

COST Action IS1303 - Citizen's Health through public-private Initiatives:  
Public health, Market and Ethical perspectives (CHIP ME)

**STSM**

Visiting researcher – Amnon Shabo (Shvo), University of Haifa

Hosting researcher - Angela Brand, University of Maastricht

**STSM Report: Public-private translational health record endeavors could address challenges in the interpretation of privately generated genetic data / Written by Amnon Shabo (Shvo)**

**Abstract**

This report focuses on how innovative health record archiving endeavors could address challenges in the interpretation of data aggregated in the patient's health record. The patient-centric lifetime health record should hold all new types of data generated about an individual such as genomics and sensors data. In particular, privately generated genetic data should be aggregated into the same lifetime health record and thus such a record serves as a bridge between healthcare providers and private facilities such as genetic labs that provide Direct-To-Consumer services, including whole genome sequencing or streaming of sensor data from wearable devices.

It is proposed that the organizations responsible for sustaining the lifetime health records of citizens will be independent of all current healthcare stakeholders. To ensure independency as well as full control of the individuals over their data, it could be beneficial to have those health record organizations incorporated as cooperatives where each individual is an owner of the cooperative and has an account where his/her record is sustained.

True personalized medicine is not possible today as most of the newly generated data items have unknown significance. However, complementing rule-based clinical decision support methods with case-based reasoning techniques could provide personalization of care through insights gained in similar cases, which are identified using ontological means. This report lists health informatics requirements needed in order to successfully run case-based reasoning, where the case is the independent lifetime health record.

## Rationale

This COST MoU Technical Annex asserts the following hypothesis:

***"The public health care system will increasingly be asked to provide interpretation and counseling relating to genetic information that has been generated privately and to satisfy the legitimate curiosity of participants in large-scale population genetic research. Existing ethical and regulatory frameworks may not be suitable to allow an efficient and ethical meeting of demand and supply of genetic knowledge and health, as well as a virtuous interaction between public and private actors. This COST Action aims to improve... with a new focus on new public-private interactions and consumer genetic testing."***

A main issue is therefore the interpretation of "***genetic information that has been generated privately***". In this regard, it's worth noting that this type of information tend to be more massive in volumes and much less understood than similar data used in clinical environments, as well as in well-controlled research assays and clinical trials.

The issue is that most of the discrete data items out of the new types of mass data sets are of 'unknown significance' and therefore are often filtered out in computing processes that aim to produce rule-based knowledge, e.g., through machine learning techniques at the feature selection phase.

In the recent years, a major theme of improving the quality of care has been the personalization of care provided to persons based on the individual makeup of each person, including genetic makeup but also based on streams of sensor data monitoring health conditions of the individual as well as the surrounding environment.

However, personalized care cannot be accomplished due to lack of published evidence on how to analyze the individual makeup in order to truly 'personalize' the care offered to that individual. In order to cope with this challenge, a powerful approach could be to complement rule-based clinical decision making with case-based reasoning, where the case being treated is compared with other cases. Very similar cases could shed light on uncertainty in difficult decision points along the execution of clinical guidelines used in practice, as published by professional medical societies.

Mass data such as genomics and sensor data are more likely to be generated privately and could help out in the personalization of care provided to a person as aforementioned. Due to the current challenging healthcare economic situation, privately generated genetic data is in fact the only practical way to get a richer description of an individual case so that case-based reasoning can become really effective. Thus, public-private collaboration along these lines is the way forward in the realization of personalized medicine.

## **Translational Health Informatics**

This COST Action strives to develop "***Consumers' desiderata with respect to genetic diagnosis, personalized medicine, life-style genomics, and genetic counseling (for individuals)***".

As aforementioned in the rationale section, personalization of care could be truly realized only with the addition of case-based reasoning (CBR) to the traditional rule-based reasoning. In order to effectively employ CBR techniques for patient data, there are several principles in health informatics that need to be implemented. These principles can be generally described as ontological, as opposed to learning and simulation means of computation used in rule-based reasoning.

The successful execution of CBR requires ontological comparison of cases, where the construction of a case should follow the following key imperatives:

### ***Explicitly represent contextual data***

Health data semantics and context cannot be faithfully represented using flat structures (e.g., a list of disconnected entries such as diagnosis, allergies, medications, procedures, etc.), rather, it requires a compositional language that associates data entries into a meaningful statement, pretty much like we associate words to sentences in natural languages.

To this end, it is important to use the new generation of health information standards that are object-oriented in nature and lend themselves to the composition of objects into clinical statements and higher level compositions such as documents and records.

### ***Strike a balance between Narrative and Unstructured data***

Health information representations need to accommodate unstructured data (e.g., clinician's narrative or patient's story), while maintaining interlinks to structured data entries corresponding to contents that have been structured.

Structuring narratives is typically done through human transcription or natural language processing. Either way, those workflows should not leave behind any type of data and strive to maintain both the narrative and its corresponding structured data in the same representational framework. For example, the HL7 Clinical Document Architecture (CDA, [1]) is one such framework that is used in the European Patients Smart Open Services (epSOS, see [www.epsos.eu/](http://www.epsos.eu/)) project for cross-border exchange of patient data. CDA is used in this project to exchange medical summary and medication documents when patients are seeking care not in their home state.

### ***Use base language in a model-driven approach***

To avoid divergence in the implementation of internationally-recognized data standards, it's extremely important for implementers to have the base standards already embedded in model-driven tools such as the MDHT (<https://www.projects.openhealthtools.org/sf/projects/mdht/>) open source tool that allows for the constraining of the HL7 CDA generic standard. The constraining doesn't allow the extension of the base model, rather merely its restriction, e.g., in aspects of cardinality, vocabulary binding, interdependency conditions, etc.

Furthermore, model-driven approach has been proven to be more effective in terms of the time it takes to complete a basic implementation using standards, as well as in terms of agile development of code that the tools support (e.g., automatically generating software components that can be embedded in the overall solution).

### ***Encapsulate raw data into clinical structures***

Key data out of raw/mass data sets pertaining to an individual (e.g., gene sequences) should be encapsulated in its native format into clinical data structures, where 'bubbled-up' items (e.g., DNA variants) out of the key data could be associated with phenotypic data (e.g., drug responsiveness) using clinical data standards that can be consumed by operational information systems of clinical environments.

Some raw data is leading to an interpretation, e.g., Oncotype DX testing kit ([www.oncotypedx.com](http://www.oncotypedx.com)) uses raw data of expression levels of several dozens of genes to calculate a single score that indicates the chances of reoccurrence of breast cancer tumors. Such raw data should be encapsulated in the message that conveys the single score interpretation, so that re-analysis of the raw data is possible. Re-analysis might be needed when there are multiple algorithms to interpret the raw data or when new knowledge becomes available, either as evidence or through translational processes and data-driven analytics.

### ***Use EHR as best organizer of patient (case) data***

There are different approaches on how best to create and maintain the patient-centric electronic health record (EHR) information entity. Some take the document approach where efforts is made to collect all available clinical documents created on a patient and present that pile of documents to the user.

Others take a record approach where all data sets created on a patient are collected and harmonized into a single record information entity. The latter is a better organizer of data as it can have a summary layer where overlaps and contradictory data in the temporal layer (i.e., the episodic medical records) are being reconciled,

and more importantly, topical summaries could be created around diseases, problems, events, goals, etc.

The document approach could be seen as step towards the record approach where documents collected can be included in the comprehensive and longitudinal EHR.

### ***Include family health history as part of the EHR***

The family health history (FHH) is an extension of the entire health history of the individual and could provide hints to a better interpretation of the health record. Even in the 'omics era', the family health history should still be a crucial tool for assessing genetic, environmental, and behavioural risks to health, as well as allowing for smarter use of screening strategies and thus leading to early and even prophylactic treatment at the point of care. FHH is considered the most unused resource in clinical practice and thus the case is incomplete, hindering CBR (case-based reasoning) power.

A new way to personalize preventive and early detection measures could be accomplished through the use of FHH acquired by both the individual as well as healthcare professionals.

There are many ethical challenges when dealing with genetic testing results in the context of FHH. For example, consider the following story published in a daily newspaper:

A young man was diagnosed with colon cancer and was tested positive for a mutation associated with risk for colon and uterine cancer (it is common to his ethnicity; 80% will get either of the diseases). His mother and grandmother were invited to be tested for the above mutation. Both were positive and thus were sent to cancer screening. The mother was diagnosed with uterine cancer but in early stage and the treatment was successful. The grandmother already had colon cancer and was diagnosed with uterine cancer as well; she started treatment. Story was titled "A man saved the lives of his mother and grandmother through his genetic testing".

When genetic tests involve family members, a few questions may arise:

If an individual becomes aware of a risk to his/her family members, should he/she inform their relatives? Who should inform? What if they do not want to know? In such case, could patients have access to medical records of their relatives without their consent? What type of consent would be needed?

To answer such questions, it is important to bring in social technologies that could help out mitigating those challenges and come up with innovative solutions. Since social technologies are typically private, this calls for public-private partnership in order to better utilize the potential of family health history.

### ***Warehouse health data by richest & standard form***

This COST Action states as one of its goals to have "***harmonized measures and standardized computing infrastructures enabling the effective pooling of data and key measures of life-style, social circumstances and environment.***"

With regard to developing "***standardized computing infrastructure***", it has been proposed [2] to have such infrastructure accomplished through an ICT architecture approach consisting of a data warehouse coupled with multiple information marts optimized for specific use cases (e.g., specific type of analytics). The data warehouse is governed by models created based on selection of generic health data standards such as HL7 CDA, which in turn are (1) constrained through model-driven tools that have the base standards and (2) interrelated to convey aspects such as cardinality and touch points centered on an individual subject (e.g., patient, clinical trial subject, healthy individual, etc.).

While the warehouse data models ensure the richest data representation possible (yet aligned with the semantics represented in internationally-recognized health data standards), information marts are created over the warehouse by means of semantic computation of data items and relationships in the warehouse in a way that can be organized in any structure that fits the purpose of the information marts (e.g., relational, RDF, etc.). However, the warehouse remains the ultimate context of any information mart created on top of it, allowing the expansion of context for any item in the mart, as well as the re-computation of data in the warehouse, generating different information that might lead to better or different analysis results.

The proposed architecture has been tested in an EC FP7 project called Hypergenes ([www.hypergenes.eu](http://www.hypergenes.eu)).

As aforementioned, the data 'warehouse  $\leftrightarrow$  information marts' architecture could better support the evolvement of a patient-centric EHR [3-4] and thus enable case-based reasoning for further personalization of care.

### **Give rise to independent health record banks**

The Independent Health Record Banks (IHRB) vision [7-9] suggests that medical records will not be kept anymore by healthcare providers; rather they will be sustained for the entire lifetime of an individual by new players in the healthcare industry - IHRBs - which will be (1) independent of healthcare providers / insurers / government-agencies / patients and (2) regulated by new legislation.

The fundamental principle of the new legislation is that the copy of a medical record stored in such an EHR bank is the only medico-legal copy. The record is sustained objectively by an EHR bank and all authorized parties can have access to it. Such a

bank acts as a custodian/trustee. Multiple competing banks will be established by private enterprises (once the appropriate legislation will be in place).

Healthcare providers could reduce their costs of medical records archiving as this function will be carried out by the EHR banks (sustaining it for lifetime, which is not done today by healthcare provides). Insurers will support it as it will improve the quality of care their customers get. Privacy will be better protected as no global patient identifiers will be needed since a bank account number will be the only access key that the individual needs.

And most importantly, true longitudinal EHRs will finally emerge out of the raw attested medical records by advanced information technologies employed by the EHR banks, and serve as the ultimate case for case-based reasoning in personalized medicine.

#### Use of the term 'bank'

The term banking is used in the context of a 'data bank' and it's important to note that current HRB efforts are not initiated or funded by the financial banking industry. HRB organizations should mainly take on the role of 'Trusted Third Parties' in the healthcare arena where existing parties not always trust each other as manifested by unavailability of complete & coherent patient health history at the point of care.

Financial banks have gradually gained our trust along the past centuries but also created antagonism due to greedy and corrupted managerial attitude in some cases. For HRB organizations, the issue could be addressed if HRBs are defined as not-for-profit organizations where additional revenue is used for purposes such as public health [5] or even cooperatives where each individual owns the HRB cooperative and equally enjoys its profit [6].

In the life sciences world, the term banking is used quite commonly, e.g., the term bio-banking is the default name (469,000 results in Google search in Nov.2014...) for banking of human-derived specimens, or the Protein Data Bank (<http://www wwpdb.org/>) that is funded by the US NSF, NIH and other federal agencies.

#### Social networking aspects of IHRB

When using CBR to help personalize the care of a patient, similar cases found in IHRBs could remain anonymous as we only use the recorded outcome along the timeline of that case to inform on the best action for the case being treated now.

However, the subject of a similar case could be contacted (if consented in advance) in request to either get in touch with the subject of the treated case, or with the

clinicians involved in treating the similar case, by either the current patient or his/her clinicians. Note that these requests could be declined or ignored.

Such workflows implemented by a network of IHRBs are related to another focus of this COST Action, that is, the core question: "**Will social web applications revolutionize the way genetic and other biological information is accessed and used?**" IHRBs will also function as social networks, but unlike current networks that typically exchange non-professional and subjective impressions of individuals about their health conditions, IHRBs will have the full-blown, curated and coherent health record, which then make CBR most effective.

## References

1. Dolin, R.H., Alschuler, L., Boyer, S. Beebe, C., Behlen, F.M., Biron, P.V., Shabo, A. (2006). HL7 Clinical Document Architecture, Release 2. Journal of the American Medical Informatics Association. 13(1): 30-39.
2. Bellazzi R, Masseroli M, Murphy S, Shabo A and Romano P. Clinical Bioinformatics: challenges and opportunities. BMC Bioinformatics. 2012; 13 Suppl 14:S1. doi: 10.1186/1471-2105-13-S14-S1. Epub 2012 Sep 7.
3. Shabo A. Meaningful use of patient-centric health records for healthcare transformation. IBM Journal of Research & Development. Vol. 56, No. 5 September / October 2012.
4. Shabo, A. (2013). The patient-centric translational health record. Pharmacogenomics. March 2013, Vol. 14, No. 4, Pages 349-352.
5. Yasnoff, W. A., Shortliffe, E. H., & Shortell, S. M. (2014). A Proposal for Financially Sustainable Population Health Organizations. Population health management, 17(5), 255-256.
6. Hafen, E., Kossmann, D., & Brand, A. (2014). Health Data Cooperatives-Citizen Empowerment. Methods of information in medicine, 53(2).
7. Shabo, A. (2014). It's Time for Health Record Banking! Editorial of the HRB Focus Theme. Methods Inf Med, 46(5), 601-607.
8. Shabo, A. (2007). Independent Health Record Banks – Integrating Clinical and Genomic Data into Patient-Centric Longitudinal and Cross-Institutional Health Records. Personalized Medicine. November 2007, 4(4): 453-455.
9. Shabo, A. (2010). Independent health record banks for older people – The ultimate integration of dispersed and disparate medical records. Informatics for Health and Social Care. September–December 2010, Vol. 35, No. 3–4, Pages 188-199.