



Big Data as a Driver for Clinical Decision Support Systems: A Learning Health Systems Perspective

Arianna Dagliati^{1,2}, *Valentina Tibollo*¹, *Lucia Sacchi*³, *Alberto Malovini*¹, *Ivan Limongelli*⁴, *Matteo Gabetta*⁵, *Carlo Napolitano*⁶, *Andrea Mazzanti*⁶, *Pasquale De Cata*⁷, *Luca Chiovato*⁷, *Silvia Priori*⁶ and *Riccardo Bellazzi*^{1,3*}

¹ Laboratorio Informatica Sistemistica Ricerca Clinica, Istituti Clinici Scientifici Maugeri, Pavia, Italy, ² Division of Informatics, Imaging & Data Sciences, Manchester Molecular Pathology Innovation Centre, University of Manchester, Manchester, United Kingdom, ³ Dipartimento di Ingegneria Industriale e dell'Informazione, Università degli Studi di Pavia, Pavia, Italy, ⁴ Engenome s.r.l., Pavia, Italy, ⁵ Biomeris s.r.l., Pavia, Italy, ⁶ Molecular Cardiology, Istituti Clinici Scientifici Maugeri, Pavia, Italy, ⁷ UO di Medicina Interna e Endocrinologia, Istituti Clinici Scientifici Maugeri, Pavia, Italy

OPEN ACCESS

Edited by:

Pierpaolo Cavallo,
Università degli Studi di Salerno, Italy

Reviewed by:

Gokarna Sharma,
Kent State University, United States
Amar Koleti,
University of Miami, United States

*Correspondence:

Riccardo Bellazzi
riccardo.bellazzi@unipv.it

Specialty section:

This article was submitted to
Big Data Networks,
a section of the journal
Frontiers in Digital Humanities

Received: 30 January 2018

Accepted: 09 April 2018

Published: 01 May 2018

Citation:

Dagliati A, Tibollo V, Sacchi L, Malovini A, Limongelli I, Gabetta M, Napolitano C, Mazzanti A, De Cata P, Chiovato L, Priori S and Bellazzi R (2018) Big Data as a Driver for Clinical Decision Support Systems: A Learning Health Systems Perspective. *Front. Digit. Humanit.* 5:8. doi: 10.3389/fdigh.2018.00008

Big data technologies are nowadays providing health care with powerful instruments to gather and analyze large volumes of heterogeneous data collected for different purposes, including clinical care, administration, and research. This makes possible to design IT infrastructures that favor the implementation of the so-called “Learning Healthcare System Cycle,” where healthcare practice and research are part of a unique and synergic process. In this paper we highlight how “Big Data enabled” integrated data collections may support clinical decision-making together with biomedical research. Two effective implementations are reported, concerning decision support in Diabetes and in Inherited Arrhythmogenic Diseases.

Keywords: big data, learning health care cycle, data warehouses, data integration, data analytics

INTRODUCTION

Following a broadly recognized definition, 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” (Harper, 2014). This definition embraces the multifactorial nature of this kind of data, and the technological challenges implied. The integration of different sources of information, from primary and secondary care to administrative data, seems a substantial opportunity that Big Data provides to healthcare (Murdoch and Detsky, 2013; Etheredge, 2014; Halamka, 2014; Krumholz, 2014; Zillner et al., 2014). Such integration may allow depicting a novel view of patient’s care processes and of single patient’s behaviors taking into account the multifaceted aspects of clinical and chronic care.

The interest in the collection of large and heterogeneous healthcare data sources finds a distinctive application in the definition of novel data-driven Decision Support Systems (Kaltoft et al., 2014; Kohn et al., 2014; Lupse et al., 2014). Several authors (Kohn et al., 2014; Lupse et al., 2014; Zhang et al., 2016) define two main fields where researchers should address their efforts to produce valuable results in this area: (i) the secondary use of data to create new evidence and glean important insights to make better clinical decision or to reshape health care organizational components and (ii) the detection of novel correlations from asynchronous events to allow clinicians to promptly identify potential complications, timely adjust treatments, or help analyze similar manifestations in clinical diagnoses. To pledge better renewed decision making, and consequent successful clinical outcomes, Big Data enabled health care systems should effectively integrate advanced computational tools, such as novel similarity measures for patients’ stratification

and predictive analytics for risk assessment and selection of therapeutic interventions (Moghimi et al., 2013; Suresh, 2014; Yun and Hui, 2014; Wang et al., 2015).

The availability of new data sources is thus leading to the development of a novel model of healthcare, able to fully exploit the potentials of data-driven decision making. The main consequence is that Big Data will not only be an important enabler for research, but also for the clinical and organizational decision making. We will discuss this perspective within the context of the so-called “Learning Healthcare System Cycle” (Grossmann et al., 2011; Skiba, 2011; Deeny and Steventon, 2015; Budrionis and Bellika, 2016; Harle et al., 2016; Lee and Yoon, 2017). We demonstrate the importance of leveraging on “Learning Healthcare System Cycle” solutions in developing next-generation clinical and organizational decision system, as follow: (i) we describe a possible formalization for the use of the Learning Healthcare Systems, proposing a conceptual solution based on state-of-the-art technologies for data production. (ii) We present two systems implanted upon the Learning Healthcare Cycle as proof of concepts of the validity of the formalized concepts in different clinical scenarios.

BIG DATA AND THE LEARNING HEALTHCARE SYSTEM CYCLE

Novel and essential directions in the use of Big Data for health care have been recently redefined within the medical informatics community (Tenenbaum et al., 2016). Specifically, the well-known conceptual approach of the “data, information, and knowledge” continuum has been reconsidered as the so-called Learning Healthcare System Cycle (LHSC), where healthcare practice and research should be part of a unique and synergic process. The first main novelty of this approach is to emphasize that clinical practice and research are complementary agents in the generation of data and knowledge.

The role of informatics is to provide the right tools to turn data into information, and information into knowledge, helping to understand deep data relations by retrieving and extracting underlying patterns. Moreover, informatics is crucial for the deployment of the acquired knowledge to support patient care and ultimately to guide individual behavior.

Our prospective is that the use of Big Data in medical informatics will be equally important in the different phases of the LHSC: from research to data driven decision-making. LHSC is indeed based on these two complementary actions, the first one focused on the exploitation of medical generated data for research purposes (Care Informs Research), and the second one focused on the development of novel systems leveraging on Big Data to guide clinical decision making (Research Informs Care).

Care Informs Research—Research

In clinical practice, data is mostly collected from electronic health records (EHR), which recent widespread adoption has made available a unique source of clinical information for research. The EHR can be used to extract and interpret clinical data, to automatically support clinical research and improve

quality of care. Specifically, EHR-based phenotyping uses data captured in the delivery of care to identify individuals or cohorts with conditions or events relevant to clinical studies (Hripcsak and Albers, 2013; Newton et al., 2013; Richesson et al., 2013). Some distinguishing aspects of the current literature include: (i) considering the temporal nature of the data, and explicitly including not only clinical information from EHR, but also process information from administrative databases; recent methods allow for example the extraction of care-flows that highlight frequent patients’ trajectories in terms of disease evolution but also in terms of patterns of care; (ii) compute patients’ similarity by resorting to advanced “multimodal” data fusion strategies, including deep learning and tensor factorization; (iii) fully apply natural language processing pipelines as enablers to integrate in the analytical process data and knowledge hidden in textual reports.

Research Informs Care—Data Driven Decision Making

Clinical decision support systems (CDSS) have been traditionally defined as software designed to aid clinical decisions making by adapting computerized clinical guidelines and protocols to individual patient characteristics (Sim et al., 2001). While it is recognized that developing and deploying CDSSs can be very beneficial in contexts that require complex decision-making, such as chronic disease management, their use in routine clinical practice is currently still limited (Belard et al., 2016). Possible causes are related to poor user interfaces, lacking integration with EHRs, and limited analytics capabilities that do not allow data-driven reasoning.

We believe that, in order to provide successful decisions support, CDSSs should comply with basic requirements including: (i) rich contents in terms of knowledge, references, and data evidences, (ii) the capability of processing huge amounts of data with fast response times, and (iii) implementations that are intuitive, appealing, and able to catch users’ attention while not delaying clinical actions. These features translate into the fundamental CDSS components: data and knowledge repositories, inference engines, and user interfaces. It is worth noticing that IT infrastructures designed to support research can be also used to assist clinical decision-making. An interesting paradigm is represented by the so-called “sidecar” approach, where the same data warehouse is used to analyze patients’ cohorts at a population level and as an instrument to enable “case-based” reasoning in front of a complex clinical case by extracting similar patients and potential treatments.

USE OF BIG DATA FOR CLINICAL DECISION SUPPORT: AVAILABLE SOLUTIONS AND SYSTEMS

Several conceptual design elements and software components are nowadays available to support the construction of systems for implementing LHSC.

Some well-known current initiatives and networks to support Big Data research include NIH’s Big Data to Knowledge

(BD2K) projects (Bourne et al., 2015), eMERGE (McCarty et al., 2011), and PCORNet (Fleurence et al., 2014). BD2K is an extensive funding initiative that encompasses several aspects of the enhancement of Big Data in biomedical research: from accessibility and reusability of data, to the development of novel methodologies and tools for analyzing Big Data. The eMERGE network serves to develop and share high-throughput clinical phenotyping algorithms in support of precision medicine. It includes several tools, like PheKB, a collaborative knowledgebase for phenotype discovery and validation. In the light of clinical decision support, the eMERGE network proposed the use of infobuttons (Cimino et al., 2007, 2013) as a decision support tool to provide context specific links within electronic health records to relevant genomic medicine content. PCORNet is aimed at improving the capacity to conduct comparative clinical effectiveness research thanks to patient-centered common data models. This data model leverages standard terminologies and coding systems for healthcare (including ICD, SNOMED, CPT, HCPSC, and LOINC) to enable interoperability with and responsiveness to evolving data standards. Examples of applications to chronic diseases include the use of PCORNet (McGlynn et al., 2014) to create a common data model for patients affected by metabolic diseases, or of eMERGE to secondary data analysis for personalized medicine and phenotype definition in Type 2 Diabetes (Yazdanpanah et al., 2013; Hall et al., 2014).

One of the most widely used open source tools to collect multidimensional data by aggregating different sources is the Informatics for Integrating Biology and the Bedside (i2b2) framework (<https://www.i2b2.org>).

i2b2 is one of the seven centers funded by the U.S. National Institute of Health (NIH) Roadmap for Biomedical Computing (<http://www.ncbcs.org>). The mission of i2b2 is to provide clinical investigators with a service-based software infrastructure able to integrate clinical records and research data, and easily query them. To facilitate the query process, data are mapped to concepts organized in an ontology-like structure. i2b2 ontologies aim at organizing concepts related to each data stream in a hierarchical structure. For example, drug prescriptions can be represented through their ATC drug codes in the Drug Ontology (Dagliati et al., 2014), or subset of Laboratory test from Anatomy Pathology can be linked to the SNOMED (Systematized Nomenclature of Medicine) Ontology (Segagni et al., 2012a). Furthermore, i2b2 is linked to ontologies available from BioPortal (<http://i2b2.bioontology.org/>) in order to integrate the most common medical ontologies into the system.

Since it was developed, i2b2 framework involved other parallel projects. The interoperability project “Substitutable Medical Applications and Reusable Technologies” (SMART) was devoted to develop a platform that allows medical applications to be written once and then run across different healthcare IT systems (Mandl et al., 2012). SMART was lately updated to take advantage of the clinical data models and the application-programming interface described in a new, openly licensed Health Level Seven (HL7) draft standard called Fast Health Interoperability Resources (FHIR). The new platform is called SMART on FHIR

(Mandel et al., 2016), and it has been recently exploited to build an interface that serves patient data from i2b2 repositories (Waghlikar et al., 2016). I2b2/SMART can thus effectively implement the sidecar approach, which allows to continue using existing clinical system (EHR) as-is while resorting to a secondary database (the i2b2 instance) for decision making.

When Big Data are specifically exploited for CDS, visual analytics enables hypothesis generation and facilitates real-time clinical decisions (Ola and Sedig, 2014; Vaitsis et al., 2014; Simpao et al., 2015). Visual analytics can be a powerful tool if used in combination with longitudinal models to analyze long time series (Mane et al., 2012; Gálvez et al., 2014) and to enhance pattern visualization to focus attention in monitoring clinical actions (Simpao et al., 2014; Cánovas-Segura et al., 2016), or to detect and show patients’ behaviors to identify health-risk scenarios (Juarez et al., 2015).

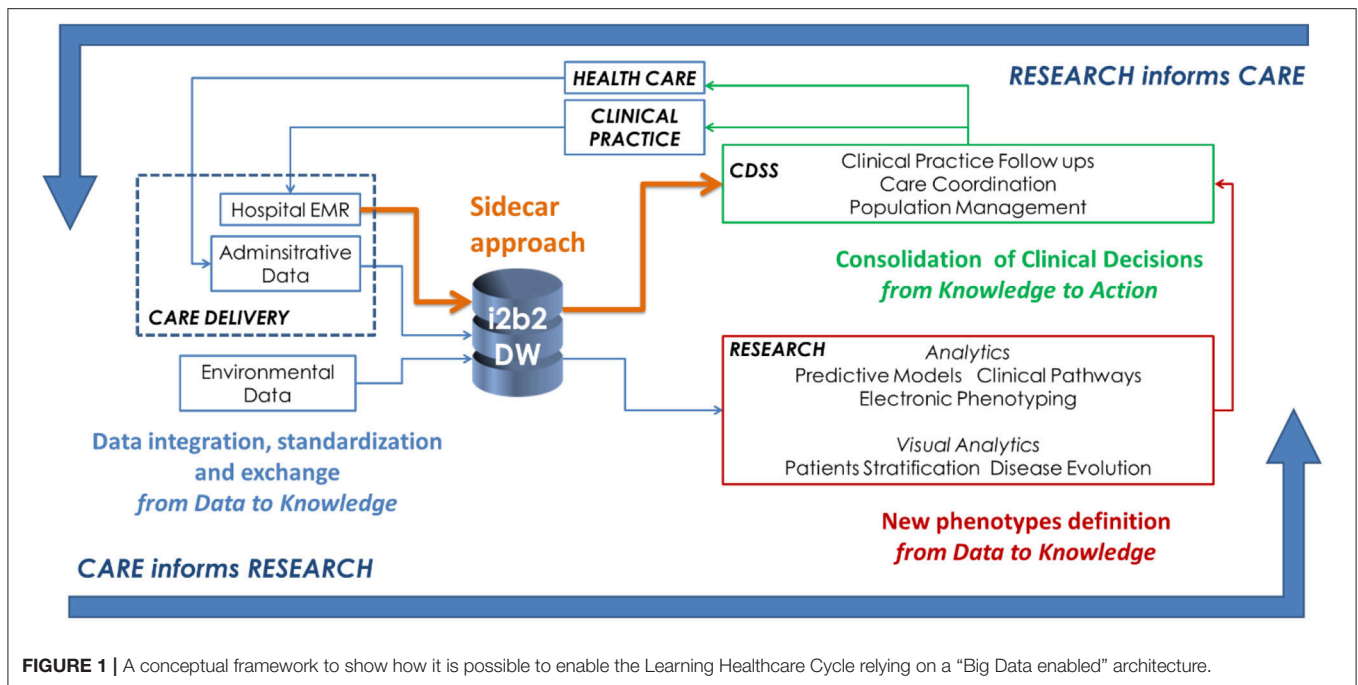
There are also several examples of CDSSs where visual analytics methods combine evidence-based and data-driven approaches to improve clinical performances, for example by retrieving drug interactions (Resetar et al., 2005; Simpao et al., 2014), by combining analytics and electronic guidelines (Slonim et al., 2012) or by gathering EHR data and entering them into models able to perform risk stratification (Gotz et al., 2012). There are attempts to use visual analytics into field of epidemiology to understand the interaction among time dependent variables (Chui et al., 2011).

BUILDING EFFECTIVE CLINICAL DECISION SYSTEMS: CONCEPTS AND METHODS TO “CLOSE” THE LOOP OF THE LEARNING HEALTH CYCLE

Leveraging on the existing components described in the previous section, an effective infrastructure to implement LHSC is shown in **Figure 1**. There are three main activity classes:

1. The activities associated to gathering data from healthcare delivery actions are represented in blue. This step encompasses activities related to data integration, standardization, and exchange, to support the translation of the information gathered during clinical practice into relevant knowledge that supports new scientific evidence.
2. Research efforts to develop analytical methods for knowledge discovery from heterogeneous data are represented in red. In particular, within the Big Data context, we suggest that one of the main focus is the definition of new phenotypes which are computationally manageable and able to describe disease behaviors.
3. Informatics activities related to CDSSs implementation to transfer research findings into medical and organizational actions are represented in green. In this schema, we show how is it possible to use the sidecar approach (in orange) as suggested by the SMART on FHIR example.

Data should be gathered from heterogeneous sources and contexts such as hospital EHRs and local health care agency



information systems. The data warehouse structures have to be implemented taking into account also clinical practice, since it can also be exploited by the CDSS. Focusing on chronic disease management, important steps include:

- the definition of a data model (e.g., an ontology) able to represent the most important features of chronic populations (i.e., diseases profiles, environmental, and behavioral factors);
- the secondary use of data originally collected for administrative purposes (e.g., patients’ drug purchases);
- the retrospective collection of multivariate longitudinal data;
- the inclusion of modules for managing patients’ generated data, provided by telecare services;
- the capability of providing decision support to both patients and caregivers.

Researchers should focus on analytical methods able to reconstruct diseases evolution from longitudinal and sparse data. This underlines the importance of defining new phenotypes which are computationally manageable and able to describe disease behaviors. The main directions are focused on the use of heterogeneous data structured data, text, images, and signals in electronic phenotyping and on the exploitation of these findings into a CDSS.

The consolidation of clinical decision encompasses two main topics. The first one is related to the need of informatics applications able to deliver comprehensive knowledge about disease subtypes, diagnosis, therapies. The second topic is about the development of tools that support custom workflows, novel analytics, data visualization, and data aggregation. It regards all the activities that allowed closing the Learning Health Care System cycle while exporting the knowledge derived from the research activities into real clinical actions. To this end, it would

be important to implement of visual analytics strategies that enable fast patient stratification. Mining algorithms and their results have to be integrated into a Clinical Decision Support System (CDSS) to make the concept of “Research informs Care” effective.

The final delivered systems should fully exploit Big Data technologies, ranging from distributed storage and computation to schema-less data models, to allow novel decision-making models. CDSS user’s actions to extract new insights on patients’ care flows must be conveyed by a process that integrates methodological novelties into established clinical approaches. For example, the usability experience might be structured to guide temporal data exploration, where visual analytics solutions, together with medical knowledge, facilitate the detection of risk profiles.

IMPLEMENTATIONS OF CDSS, TWO EXAMPLES BASED ON THE LEARNING HEALTHCARE CYCLE

As effective examples of the implementation of LHSC, we illustrate two CDSSs designed and implemented by the University of Pavia and the IRCCS Istituti Clinici Scientifici (ICSM) of Pavia: one to prevent Diabetes complications, and one to support arrhythmogenic diseases research and clinical care.

The Mosaic Project

The system has been developed within the EU “MOSAIC” project, aimed at supporting Diabetes management by resorting to advanced mathematical modeling solutions. As described in (Dagliati et al., 2018), we have developed a system that integrates data coming from hospitals and public health repositories. The

data is collected into an i2b2 data warehouse, and exploited via advanced temporal analytics tools focused on Diabetes complications. Such tools include risk prediction models, temporal abstractions, careflow mining and drug exposure patterns detection. Different users have access to the model results through a “dashboard” interface that allows: (i) clinical decision support during follow-up encounters, and (ii) periodic outcome assessment of the whole cohort.

The system has been validated in a pilot study on more than 700 ICSM patients, showing a reduction in visit duration ($p \ll 0.01$), an increased number of screening exams for complications ($p < 0.01$), and an increase in lifestyle interventions (from the 69% to the 77% of the visits).

Integrated Molecular Cardiology System

LHSC is the basis of the current implementation of an integrated system to support research and clinical care in arrhythmogenic diseases running at ICSM. Such system is based on a EHR linked with a CDSS for genetic variant classification and on a registry semiautomatically synchronized with the EHR.

Mantra is the EHR that collects the molecular and clinical data about patients of the molecular cardiology unit at ICS Maugeri (Segagni et al., 2012b). It collects information on more than 20,000 individuals, mainly affected by Long QT Syndrome (>9,000 patients), Brugada Syndrome (>6,000 patients), Arrhythmogenic Right Ventricular Cardiomyopathy (>900 patients), and Catecholaminergic Ventricular Tachycardia (>900 patients).

Mantra exports data to the Transatlantic Registry of Inherited Arrhythmogenic Diseases (TRIAD). TRIAD is a prospective registry of inherited arrhythmogenic diseases active at ICSM since 2000. It serves as a platform to improve the knowledge on genetic diseases causing life-threatening arrhythmias in the structurally normal heart. Currently, the registry counts 9,700 patients and 29,000 visits. The main considered diseases are: Long QT Syndrome (5,000 patients), Brugada Syndrome (3,000 patients), Catecholaminergic Ventricular Tachycardia (550 patients), Arrhythmogenic Right Ventricular Cardiomyopathy (350 patients). The data stored in TRIAD are collected during outpatient visits and phone follow-ups, and include the results of instrumental exams and cardiac events related to the diseases. An i2b2 instance of TRIAD has been implemented, too, in order to allow fast querying and retrieval of the collected data (Segagni et al., 2012b).

Mantra has been recently linked with the Variant Interpreter software Cardiovai (<http://cardiovai.engenome.com>), which implements a systematic approach to the classification of variants according to American College of Medical Genetics and Genomics–Association for Molecular Pathology (ACMG/AMP) guidelines. Most of the ACMG/AMP criteria are implemented relying on data integration of different omics-resources such as ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar>), MedGen (<http://www.medgen.co.uk>), ExAC (<http://exac.broadinstitute.org>), and PaPI (<http://papi.unipv.it>). However, other criteria were tailored to cardiovascular diseases. The software was tested on benchmark datasets reporting high concordance both for pathogenic and benign variants (Limongelli et al., 2017).

CONCLUSIONS

In medical informatics, Big Data technologies are providing new powerful instruments to gather and jointly analyze large volumes of heterogeneous data collected for different purposes, including clinical care, administration, and research. This makes possible the effective implementation of the “Learning Healthcare System Cycle,” where healthcare practice and research play a synergistic role. In particular, clinical decision support can be strongly enabled by providing fast access to the same set of heterogeneous data available also for research purposes. The proper design of dashboard-based tools may enable precision medicine decision-making and case-based reasoning.

In this paper we have shown two successful examples of the LSHC. The MOSAIC project has shown that decision support tools can be effectively implemented by integrating multiple-sources of data and by resorting to big-data oriented visual and predictive analytics. The temporal dimension of data has been used to deepen the insights on Diabetes monitoring, allowing a better understanding of clinical phenomena, recognizing novel phenotypes, and triggering suitable clinical actions. The second example concerns a molecular cardiology integrated system, where the combination of different software tools is exploited to translate as fast as possible the results of molecular research into clinical decisions.

CDSSs that embed Big Data represent a novel opportunity to support clinical diagnostics, therapeutic interventions, and research. When information is properly organized and displayed, it may highlight clinical patterns not previously considered. This generates new reasoning cycles where explanatory assumptions can be formed and evaluated. Therefore, the future design of such novel CDSSs needs to support entailments among events by properly modeling and updating the different aspects of clinical care. Formal models of clinical guidelines and care pathways can be very effective tools to compare the analytics results with expected behaviors. This may permit to effectively revise routinely collected data, to get new insights about patients’ outcomes, and to explain clinical patterns. All these actions are the essence of a *Learning Health Care System*.

ETHICS STATEMENT

Two effective implementations are reported, concerning decision support in Diabetes (the study protocol was approved by the ICSM Ethics Committee - ID# 2100 CE) and in Inherited Arrhythmogenic Diseases (the electronic registry was approved by the ICSM Ethics Committee - ID# 911 CEC). Written informed consent was obtained from all patients.

AUTHOR CONTRIBUTIONS

AD wrote the first draft of the paper and contributed to its conception. RB had the original idea and revised

the paper. LS revised the paper and contributed to its conception. VT contributed to the software projects. AMal, AMaz, CN, and SP worked on the arrhythmogenic disease project. PD and LC worked on the Mosaic project. MG developed Mantra. IL developed eVai.

REFERENCES

- Belard, A., Buchman, T., Forsberg, J., Potter, B. K., Dente, C. J., Kirk, A., et al. (2016). Precision diagnosis: a view of the clinical decision support systems (CDSS) landscape through the lens of critical care. *J. Clin. Monit. Comput.* 31, 261–271. doi: 10.1007/s10877-016-9849-1
- Bourne, P. E., Bonazzi, V., Dunn, M., Green, E. D., Guyer, M., Komatsoulis, G., et al. (2015). The NIH big data to knowledge (BD2K) initiative. *J. Am. Med. Inform. Assoc.* 22, 1114–1114. doi: 10.1093/jamia/ocv136
- Budrionis, A., and Bellika, J. G. (2016). The learning healthcare system: where are we now? A systematic review. *J. Biomed. Inform.* 64, 87–92. doi: 10.1016/j.jbi.2016.09.018
- Cánovas-Segura, B., Campos, M., Morales, A., Juárez, J. M., and Palacios, F. (2016). Development of a clinical decision support system for antibiotic management in a hospital environment. *Prog. Artif. Intell.* 5, 181–197. doi: 10.1007/s13748-016-0089-x
- Chui, K. K., Wenger, J. B., Cohen, S. A., and Naumova, E. N. (2011). Visual analytics for epidemiologists: understanding the interactions between age, time, and disease with multi-panel graphs. *PLoS ONE* 6:e14683. doi: 10.1371/journal.pone.0014683
- Cimino, J. J., Friedmann, B. E., Jackson, K. M., Li, J., Pevzner, J., and Wrenn, J. (2007). Redesign of the Columbia university infobutton manager. *AMIA Annu. Symp. Proc.* 11, 135–139.
- Cimino, J. J., Overby, C. L., Devine, E. B., Hulse, N. C., Jing, X., Maviglia, S. M., et al. (2013). Practical choices for infobutton customization: experience from four sites. *AMIA Annu. Symp. Proc.* 2013, 236–245.
- Dagliati, A., Sacchi, L., Bucalo, M., Segagni, D., Zarkogianni, K., Millana, A. M., et al. (2014). “A data gathering framework to collect Type 2 diabetes patients data” in *2014 IEEE-Embs International Conference on Biomedical and Health Informatics (Bhi)* (Valencia), 244–247.
- Dagliati, A., Sacchi, L., Tibollo, V., Cogni, G., Teliti, M., Martinez-Millana, A., et al. (2018). A dashboard-based system for supporting diabetes care. *J. Am. Med. Inform. Assoc.* 25, 538–547. doi: 10.1093/jamia/ocx159
- Deeny, S. R., and Steventon, A. (2015). Making sense of the shadows: Priorities for creating a learning healthcare system based on routinely collected data. *BMJ Qual. Saf.* 24, 505–515. doi: 10.1136/bmjqs-2015-004278
- Etheredge, L. M. (2014). Rapid learning: a breakthrough agenda. *Health Aff.* 33, 1155–1162. doi: 10.1377/hlthaff.2014.0043
- Fleurence, R. L., Curtis, L. H., Califf, R. M., Platt, R., Selby, J. V., and Brown, J. S. (2014). Launching PCORnet, a national patient-centered clinical research network. *J. Am. Med. Inform. Assoc.* 21, 578–582. doi: 10.1136/amiajnl-2014-002747
- Gálvez, J. A., Ahumada, L., Simpaio, A. F., Lin, E. E., Bonafide, C. P., Choudhry, D., et al. (2014). Visual analytical tool for evaluation of 10-year perioperative transfusion practice at a children’s hospital. *J. Am. Med. Informatics Assoc.* 21, 529–534. doi: 10.1136/amiajnl-2013-002241
- Gotz, D., Stavropoulos, H., Sun, J., and Wang, F. (2012). ICDA: a platform for Intelligent Care Delivery Analytics. *AMIA Annu. Symp. Proc.* 2012, 264–273.
- Grossmann, C., Goolsby, W., Olsen, L., and McGinnis, J. (2011). *Engineering a Learning Healthcare System: A Look at the Future*. Workshop summary. Washington, DC: National Academies Press.
- Halamka, J. D. (2014). Early Experiences with big data at an academic medical center. *Health Aff.* 33, 1132–1138. doi: 10.1377/hlthaff.2014.0031
- Hall, M. A., Dudek, S. M., Goodloe, R., Crawford, D. C., Pendergrass, S. A., Peissig, P., et al. (2014). Environment-wide association study (ewas) for type 2 diabetes in the marshfield personalized medicine research project biobank. *Pac. Symp. Biocomput.* 19, 200–211.
- Harle, C. A., Lipori, G., and Hurley, R. W. (2016). Collecting, integrating, and disseminating patient-reported outcomes for research in a learning healthcare system. *eGEMs* 4:1240. doi: 10.13063/2327-9214.1240
- Harper, E. (2014). Can big data transform electronic health records into learning health systems? *Stud. Health Technol. Inform.* 201, 470–475. doi: 10.3233/978-1-61499-415-2-470
- Hripcsak, G., and Albers, D. J. (2013). Next-generation phenotyping of electronic health records. *J. Am. Med. Inform. Assoc.* 20, 117–121. doi: 10.1136/amiajnl-2012-001145
- Juarez, J. M., Ochotorena, J. M., Campos, M., and Combi, C. (2015). Spatiotemporal data visualisation for homecare monitoring of elderly people. *Artif. Intell. Med.* 65, 97–111. doi: 10.1016/j.artmed.2015.05.008
- Kaltoft, M. K., Nielsen, J. B., Salkeld, G., and Dowie, J. (2014). Enhancing informatics competency under uncertainty at the point of decision: A knowing about knowing vision. *Stud. Health Technol. Inform.* 205, 975–979. doi: 10.3233/978-1-61499-432-9-975
- Kohn, M. S., Sun, J., Knoop, S., Shabo, A., Carmeli, B., Sow, D., et al. (2014). IBM’s health analytics and clinical decision support. *Yearb. Med. Inform.* 9, 154–162. doi: 10.15265/IY-2014-0002
- Krumholz, H. M. (2014). Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system. *Health Aff.* 33, 1163–1170. doi: 10.1377/hlthaff.2014.0053
- Lee, C. H., and Yoon, H.-J. (2017). Medical big data: promise and challenges. *Kidney Res. Clin. Pract.* 36, 3–11. doi: 10.23876/j.krcp.2017.36.1.3
- Limongelli, I., Nicora, G., Gambelli, P., Memmi, M., Napolitano, C., Malovini, A., et al. (2017). “An automated guidelines-based approach for variants pathogenicity assessment in the diagnosis of genetic cardiovascular diseases.” in *Proceedings XX SIGU Conference*, 112.
- Lupse, O. S., Crisan-Vida, M., Stoicu-Tivadar, L., and Bernard, E. (2014). Supporting diagnosis and treatment in medical care based on Big Data processing. *Stud. Health Technol. Inform.* 197, 65–69.
- Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S., and Ramoni, R. B. (2016). SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. *J. Am. Med. Inform. Assoc.* 23, 899–908. doi: 10.1093/jamia/ocv189.
- Mandl, K. D., Mandel, J. C., Murphy, S. N., Bernstam, E. V., Ramoni, R. L., Kreda, D. A., et al. (2012). The SMART platform: early experience enabling substitutable applications for electronic health records. *J. Am. Med. Informatics Assoc.* 19, 597–603. doi: 10.1136/amiajnl-2011-000622
- Mane, K. K., Bizon, C., Schmitt, C., Owen, P., Burchett, B., Pietrobon, R., et al. (2012). VisualDecisionLinc: A visual analytics approach for comparative effectiveness-based clinical decision support in psychiatry. *J. Biomed. Inform.* 45, 101–106. doi: 10.1016/j.jbi.2011.09.003
- McCarty, C. A., Chisholm, R. L., Chute, C. G., Kullo, I. J., Jarvik, G. P., Larson, E. B., et al. (2011). The eMERGE Network: a consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Med. Genomics* 4:13. doi: 10.1186/1755-8794-4-13
- McGlynn, E. A., Lieu, T. A., Durham, M. L., Bauck, A., Laws, R., Go, A. S., et al. (2014). Developing a data infrastructure for a learning health system: the PORTAL network. *J. Am. Med. Inform. Assoc.* 21, 596–601. doi: 10.1136/amiajnl-2014-002746
- Moghimi, F. H., Cheung, M., and Wickramasinghe, N. (2013). Applying predictive analytics to develop an intelligent risk detection application for healthcare contexts. *Stud. Health Technol. Inform.* 192, 926. doi: 10.3233/978-1-61499-289-9-926
- Murdoch, T. B., and Detsky, A. S. (2013). The inevitable application of big data to health care. *JAMA* 309, 1351–1352. doi: 10.1001/jama.2013.393

ACKNOWLEDGMENTS

This work has been supported by the ICS Maugeri and by the Mosaic project, funded by the European Union. We gratefully acknowledge Federico Dagostin for revising an early draft of the paper.

- Newton, K. M. Peissig, P. L., Kho, A. N., Bielinski, S. J., Berg, R. L., Choudhary, V., et al. (2013). Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *J. Am. Med. Inf. Assoc.* 20, e147–e154. doi: 10.1136/amiajnl-2012-000896
- Ola, O., and Sedig, K. (2014). The challenge of big data in public health: an opportunity for visual analytics. *Online J. Public Health Inform.* 5, 1–21. doi: 10.5210/ojphi.v5i3.4933
- Resetar, E., Reichley, R. M., a Noirot, L., Dunagan, W. C., and Bailey, T. C. (2005). Customizing a commercial rule base for detecting drug-drug interactions. *AMIA Annu. Symp. Proc.* 2005, 1094.
- Richesson, R. L., Hammond, W. E., Nahm, M., Wixted, D., Simon, G. E., Robinson, J. G., et al. (2013). Electronic health records based phenotyping in next-generation clinical trials: a perspective from the NIH Health Care Systems Collaboratory. *J. Am. Med. Inform. Assoc.* 20, e226–e231. doi: 10.1136/amiajnl-2013-001926
- Segagni, D., Tibollo, V., Dagliati, A., Napolitano, C., Priori, S., and Bellazzi, R. (2012b). CARDIO-i2b2: Integrating arrhythmogenic disease data in i2b2. *Stud. Health Technol. Inform.* 180, 1126–1128.
- Segagni, D., Tibollo, V., Dagliati, A., Zambelli, A., Priori, S. G., and Bellazzi, R. (2012a). An ICT infrastructure to integrate clinical and molecular data in oncology research. *BMC Bioinformatics* 13:S5. doi: 10.1186/1471-2105-13-S4-S5
- Sim, I., Gorman, P., Greenes, R. A., Haynes, R. B., Kaplan, B., Lehmann, H., et al. (2001). Clinical decision support systems for the practice of evidence-based medicine. *J. Am. Med. Inform. Assoc.* 8, 527–534. doi: 10.1136/jamia.2001.0080527
- Simpao, A. F., Ahumada, L. M., Desai, B. R., Bonafide, C. P., Gálvez, J. A., Rehman, M. A., et al. (2014). Optimization of drug-drug interaction alert rules in a pediatric hospital's electronic health record system using a visual analytics dashboard. *J. Am. Med. Informatics Assoc.* 22, 361–369. doi: 10.1136/amiajnl-2013-002538
- Simpao, A. F., Ahumada, L. M., and Rehman, M. A. (2015). Big data and visual analytics in anaesthesia and health care. *Br. J. Anaesth.* 115, 350–356. doi: 10.1093/bja/aeu552
- Skiba, D. J. (2011). Informatics and the learning healthcare system. *Nurs. Educ. Perspect.* 32, 334–336. doi: 10.5480/1536-5026-32.5.334
- Slonim, N., Carmeli, B., Goldstein, A., Keller, O., Kent, C., and Rinott, R. (2012). Knowledge-analytics synergy in clinical decision support. *Stud. Health Technol. Inform.* 180, 703–707. doi: 10.3233/978-1-61499-101-4-7031
- Suresh, S. (2014). Big data and predictive analytics applications in the care of children. *IT Prof.* 16, 13–15. doi: 10.1109/MITP.2014.3
- Tenenbaum, J. D., Avillach, P., Benham-Hutchins, M., Breitenstein, M. K., Crowgey, E. L., Hoffman, M. A., et al. (2016). An informatics research agenda to support precision medicine: seven key areas. *J. Am. Med. Informatics Assoc.* 23, 791–795. doi: 10.1093/jamia/ocv213
- Vaitsis, C., Nilsson, G., and Zary, N. (2014). Big data in medical informatics: improving education through visual analytics. *Stud. Health Technol. Inform.* 205, 1163–1167. doi: 10.3233/978-1-61499-432-9-1163
- Waghlikar, K. B., Mandel, J. C., Klann, J. G., Wattanasin, N., Mendis, M., Chute, C. G., et al. (2016). SMART-on-FHIR implemented over i2b2. *J. Am. Med. Inform. Assoc.* 24, 398–402. doi: 10.1093/jamia/ocw079
- Wang, L.-Y., Liu, J., Li, Y., Li, B., Zhang, Y. Y., Jing, Z. W., et al. (2015). Time-dependent variation of pathways and networks in a 24-hour window after cerebral ischemia-reperfusion injury. *BMC Syst. Biol.* 9:11. doi: 10.1186/s12918-015-0152-4
- Yazdanpanah, M., Chen, C., and Graham, J. (2013). Secondary analysis of publicly available data reveals superoxide and oxygen radical pathways are enriched for associations between type 2 diabetes and low-frequency variants. *Ann. Hum. Genet.* 77, 472–481. doi: 10.1111/ahg.12035
- Yun, C., and Hui, YH. (2014). Heterogeneous postsurgical data analytics for predictive modeling of mortality risks in intensive care units. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2014, 4310–4314. doi: 10.1109/EMBC.2014.6944578
- Zhang, Y., Guo, S. L., Han, L. N., and Li, T. L. (2016). Application and exploration of big data mining in clinical medicine. *Chin. Med. J.* 129, 731–738. doi: 10.4103/0366-6999.178019
- Zillner, S., Lasierra, N., Faix, W., and Neururer, S. (2014). User needs and requirements analysis for big data healthcare applications. *Stud. Health Technol. Inform.* 205, 657–661. doi: 10.3233/978-1-61499-432-9-657

Conflict of Interest Statement: Some of the software solutions mentioned in the paper are proprietary software (eVai and Mantra). IL is shareholder of Engenome s.r.l., RB is shareholder of Biomeris s.r.l. and Engenome s.r.l.

The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Dagliati, Tibollo, Sacchi, Malovini, Limongelli, Gabetta, Napolitano, Mazzanti, De Cata, Chiovato, Priori and Bellazzi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.