MAGE-ML file exchange jamboree, EBI, December 1-6, 2003

AGENDA

December 1

9:00 - 9:30	Registration
9:30 - 10:00	Introduction (in the James Watson pavilion)
10:00 - 12:30	Tutorial in the James Watson pavilion (*Michael Miller, Charles Troup, Angel Pizarro*) *IT training room is available for those not wishing to attend the tutorial.*
12:30 - 13:30	Lunch
13:30 - 14:30	Tutorial continues
14:30 - 16:30	Identifiers (from this point onwards in the IT training room) - how to ensure uniqueness/reusability (*Joe White, Michael Miller*) - LSID (*Ugis Sarkans*) - assigning namespaces to pipeline submitters
16:30 - 18:00	Policies for using the MGED Ontology in MAGE: - presentation - rules, examples, query examples (*Angel Pizarro*) - questions/discussion

December 2

9:00 - 11:30	QuantitationType usage - Special/Standard… - common reusable definitions of QTDimensions for various platforms - ArrayVision, GenePix, Affy, Agilent,.. (*Philippe Rocca-Serra, Ele Holloway*)
11:30 - 12:30	Future of MAGE: reduced representations of raw and normalized array data (*Peter Eastman*)
12:30 - 13:30	Lunch
13:30 - 16:00	How to use MAGE in a uniform way: short presentations of what structures are used for data capture/export (*Joe White, Michael Miller, Thessa Kockelkorn, Mohammad Shojatalab, David Craigon*)
16:00 - 17:30	How to use MAGE in a uniform way: short presentations of how MAGE is expected to look like for import (*Helen Parkinson, Steffan Durinck, Kjell Petersen, Michael Miller*)
17:30 - 18:00	OntologyEntry usage for BioMaterials (*Susanna Sansone*)

December 3

9:00 - 12:30	How to use MAGE in a uniform way: discussion/setting common policies (*Paul Spellman, Ugis Sarkans*)

Sources of non-uniformity:

- more than one MAGE structure possible for encoding conceptually same information
- is this a problem, do we want uniformity?
- if yes, do we want to
  - recommend one possibility
  - enforce one possibility
  - admit that with current MAGE this cannot be solved

Nonexclusive list of issues:

- Experiment package
  - ExperimentalFactor usage
    - FactorValue – Channel
    - null Factor values
  - Experiment - BioAssay links
  - Experiment - BioAssayData links
  - ExperimentDesign - BioAssay links
- BioMaterial package
  - BioMaterial - Treatment cycle, BioMaterial subclasses used
  - Treatment/action/compound vs. ParameterValues
  - 1 Measurement per Treatment
  - 1 Protocol per Treatment
  - what are "important" properties of BioMaterials/Treatments (for data mining)
- BioAssay package
  - structure (Physical/Measured/Derived, links to BioAssayData objects)
  - BioAssayTreatment - pointer back to the same BA
  - links to ExperimentalFactors
- BioAssayData package
  - Redundant DesignElementDimensions
  - DesignElementDimension descriptions
  - Channel – which is reference?
  - SummaryStatistics
- ArrayDesign
  - how Features/Reporters/CompositeSequences are used
    - primers, genes, …
  - their grouping into DesignElementGroups
  - DesignElementMaps
    - protocolApplications
  - controlType
  - species – for all DesignElementGroups?
  - Grouping DesignElementMaps (e.g. for different versions of array annotation)
  - DatabaseEntry/OntologyEntry usage
  - what are "important" DesignElement properties for data mining
  - BioSequence_ref pointers - using only ids, no explicit MAGE constructs
  - failures → a new ArrayDesign?
- Array package
  - batch - ArrayManufacture
- Protocol
- Other
  - auditTrail, Security - when to use, granularity
  - HigherLevelAnalysis
  - Identifiers – case sensitive
  - updates

- Reusable objects
  - Database, Channel, QuantitationType, Contact, ...
  - How to maintain, how to facilitate re-using
  - is it important at all?

12:30 - 13:30 Lunch

13:30 - 18:00 Recommendations for minimal MIAME-to-MAGE mapping (*Ugis Sarkans*)

December 4

9:00 - 12:30 Future of MAGE / SystemsBiology-OM
- CEBS model (*John Yost*)
- proteomics models (*Andrew Jones, Chris Taylor*)
- lightweight array design description framework (ADF) (*Philippe Rocca-Serra*)
- discussion
  - layered design
  - easy separation of "interesting" data from evidence (protocols etc.)
  - ……

12:30 - 13:30 Lunch

13:30 - 18:00 Work on MAGE-ML export/import
From ArrayExpress point of view the ultimate goal is to achieve that pipeline submitters' data can be loaded in AE without manual curation work required.

13:30 - 18:00 (For those who are happy with their export/import) - further work on MAGEstk
- needed leader (per language?)
- MAGEstk installed (before coming) on laptops of those planning to participate in this
- task list
- everybody should know what they would like to/need to do

December 5

9:00 - 18:00 Work on MAGE-ML export/import continues

9:00 - 18:00 MAGEstk work continues

9:00 - 11:00 MAGE homepage

11:00 → OMG activities
- MAGE RTF, documentation
- queries RFP (*Michael Miller*)
- discussion (*Kjell Petersen, others*)

12:30 - 13:30 Lunch

December 6

aim to finish before 12:30
Sum up what has been done:
- data producers - changes made/planned
- data consumers - modified assumptions on how MAGE looks like
- documented best practices

- "minimal" MIAME-to-MAGE mapping
- MAGEstk advances
- ideas for MAGEv2, SystemsBiology-OM, Proteomics-OM

12:30 - 13:30 Lunch

13:30 - 18:00 (For those who have not yet left) - room available

The jamboree will take place in the Wellcome Trust Conference Centre IT training room, which is equipped with 20 PCs and network connections. PCs are dual boot Windows2000/Linux, I can provide more detailed description if there is need. Own laptops are preferred in the sense that everything needed can be (and should be) preinstalled.