

CrowdHEALTH: Holistic Health Records & Big Data Analytics for Health Policy Making and Personalized Health

Dimosthenis KYRIAZIS ^{a,1}, Serge AUTEXIER ^b, Iván BRONDINO ^c, Michael BONIFACE ^d,
Lucas DONAT ^e, Vegard ENGEN ^d, Rafael FERNANDEZ ^f, Ricardo JIMENEZ-PERIS ^c, Blanca
JORDAN ^l, Gregor JURAK ^g, Athanasios KIOURTIS ^a, Thanos KOSMIDIS ^h, Mitja LUSTREK
ⁱ, Ilias MAGLOGIANNIS ^a, John MANTAS ^j, Antonio MARTINEZ ^e, Argyro
MAVROGIORGOU ^a, Andreas MENYCHTAS ^k, Lydia MONTANDON ^l, Cosmin-Septimiu
NECHIFOR ^m, Sokratis NIFAKOS ⁿ, Alexandra PAPAGEORGIOU ^o, Marta PATINO-
MARTINEZ ^f, Manuel PEREZ ^l, Vassilis PLAGIANAKOS ^o, Dalibor STANIMIROVIC ^p,
Gregor STARCAND ^g, Tanja TOMSON ⁿ, Francesco TORELLI ^q, Vicente TRAVER-
SALCEDO ^e, George VASSILACOPOULOS ^a, Usman WAJID ^r

^a *University of Piraeus, Piraeus, Greece*

^b *Deutsches Forschungszentrum für Künstliche Intelligenz, Kaiserslautern, Germany*

^c *LeanXcale, Madrid, Spain*

^d *University of Southampton, IT Innovation Centre, Southampton, United Kingdom*

^e *Fundación para la Investigación del Hospital Universitario la Fe, Valencia, Spain*

^f *Universidad Politécnica de Madrid, Madrid, Spain*

^g *University of Ljubljana, Ljubljana, Slovenia*

^h *CareAcross Ltd, London, United Kingdom*

ⁱ *Institut Jozef Stefan, Ljubljana, Slovenia*

^j *European Federation for Medical informatics, Lausanne, Switzerland*

^k *BioAssist SA, Athens, Greece*

^l *ATOS Spain SA, Madrid, Spain*

^m *Siemens SRL, Brasov, Romania*

ⁿ *Karolinska Institutet, Stockholm, Sweden*

^o *National Organization for Health Care Services Provision, Athens, Greece*

^p *Nacionalni Institut za Javno Zdravje, Ljubljana, Slovenia*

^q *Engineering Ingegneria Informatica, Rome, Italy*

^r *Information Catalyst, London, United Kingdom*

Abstract. Today's rich digital information environment is characterized by the multitude of data sources providing information that has not yet reached its full potential in eHealth. The aim of the presented approach, namely CrowdHEALTH, is to introduce a new paradigm of Holistic Health Records (HHRs) that include all health determinants. HHRs are transformed into HHRs clusters capturing the clinical, social and human context of population segments and as a result collective knowledge for different factors. The proposed approach also seamlessly integrates big data technologies across the complete data path, providing of Data as a Service (DaaS) to the health ecosystem stakeholders, as well as to policy makers towards a "health in all policies" approach. Cross-domain co-creation of policies is feasible through a rich toolkit, being provided on top of the DaaS, incorporating mechanisms for causal and risk analysis, and for the compilation of predictions.

Keywords. Big data, health promotion, disease prevention, health analytics, public health policy making

¹ Corresponding Author: Assistant Professor, Department of Digital Systems, University of Piraeus Gr. Lampraki 126, 185 32, Piraeus, Greece; E-mail: dimos@unipi.gr.

Introduction

The explosion of ICT services led to several devices and platforms providing additional data (e.g. monitoring blood pressure, heart rate, sugar levels, etc). However, these data and the respective services operate independently, and due to this inadequate integration, it is getting increasingly common for important events to be missed (e.g. early indications of spatiotemporal development of diseases). On the other hand, the multitude of data sources highlights a unique opportunity: data to be exploited for effective and targeted policy making, development of personalised medicine, prevention of diseases and health promotion in general. Additional factors are health determinants and should also be considered, as highlighted by the WHO [1], including the physical, social and economic environment, genetics, and relationships with friends and family. Additionally, today's health records (EHRs and PHRs) *are far from being what the citizens consider as of value to their health*. This is consistent with the beliefs of 80% of the public regarding health as more than being disease-free [2] and includes a *variety of everyday aspects, such as the environment, the fit lifestyle, the nutrition, the mental and emotional health*. Capturing and linking this information with other data in EHRs and PHRs would be of benefit for learning about outcomes of prevention strategies, health policies, diseases, and efficiency of clinical pathways. HITECH and PPACA consider the incorporation of this information into EHRs and the meaningful adoption of such enhanced records of major importance [3]: Records would become placeholders of all types of *multi-information*: data from multiple sources, incorporating multi-discipline knowledge, facilitating multi-stakeholder collaboration, capturing multi-morbidity cases.

Furthermore, *collective community knowledge* could play a significant dual goal: (i) collect, fuse and analyse information from different entities to extract valuable information for the provision of actionable insights, (ii) provide the ground for targeted health policy making. The impact is apparent: 46% of the respondents in [5] highlighted that information sharing has changed their overall approach towards healthy life. Another survey [6] shows the need and value for sharing health information with others and communities: 70% tracked a health indicator for themselves / for someone else and 34% of those share their health tracking records or notes with another person or group. With respect to data sharing concerns, the acceptance of online platforms (e.g. PatientsLikeMe [7]) and in general health-oriented social networks (as a venue for sharing ideas, debating treatment options, etc) showcases that these concerns are increasingly limited. Thus, the challenge is not only how to *utilize the data from different sources*, but also how to *combine it to exploit the community knowledge benefits*. To this end, management of big data is key, which provides significant added-value when combined with eHealth tools (e.g. causal analysis, evidence-based evaluation of strategies, risk stratification, etc).

In this paper, we present CrowdHEALTH as an *integrated platform that incorporates big data management mechanisms addressing the complete data path: from acquisition, cleaning, to data integration, modelling, analysis, information extraction and interpretation*. The goal of CrowdHEALTH is to enable proactive and personalized disease prevention and health promotion, while providing decision support to authorities for policy creation, through the exploitation of collective knowledge and intelligence, following paradigms such as [] that emerges from multiple information sources and its combination with situational awareness artefacts.

1. Holistic Health Records

CrowdHEALTH explores mechanisms that can be clustered across three main areas: (i) extended health records, (ii) collective health knowledge (i.e. clustered records), both produced and exploited by (iii) big data techniques. As highlighted by CISCO [9]: “Humans evolve because they communicate, creating knowledge out of data and wisdom based on experience”. Our hypothesis is that the “extended” health records can be exploited to a greater degree if they can evolve by following the human communication paradigm. This metaphor means enhancing records with technologies to exploit the knowledge and experience derived from other records (i.e. from patients in the same medical, social and environmental situations). Thus, we propose the evolution of health records in two stages: (i) towards *Holistic Health Records - HHRs* providing a complete view of the citizen including all health determinants, (ii) towards *HHRs Clusters* to extract collective knowledge. As depicted in the following figure, an HHR contains *several components*: (a) the personal component containing health, social and lifestyle data (such as nutrition or physical activities) collected by either the citizen, her family, friends, etc., (b) the social component containing social care data collected from social care providers, (c) the medical device component containing health data from medical devices (e.g. home care systems or wearables), (d) the healthcare component containing data (e.g. clinical data, diagnoses, medication, etc) obtained by healthcare providers (e.g. primary care systems) and (e) laboratory medical data.

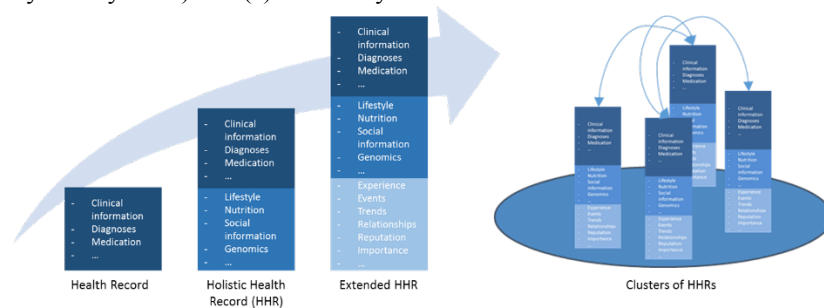


Figure 1. Holistic Health Records and Clusters of Records.

The HHRs clusters act as living entities, including properties such as experience (i.e. medication experiences of patients), relationship with other HHRs (i.e. relationships with friends and family, and “classification” of relationships as for example patients with the same disease), reputation, events and trends that affect the citizen or similar citizens. The latter means that HHRs could form networks in an automated way based on a variety of criteria (such as lifestyle choices or disease symptoms) and exchange information as experiences. Note that we are not proposing yet another social health platform for information sharing amongst individuals but the realization of HHRs clusters allowing records to form communities, share information, identify trends that impact patients.

2. Efficient health services through big data analytics

As depicted in the following figure, the overall architecture of CrowdHEALTH platform consists of three main pillars: Data & structures, Health analytics, and Policies. In the context of *Data & structures*, *Plug'n'Play* enables the integration of unknown sources to avoid the manual and ad-hoc integration of these sources. As it is important that the

collected data emerge from reliable sources to ensure information “fresh-ness” and appropriateness, *Sources Reliability* enable adaptive selection of sources based on the corresponding availability patterns and volatility levels. A *FHIR-compliant API* enables connectivity and communication, ensuring meaningful interpretation of the acquired data and the feasibility of their incorporation into HHRs. *Data Anonymization* facilitates adherence to privacy requirements, while *Data Quality Assessment* techniques are performed to ensure that different sources’ data that are fed in the constructed HHRs actually originate from the expected sources and are of the required quality. Regarding HHRs, 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 captures correlations among similar HHRs. The HHRs and HHRs clusters are stored into the *Data Store*, along with the data derived from the quality tools, which are made interoperable through the *Interoperability Layer* in order to address heterogeneity of health data. These data are aggregated into HHRs through different data models and query languages. *Real-time big data analytics* are performed on the stored data (i.e. HHRs, clusters, historical citizen, health analytics results) enabling correlations and extraction of situational factors between biosignals, physical activities, medical data patterns, clinical assessment and laboratory exams. The big data approach is able to process millions of events per second allowing the exploitation of (often critical) medical data from different sources as things happen.

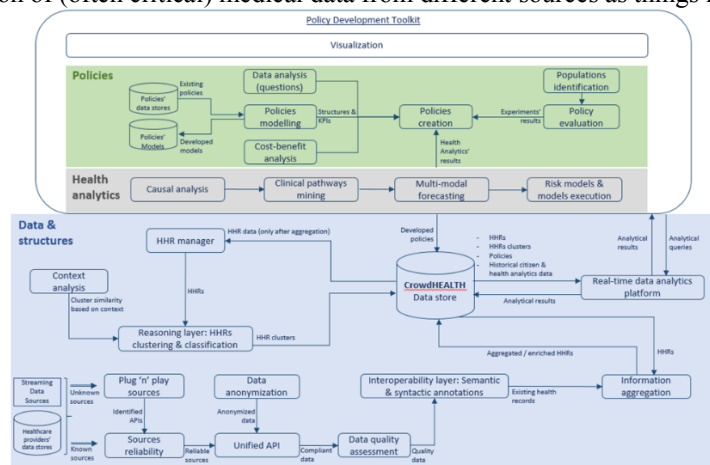


Figure 2. CrowdHEALTH Conceptual Architecture.

In the context of the *Health Analytics*, analytics techniques are utilized for carrying out *Risk Models & Models Execution*, *Causal Analysis*, *Multimodal Forecasting*, and *Clinical Pathway Mining*, upon all the gathered data. Regarding *Risk Models & Models Execution*, class-specific care plans and policies are compiled according to data-driven models. *Causal Analysis* allows the identification of the properties that affect the performance of policies and care plans. *Clinical Pathway Mining* supports data analysis to identify similarities or differences in treatment among groups of patients, indicate major effective factors that affect several treatments and establish a supporting framework for improving 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 consideration social information and spatiotemporal properties.

In the *Policies* pillar, *Policies Creation* obtains the modelled policies and the *Cost-benefit Analysis* outcomes, and in combination with the health analytics results proposes policies that are *Evaluated* based on the experiments on *Identified Populations*. With regards to the first input, health and disease prevention policies are analyzed, while the integrated health policy making paradigm is refined and updated with data analytics outcomes and experiences. Regarding the second input, an evidence based framework produces guides and identifies indicators in the development of public health policies, combined with leveraged knowledge from existing public health policies.

All the information is provided to different entities in the ecosystem (e.g. healthcare providers, policy makers, care professionals, nutrition experts, etc) via a *Visualization* environment that enables stakeholders to interact with the platform through analytical queries, while manipulating the results and visualizing them in an adaptive way. Visualization is part of a *Policy Development Toolkit* that exploits created policies, and health analytics results to advance the processes of policies co-creation and evaluation.

3. Conclusions

The health data from multiple information sources constitute a computable collection of fine-grained longitudinal phenotypic profiles that may facilitate knowledge discovery and cohort-wide investigations on an unprecedented scale, which is the prerequisite for patient-centered care [10]. To this end, in this paper we propose a complete approach for capturing all health determinants in new structures, the HHRs, while creating groupings of them (i.e. clusters) to provide the ground for collective knowledge and provision of insight for different population segments according to various criteria (e.g. location, occupation, medication status, emerging risks, etc). This opens opportunities for disease prevention, personalised medicine, and may lead to reduction in readmission rates [11].

The proposed 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 environments (care centers, social networks, public environments, living labs, etc). Exploiting 2 million records and 700.000 streams of everyday activities, while engaging 200.000 users, the platform is expected to exploit the current 7.5 million measurements from 1 million people with additional 200.000 / year being also analysed.

References

- [1] World Health Organization, *The determinants of health*, <http://www.who.int/hia/evidence/doh/en>
- [2] Edelman, *Health Barometer 2011: Global Findings*, <http://healthbarometer.edelman.com/2011/10/health-barometer-2011-global-findings>
- [3] Institute of Medicine, *Capturing Social and Behavioral Domains in Electronic Health Records*, 2014
- [4] Pew Internet Project survey of US citizens (conducted in September 2012), www.pewinternet.org
- [5] Flash Eurobarometer 404, European citizens' digital health literacy, http://ec.europa.eu/public_opinion/flash/fl_404_en.pdf
- [6] M.C. Domingo, Managing Healthcare through Social Networks, IEEE Computer, vol. 43, no. 7, 2010
- [7] Patientslikeme Online Platform, <http://www.patientslikeme.com/>
- [8] CrowdMed Crowdsourcing Platform, <https://www.crowdmed.com/>
- [9] Cisco Internet Business Solutions Group (IBSG), *The Internet of Things - How the Next Evolution of the Internet Is Changing Everything*
- [10] Chawla NV, Davis DA, *Bringing Big Data to Personalized Healthcare: A Patient-centered Framework*, Journal of General Internal Medicine, 28(3), 660-665, 2013
- [11] Joshi K., Yesha Y., *Analytics for Big Data Generated by Healthcare and Personalized Medicine Domain*, Conference of the Center for Advanced Studies on Collaborative Research, pp. 267-269, 2012