

Big³. Editorial

Christoph U. Lehmann¹, Brigitte Séroussi^{2,3}, Marie-Christine Jaulent³

¹ Departments of Pediatrics and Biomedical Informatics, Vanderbilt University, Nashville, TN, USA

² Sorbonne Universités, UPMC Université Paris 06, UMR_S 1142, LIMICS, Paris, France; AP-HP, Hôpital Tenon, DSP, Paris, France

³ INSERM, UMR_S 1142, LIMICS, Paris, France

Summary

Objectives: To provide an editorial introduction into the 2014 IMIA Yearbook of Medical Informatics with an overview of the content, the new publishing scheme, and upcoming 25th anniversary.

Methods: A brief overview of the 2014 special topic, Big Data - Smart Health Strategies, and an outline of the novel publishing model is provided in conjunction with a call for proposals to celebrate the 25th anniversary of the Yearbook.

Results: 'Big Data' has become the latest buzzword in informatics and promise new approaches and interventions that can improve health, well-being, and quality of life. This edition of the Yearbook acknowledges the fact that we just started to explore the opportunities that 'Big Data' will bring. However, it will become apparent to the reader that its pervasive nature has invaded all aspects of biomedical informatics – some to a higher degree than others. It was our goal to provide a comprehensive view at the state of 'Big Data' today, explore its strengths and weaknesses, as well as its risks, discuss emerging trends, tools, and applications, and stimulate the development of the field through the aggregation of excellent survey papers and working group contributions to the topic.

Conclusions: For the first time in history will the IMIA Yearbook be published in an open access online format allowing a broader readership especially in resource poor countries. For the first time, thanks to the online format, will the IMIA Yearbook be published twice in the year, with two different tracks of papers. We anticipate that the important role of the IMIA yearbook will further increase with these changes just in time for its 25th anniversary in 2016.

Keywords

Editorial, 2014 IMIA Yearbook of Medical Informatics, open access online format, Big Data, survey of biomedical informatics, IMIA and its societies

Yearb Med Inform 2014: 6-7

<http://dx.doi.org/10.15265/IY-2014-0030>

Published online May 22, 2014

Big Changes

In the history of the IMIA Yearbook, 2014 will be a special year with a number of highly anticipated changes and one farewell. When you are reading this, you probably already know that we are publishing the IMIA Yearbook in an online open access format. For the first time in more than two decades of publishing the Yearbook, all papers are accessible to anyone with an Internet browser – free of charge and courtesy of the International Medical Informatics Association.

The IMIA Yearbook had been published in an electronic version since 2006, but access was restricted to member societies and institutional members, who paid an extra fee for the privilege of accessing cutting edge keynotes, surveys, reviews, and history of Medical Informatics, research and education papers. Limiting access to paying members only had significant disadvantages. Especially in resource poor member countries, the ability to access the content of the Yearbook was severely limited. Because of access restrictions, less researchers and scientists were able to read the Yearbook papers, and less cited them. Offering open access enables us to provide the valuable content to all IMIA member countries (and others) and will allow contributors to the Yearbook see an increase in readership and citations. The Yearbook editors applaud and thank the IMIA General Assembly for their wise decision in this matter.

With open access, the IMIA Yearbook loses a valuable source of income – the contribution of member societies, who had subscribed to the Yearbook. To make up for some of those losses, IMIA decided

to no longer publish the Yearbook in a paper version. We are a little sad to see the print version go, but it was an acceptable sacrifice in order to provide open access to the Yearbook.

Another change in 2014 will be a staged publication of the Yearbook. No longer encumbered by the process of printing the Yearbook, the editors have planned to make content available to you as it becomes available. This will lead to a more timely publication and earlier access to the content of the Yearbook. So, what will you find in this early partial release of the Yearbook? We are excited to publish the lecture that Prof. Reinhold Haux gave to the IMIA community in connection with receiving the IMIA Award of Excellence at MedInfo 2013 (Copenhagen). We were able to solicit and publish responses to his lecture by leading biomedical informatics researchers, who insightfully reviewed and commented the document. In this early edition we also publish two History-of-Medical-Informatics papers, and the contribution by Turkey on healthcare information technology infrastructures as a research-and-education paper. These first contributions are accompanied by a keynote authored by Prof. Riccardo Bellazzi, which discusses the fundamentals of how Big Data is enabling unprecedented research studies and new models of healthcare delivery. The second track of articles including surveys, best papers, and working group contributions distributed among the ten usual sections (Special section Big Data – Smart health strategies, Health and Clinical Management, Human Factors and Organizational Issues, Health Information Systems, Sensor, Signal and Imaging Informatics, Deci-

sion Support, Knowledge Representation and Management, Education and Consumer Health Informatics, Bioinformatics and Translational Informatics, Clinical Research Informatics) will be available along with IMIA regional information in August 2014.

Big Data

In the 1992 book “Snowcrash”, a cyberpunk novel by Neal Stephenson, a future is described where a company evolved into the data business – it collects any data that can be obtained, stored, analyzed, and sold. Twenty-two years after the first publication of this novel, this science fiction company has found multiple Doppelgängers in reality. No matter, where we look, companies (and government agencies) are collecting, storing, and analyzing data to better understand our choices and preferences, better predict or affect our behavior.

“There was [sic] 5 exabytes of information created between the dawn of civilization through 2003,” Google CEO Eric Schmidt said at the 2010 Technonomy conference, “but that much information is now created every 2 days, and the pace is increasing...”. We can no longer deny that we are witnessing an explosion of data, mostly user-generated, and are entering the era of Big Data. Big Data promises that exploiting vast amounts of information will help us discover new knowledge, and better understand the world. Big Data is usually characterized by four Vs. The first one is Volume. Statisticians “like to argue, big data isn’t much more than a sexier version of statistics”, as Samuel Arbesman reported in the Washington Post of August 16th 2013. But there is much more in Big Data because of Velocity, making classical data analyses irrelevant and creating a need for new software, algorithms, and architectures to better exploit computing power. The third V is Variety. Data have become more diverse including motion, location,

physiologic, and behavioral data as well as data from devices we are interacting with like refrigerators and gym equipment in addition to “traditional” big data from social networks, online shopping, and browsing. In EHRs, data may be expressed either as numerical values, structured or unstructured texts, images, videos. This variety of formats needs proper management for the data to be correctly extracted and analyzed. The fourth V is Veracity or the question of data reliability. Can the probably poor quality of data collected (always incomplete, and sometimes erroneous) be balanced by the quantity? Big Data is sometimes considered profoundly different, and even antagonistic, to science, including epidemiology. Epidemiology works on assumptions that clinical trials seek to reject. Big Data works on existing data, usually collected with no research objective, sometimes without any health care objective, and even without the notion that they would be collected and analyzed by someone one day. One example is the realtime analysis of Google queries used in Google Flu Trends to precociously detect and monitor the evolution of influenza epidemics. In Big Data, the principle is to mine data for relevant correlations, interpret them, and provide the “why”, and then conduct the survey that should bring the “because”.

Big Data promises the detection of weak signals: emerging infectious diseases, drug side effects and other undesired outcomes, where classical epidemiology proved to be inefficient. Big Data should also be effective to process mass data accumulated by genomics and proteomics, where genetic epidemiology had prudently given way to the new bioinformatics. Finally, Big Data would probably help to process data collected in the field of environmental health (climate, pollution, pollen, satellite data, etc.). It requires little imagination to predict that amidst the cacophony and flood of data, valuable conclusions can be made, behavioral and health behaviors be predicted with the goal of devising new approaches and interventions that can improve

health, wellbeing, and quality of life. But all this can happen only on the condition that Big Data can be deployed in conjunction with Open Data. By opening the access to medical databases to the greatest number of researchers, with guaranteed de-identification, we could save precious time on early warning and safety improvements.

This edition of the Yearbook acknowledges the fact that we just started to explore the opportunities that ‘Big Data’ will bring, however it will become apparent to the reader that its pervasive nature has invaded all aspects of biomedical informatics – some to a higher degree than others. It was our goal to provide a comprehensive view at the state of ‘Big Data’ today, explore its risks and weaknesses, discuss emerging trends, tools, and applications, and stimulate the development of the field through the aggregation of the excellent contributions to this Yearbook. We hope that you will enjoy reading the manuscripts as much as we did in preparing for this edition.

Big Birthday

In two years, the IMIA yearbook will celebrate its 25th edition. Since its beginnings in 1992 with Jan van Bommel and Alexa McCray at the helm a number of prestigious editors have shaped the content and the style of the Yearbook and this editorial team is proud to follow in the footsteps of these giants (Reinhold Haux, Casimir Kulikowski, Antoine Geissbuhler). We invite our readers to mark this silver anniversary through contributions and suggestions. While the Yearbook usually does not publish letters to the editor, we will entertain publishing your stories about the impact that the Yearbook had on your life or what happened when you contributed to it. We want to hear from you and learn what has mattered to you over the years and how you envision the next 25 years of the Yearbook. Come help us prepare for a big birthday.