# Perspectives on Big Data applications of health information

Isaac Cano[a], Akos Tenyi[a], Emili Vela[b], Felip Miralles[c] and
Josep Roca[a]

## Abstract

Recent advances on prospective monitoring and retrospective
analysis of health information at national or regional level are
generating high expectations for the application of Big Data
technologies that aim to analyze at real time high-volumes and/or
complex of data from healthcare delivery (e.g., electronic health
records, laboratory and radiology information, electronic pre-
scriptions, etc.) and citizens' lifestyles (e.g., personal health re-
cords, personal monitoring devices, social networks, etc.). Along
these same lines, advances in the field of genomics are revolu-
tionizing biomedical research, both in terms of data volume and
prospects, as well as in terms of the social impact it entails.
The potential of Big Data applications that consider all of the
above levels of health information lies in the possibility of
combining and integrating de-identified health information to
allow secondary uses of data. This is the use and re-use of
various sources of health information for purposes in addition
to the direct clinical care of specific patients or the direct
investigation of specific biomedical research hypotheses.
Current applications include: epidemiological and pharmaco-
vigilance studies, facilitating recruitment to randomized
controlled trials, carrying out audits and benchmarking studies,
financial and service planning, and ultimately supporting the
generation of novel biomedical research outcomes.

## Addresses

[a] Hospital Clinic de Barcelona, Institut d'Investigacions Biomèdiques
August Pi i Sunyer (IDIBAPS), CIBERES, Universitat de Barcelona,
Barcelona, Catalonia, Spain
[b] CatSalut, Servei Català de la Salut, Barcelona, Catalonia, Spain
[c] Technology Centre of Catalonia (Eurecat), Barcelona, Catalonia,
Spain

Corresponding author: Cano, Isaac, Hospital Clínic de Barcelona,
IDIBAPS, Villarroel 170, 08036 Barcelona, Spain. Fax: +34 932 275
455. (iscano@clinic.cat)

## Introduction

The high prevalence of chronic patients with one or
more associated disorders, known as multi-morbidity, is
the main source of dysfunctions and avoidable costs in
conventional health systems worldwide [1,2]. In this
scenario, health risk assessment and stratification are
widely accepted tools facilitating large-scale adoption of
integrated care of chronic patients [3,4] while gener-
ating efficient healthcare and supporting the vision of
personalized medicine. However, only a small propor-
tion of the huge potential of risk predictive modeling is
being applied [5] for health forecasting of chronic pa-
tients due to the lack of in-place procedures for
accessing and mining health information from daily
clinical practice.

Applying holistic strategies for subject-specific risk
prediction and stratification, that consider multilevel
covariates influencing patient health, would increase the
predictive accuracy and facilitate clinical decision-
making based on sound estimates of individual prog-
nosis [6]. For instance, on a daily clinical setting early
identification of patient susceptibility to multi-
morbidity might enable cost-effective preventive stra-
tegies (pharmacological and non-pharmacological) and
enhance management of chronic patients [7].

Such strategies require dealing with highly complex data
and creating new biomedical knowledge, which opens
entirely new translational medicine scenarios and re-
quires interplay between clinical practice and biomed-
ical research. This holistic approach generates novel
requirements to be adopted by the field. Firstly, the
need for multilevel integration of heterogeneous patient
information, namely: socio-economical, life-style,
behavioral, clinical, physiological, cellular and "omics"
data [8], and their use for the study of disease mecha-
nisms. Secondly, the need to extend current trends on
open data from the biomedical community [9] to the
clinical practice and the whole society, by engaging cit-
izens and solving privacy and regulatory constraints.

A core element for addressing current unmet needs in
any given healthcare setting is the deployment of a
Digital Health Framework (DHF), as displayed in
Figure 1 and extensively described in [10]. A DHF aims
at fostering communication among health registries

Figure 1
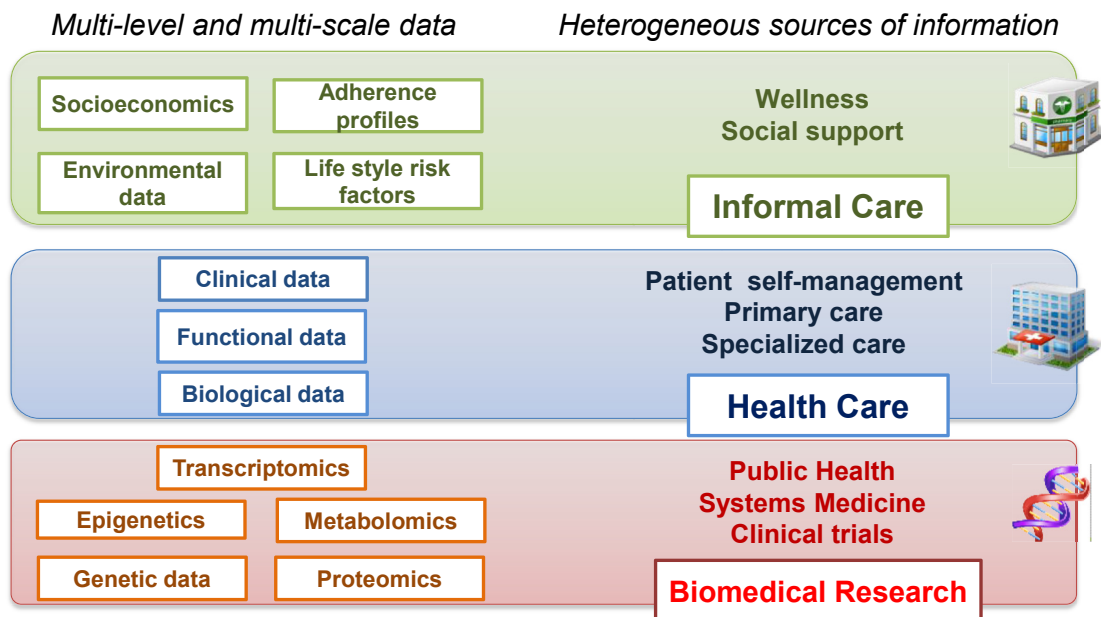


Key dimensions of a digital health framework for enhancing communication among informal care, health care and biomedical research, as first described at [10].

containing health information from various sources, namely: (i) healthcare; (ii) informal care, with special emphasis on environmental and self-management information potentially gathered via personal health folders; and (iii) biomedical research.

Driven by recent advances on big data applications [11–17] and the potentials of the vast amount of accumulated patient data, the scope of this manuscript is to firstly report on the key dimensions of a DHF to provide unified access to health registries with all information about the patient's health determinants. Then, we report on current and future potential applications to gain new understanding about the patient's health through big data and modeling tools [15]. It shall ultimately contribute to enhance dynamic health risk assessment and patient stratification, as well as a technological facilitator to support collaborative case management [10].

## Health information within a digital health framework

Worldwide, most national and/or regional health services have positioned themselves to allow secondary uses of digitalized real-world data for quality and safety of care, financial management and most recently for research purposes [18]. Examples include, the Clinical Practice Research Datalink health registry [19,20] in UK, the National population-based registries in the Nordic countries [21] or the Medicare registries from the Centers for Medicare & Medicaid Services (www.cms.

gov) in US. However, most of these health registries are being adapted and expanded in order to include necessary information not being captured by formal care providers, namely: i) informal care, and, ii) biomedical research. Briefly, informal care includes any aspect with impact on health (e.g. life style, environmental and behavioral aspects, etc.) occurring in the community, whereas biomedical research refers to all research levels from bench to clinical and to public health. This requires development of policies and software solutions [11,12,15,16] that enable smooth data collection and storage as well as data linkage in order to facilitate the extraction of relevant data, its analysis and the communication of findings to relevant parties [10].

### Formal care information

Apart from administrative and reimbursement needs, formal care registries aim to focus on health care of the patient, so they contain information from all clinicians involved in the patient's care. This information is mainly captured through Electronic Health Records (EHRs). EHRs refer to the electronic systems that health care professionals use to manage, store, share, and increasingly to analyze heterogeneous health information from emergency room visits and hospitalizations, primary care visits, mental health centers, socio-sanitary centers, drug prescriptions, etc. EHRs allow doctors to better keep track of patients' health information in structured (e.g., ICD [22], HL7-CDA [23], etc.) or non-structured (e.g., free text, pdf, etc.) formats and make it easier to ensure privacy and security of patients' health