The
EPMA
Journal

# Towards a translational health information language

Amnon Shabo

## Scientific objectives

Recent calls in the EU and US for the creation of a "universal exchange language" for health information representation [1,2] seemed to overlook the contribution of translational medicine in general, and translational informatics in particular. This paper calls for the development of a 'translational health information language' (THIL), which could serve the translational information continuum. Its backbone spans from the biological research results and new types of raw data coming from sources like omics assays, sensors data and imaging techniques, and then clinical trials data and on to clinical data. Alongside this backbone of information there are also contributions of economic, social and psychological considerations that quite often prevent a new and successful intervention at the bedside from scaling out to the community and policy. The main assumption of this paper is that given the current diversity of such a continuum, it is not reasonable to expect a simplified exchange language that covers all portions of that continuum. Instead, THIL strives to mix & match existing and emerging languages through fundamental touch-points in order to enable the integration of patients' data through a conceptual workflow of continuous data encapsulation & bubbling up loop [3]. These processes lead to gradual distillation of the raw data that makes the data usable and useful at the point of care, or for early detection, prevention and well-being.

## Technological approaches

The aforementioned 'translational health information language' (THIL) should draw on the following major efforts:

• On the research side, ISA [4] and Nano-publication [5] could be matched to provide expressive representation of research metadata and the resulting published knowledge;

• On the omics front, a paper recently published in Cell [6], presented the iPOP format that includes data from a variety of omics technologies that were successfully used for early detection and disease prevention;

• When raw data is integrated into clinical structures, it is important to include representational languages for the three pillars of structuring health information:

○ compositional language
○ constraining syntax
○ profiling methodology

Through profiling, it is possible to put together composed & constrained representational artifacts that can be sent across the wire for semantic interoperability purposes or for incorporation into the patient-centric electronic health record [8].

Regarding terminologies, it is beneficial to draw on results of the FP7 SemanticHealthNet [7] project that attempts to resolve the fuzzy relationship and overlaps between information models and ontologies, eventually harmonizing them into a coherent eHealth infostructure.

## Outlook and expert recommendations

Figure 1 depicts the languages assembly proposed to be the roadmap to creating a 'translational health information language' (THIL). The challenges are in achieving effective touch points between the candidate schemas, for example, the Nano-publication format includes provenance as essential part of its schema, where provenance refers to a list of underpinning articles and other data sources supporting the assertion at stake. It could be much more effective if the referenced data sources are described through the ISA standard, which provides structured and unified description of the research and data used for the investigations that eventually led to the published assertion.

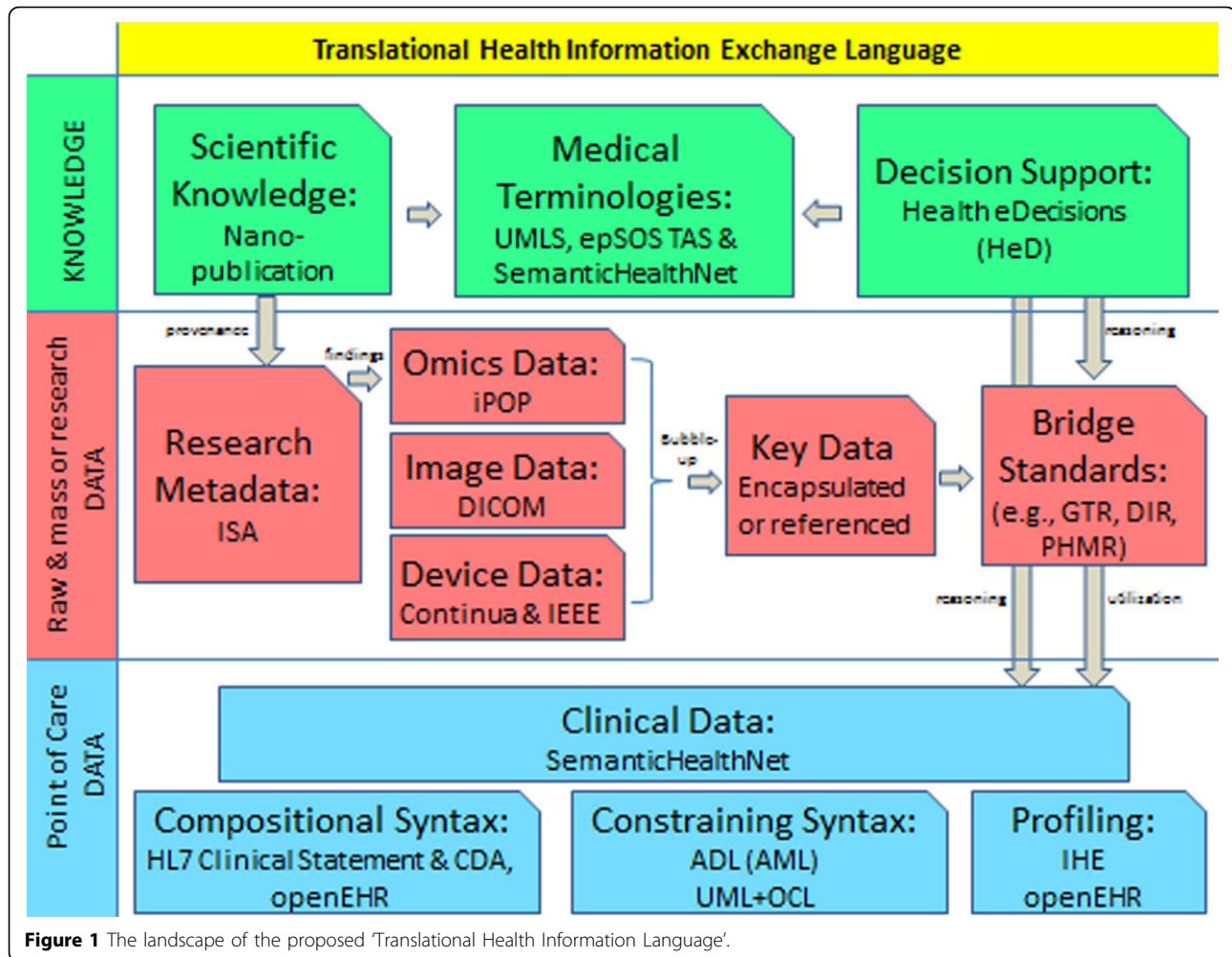Correspondence: shabo@il.ibm.com
IBM Haifa Research Lab, Israel

**BioMed** Central

**Figure 1** The landscape of the proposed 'Translational Health Information Language'.

### References
1. European Commission C: WORK PROGRAMME 2013, COOPERATION, THEME 3, ICT – INFORMATION AND COMMUNICATIONS TECHNOLOGIES, 4536 of 09 July 2012. See Coordination and Support Action e5 in Objective ICT-2013.5.1. 2012, 56.
2. Executive Office of the USA President, President's Council of Advisors on Science and Technology (PCAST): Report To The President Realizing The Full Potential Of Health Information Technology To Improve Healthcare For Americans: The Path Forward. 2010, [Online]. http://www.whitehouse.gov/ostp/pcast. Access 31 May 2013.
3. Shabo A, Dotan D: The Seventh Layer of Clinical Genomics. Special Issue on Information-Based Medicine. *IBM Systems Journal* 2007, **46(1)**:1-11.
4. Sansone SA, et al: Toward interoperable bioscience data. *Nat Genet* 2012, **44(2)**, [http://www.nature.com/ng/journal/v43/n4/abs/ng0411-281.html] Access 31 May 2013.
5. Mon B, et al: The value of data. *Nat Genet* 2011, **43**:281-283, [http://www.nature.com/ng/journal/v43/n4/abs/ng0411-281.html] Access 31 May 2013.
6. Chen R, et al: Personal Omics Profiling Reveals Dynamic Molecular and Medical Phenotypes. *Cell* 2012, **148**:1293-1307.
7. SemanticHealthNet (FP7-ICT-2011-7):[http://www.semantichealthnet.eu/], Access 31 May 2013.
8. Shabo A: The patient-centric translational health record. *Pharmacogenomics* 2013, **14(4)**:349-352.