

PSI-PI Teleconference 6 June 2006

Attendees:

Angel Pizarro (Chair)
David Creasy (Chair)
Phil Jones (Minutes)
Ruth McNally
Pierre-Alain Binz
Trish Whetzel

1. Logistics for teleconferences

EBI can keep the teleconferences going for the time being. The ProDac funding (starting 1 October 2006) may be able to cover this in the longer term. To be followed up with Rolf Apweiler.

2. Feedback and additions to the use cases for analysisXML

* Please see the current version of the use case document

The list on the spreadsheet is possibly not complete, although the conference call is not the place to discuss this in detail. Please take a good look and add as appropriate. Suggestions sent to David Creasy.

Phil – Use case being the ability to link to a spot on a gel – i.e. cross references to other formats. Agreed that this was required.

Up-coming conference call (12 June) to discuss the gel based formats that Angel will join, which will cover current status of GelML and GinML.

Need identified to be able to link to LC-MS runs / fractions / retention times etc.

Need highlighted to be able to indicate quantitation information.

Support of row data, i.e. row traces or images from gels as part of data section of analysisXML? To revisit in UML?

Use case should allow a reference to raw format etc.

Perhaps need use case to allow questions such as - return identifications containing a particular PTM.

PSI-MOD format for annotating PTMs? What is the relationship to PSI-MOD?

Do we need to explicitly indicate the nature of molecules that can be identified?

Any more feedback? - To be sent to David Creasy.

Need to explicitly state what the format does not and should not cover.

PSI-PI Teleconference 6 June 2006

3. Feedback and status of the spreadsheet with requirements

* Please see the current version of the search engine output parameter analysis spreadsheet

Waiting for SpectrumMill feedback. This is important before we can move forwards with this.

Still need to ask Waters for feedback.

Still need to consider the relationship to MIAPE and MCP.

Do we need to include MCP guidelines? Are they subsumed by MIAPE? Pierre-Alain attempted to map MCP to MIAPE-MS, but some difficulties in matching the two. Some conditional requirements in MCP guidelines that are complex to control, e.g. for quantitation data, there are some data points that need to be provided sometimes but not always. Another important issue in MCP is the reporting of upstream and down flanking sequences. Pierre-Alain to send in the mapping between MIAPE and MCP.

Yellow blocks on the spreadsheet are additional rows that have not been considered for the original contributors.

Problem - supporting de novo searching, this makes almost everything in the spreadsheet optional. Is this desirable? Need to specify the type of analysis that will then control the requirement for data? In MAGE / MIAME, the mechanism allowed most fields to be optional, but MIAME was the main source of specifying requirements.

De novo will necessarily be supported by the analysisXML format - sub classing used to get around problem of different required fields.

Tag searches where there is no peak list. E.g. looked at spectrum - picked out run of residues - start mass, end mass, peptide mass. Method still used in Mascot, Spectrum Mill etc. Therefore needs to be supported. Often results from (for example) manual interpretation of spectrum.

Also need to handle searches where the database contains annotated spectra.

Rules for what to do with spreadsheet data -

Suggested:

1. In 2 or more columns, should be a node in analysisXML. (e.g. - taxonomy data should be a node)
2. Anything require by MIAPE or MCP should be required in analysisXML. - Experience from PRIDE and MAGE suggest that making complete MIAPE or MCP

PSI-PI Teleconference 6 June 2006

guideline compliance necessary for XML validation will prevent the format being useful.

So – do we need to produce a tool to 'validate' against MIAPE?

Also need to include some additional data such as reading frame for nucleic acid data etc. (optional parameters).

"Number of peptide sequences compared to each spectrum" - for where search sequences

4. Status of sample AnalysisXML in UML format.

Angel - gave explanation of UML (inheritance / cardinality / aggregation).
SpectraDataPartition - spectrum that is used to identify within a large file. (e.g. which spectrum in the related mzData file)

Pierre-Alain requested a concrete example of analysisXML (i.e. an XML file) to clarify the model and its use.

5. AOB

Charter needs to be made more specific for PSI-PI - currently virtually identical to the PSI charter.

Next conference call for PSI-PI in four weeks time – *Tuesday 4th July 2006.*