

Title: MGED ontologies for consistent annotation of micro-array experiments.

Authors: Philippe Rocca-Serra *, Helen Parkinson *, Susana A. Sansone *, Ele Holloway *, Chris Stoeckert ? and Alvis Brazma *.

*. European Bioinformatics Institute, EMBL-EBI

?. Center for Bioinformatics, University of Pennsylvania

Address: EBI- The European Bioinformatics Institute
EMBL Outstation Hinxton

Wellcome Trust Genome Campus
CB10 1SD Cambridge, UK

Abstract:

The ArrayExpress database is a repository for gene expression experiment data. A critical issue in database management is data quality control and annotation. In order to enable efficient querying and mining, data should not only be sufficiently but also consistently annotated. Minimum Information About a Microarray Experiment requirements (MIAME) (1) and use of ontologies are the tools to address these issues. The first sets the limit to how extensive the annotation should be. Then, ontologies are used as a solution to annotation heterogeneity. Indeed, by capturing semantics, they offer an efficient means for standardizing descriptions. MGED Ontology Working Group is currently developing specific ontologies and CVs for describing experimental conditions and nucleic acid sources in gene expression experiments. This includes implementation of existing ontologies through referencing, avoiding redundancy. An example of the MGED ontology is available from <http://mged.sourceforge.net/Ontologies.shtml>.

This work will evolve, as users will provide us with comments and suggestions. For instance, should the future version of MGED ontology be a unique, global tool, providing internal specifications through concepts (e.g. age, treatment) or kept as separated, cross referenced ontologies ? This topic is actually tightly linked to the integration of ontologies into annotation tools. These issues will be discussed.

(1). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat. Genet. 2001 Dec;29(4):365-71.