

AnalysisXML Development Workshop, 25-27 June 2007

A 3 day satellite workshop to develop AnalysisXML at the EBI in Hinxton, Cambridge.

Attendees

- Angel Pizarro (PSI-PI Co-chair - University of Pennsylvania)
- David Creasy (PSI-PI Co-chair - Matrix Science)
- Jennifer Siepen (Manchester University)
- Martin Eisenacher (Medizinisches Proteom Center)
- Philip Jones (PSI-PI Secretary – EMBL-EBI)
- Richard Côté (EMBL-EBI)

Minutes / Record of Activities

During the course of the workshop the following activities were completed:

- A review of the object model, XML schema and available instance documents was performed.
- Several iterations of development were completed, in which problems or missing functionality were identified, the UML model was updated, a new XML schema was generated and instance documents were developed that validate against the schema.
- Specific problems were documented and are being tracked on the PSI case tracker system. These issues include:
 - The need to be able to annotate the ion series for each identification. A compact, usable format for this was developed during the workshop.
 - The relationship between protein and peptide identifications was moved from the 'molecule' description section to the 'analysis' section, to reflect the fact that this is an analysis, not a statement of fact.
 - Considerable discussion of the scope of analysisXML took place. It was suggested that analysisXML should handle any kind of molecule (identified from mass spectrometry). This case was dismissed – the scope of analysisXML will remain limited to polypeptide identification.
 - Where identification is completed against nucleic acid sequences, the need to handle different translation tables for different species was flagged.
- By the end of the three day meeting, the following work had been completed:
 - The development of UML, XML schema and instance documents that are able to represent (All items being new or significantly modified in this workshop):
 - search protocols
 - the identifications of proteins, peptides and protein modifications
 - the use of controlled vocabularies throughout the schema was partially remodelled and made more robust / efficient.
 - the model now allows a complete representation of multiple analysis steps in a single XML file, e.g. search engine -> protein determination algorithm.
 - an explicit representation of quality measurement is included.

- In addition, candidate structures were developed for representing:
 - ion series matching
- Documentation was developed, including the template for the final specification document. (Note – this is the specification of the format, not the requirements document).
- Documentation was added to the UML model and XML schema.
- The following commitments were agreed to by members of the working group:
 - Complete UML / XML schema documentation: *Martin Eisenacher*
 - Complete specification document: *Phil Jones & Richard Côté*
 - Further development of ontology: *David Creasy & Zsuzsanna Bencsath-Makkai*
 - Validation Tool (based upon MI-Check code): *EBI (PRIDE team etc.)*
- It became clear that a great deal more input is required from a range of domain experts to expedite the development of a complete model for the encoding of quantitation information.