

Supporting Bio-Medical Knowledge Discovery: the Archetype-Based Electronic Bio-Medical Research Record (eBMRR)

Melanie Bettina Späth, Ph.D.

University of Dublin, Trinity College, 2013

To support bio-medical knowledge discovery, clinical patient information, such as medical history and lifestyle information, needs to be combined with results from molecular research on patient samples. While the former can be found in the patients' medical records, the latter may be scattered over researchers' computers in a variety of proprietary data formats, hampering the integration and the sharing of this information between collaborating researchers and its preservation for future integrative research. Biobanks play an important role at the intersection of these two streams of information. However, current database solutions for biobank information management systems (BIMS) tend to be proprietary, disease and study-specific, so that major resources need to be invested in setting up a BIMS for a new study or biobank and information is not easily shareable or available for reuse in future research.

This thesis proposes the development of a novel type of record, the electronic Bio-Medical Research Record (eBMRR), to support the sharing and integration of molecular research and clinical patient information within and across biobanks to facilitate bio-medical knowledge discovery. The eBMRR aims to generically integrate molecular research and clinical patient data in an interoperable manner and be easily adaptable to any biobank or study, thus easing the process of setting up a BIMS.

This thesis evaluates the feasibility of using the *openEHR* archetype approach as the basis for the development of the proposed eBMRR. The *openEHR* archetype approach is a novel, standard-based approach to generically model clinical information in reusable archetypes to support the development of syntactic and semantic interoperable EHRs. This research shows that clinical archetypes can be reused in a research environment to record clinical patient information for research purposes, using the database model of the biobank of the Irish Prostate Cancer Research Consortium as the test database, although the integration of legacy data can be challenging. In addition, an approach to develop archetypes modelling omic concepts from standard-based omic data formats is suggested and applied. This resulted in the successful development of sets of generic, reusable omic archetypes that can be used to record information produced from omic experiments on patient samples. The syntactic correctness of these newly developed archetypes was evaluated with available archetype tools, whereas the semantic correctness was evaluated by arranging the archetypes in templates and mapping them to omic instance data provided in the original data formats. A number of limitations using the *openEHR* approach to model omic information are identified. The eBMRR can thus be constructed from clinical and omic archetypes to integrate clinical and research information in a generic and safely shareable manner.

This research is promising towards the creation of an interoperable eBMRR for biobanks to facilitate biomedical knowledge discovery by allowing for the safe sharing of information between researchers, while reducing modelling efforts in biobanks.