

Place: Biomedicum, Helsinki

Time: 5.5.2004 10:15- 16

Attendees: Juha Muilu, Jan-Eric Litton, Pontus Lindqvist, George Ölund, Kauko Heikkilä, Juha Saharinen, Juri Ahokas, Anthony Metzidis, Markus Perola, Anne Leinonen
Martin Senger (EBI), Heikki Lehväslaiho (EBI)

Data Modelling Meeting

Juha Muilu introduced GenomEUtwin project in data structure and modelling point of view. Purpose of this meeting was to consider for participation on international standardization efforts. There are many good reasons to favour this decision (e.g. joint meta-studies). At least, common XML schema for geno- and phenotype data should be achieved

Object Management Group and standards

Martin Senger told us about function of standards and introduced OMG as one of the standardization consortium. Life Sciences Research (LSR) is one of the domain task forces in Domain Technology Committee in OMG. OMG technology adoption process works through submissions for aiming to implement a product. Implementation strategy is based on Model driven architecture (MDA) where the domain problem is solved once (platform independent model (PIM)). From the PIM several different platform specific models (PSM) and implementations can be derived. For example XML schema and database structure can be implemented from same PIM model, reusing the design and making implementations more interoperable. Examples of OMG works are LSID (Life Science Identifier) and Gene Expression specification (MGED).

Using UML and XML tools for modeling, there is design patterns for revising it into relational model.

Discussion

GenomEUtwin project schedule is tied to strict deliverables. There is not enough time for this kind of process within the deadlines. Also, there is a challenge of convincing the importance of this to steering committee. It would be worthwhile of starting this thread as side of these

big studies. For covering the cost of OMG memberships, separate money needs to be applied. An alternative is to use existing members, like European Bioinformatics Institute, as a mediator as done with the gene expression data.

The OMG provides consensus driven standardization process, with well defined practices. The lack of proper process in some other standardization efforts have made standardization difficult. As an example Marting gave LSID which started in I3C and never really converged until overtaken by the OMG/LSR. Other benefits are: hosting of specifications (all the specs are freely available), visibility (OMG is world largest software consortium), interaction possibilities between academia and commercial companies, expertise on the standardization work.

Full membership is \$4000, which allows to make submissions. University membership is \$500/yr. This allows everyone associated with the university (faculty, staff, students) to sign up for as many OMG email lists as they want, and attend OMG meetings. The university gets one vote in each Task Force (TF) that at least one person attends.

Harmonization of data in world wide prospective population studies

Jan-Eric Litton went through the goals and objectives for data harmonisation of GenomEUtwin. P3G consortium with CARTaGENE and Estonian Genome Project, and LifeGene, PRECISE and UK Biobank projects are examples of collaboration projects. In GenomEUtwin, database core has already done some standardization; EUIDNUM. And it took quite a long time to make a code for identifying twins. Compared to LSID, it is not so long, but it includes some attributes.

There is a urgent need for the global ontology engineering and data modelling to achieve collaboration population studies successfully.

PhenoGeno Model

Building a PhenoGeno model started in database core meeting at Helsinki in March 2004 by working group 2. Juha Muilu have developed it further. It considers phenotype as life event (e.g. even color of eyes can change). All phenotypes can depend on time. Phenotype can be described based on one or many measurements or other phenotypes/life events. The model can adopt multiple classification systems like ICD 9 and SNOMED.

It was decided to document well-defined entities and present it to other groups of GenomEUtwin. Ontology for measurements is also needed.

The scope of the problem was also pointed out. Meaning that problem descriptions should not include implementation. Heikki advised not to do too detailed model on phenotypes that we already know that will change later. Instead devise modelling for higher and lower level modelling.

It was decided to concentrate on Humans. Model is general enough to apply other species and biological systems as well.

Juha Muilu will update this model and circulate it via email to WG2. After reviewing it locally, phone meeting will be arranged. Migraine and CHD will be used as test cases.