Nursing Informatics 2014
K. Saranto et al. (Eds.)

doi:10.3233/978-1-61499-415-2-470

Can Big Data Transform Electronic Health Records Into Learning Health Systems?

Ellen HARPER
Cerner Corporation, Kansas City, MO, USA

Abstract. In the United States and globally, healthcare delivery is in the midst of an acute transformation with the adoption and use of health information technology (health IT) thus generating increasing amounts of patient care data available in computable form. Secure and trusted use of these data, beyond their original purpose can change the way we think about business, health, education, and innovation in the years to come. "Big Data" is data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it.

Keywords. Big data, Health IT, EHR, electronic health record, learning health system, digitization, datafication

Introduction

In the United States (U.S.) and globally, healthcare delivery is in the midst of an acute transformation with the adoption and use of health information technology (health IT) thus generating increasing amounts of patient care data available in computable form. Secure and trusted use of these data, beyond their original purpose can change the way we think about business, health, education, and innovation in the years to come. "Big Data" is data whose scale, diversity, and complexity require new architecture, techniques, algorithms, and analytics to manage it and extract value and hidden knowledge from it [1]. Big Data refers to the ability to process mountains of data, provide instantaneous knowledge to the clinicians at the point of care and draw sometimes astonishing conclusion from it. The Institute of Medicine published on the characteristics and journey of a "Learning Healthcare System". The quick summary is a system that is continuously innovating and improving in its ability to deliver high value to patients with continuous learning, best evidence based care, and lower costs [2]. Innovations are needed in hardware, software, technology and expertise.

Big Data is all about seeing and understanding the relations within and among pieces of information that, until recently, we struggled to fully grasp. The ability to learn from the data is dependent on our ability to turn the data into a searchable form and use computing power to discover insights that we never could have seen before. Most countries, including the US, lack integrated electronic health records (EHR) [3-5]. Efficient and seamless access to standards-based clinical information is important to

Corresponding Author: Ellen Harper, Cerner Corporation, 2000 Rockcreek Parway, Kansas City, MO 64117: eharper@cerner.com

support informed clinical decisions at the point of care, while correspondingly large data sets can provide the knowledge to continuously improve and innovate [6]. The use of standards facilitates use of clinical decision support capabilities, which are often triggered based on structured and coded information such as the assessment data, diagnosis/problem list, medication list, and lab results to name a few. Using data elements from the EHR for purposes beyond clinical documentation, billing, and administration is a rapidly growing practice [7]. Secondary use of data initially meant for clinical documentation, billing, or administrative purposes could become a healthcare system's most valuable asset [8,9]. We are in the beginning stages of transforming EHR data into learning health systems. The discovery and predictive analytics of modeling and forecasting may explain why the value of Big Data for practice, economics, and research tend to be under-recognized today. This paper will focus on the elements that distinguish Big Data and some practical uses currently being implemented today.

1. Elements to Distinguish Big Data

Data are no longer regarded as static or stale, whose usefulness if finished once the purpose for which it was collected was achieved. Rather, data has become a raw material used to create innovation. Historically Big Data tended to be generated as a result of economic activities. Today we focus on the relationship of data to individuals (person centered) and individual's movements, preferences, sentiments, and decisions. The individualization of Big Data stems from self-measurement mobile devices for health and other purposes, as well as societal data from credit/debit cards and social media. Property rights and ownership of data that are the by-product or residue of people's ordinary activities and that have financial value are a new frontier.

Raw data can alternately be viewed like clean air and, in many respects, they are public goods. But portions of them can also be owned, just like other commodities can be owned. Those whose data they are should be fairly compensated when others use the data and derive financial value from that usage.

Volume – Moving From	**Value – Moving to**
Retrospective	Prospective
Individuals	Groups, Population
Individual Data Capture	Meta, Relational
Acute Diseases	Chronic Diseases
Phenotype	Genotype
Simple linear (chief complaint)	Complex Factors

Figure 1. Demonstrates the change in focus from volume to value.

Computer systems build their clinical decision support (CDS) using rules that have been explicitly programed to run a program. Big Data operates at a scale that transcends our ordinary thinking. For example, individual patient elements are recorded when the patient is admitted and ongoing data is entered by nurses as part of their EMR documentation of nursing assessments and observations. Each individual patient's elements can be matched to a computerized database and software is programmed to run an algorithm that examines, processes, and relates this data providing specific recommendations for treatments, interventions, and care. Similar to rules built to detect a patient's early signs and symptoms of sepsis and alerting the rapid response team, there is the potential to run hundreds of rules simultaneously to identify patients "like this" also have potential for kidney shutdown. Demonstrating the importance of mathematical analysis of data in clinical computing, machine learning has made both a quantitative and a qualitative difference in how and when data are used and interpreted. Increasingly, the uses pertain to forecasting or predicting what is likely to happen, and altering decisions at the point of care in light of that. Similar to accountants managing the financials, there will be "algoroithmists" who are experts in computer science, mathematics, and statistics who will act as reviewers of the Big Data analysis and predictions [1].

Big Data also has to do with the relationship of data to groups and its utilization for population level policy-setting and services provisioning [10]. Governance of management of data that are sourced by various individuals and groups in different legal jurisdictions presents some new issues with regard to justice and ethics, plus many old and familiar issues that are massively more frequent than before. Surveillance and data mining that are nominally for the purpose of research and public health can potentially be subverted to other unstated purposes that potentially can infringe on privacy [11].

2. Digitization verse Datafication

There has been a multi-decade journey to digitize all of the health information. A common miss-understanding is that "digitize" means adoption to automate the paper to achieve a paperless EHR environment. Yet this has been a Health IT oriented target, with no mention of quality, cost, efficiency, and elimination of errors. Health IT involves more than storing the record, measurements and getting paid. Just because information can be exchanged electronically using standards like HL7, it does not mean it can be digitally analyzed. Data must be structured using electronic standards for health information management professionals and clinicians to analyze and use data for critical decision making [9]. Digitization is scanning a paper document into a high resolution digital file that can be stored and accessed electronically. Figure 2 shows an ICU flowsheet that was scanned into the EHR. The paper chart has been digitized, as an attempt to get the entire patient's information into one place and remove the paper record. This is beneficial to have a single source of truth yet it was only as good as the clinician reading it to uncover the question they seeking to answer.

Datafication is when words become data that is machine readable [1]. Once the data are readable, it unleashes the possibility to transform their purpose and turn the information into new forms of knowledge and value. Once every assessment, dictated note or scanned document is processed through a natural language processor and stored with machine readable formats the possibilities are only limited by our imagination.

Figure 3 demonstrates how a patient's entire EHR can be searched for hypertension. Not only by simple text search on the term hypertension or the concepts (high blood pressure) but also by the related concepts such as treated by (ace inhibitor, low salt diet) , monitored by (pulse pressure), and measured by (orthostatic blood pressure). Industry has been doing this for years and now healthcare is slowing moving in this direction.

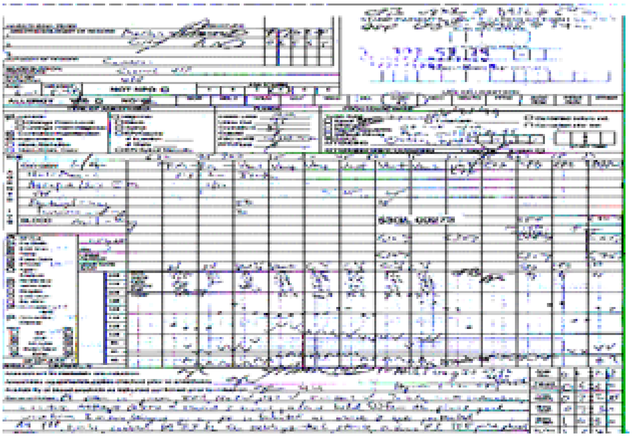

Figure 2. Scanned ICU Flowsheet document

Figure 3. Google-like search of a patient's electronic health record for hypertension

3. Scale (Volume), Complexity (Varity) and Speed (Velocity)

Big Data processes must address the structure of the data, with fusion of different data types from disparate sources. Prior to 1999, most of the data available for analysis was discreet and structured. Now it is substantially unstructured or hybrid/semi-structured and varies in terms of format and dimensionality [10]. The use of standards and nomenclature mapping process with a high degree of quality assurance or, more broadly, ontology management is far more important than it historically was given (a) the disparities in structure and content of the data now and (b) the fact that the use-cases are now predominantly decision-support oriented, many of them real-time [10].

Much work has been done to scale to the processes and storage needs of Big Data, easily moving from dozens of terabytes into the petabyte range. Scaling from a sample size sufficient to support point of care decision-support discovery analyses, while also providing sufficient power to statistically power actionable accurate predictive models for reliable forecasting and planning models for populations, is one of the main benefits of Big Data in health care. But there is a point at which conventional approaches to storage (e.g., copies of files) may stop scaling. Novel approaches to storage replication and smart storage-retrieval engineering may help, such as the use of mathematical techniques for distributing elements of data sets and then transporting them or recreating them as needed. Challenges are: real-time decision-support use-cases (sub-second timescale response) and retrospective analysis of an increasing amount of data over a larger and larger period of historical time. The velocity or rate of accumulation of new data is as technically challenging. Which is why parallelization (spreading the load) among dozens to thousands of servers is necessary, along with hybrid index preprocessing and other techniques.

4. Conclusion

There is a quiet confidence that nurses' use of Big Data will continue to improve as a collective, not only for themselves or the communities they serve, but as a model for the industry to learn from. This is inherent in the profession of nursing's mission. Nurses will be at the front of the line, facing obstacles and creating solutions. Understanding that data from an assessment is important to the well-being of patient care, the greater potential of a learning health system will provide new knowledge to the management of populations and should receive the attention to detail that the nursing profession is known for. All the digital bits that have been gathered can now be harnessed in a novel ways to serve new purposes and unlock new forms of value. This will require a new way of thinking and will challenge our institutions and even our sense of identity. The one certainty is that the amount of data will continue to grow, as will the power to process it all. But where most people have considered Big Data as a technological matter focusing on the hardware or the software, the emphasis needs to shift to what happens when the data speaks.

References

[1] V. Mayer-Schonberger, K. Cukier, *Big Data a Revolution That Will Transform How We Live, Work and Think*. Boston: Houghton Mifflin Harcourt, 2013.

[2] Institute of Medicine. *The learning health care system in America*. (2012). Retrieved from http://www.iom.edu

[3] A.K. Jha, C.M. DesRoches, G. Compbell, et. al., Use of electronic health records in U.S. hospitals. *N Eng J Med* **306** (2009). Retrieved from http://www.evernote.com

[4] A. Samarth, *Interoperability for Dummies*. (IHE ed.), 2013. Retrieved from http://www.iheusa.org/ebook.aspx

[5] J. Walker, E. Pan, D. Johnston, J. Adler-Milstein, e.al., The value of health care information exchange and interoperability. *Health Affairs*. doi: 10.1377/hlthaff.w5.10 (2005, January 19).

[6] K.H. Bowles, S. Potashnik, S.J. Ratcliffe, M. Rosenberg, Conducting Research Using the electronic health record across multi-hospital systems [Semantic harmonization implications for administrators]. *J Nurs Admin* 43:6 (2013), 355-360.

[7] K. Hayrinen, K. Saranto, P. Nykanen, Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform*, **77**:5(2008), 291-304.

[8] C. Safran, M. Bloomrosen, W.E. Hammond, et al., Toward a national framework for the secondary use of health data: an American Medical Informatics Association white paper. *JAMIA* **14**:1(2007), 1-9.

[9] E.M. Harper, The economic value of healthcare data. *Nursing Administration*, **37**:2(2013), 105-108.
[10] D. McNair, Six elements that distinguish 'Big Data' from earlier eras' data gathering and utilization. 2013. [Discussion comment]. Retrieved from Cerner.com
[11] C.P. Friedman, A.K. Wong, D. Blumenthal, Achieving a nationwide learning health system. Science Translational Medicine, **2**:57 (2010, November 10). Retrieved from http://rds.epi-ucsf.org