

CASIMIR WP4

Data Representation

John Hancock



WP Objectives

- Assessment of technical aspects of database interoperability as a barrier to scientific and financial sustainability
- Assessment of the variability of practice in the semantics of biological data representation, e.g. genotype, gene expression
- Assessment of emerging standards and current practice for data representation, annotation and ontologies



WP4 Meetings

We have held two WP networking meetings:

Heathrow (July 2007) (*in conjunction with the Euromouse meeting*).

Aim: to introduce CASIMIR and the questionnaire to Euromouse participants and canvass responses

Corfu (October 2007)

- Discuss results of survey
- Head up on emerging ontologies
- Discuss the relationship between phenotype ontologies and human disease description



Questionnaire: Conclusions

- Good response
- Most DBs now using RDBMS - mostly MySQL or PostgreSQL
- Web services are a popular option amongst the “EU” databases although less than half have implemented them so far. The other DBs show less penetration of web services [Direct SQL access?]
- Ontologies are in wide use but relatively poorly used in “EU” databases. Supplementation with CVs more common in non-”EU” databases
- Many, but not all, DBs not using ontologies will or are considering using them. Many DBs currently using ontologies will also extend their ontology use
- Although OBO ontologies are used by most ontology users these only make up a slim majority [Why?; What are perceived advantages of OBO?]
- Most DBs see a need for new ontologies (often undefined)
- Minimum standards are currently in use by only c. 25% of DBs - the most commonly used is MAGE/MIAME [Why?]



Deliverables and Actions



	0-3	3-6	6-9	9-12	12-15	15-18	18-21	21-24	24-27	27-30	30-33	33-36
WP4:												
4. Data Representation												
4.1 Web publication of range of data representations in European mouse centric databases												
4.2 Report to Commission on data representation												
4.3 Publication of major findings of 4.2 in then scientific literature												
4.4 Network meetings												
4.5 Three Network meeting reports												
4.6 Presentations at the GA and AM												

- 4.1 - D9 - Web publication of classified list of data representations in European mouse-centric and related databases - for early 08
- 4.2 - Report to Commission
- 4.3 - Publication
- ✓ 4.4 - D5 - Network meeting 1 - June-Sep 07
- ✓ 4.5 - Sep - Dec 07 - Report of network meeting
- ✓ 4.6 - Present conclusions at meetings



Deliverables and Actions

Actions arising from Corfu meeting:

- Submit additional questions to questionnaire respondees on OBO ontology and minimum standards usage
- Attempt to get responses from two outstanding projects
- Produce report on questionnaire
- Sponsor workshop on annotation of phenotype data using PATO during 2008
- Discuss further the potential to implement ideas on phenotype/disease ontologies



Thanks To

- All work package participants
- All those who responded to the questionnaire
- Non-CASIMIR speakers at our meetings
- Paul Schofield
- Dawn Muddyman

