research [14] is central to the notion of the LHS by capturing, analyzing, and making actionable data from the entire spectrum of care. The IOM report [2] provides a schematic of the healthcare system that allows all critical informatics challenges and opportunities to be enumerated, showing that the overall patient care experience emanates from underlying biomedical science, moving to generating the evidence of how to manage the patient, followed by the delivery of that best care that will ideally result in the optimal patient outcomes and satisfaction. If any of these elements is carried out sub optimally, there are missed opportunities, waste, and harm. Informatics plays a role in each of these elements as well as the transitions between them. Starting with science, informatics increasingly plays a role in both driving and facilitating science. Informatics allows the science to learn from new discoveries in the data and also helps the scientist manage and analyze that data. It helps clinical researchers select the best science to evaluate for evidence. Informatics also allows the best evidence to get implemented as care through

methods such as clinical decision support. It also optimizes the care experience through quality measurement and improvement. In addition, informatics engages not only patients and caregivers but also other providers through health information exchange. Informatics also provides “safety rails” of sorts through maintaining safety, reducing errors, facilitating privacy and security, and promoting adherence to standards. There is really no aspect of informatics that cannot be connected to the IOM LHS schematic.

Of course, vision alone is not enough, and attention must be turned to implementing the LHS. Encouraging studies and reports are already coming forth, such as the learning healthcare system operationalized at Group Health in Seattle [15], coordinated care projects implemented by Medicare to reduce hospital readmissions [16], the “Choosing Wisely” initiative to reduce unnecessary and potential harmful tests and treatments [17], and recent scientific initiative making the vast findings of genomics clinically “actionable” [18].

## Big Data Workforce Development

Big Data needs “Big Data experts” that can provide the integration of data to create value. As noted above, most data at the present time is electronically generated. This requires experts to develop tools that can integrate it from whatever source it comes. Data adhering to standards can help in consolidating the information and natural language processing techniques can identify health-related information from diverse sources that may not be health-related in origin but are part of a patient’s daily life. Finally once all this data is processed, it can provide tailored reports and/or recommendations that can lead to new or improved knowledge that can transform healthcare [19].

Although much has been written extolling the virtues of analytics and Big Data, much less has been said about the human experts who will carry out the work, to say nothing, of those who will support the efforts for building systems to capture data, put it into usable form, and apply the results of analyses.

Many of those who collect, analyze, use, and evaluate data will come from the workforce of biomedical and health informatics. To this end, we must ask questions about the job activity as well as the education of those who work in this emerging area that some call “data science” [20]. Davenport asserts that data analytics is the “sexiest job of the 21st century,” meaning that those who perform it have rare qualities in high demand [21].

In healthcare and biomedicine, the field poised to lead in data science is informatics. After all, informatics has led the charge in implementing systems that capture, analyze, and apply data across the biomedical spectrum from genomics to health care and public health. From basic biomedical scientists to clinicians and public health workers, those who are researchers and practitioners are drowning in data, needing tools and techniques to allow its use in meaningful and actionable ways.

Data science is more than statistics or computer science applied in a specific subject domain. Dhar notes that a key aspect of data science, in particular what distinguishes it from statistics, is an understanding of data, its varying types, and how to manipulate and leverage it [20]. He notes that skills in machine learning are key, based upon a foundation of statistics (especially Bayesian), computer science (representation and manipulation of data), and knowledge of correlation and causation (modeling). Dhar noted a challenge to organizational culture that might occur as organizations moved from “intuition-based” to “fact-based” decision-making.

It is also clear that there are two types of individuals working with analytics and Big Data. As noted in a report by the McKinsey consulting firm, there will soon be a need in the US for 140,000-190,000 individuals who have “deep analytical talent” [22]. Furthermore, the report notes there will be need for an additional 1.5 million “data-savvy managers [needed] to take full advantage of Big Data” [22]. Analyses from the UK find similar results. An analysis by SAS estimated that by 2018, more than 6,400 organizations will hire 100 or more analytics staff [23]. Another report found that data scientists currently comprise less than 1% of all Big Data positions, with more common job roles consisting of developers (42% of

advertised positions), architects (10%), analysts (8%) and administrators (6%) [24]. It was also found that the technical skills most commonly required for Big Data positions as a whole were NoSQL, Oracle, Java and SQL. While these estimates are not limited to health care, they also do not include other countries that have comparable needs to the US and the UK for such talent.

A report from IBM Global Services notes healthcare organizations are lagging behind in hiring individuals who are proficient in both “numerate” and business-oriented skills [25]. An additional report from IBM Global Services lists “expertise” among the critical attributes that are needed in organizations to complement technology. This expertise includes the supplementation of business knowledge with analytics knowledge, establishing formal career paths for analytics professionals, and tapping partners to supplement skills gaps that may exist [26]. Another US-based report by PricewaterhouseCoopers on health IT talent shortages notes that healthcare organizations wanting to keep ahead need to acquire talents in systems and data integration, data statistics and analytics, technology and architecture support, and clinical informatics [27].

The US National Institutes of Health (NIH) also recognizes that Big Data skills will be important for conducting biomedical research. In 2013, NIH convened a workshop on enhancing training in Big Data among researchers [28]. Similar to the healthcare domain, participants called for skills in quantitative sciences, domain expertise, and ability to work in diverse teams. The workshop also noted a need for those working in Big Data to understand the concepts of managing and sharing data. Trainees should also have access to real-world data problems and real-size data sets to solve them. Longer-term training would be required for those becoming experts and leaders in data science.

Data scientists involved in Big Data analytics deal with managing data from heterogeneous sources (aggregation) and preparing customized analyses over massive volume of data. Moving along the adoption curve for Big Data will require development in experience and expertise among users whose skill sets need to cover a wide range from deep technical to deep analytical and their combinations.

The American Society for Training & Development (ASTD) defines a skills gap as a significant gap between an organization's current capabilities and the skills it needs to achieve its goals. It is the point at which an organization can no longer grow or remain competitive because it cannot fill critical jobs with employees who have the right knowledge, skills, and abilities [29].

CompTIA [30] conducted a state of the IT skills gap study among 1061 IT and business managers involved in managing IT or IT staffs for their organizations in Canada, Japan, South Africa, United Kingdom and the United States to gain a better understanding of the IT skills in demand and identify any existing or forthcoming IT skills shortages. The United States portion of the survey on 502 US IT and business managers reveals that eight in ten organizations say their business operations are impacted by gaps in the skill sets of their IT staff, and the dynamic, fast-changing nature of technology, and a lack of training resources is the biggest factor contributing to the skills gap. It further notes that nearly six in ten organizations (57 percent) intend to address their IT skills gap challenges by training or retraining existing staff in areas where skills are lacking.

A more recent version of the study from CompTIA [31] presents an interesting question from a workforce perspective on how best to develop staff equipped to drive Big Data initiatives. Does it make more sense to cross-train IT-centric staff on the business intelligence/analytics/interpretation side of Big Data, or cross-train business-centric staff on the technical side of data management and utilization? The study notes that the ideal position incorporates skills across IT-centric (Data Infrastructure), Business-centric (Interpretation, Visualization and Presentation) as well as Dual IT & Business-centric (Data Management, Processing and Analytics) functional areas.

As part of IBM's Academic Initiative [32], the company is collaborating with more than 1,000 academic partners around the globe to develop curriculum that reflects the mix of technical and problem-solving skills considered necessary to prepare students for Big Data and analytics careers, across industries including health services. As part of this collaboration, IBM would provide schools

with access to IBM Big Data and analytics software, curriculum materials, case study projects, and IBM data scientists to visit classes as guest lecturers. The company also announced awards for Big Data curricula.

The biomedical and health informatics training programs being offered by various universities should explore partnerships and collaborations with industry towards providing electives, modules and specialization tracks on data analytics as part of their coursework. Although it is unknown how many informatics educational programs cover Big Data or its underlying topics, Table 1 shows that a number of recommendations in the Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics cover foundational aspects of Big Data [33]. Clearly educational programs must add coursework in Big Data and related areas, such as data science and data analytics. Outside of healthcare, some

information science programs have begun to chart a course toward this [34].

What do biomedical and health informaticians working in analytics and Big Data need to know? We reviewed the reports cited above in an effort to distill the key skills that are required for informaticians to work with Big Data. Many educational programs already teach aspects of these, although they must be integrated and then combined into a capstone experience that applies them to Big Data problems. We set forth this list of skills and encourage the informatics community to begin discussion on whether they are correct and complete, and what depth is required:

1. **Programming** - especially with data-oriented tools, such as SQL and statistical programming languages
2. **Statistics** - working knowledge to apply tools and techniques
3. **Domain knowledge** - depending on one's area of work, bioscience or health care

**Table 1** Learning outcomes from Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics that cover some aspects of the foundations of Big Data (33)

Knowledge/Skill – Domain	Learning outcomes
(1) Biomedical and Health Informatics Core Knowledge and Skills	1.11 Appropriate documentation and health data management principles including ability to use health and medical coding systems, construction of health and medical coding system 1.12 Structure, design and analysis principles of the health record including notions of data quality, minimum data sets, architecture and general applications of the electronic patient record/electronic health record 1.14 Principles of data representation and data analysis using primary and secondary data sources, principles of data mining, data warehouses, knowledge management 1.16 Ethical and security issues including accountability of health care providers, managers and BMHI specialists, and the confidentiality, privacy and security of patient data 1.17 Nomenclatures, vocabularies, terminologies, ontologies and taxonomies in BMHI
(2) Medicine, Health and Biosciences, Health System Organization	2.7 Health administration, health economics, health quality management and resource management, patient safety initiatives, public health services and outcome measurement
(3) Informatics/Computer Science, Mathematics, Biometry	3.1 Basic informatics terminology like data, information, knowledge, hardware, software, computer, networks, information systems, information systems management 3.3 Ability to communicate electronically, including electronic data exchange, with other health care professionals, internet/intranet use 3.4 Methods of practical informatics/computer science, especially on programming languages, software engineering, data structures, database management systems, information and system modeling tools, information systems theory and practice, knowledge engineering, (concept) representation and acquisition, software architectures 3.5 Methods of theoretical informatics/computer science, e.g. complexity theory, encryption/security 3.10 Mathematics: algebra, analysis, logic, numerical mathematics, probability theory and statistics, cryptography 3.11 Biometry, epidemiology, and health research methods, including study design

4. **Communication** - being able to understand needs of people and organizations and articulate results back to them

To be relevant, informatics educational programs will need to introduce concepts of analytics, Big Data, and the underlying skills to use and apply them into their curricula. There will be a need for appropriate coursework for those who will become the “deep analytical talent” as well as higher breadth, perhaps with lesser depth, for the order of magnitude more individuals who will apply the results of Big Data analytics in healthcare and biomedical research. Health informatics training programs need to identify the knowledge, skills, and resources based on the potential impact that Big Data will have in healthcare as a whole and in biomedical research.

## Conclusion

Advances in biomedical and health informatics provide unprecedented data to inform a health care system that learns from its day-to-day work. The growing quantity and diversity of this data put it into category of Big Data. To effectively use this data, there will be a need for experts trained in biomedical and health informatics who will both perform the analytic work with the data as well as support its collection, optimize its quality, and apply its results to improve health. The informatics field will not only need to develop systems and methods to best utilize this data, but also train the professionals who will perform and lead this work.

## References

- Kayyalil B, Knott D, Van Kuiken S. The big-data revolution in US health care: Accelerating value and innovation. Mc Kinsey & Company; 2013.
- Smith M, Saunders R, Stuckhardt L, McGinnis J. Best Care at Lower Cost: The Path to Continuously Learning Health Care in America: The National Academies Press; 2013.
- Grossman C, McGinnis JM. Digital Infrastructure for the Learning Health System: The Foundation for Continuous Improvement in Health and Health Care: Workshop Series Summary: National Academies Press; 2011.
- Jee K, Kim G-H. Potentiality of big data in the medical sector: Focus on how to reshape the healthcare system. *Healthc Inform Res* 2013;19(2):79-85.
- Feldman B, Martin EM, Skotnes T. Big Data in Healthcare Hype and Hope 2012 [cited 2013 28/11/2013]. Available from: <http://www.west-info.eu/files/big-data-in-healthcare.pdf>.
- Miller AR, Tucker C. Health Information Exchange, System Size and Information Silos. *J Health Econ* 2014;33(1):28-42.
- Ward J, Barker A. Undefined By Data: A Survey of Big Data Definitions. *Databases (csDB)* [Internet]. 2013. Available from: <http://arxiv.org/abs/1309.5821>.
- Bourne P. What Big Data means to me. *J Am Med Inform Assoc* 2014 Mar 1;21(2):194.
- Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA* 2013 Apr 3;309(13):1351-2.
- Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. *Med care* 2013 Aug;51(8 Suppl 3):S30-7.
- Haynes B. Can it work? Does it work? Is it worth it? The testing of healthcare interventions is evolving. *BMJ* 1999 Sep 11;319(7211):652-3.
- Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results From the Women's Health Initiative randomized controlled trial. *JAMA* 2002 Jul 17;288(3):321-33..
- Hersh W, Cimino J, Payne P, Embi P, Logan J, Weiner M, et al. Recommendations for the Use of Operational Electronic Health Record Data in Comparative Effectiveness Research, eGEMs (Generating Evidence & Methods to improve patient outcomes). 2013;1(1).
- Hersh W. A stimulus to define informatics and health information technology. *BMC Med Inform Decis Mak* 2009;9:24.
- Greene SM, Reid RJ, Larson EB. Implementing the learning health system: from concept to action. *Ann Intern Med* 2012 Aug 7;157(3):207-10.
- Brown RS, Peikes D, Peterson G, Schore J, Razafindrakoto CM. Six features of Medicare coordinated care demonstration programs that cut hospital admissions of high-risk patients. *Health Aff* 2012 Jun;31(6):1156-66.
- Cassel CK, Guest JA. Choosing wisely: helping physicians and patients make smart decisions about their care. *JAMA* 2012 May 2;307(17):1801-2.
- Feero W. Determining actionability of genetic findings in clinical practice. *ACP Internist*. 2012 (July/August).
- Larson EB. Building trust in the power of “big data” research to serve the public good. *JAMA* 2013 Jun 19;309(23):2443-4.
- V. D. Data science and prediction. *Communications of the ACM* 2013;56(12):64-73.
- Davenport TH, Patil DJ. Data Scientist: The Sexiest Job of the 21st Century 2012. Available from: <http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>.
- Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, et al. Big data: The next frontier for innovation, competition, and productivity. May 2011. Available from: [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation).
- Big Data Analytics. An assessment of demand for labour and skills, 2012-2017 2013 [cited 2013 9/12/2013]. Available from: [http://www.e-skills.com/Documents/Research/General/Big-DataAnalytics\\_Report\\_Jan2013.pdf](http://www.e-skills.com/Documents/Research/General/Big-DataAnalytics_Report_Jan2013.pdf).
- Big data analytics. Adoption and employment trends, 2012-2017. 2013 [cited 2013 9/12/2013]. Available from: [http://www.e-skills.com/Documents/Research/General/BigDataAnalytics\\_Report\\_Nov2013.pdf](http://www.e-skills.com/Documents/Research/General/BigDataAnalytics_Report_Nov2013.pdf).
- Fraser H, Jayadewa C, Mooiweer P, Gordon D, Piccone J. Analytics across the ecosystem - A prescription for optimizing healthcare outcomes Somers, NY 2013.
- Balboni F, Finch G, Rodenbeck-Reese C, Shockley R. Analytics: A blueprint for value. Converting big data and analytics insights into results 2013 [cited 2013 28/11/2013]. Available from: <http://www-935.ibm.com/services/us/gbs/thoughtleadership/ninelevers/>.
- Solving the talent equation for health IT: PwC Health Research Institute 2013. Available from: <http://pwchealth.com/cgi-local/hregister.cgi/reg/pwc-hri-healthcare-it-staffing-strategies.pdf>.
- Workshop on Enhancing Training for Biomedical Big Data - Big Data to Knowledge (BD2K) Initiative: National Institutes of Health; 2013. Available from: [http://bd2k.nih.gov/pdf/bd2k\\_training\\_workshop\\_report.pdf](http://bd2k.nih.gov/pdf/bd2k_training_workshop_report.pdf).
- ASTD. Bridging the Skills Gap 2012 [cited 2014 01/24/14]. 52]. Available from: [http://nist.gov/mep/upload/Bridging-the-Skills-Gap\\_2012.pdf](http://nist.gov/mep/upload/Bridging-the-Skills-Gap_2012.pdf).
- CompTIA. State of the IT Skills Gap 2012 [cited 2014 01/24/2014]. Available from: [http://www.wired.com/wiredenterprise/wp-content/uploads/2012/03/Report\\_-\\_CompTIA\\_IT\\_Skills\\_Gap\\_study\\_-\\_Full\\_Report.sflb.pdf](http://www.wired.com/wiredenterprise/wp-content/uploads/2012/03/Report_-_CompTIA_IT_Skills_Gap_study_-_Full_Report.sflb.pdf).
- CompTIA. Big Data Insights and Opportunities 2013 [cited 2014 01/24/2014]. Available from: <http://www.euroanresearch.com/otherUploaddocs/2ndBigData.pdf>.
- IBM. IBM Narrows Big Data Skills Gap By Partnering With More Than 1,000 Global Universities 2013. Available from: <http://www-03.ibm.com/press/us/en/pressrelease/41733.wss>.
- Mantas J, Ammenwerth E, Demiris G, Hasman A, Haux R, Hersh W, et al. Recommendations of the International Medical Informatics Association (IMIA) on Education in Biomedical and Health Informatics. First Revision. *Methods Inf Med* 2010 Jan 7;49(2):105-20.
- Dumbill W, Liddy E, Stanton J, Mueller K, Farnham S. Educating the next generation of data scientists. *Big Data* 2013;1(1): BD21-BD7.

### Correspondence to:

Dra. Paula Otero  
 Department of Health Informatics  
 Hospital Italiano de Buenos Aires  
 Peron 4190  
 (1199) Ciudad Autonoma de Buenos  
 Argentina  
 E-mail: [paula.otero@hiba.org.ar](mailto:paula.otero@hiba.org.ar)