ORIGINAL PAPER

The CrowdHEALTH project and the Hollistic Health Records: Collective Wisdom Driving Public Health Policies

Dimosthenis Kyriazis[1], Serge Autexier[2], Iván Brondino[3], Michael Boniface[4], Lucas Donat[5], Vegard Engen[4], Rafael Fernandez[6], Ricardo Jimenez-Peris[2], Blanca Jordan[1,2], Gregor Jurak[7], Athanasios Kiourtis[1], Thanos Kosmidis[8], Mitja Lustrek[9], Ilias Maglogiannis[1], John Mantas[10], Antonio Martinez[5], Argyro Mavrogiorgou[1], Andreas Menychtas[11], Lydia Montandon[12], Cosmin-Septimiu Nechifor[13], Sokratis Nifakos[14], Alexandra Papageorgiou[15], Marta Patino-Martinez[6], Manuel Perez[12], Vassilis Plagianakos[15], Dalibor Stanimirovic[16], Gregor Starc[7], Tanja Tomson[14], Francesco Torelli[17], Vicente Traver-Salcedo[5], George Vassilacopoulos[1], Andriana Magdalinou[10], Usman Wajid[18]

[1]University of Piraeus, Piraeus, Greece
[2]Deutsches Forschungszentrumfür Künstliche Intelligenz, Bremen, Germany
[3]LeanXcale, Madrid, Spain
[4]University of Southampton, IT Innovation Centre, Southampton, United Kingdom
[5]Fundación para la Investigación del Hospital Universitario la Fe, Valencia, Spain
[6]Universidad Politécnica de Madrid, Madrid, Spain
[7]University of Ljubljana, Ljubljana, Slovenia
[8]CareAcross Ltd, London, United Kingdom
[9]InstitutJozef Stefan, Ljubljana, Slovenia
[10]European Federation for Medical informatics, Lausanne, Switzerland
[11]BioAssist SA, Athens, Greece
[12]ATOS Spain SA, Madrid, Spain
[13]Siemens SRL, Brasov, Romania
[14]Karolinska Institutet, Stockholm, Sweden
[15]National Organization for Health Care Services Provision, Athens, Greece
[16]Nacionalni Institut za Javno Zdravje, Ljubljana, Slovenia
[17]Engineering Ingegneria Informatica, Rome, Italy
[18]Information Catalyst, London, United Kingdom

Corresponding author. Dimosthenis Kyriazis, Assistant Professor, Department of Digital Systems, University of Piraeus, Gr. Lampraki 126, 185 32,Piraeus, Greece; E-mail: dimos@unipi.gr. ORCID ID: http//www.orcid.org/0000-0000-0000-0000.

doi: 10.5455/aim.2019.27.369-373
ACTA INFORM MED. 2019 DEC 27(5-6): 369-373
Received: Nov 25, 2019
Accepted: Dec 30, 2019

ABSTRACT

Introduction: With the expansion of available Information and Communication Technology (ICT) services, a plethora of data sources provide structured and unstructured data used to detect certain health conditions or indicators of disease. Data is spread across various settings, stored and managed in different systems. Due to the lack of technology interoperability and the large amounts of health-related data, data exploitation has not reached its full potential yet. **Aim:** The aim of the CrowdHEALTH approach, is to introduce a new paradigm of Holistic Health Records (HHRs) that include all health determinants defining health status by using big data management mechanisms. **Methods:** HHRs are transformed into HHRs clusters capturing the clinical, social and human context with the aim to benefit from the collective knowledge. The presented approach integrates big data technologies, providing Data as a Service (DaaS) to healthcare professionals and policy makers towards a "health in all policies" approach. A toolkit, on top of the DaaS, providing mechanisms for causal and risk analysis, and for the compilation of predictions is developed. **Results:** CrowdHEALTH platform is based on three main pillars: Data & structures, Health analytics, and Policies. **Conclusions:** A holistic approach for capturing all health determinants in the proposed HHRs, while creating clusters of them to exploit collective knowledge with the aim of the provision of insight for different population segments according to different factors (e.g. location, occupation, medication status, emerging risks, etc) was presented. The aforementioned approach is under evaluation through different scenarios with heterogeneous data from multiple sources.

Keywords: Holistic Health records, Health Analytics, Public Health Policy Making.

1. INTRODUCTION

A plethora of sensors and applications for individualized care is widely available providing valuable information for certain health conditions and early detection of disease (1). Nonetheless, data is spread across different settings and stored in different systems. Due to the lack of technology interoperability, large amounts of health-related data cannot be processed, and important health events can be missed (2). On the other hand, the large amount of data sources opens a window for opportunities in public health policy making, personalised medicine, prevention of diseases and health promotion. WHO mentions several health determinants that need to be considered (3) including the physical condition, socio-economic status, occupational environment, genetics, and family relationships. According to a study (4), almost 80% of people believe that health is not only the absence of physical disease but relies also on multiple everyday aspects, such as fit lifestyle, nutrition and mental and emotional status. To gain deep knowledge about outcomes of prevention strategies, health policies, and efficiency of care, accurate information deriving from Electronic Health Record (HER) and Personal Health Record (PHR) should be captured and linked with data from other sources. Hitech and Patient Protection and Affordable Care Act (PPACA) consider the adoption of such enhanced records of major importance (5). Records would become placeholders of all types of information coming from multiple sources, including multi-disciplinary knowledge with the aim of facilitating interdisciplinary collaboration and capturing multi-morbidity cases that may remain undetected.

Furthermore, collective community knowledge could be significant in two phases: Collect, aggregate and analyse information from different sources to extract and exploit useful information for the provision of useful insights, and provide the ground for targeted health policy making (5). Surveys (6, 7) also highlight the need and value of accurate information and efficient health information exchange with stakeholders and communities. With respect to data sharing concerns, the acceptance of online platforms e.g. PatientsLikeMe (8) suggests that concerns are deacreasing steadily. Thus, the challenge is to combine data from various sources in order to benefit from community knowledge. To this end, big data management can be combined with eHealth tools (e.g. causal analysis, evidence-based evaluation of strategies, risk stratification, etc) to achieve optimal results. The CrowdHEALTH project, proposes an integrated holistic platform that adopts big data management mechanisms including data acquisition, cleaning, integration, modelling, analysis, information extraction and interpretation (9). CrowdHEALTH explores mechanisms to provide extended health records and exploit collective health knowledge (i.e. clustered records) produced by big data techniques (10). The CrowdHEALTH vision is to make feasible proactive and individualized disease prevention and health promotion, while enhancing policy making, through the provision of collective knowledge and intelligence, following relevant paradigms (11). The outcomes of such milestone projects affect also health policy making, healthcare personnel (13, 14), and educational curricula developments (15, 16) in Biomedical and Health Informatics.

2. AIM

The aim of this paper is to introduce a new paradigm of Holistic Health Records (HHRs) that include all health determinants defining health status by using big data management mechanisms.

3. METHODS

CrowdHEALTH follows a hybrid research and innovation methodology, building upon a well-known iterative approach used in several complex R&D projects. The CrowdHEALTH platform includes several components providing solutions for data manipulation, health analytics, and health policies. In the context of data manipulation components such as the "Data Converter", the "Data Cleaner", the "Data Aggregator", or the "Data Anonymizer" have been implemented, each one serving different data processing purposes, including state-of-the-art technologies. "Data Storage" has also been included, because an ultra-scalable database management system has been integrated, enabling to blend Online Transaction Processing (OLTP) and Online Analytical Processing (OLAP) workloads without compromising performance and enabling real-time analytical queries on operational data without the delay and cost of Extract-Transform-Load (ETL) from siloed operational and analytics data stores. Concerning the health analytics part, algorithms implementing "Clinical Pathway Mining", "Multimodal Forecasting", "Causal Analysis", or "Risk Stratification" have also been designed and integrated, serving multiple needs accordingly. As long as health policies are concerned, components including "Results and Data Visualization", "Policies Modelling" and "Policies Creation" have also been implemented to meet the needs of different stakeholders.

4. RESULTS

Holistic Health Records

CrowdHEALTH explores mechanisms to create extended health records, and benefit from collective health knowledge (i.e. clustered records) produced by big data techniques. As highlighted by Cisco (9): "Humans evolve because they communicate, creating knowledge out of data and wisdom based on experience". "extended" health records can evolve by following the human communication pathway using technologies to in-

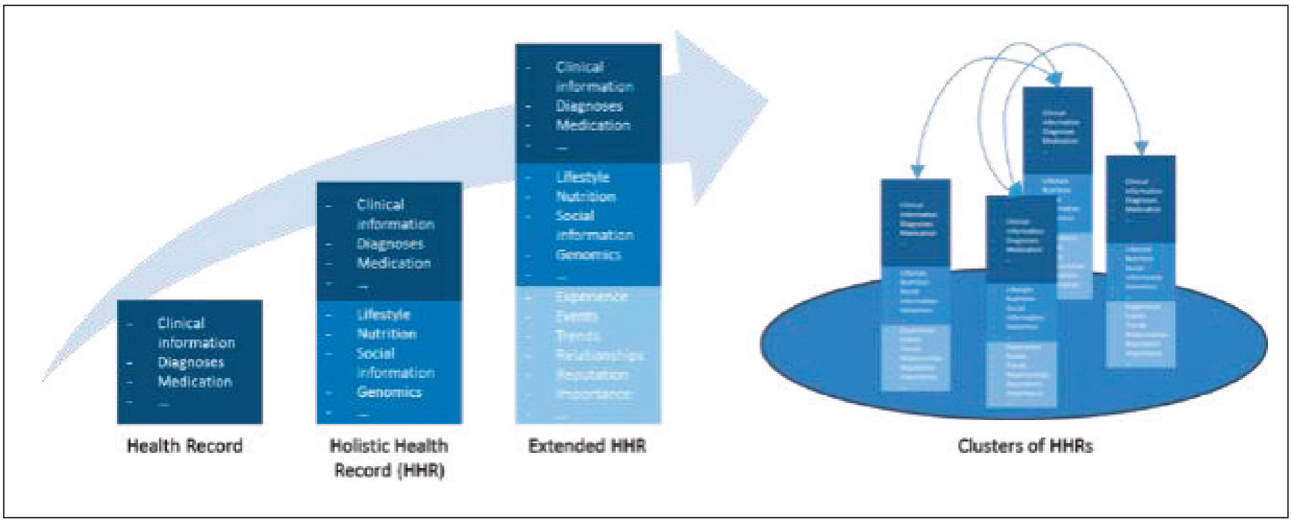

Figure 1.. Holistic Health Records and Clusters of Records.

corporate knowledge derived from other records. Therefore, the Holistic Health Records that CrowdHEALTH proposes can provide a reflection of the citizen capturing all health determinants, and the HHRs Clusters can extract collective knowledge. As illustrated in Figure 1, an HHR consists of several components: (a) the personal component with health, social and lifestyle data collected by patients, relatives or informal carers, (b) the social component containing social care data collected from social care providers, (c) the medical device component containing health data from medical devices or sensors (e.g. home care systems or wearables), (d) the healthcare component containing health-related data such as clinical observations, treatment, medication, diagnosis obtained by healthcare providers and (e) laboratory medical data. The HHRs clusters act as living entities, including properties such as experience (i.e. medication experiences of patients), relationships with other HHRs and classification of relationships (i.e. relationships with friends and family, and "classification" of relationships as for example patients with the same disease), events and trends that affect the citizen or citizens with similar characteristics with the aim of forming networks in an automated way based on specific criteria (such as lifestyle choices or disease symptoms) and exchange information as experiences. Up until now, the project has achieved an HHR structure for the next generation of health records, Fast Healthcare Interoperability Resources (FHIR) compliant (17), ready to include additional properties in health records.

Efficient health services through big data analytics

As depicted in Figure 2, the overall CrowdHEALTH platform architecture presents three main pillars: Data and structures, Health analytics, and Policies. Plug'n'Play (18) is implemented in the context of Data and Structures as an approach for dynamic data acquisition from unknown heterogeneous sources in order to avoid the manual and ad-hoc integration of these sources. Sources Reliability enables adaptive selection of sources based on the corresponding availability patterns and volatility levels. A FHIR-compliant Application Programming Interface (API) enables connectivity and communication, ensuring meaningful interpretation of the acquired data and the feasibility of their incorporation into HHRs. Data Anonymization facilitates compliance with privacy and confidentiality regulations and requirements. The project has implemented Data anonymization techniques, trust and reputation modelling offering all the required mechanisms for enabling data anonymization, data protection and access control, while also ensuring that data sources/entities have the required profiles to account for their datasets.

Data Quality Assessment techniques are implemented to improve the quality of the different sources' data. Regarding HHRs, the project has implemented the HHR manager to facilitate the development and support of the HHR model. the HHR Manager provides the new structures as a basis for the compilation of the HHR. Context Analysis enables the identification of cluster similarities based on the context obtained from the compiled HHRs. The HHR clusters are defined through the Clustering and Classification mechanism that detect correlations among similar HHRs. The HHRs and HHR clusters are stored into the Data Store, together with the data deriving

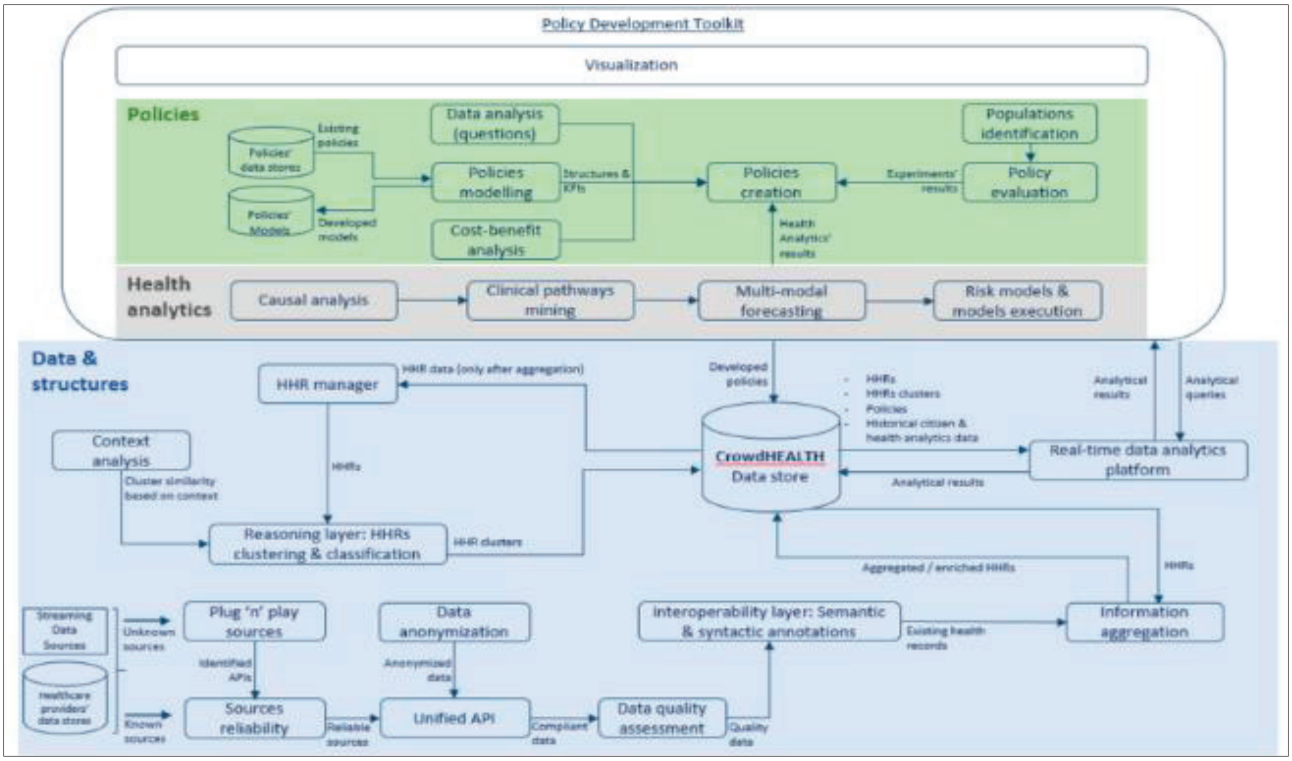

Figure 2. CrowdHEALTH Conceptual Architecture.

from the quality tools, which are made interoperable through the Interoperability Layer to prevent heterogeneity issues of health data. These data are aggregated into HHRs through different data models and query languages. On the stored data (i.e. HHRs, clusters, historical citizen, health analytics results) real-time big data analytics are performed, in order to enable correlations and extraction of situational factors between biosignals, physical activities, medical data patterns, clinical assessment and laboratory tests. The big data approach is able to process millions of events per second allowing the exploitation of real-time medical data from multiple sources.

In the context of the Health Analytics, the Risk Models and Models Execution, Causal Analysis, Multimodal Forecasting, and Clinical Pathway Mining have been implemented. Regarding Risk Models and Models Execution, class-specific care plans and policies are created based on data-driven models. Causal Analysis is deployed for the identification of the properties that affect policies performance and care plans appropriateness. Clinical Pathway Mining supports data analysis to identify similarities or differences in treatment options among groups of patients, reveal factors that affect several treatments and establish a supporting framework for ameliorating the treatment of patients with different diseases. Multimodal Forecasting estimates the applicability and effectiveness of health policies, their variations and combinations to particular population segments taking into account social issues and spatiotemporal properties. As long as the Policies pillar is concerned, Policies Creation obtains the modelled policies and the Cost-benefit Analysis outcomes and proposes evaluated policies based on results obtained by experiments on Identified Populations. Health promotion and disease prevention policies are analyzed, while the integrated health policy making paradigm is refined and updated with data analytics outcomes and experiences. Meanwhile an evidence-based framework creates policy guides and detects indicators in the development of public health policies, combined with leveraged knowledge from existing public health policies. All information is offered to support different actors in the healthcare network (e.g. healthcare providers, policy makers, care professionals, nutrition experts, etc) via the Visualization environment that enables stakeholders to interact with the platform through analytical queries, while manipulating the results and visualizing them. Visualization is part of a Policy Development Toolkit that is created to exploit policies, and health analytics results to advance the processes of policies co-creation and evaluation. Policies models reflecting a structural representation of policies including KPIs as parameters and outcomes what will be monitored, evaluated, adapted etc. Up until now, the project has achieved a policy development toolkit (PDT) serving as a unique point of policy makers to visualize existing data in an interactive way, trigger health analytics mechanisms on different datasets and obtain results, models and create policies. The Data visualization environment integrated into the PDT offers adaptive and incremental visualization of the data to facilitate the analysis of the data by policy makers. Data cleaning approaches are also implemented to address cases of missing values in the datasets, or incorrect or incomplete Information from different scenarios /use cases.

5. CONCLUSIONS

The health data coming from various information sources makes a collection of patient profiles that may facilitate knowledge discovery and support research. To this end, a holistic approach for capturing all health determinants in the proposed HHRs, while creating clusters of them to exploit collective knowledge with the aim of the provision of insight for different population segments according to different factors (e.g. location, occupation, medication status, emerging risks, etc) was presented. The aforementioned approach is under evaluation through scenarios with heterogeneous data sources / devices, data to be included in HHRs, target groups (e.g. chronic diseases or youth obesity), and different environments (care centres, social networks, public environments, living labs, etc). The platform is expected to exploit the current 7.5 million measurements from 1 million people with additional 200.000 / year being also analysed.

- **Acknowledgements:** CrowdHEALTH project is co-funded by the Horizon 2020 Programme of the European Commission Grant Agreement number: 727560 – Collective wisdom driving public health policies.
- **Author's contribution:** Each author gave substantial contribution in acquisition, analysis and data interpretation. Each author had a part in preparing article for drafting and revising it critically for important intellectual content. Each author gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.
- **Conflict of interest:** None declared.

REFERENCES

1. World Health Organization. The determinants of health. 2017. Available at: http://www.who.int/hia/evidence/doh/en.
2. Mantas J. Future trends in Health Informatics - theoretical and practical. Studies in health technology and informatics. 2004; 109: 114-127:
3. Edelman, Health Barometer 2011: Global Findings. Available at: http://healthbarometer.edelman.com/2011/10/health-barometer-2011-global-findings.
4. Institute of Medicine. Capturing Social and Behavioral Domains and Measures in Electronic Health Records: Phase 2. Washington, DC: The National Academies Press. 2014. doi: 10.17226/18951.
5. Pew research Centre. Internet Project survey of US citizens. pewinternet.org 2012. Available at: www.pewinternet.org.
6. Flash Eurobarometer 404. European citizens' digital health literacy. 2014. Available at: http://ec.europa.eu/public_opinion/flash/fl_404_en.pdf
7. Domingo Aladren MC. Managing Healthcare through Social Networks, IEEE Computer. 2010; 43(7): 20-25.
8. Patientslikeme Online Platform. 2017. Available at: http://www.patientslikeme.com.
9. Kyriazis D, Autexier S, Brondino I, Boniface M, Donat L, Engen V. et al., CrowdHEALTH: Holistic Health Records and Big Data Analytics for Health Policy Making and Personalized Health, Stud Health Technol Inform. 2017; 238: 19-23. doi: 10.3233/978-161499-781-8-19
10. Montandon L, Kyriazis D, Ramon V Z, Llatas F,Traver V. CrowdHEALTH- Collective wisdom driving public health policies. 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems CBMS. 2019; 1-3. doi: 10.1109/CBMS.2019.00010
11. CrowdMed Crowdsourcing Platform. 2017 Available at https://www.crowdmed.com
12. Mena LJ, Felix VG, Ostos R, Gonzalez JA, Cervantes A, Ochoa A, et al. Mobile personal health system for ambulatory blood pressure monitoring. Computational and mathematical methods in medicine. 2013: 598196,13.doi: 10.1155/2013/598196.
13. Liaskos J, Frigas A, Antypas K, Zikos D, Diomidous M, Mantas J. Promoting interprofessional education in health sector within the European Interprofessional Education Network. International Journal of Medical Informatics. 2009; 78: S43-S47.
14. Plati C, Lemonidou C, Katostaras T, Mantas J, Lanara V. Nursing manpower development and strategic planning in Greece. Image: the Journal of Nursing Scholarship. 1998; 30(4): 329-333.
15. Mantas J. Developing curriculum in nursing informatics in Europe. International journal of medical informatics. 1998; 50(1-3): 123-132.
16. Hasman A, Mantas J. IMIA accreditation of health informatics programs. Healthcare informatics research. 2013; 19(3): 154-161.
17. Kiourtis A, Mavrogiorgou A, Kyriazis D. FHIR Ontology Mapper (FOM): Aggregating Structural and Semantic Similarities of Ontologies towards their Alignment to HL7 FHIR. IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom), 2018. doi: 10.1109/HealthCom.2018.8531149
18. Mavrogiorgou A, Kiourtis A, Kyriazis D. A plug 'n'play approach for dynamic data acquisition from heterogeneous IoT medical devices of unknown nature. Evolving Systems. 2019. doi: https://doi.org/10.1007/s12530-019-09286-5