BIOINFORMATICS ORIGINAL PAPER Vol. 22 no. 7 2006, pages 866–873
doi:10.1093/bioinformatics/btl005

Databases and ontologies

The MGED Ontology: a resource for semantics-based description of microarray experiments

Patricia L. Whetzel[1], Helen Parkinson[2], Helen C. Causton[3], Liju Fan[4], Jennifer Fostel[5], Gilberto Fragoso[6], Laurence Game[3], Mervi Heiskanen[6], Norman Morrison[7], Philippe Rocca-Serra[2], Susanna-Assunta Sansone[2], Chris Taylor[2], Joseph White[8] and Christian J. Stoeckert, Jr[1,*]

[1]Center for Bioinformatics and Department of Genetics, University of Pennsylvania School of Medicine, USA, [2]European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK, [3]MRC Clinical Sciences Centre, Faculty of Medicine, Imperial College, Hammersmith Hospital Campus, DuCane Road, London W12 0NN, UK, [4]Ontology Workshop LLC, PO Box 182, Columbia, MD 21045-9998, USA, [5]NIEHS, PO Box 12233 MD F1-05, 111 Alexander Drive Research Triangle Park, NC 27709-2233, USA, [6]NCICB, NCI Center for Bioinformatics, 6116 Executive Boulevard, Rockville, MD 20852, USA, [7]Department of Computer Science, Kilburn Building University of Manchester, Oxford Road, Manchester M13 9PL, UK and [8]Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, USA

Received on July 14, 2005; revised on January 6, 2006; accepted on January 13, 2006
Advance Access publication January 21, 2006
Associate Editor: Alvis Brazma

ABSTRACT

Motivation: The generation of large amounts of microarray data and the need to share these data bring challenges for both data management and annotation and highlights the need for standards. MIAME specifies the minimum information needed to describe a microarray experiment and the Microarray Gene Expression Object Model (MAGE-OM) and resulting MAGE-ML provide a mechanism to standardize data representation for data exchange, however a common terminology for data annotation is needed to support these standards.

Results: Here we describe the MGED Ontology (MO) developed by the Ontology Working Group of the Microarray Gene Expression Data (MGED) Society. The MO provides terms for annotating all aspects of a microarray experiment from the design of the experiment and array layout, through to the preparation of the biological sample and the protocols used to hybridize the RNA and analyze the data. The MO was developed to provide terms for annotating experiments in line with the MIAME guidelines, i.e. to provide the semantics to describe a microarray experiment according to the concepts specified in MIAME. The MO does not attempt to incorporate terms from existing ontologies, e.g. those that deal with anatomical parts or developmental stages terms, but provides a framework to reference terms in other ontologies and therefore facilitates the use of ontologies in microarray data annotation.

Availability: The MGED Ontology version.1.2.0 is available as a file in both DAML and OWL formats at http://mged.sourceforge.net/ontologies/index.php. Release notes and annotation examples are provided. The MO is also provided via the NCICB's Enterprise Vocabulary System (http://nciterms.nci.nih.gov/NCIBrowser/Dictionary.do).

Contact: Stoeckrt@pcbi.upenn.edu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

INTRODUCTION

Microarray experiments are both complex and high-throughput, so data storage, management, exchange and annotation present challenges for biologists and bioinformaticians. There are a variety of academic and commercial database systems available (Gardiner-Garden, 2001) for laboratories and institutions as well as community resources such as ArrayExpress (Parkinson *et al*., 2005), the Gene Expresssion Omnibus (Barrett *et al*., 2005) and the Center for Information Biology Gene Expression Database (CiBEX) (Ikeo *et al*., 2003) that provide access to public microarray data. The development and use of the Microarray Gene Expression Object Model (MAGE-OM), and the related XML format (MAGE-ML) (Spellman *et al*., 2002) have provided a common syntactic format for data exchange and a structure that can capture data described according to the Minimum Information About at Microarray Experiment (MIAME) guidelines (Brazma *et al*., 2001). However, neither MIAME nor the MAGE-OM provides explicit terminology to annotate this complex domain. We are therefore faced with the problem of consistently describing methodology, experimental design, sequences and biological samples across diverse resources.

*To whom correspondence should be addressed.

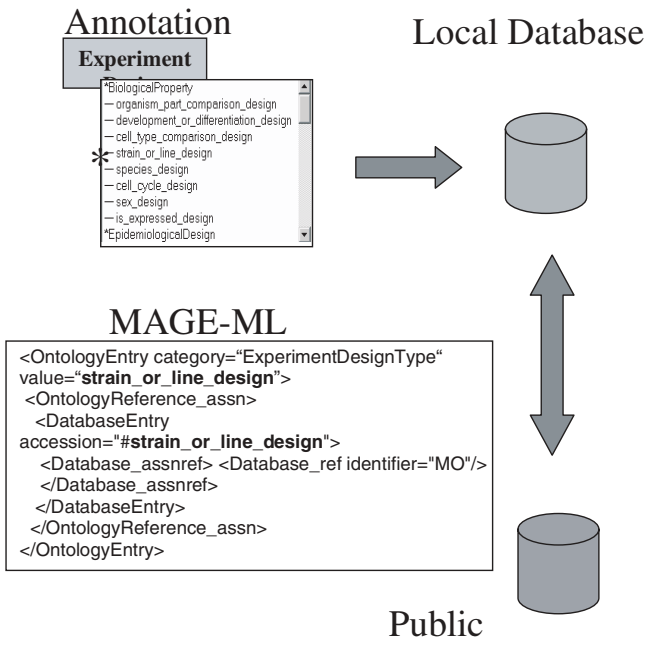


Fig. 1. Illustration of the MO usage in annotation and data transfer with MAGE-ML. Local applications (Table 1) provide terms from the MO organized by MO Classes. These are generally stored in local relational databases from which MAGE-ML can be generated. Data in the MAGE-ML can be transferred between a number of applications and databases, including microarray data repositories in the public domain such as ArrayExpress and GEO.

The MO was developed to provide the semantics required to support the MAGE-OM and as a resource for the development of tools for microarray data acquisition and query (Fig. 1). The MO is primarily an ontology used to annotate microarray experiments, however it contains concepts that are universal to other types of functional genomics experiments such as protocol and experiment design and can thus also be used for annotation of some of the data in these domains. The major component of the ontology involves biological descriptors relating to samples or their processing; it is not an ontology of molecular, cellular or organismal biology, such as the Gene Ontology (Gene Ontology Consortium, 2001).

THE MGED ONTOLOGY CONTENT AND STRUCTURE

The MGED Ontology (MO) is a semantic resource that includes terminology for all aspects of microarray experiments. It was developed by the microarray community and is a species neutral ontology that focuses on the commonalities among experiments rather than the differences between them. In building the MO, we evaluated which ontological resources were needed to describe microarray experiments and developed use cases based on queries of experimental meta-data. Many of the authors manage and/or develop microarray databases and the annotation provided by users of these resources was used as a source of concepts for the ontology in the preliminary card sorting exercise. These contributed to the biological content of the MO. Concepts were mapped between contributors, defined and properties and synonyms were created. The MO was initially released in DAML+OIL format and later in OWL. This set of classes is meant to fulfil the needs of users for annotating biological samples, experiments and sample processing during a microarray experiment.

Users of the MAGE-OM (and the related exchange format MAGE-ML) have contributed to the MO; and in part the MO was developed to support the annotation of data in MAGE-ML format (Fig. 1). The need to support MAGE has had a significant impact on the top-level structure of the MO, while the requirements of the data-generating community have largely determined the content. The impact this has had on the MO is explored below. Although the MO was primarily developed for use by the microarray gene expression community the ontology, like the MAGE-OM, can also be used to describe experiments generated on other functional genomics platforms such as array-centric comparative genome hybridization, chromatin immunoprecipitation on a chip (location analysis) or proteomics experiments and is currently being used for these purposes.

STRUCTURE OF THE MGED ONTOLOGY

The MO consists of two parts: a stable core ontology and an extended ontology . MO version 1.2 contains 229 classes, 110 properties and 658 instances (individuals). The core ontology includes a minimal semantic set that is stable for use in production software and contains all necessary MAGE classes to map the MO content to the MAGE-OM, while the extended ontology permits further development. This bipartite model is also used in the mmCIF vocabulary as part of the Protein Data Bank (Berman *et al*., 2000) and permits evolution of content while ensuring that the basic structure needed for related applications is maintained. Although subclasses are used to organize instances the MGED Core Ontology (MCO) is not highly nested so that it can readily be presented in web-based applications. MCO classes that are referenced in multiple MAGE-OM packages, such as DataType and Scale, are direct subclasses of the MCO. The MCO also contains classes to track terms that have been deprecated and the reason for deprecation.

There are four types of classes used in the MO:

(1) Instantiated MO classes are those that refer to parts of the microarray experiment and contain terms that are common to many experiments. They can be described in terms of properties, contained instances and subclasses (and their properties and values). For example SurfaceType is instantiated within the MO (Fig. 2).

(2) Abstract classes used to provide organization and structure to the MO. For example, the abstract ExperimentDesignType class provides organization to several instantiated subclasses for types of experiments addressing the effects of compounds (PerturbationalDesign class) or addressing the differences between strains (BiologicalProperty class) and instances that describe a particular type of experiment, e.g. time_series_design are provided.

(3) Abstract classes used to represent MAGE classes that have an ontology entry association to allow developers to identify which MO terms to use. For example the PhysicalArrayDesign class is a MAGE class represented in the MO as it has an ontology entry association called SurfaceType (Fig. 2).

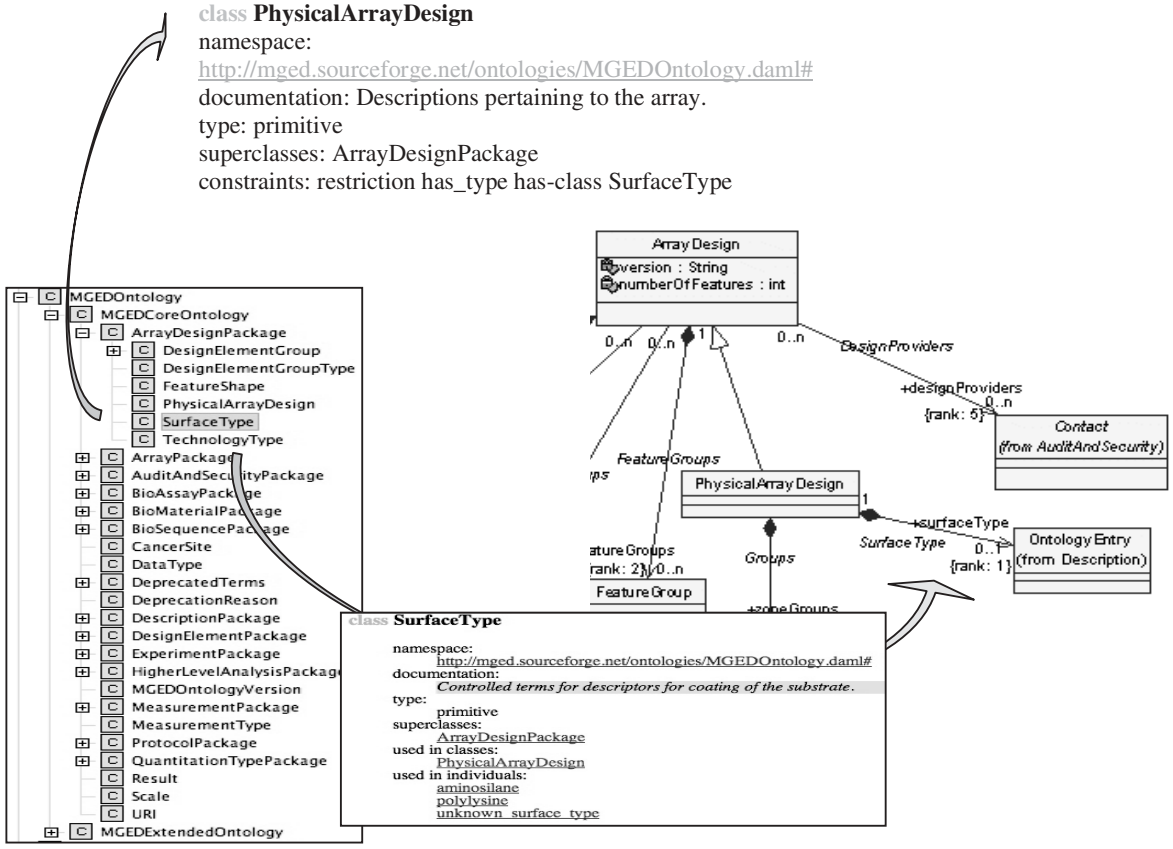


Fig. 2. Class hierarchy of the MO and relationship to the MAGE-OM. In this example, the MAGE-OM specifies a 'surfaceType' association to OntologyEntry from PhysicalArrayDesign. Terms (polylysine, aminosilane, unknown_surface_type) for surface type can be found in the MO in the class 'SurfaceType' which is located in the ArrayDesignPackage class. The relationship of SurfaceType to PhysicalArrayDesign is captured in MO: (PhysicalDesignType has_type SurfaceType).

(4) Abstract classes that are subclasses of OntologyEntry which are instantiated from some other identified resource. For example Organism, Compound, etc.

MGED CORE ONTOLOGY

The MCO hierarchy reflects the structure of the packages in the MAGE-OM and represents a set of is-a relationships in the sense that all the classes are a kind of descriptor for microarray experiments. The top-level classes mimic the MAGE-OM structure and were provided for software developers using MAGE-OM and requiring MO to annotate their MAGE-ML. The lower level classes contain the experimental details used by annotators of microarray experiments and are usually presented in the context of some annotation or query application. The top-level MCO class names therefore are the same as the packages in the MAGE-OM and the MCO instantiated classes are named after the association to the MAGE-OM OntologyEntry class. The MCO does not duplicate the entirety of MAGE-OM, but includes only those classes in MAGE-OM that have an association to the OntologyEntry class. Therefore, navigating from MAGE-OM to the MO requires no concept mapping. This decision was taken after discussion with the developers of MAGE-OM and with the input of the MGED advisory board. The alternative—to build a stand alone ontology and map it to MAGE-OM later was not practical as there was considerable demand for the MO from those using the MAGE-OM. A MAGE-OM view is therefore explicit within the MO. The MCO uses organizing subclasses so that similar types of terms are grouped together within a class, these obey the is-a heirarchy. For example, the class ExperimentDesignType contains five subclasses: PertubationalDesign, MethodologicalDesign, BiologicalProperty, EpidemiologicalDesign and BioMolecularAnnotation. The additional subclasses separate terms such as compound_treatment_design from replicate_design and reduces the list from 52 terms for all classes of ExperimentDesignType to a maximum of 16 terms within the subclass BiologicalProperty.

MO CLASSES, PROPERTIES AND ATTRIBUTES

Experimental or sample descriptors in the MO fall into one of three categories: the types of information (classes) that need to be captured, their properties (attributes) and the actual values (instances) used. All classes, properties and instances in MO are defined in natural language. Synonyms, exact and non-exact, are included in the definition for the term as OilEd, the software used for the initial development of the MO, has limited synonym handling at the instance level (Bechhofer *et al*., 2001).

For example in a hypothetical study in which mice were injected with a drug, categories or classes for 'Organism' are provided in the MCO, to indicate that mice were used, for 'Compound' to indicate which substance, drug or chemical was used, and for 'Treatment' to indicate how the compound was administered to the mice. Classes are also provided for Age, Sex, Strain and other characteristics relating to the mice. The classes from the MCO can be instantiated or abstract as described in the previous section.

Abstract classes (type 4) having instances external to the MO are all subclasses of the OntologyEntry class and inherit properties including a reference to a database and a URI. The database entry association specifies the type of semantic resource, e.g. organism database, compound database, and the URL provides the web address of the resource. This information identifies the term as being external to the MO and the class that it instantiates as internal to the MO.

Classes of this type, such as Compound, cannot easily be provided in an itemized list within the MCO as the number of terms needed is large and such terms are present in external resources. Many of these classes are the focus of efforts by other groups to generate ontologies or various types of controlled vocabularies. MO therefore provides pointers to relevant efforts, for example, in the case of 'Compound' as ChemIDplus (Tomasulo, 2002), available from the National Library of Medicine, which includes 350 000 chemical records that can be searched by CAS Registry Number.

Other examples of this type of abstract class include 'Organism', for which the taxonomy is available from the National Center for Biotechnology Information (Wheeler *et al*., 2005), and 'Disease'. For some classes multiple non-orthogonal choices are available, such as GALEN (Rogers *et al*., 2001), ICD-9 and the nascent Disease Ontology (http://diseaseontology.sourceforge.net/). It is clear that in some cases there are competing efforts, e.g. there are several mammalian anatomy ontologies. The MO does not attempt to provide mappings between synonymous terms in different ontologies, or preferentially recommend one over the other instead, it provides source information for these terms, which in turn can be queried.

On occasion, an external ontology emerges which supersedes part of the MO. The Sequence Ontology (SO) (Eilbeck, 2005) is used for semantics relating to sequence features and describes properties of the sequences represented on the array (exon, gene, etc.). The SO was found to be non-orthogonal with instances from the MO class BioSequenceType. A mapping was therefore performed between the MO terms and the SO terms. As the SO has matured the corresponding MO terms have been deprecated in favour of using the SO directly.

Where there are incomplete term lists MO can be used to extend these, e.g. instances of light units were absent from the list of terms provided by the MAGE-OM and were therefore included in the MO. The MO is extensible while the MAGE-OM is not and it is likely that future versions of the MAGE-OM will devolve all semantic content to a supporting ontology.

USING AND ACCESSING THE MO

The MO is primarily used in three ways:

(1) Embedded within an application to annotate or query microarray data, e.g. by biologists who may have little knowledge of the MO structure.
(2) Directly for annotating microarray data, e.g. by an annotator.
(3) For producing an application that uses the MO, e.g. by a software developer.

This diversity among uses and user groups is similar to that of the Gene Ontology which is used in many applications including direct use by annotators who select appropriate terms for a given gene product. Access to the MO is provided in line with the needs of each of these user groups.

(1) MO files are available in their native OWL format with release notes for developers who typically parse the OWL file and use it locally to build an application seen by biologists.
(2) Via web browser access of the NCI Metathesaurus which allows the tree structure to be visualized and navigated.
(3) Via a web page where a URL identifies each each Class, Property or instance in the ontology e.g. http://mged.sourceforge.net/ontologies/MGEDontology.php#polylysine.

In anticipation of providing MO terms through web services, the MO is registered with BioMoby.

USE OF THE MO FOR DATA ANNOTATION

Use of the MO is best demonstrated by considering an example in which the ontology is used to describe part of a microarray experiment. The information obtained from the biologist is free text:

'A murine embryo fibroblast cell line (Swiss 3T3-L1) was plated out. Two plates were treated with 10 nM insulin, two with 100 nM insulin and the other two were left untreated. The cells were harvested after 4 hours incubation.'

This description can be annotated using terms from the MO (Fig. 3).

The experiment is a kind of PerturbationalDesign, and instances from this class dose_response_design, compound_treatment_design further describe how the experiment was conducted. The cell type and cell lines are described using the MO terms 'CellLine' and 'CellType' respectively, however, the MO does not include instances that specify particular cell lines or cell types so other, domain specific, ontologies need to be referenced. Here the MO is used to refer to the terms 'Fibroblast' and '3T3-L1 Cells' from the NCI Metathesaurus. Further examples of how the MO can be used to annotate experiments can be found at http://microarray.csc.mrc.ac.uk/_private/Support/development_page.htm Systematically annotated and published experiments can also be downloaded, along with the MAGE-ML used for data transfer from public repositories such as ArrayExpress. One example of a published experiment that has been annotated using the MO and exported as MAGE-ML can be accessed at http://www.ebi.ac.uk/arrayexpress/query/result;jsessionid=7D17C32BFAAED8D3CBDC49F697582C31?queryFor=Experiment&eAccession=E-MiMR-12&eSpecies=&eAuthor=&eArrayAccession=&eExperimentType=&eLaboratory=&eArrayDesignName=&eExperimentalFactor=&ePublication=&eArrayProvider=&eDescription= (Kemp *et al*., 2003).

ENCODING THE MO IN MAGE-ML

MO concepts are typically expressed as MAGE-ML when annotated microarray data are exchanged. The MAGE-OM

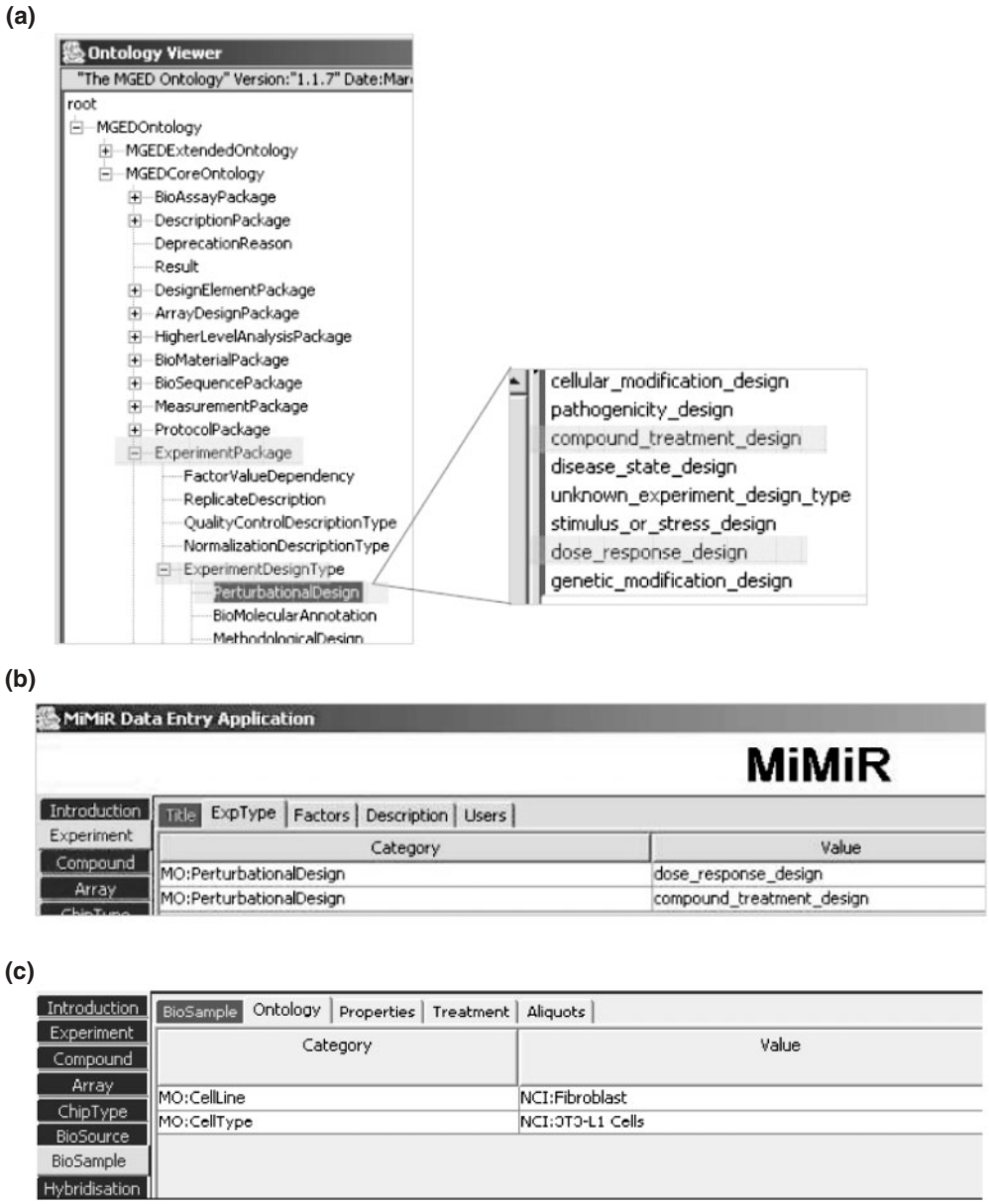


Fig. 3. Panel (**a**) shows an expanded view of the MO and the terms that are relevant for describing the design of an experiment in which cells were treated with one of two concentrations of insulin. Panels (**b**) and (**c**) illustrate how this information is represented in MiMiR (Navarange *et al*., 2005), one of the applications used for data annotation and management that incorporates the MO. Terms selected from the MGED Ontology have the prefix 'MO:' and those from the NCI Metathesaurus have the prefix 'NCI:'.

recognizes that semantics are required and provides a mechanism to provide semantic content via the MAGE-OM OntologyEntry. The MAGE-ML format was not built to express complex concepts parsimoniously and relationship types cannot currently be expressed in MAGE owing to limitations in the MAGE-OM. As a consequence, the MAGE-ML structure becomes complex when represented in MAGE (even though the ontology is not deeply nested) and leads to XML bloat and the need for a rule-based system for application-processing semantics. This has been implemented by ArrayExpress and is used to process complex MAGE-ML coding to a simpler state for local queries. The XML bloat inherent in the representation of any ontology in MAGE-ML will not be addressed completely until the next version of MAGE becomes available, so annotation examples and pseudo code have been generated to assist developers to use the MO in the context of the MAGE-OM. These examples are provided to promote consistent use of the MO. An ontology helper module for the MAGEstk (Spellman *et al*., 2002) for both Java and Perl is also under development to support coding of the MO in MAGE-ML (code available from http://cvs.sourceforge.net/viewcvs.py/mged/MAGE-Java/MGEDOntologyEntry/).

Table 1. Microarray resources that use the MGED ontology

Name	URL
BuG@Sbase	http://bugs.sgul.ac.uk/bugsbase/
CaArray	http://caarraydb.nci.nih.gov/caarray/index.jsp
Chemical effects in biological systems	http://cebs.niehs.nih.gov/
Maxd	http://bioinf.man.ac.uk/microarray/maxd/
MIAMExpress	http://www.ebi.ac.uk/miamexpress/
MiMiR	http://microarray.csc.mrc.ac.uk/_private/activities/data_warehouse_text.htm
NASCArrays	http://affymetrix.arabidopsis.info/narrays/experimentbrowse.pl
RAD StudyAnnotator	https://www.cbil.upenn.edu/RAD/StudyAnnotator/
SMD experiment set creator	http://genome-www.stanford.edu/microarray
Tox-MIAMExpress	http://www.ebi.ac.uk/tox-miamexpress/

USE OF THE MO IN APPLICATIONS FOR DATA ANNOTATION

The MO has been implemented in web-based microarray annotation applications (Table 1) such as MIAMExpress (Parkinson *et al*., 2005), Tox-MIAMExpress (Mattes *et al*., 2004), RAD Study Annotator (Manduchi *et al*., 2004) and MiMiR (Navarange *et al*., 2005). These applications provide forms for annotating the components of a microarray experiment specified by MIAME and the MO terms are typically presented in menus from which terms may be selected as part of a web interface. Different strategies have been chosen for managing the MO. RAD databases a local copy of the MO, maxdLoad2 presents a simplified abstraction of the MO graph while utilizing the full set of terms if desired, and MIAMExpress abstracts instantiated classes for local use. Tox-MIAMExpress abstracts those MO classes relevant to the description of chemical treatments and toxicological endpoints (e.g. Compound, Histology, Observation for macroscopic records, Test for clinical chemistry assays). Once the data are submitted to a public repository such as ArrayExpress, ontology-driven annotation will provide users with a powerful means to query microarray experiments. The MO has also been made available directly via the NCICB's Enterprise Vocabulary System (Covitz *et al*., 2003) and is used by NCICB applications such as caArray.

REVISING AND EXTENDING THE MO

The initial motivation for development of the MO was provided by the microarray data community who presented a real and immediate need for terms for data description and support for the MAGE-OM. Although much of the terminology needed by the community was provided in the early releases, technology is evolving rapidly and examples of novel requirements for data annotation arise continually. This however can conflict with the need to maintain the stable core structure. The MO can therefore be extended in the following two ways

(1) By adding new Classes and/or instances to the MGED Extended Ontology (MEO).

(2) By addition of new instances to existing classes according to development rules.

The MEO provides a framework for adding new classes that are not currently part of the MCO. This ensures that the wider community can identify new terms for data annotation within the MO and see the relationships among them, promotes systematic use of terminology and allows areas for further development to be readily identified for future releases. The MEO also contains classes from previous versions that represent knowledge we want to maintain, but which do not fit into the current version of the MCO.

When a term required for annotating an experiment is not available in the MO users may add their own terms and definitions using one of the applications implementing MO. User defined terms are curated by the MO developers via the MO tracker and are added to the MO provided they are (1) not domain or species specific and (2) are orthogonal (do not overlap) with existing concepts. The MO website also provides release notes for each version of the MO that represent approved changes to the MO such as corrections, or new instances. MO development and maintenance activities such as proposals for new terms or modifications to definitions are discussed via the MO tracker and curated by the MO working group (Fig. 4) (http://sourceforge.net/tracker/?atid=603031&group_id=16076&func=browse).

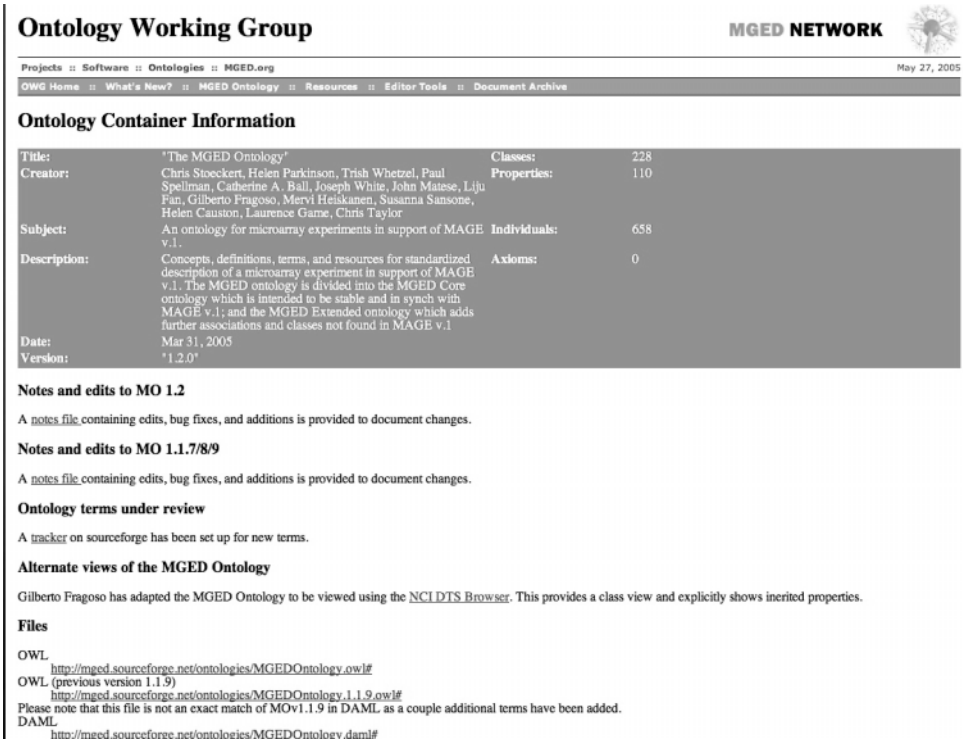
DISCUSSION

The MO supports MAGE-OM v1 and v1.1 and provides descriptors for microarray experiments for use by biologists and software developers. The MO is in active use by both of these communities of users, however, the ontology is also evolving in line with their needs. Areas for future development include the addition of terms for describing normalization and data transformation, and the review of existing term usage in resources using the MO.

Changes are also being made to leverage the improved representational power provided by OWL (the ontology was migrated from DAML+OIL to OWL representation for this reason). Changes include the use of synonyms in definitions of terms, the display of class trees (see http://mged.sourceforge.net/ontologies/MGEDontology.php for a summary of changes made) and use of Annotation properties for annotating MAGE classes explicitly.

The MO is provided as a Resource Description Framework (RDF)-based file in either the DAML or OWL formats. This format enables direct programmatic queries in the form of web services that use software libraries which parse the RDF graph from XML (e.g. http://www.redland.opensource.ac.uk/). We envision searching for MO terms via web services at central registries such as BioMOBY

(a)



Ontology Working Group

MGED NETWORK

Projects :: Software :: Ontologies :: MGED.org

May 27, 2005

OWG Home :: What's New? :: MGED Ontology :: Resources :: Editor Tools :: Document Archive

Ontology Container Information

Title:	'The MGED Ontology'	Classes:	228
Creator:	Chris Stoeckert, Helen Parkinson, Trish Whetzel, Paul Spellman, Catherine A. Ball, Joseph White, John Matese, Liju Fan, Gilberto Fragoso, Mervi Heiskanen, Susanna Sansone, Helen Causton, Laurence Game, Chris Taylor	Properties:	110
Subject:	An ontology for microarray experiments in support of MAGE v.1.	Individuals:	658
Description:	Concepts, definitions, terms, and resources for standardized description of a microarray experiment in support of MAGE v.1. The MGED ontology is divided into the MGED Core ontology which is intended to be stable and in synch with MAGE v.1; and the MGED Extended ontology which adds further associations and classes not found in MAGE v.1	Axioms:	0
Date:	Mar 31, 2005		
Version:	"1.2.0"		

Notes and edits to MO 1.2

A notes file containing edits, bug fixes, and additions is provided to document changes.

Notes and edits to MO 1.1.7/8/9

A notes file containing edits, bug fixes, and additions is provided to document changes.

Ontology terms under review

A tracker on sourceforge has been set up for new terms.

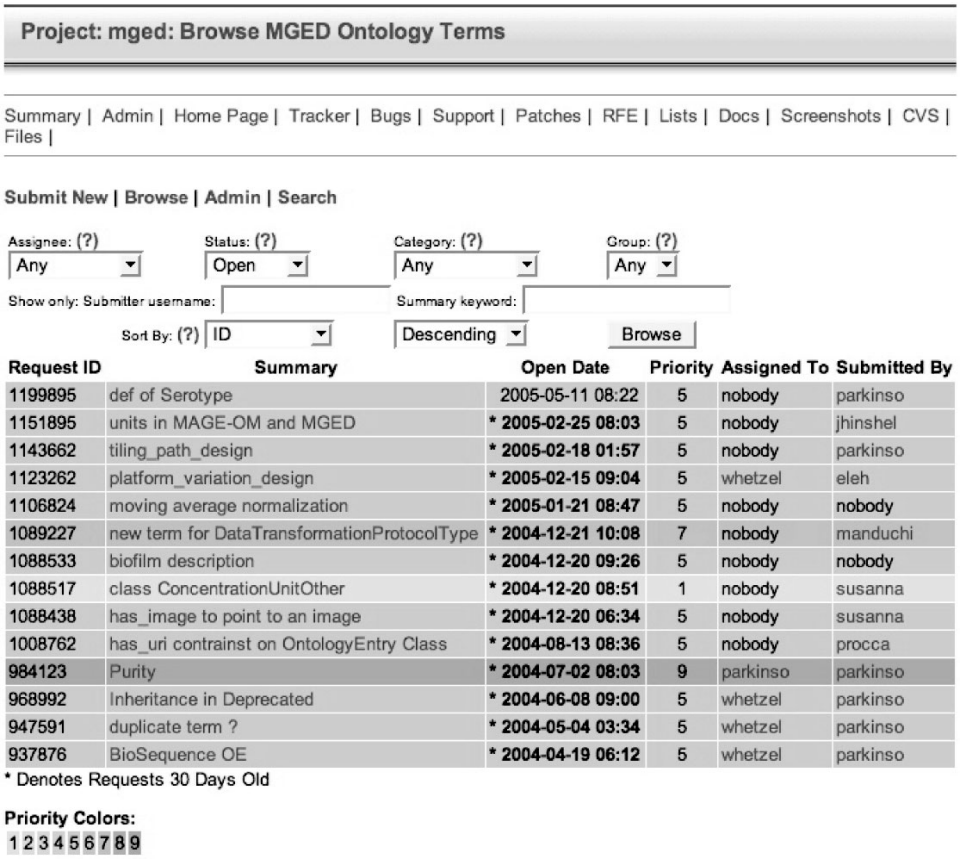
Alternate views of the MGED Ontology

Gilberto Fragoso has adapted the MGED Ontology to be viewed using the NCI DTS Browser. This provides a class view and explicitly shows inerited properties.

Files

OWL
http://mged.sourceforge.net/ontologies/MGEDOntology.owl#
OWL (previous version 1.1.9)
http://mged.sourceforge.net/ontologies/MGEDOntology.1.1.9.owl#
Please note that this file is not an exact match of MOv1.1.9 in DAML as a couple additional terms have been added.
DAML
http://mged.sourceforge.net/ontologies/MGEDOntology.daml#

(b)



Project: mged: Browse MGED Ontology Terms

Summary | Admin | Home Page | Tracker | Bugs | Support | Patches | RFE | Lists | Docs | Screenshots | CVS | Files |

Submit New | Browse | Admin | Search

Assignee: (?) Any | Status: (?) Open | Category: (?) Any | Group: (?) Any

Show only: Submitter username: | Summary keyword:

Sort By: (?) ID | Descending | Browse

Request ID	Summary	Open Date	Priority	Assigned To	Submitted By
1199895	def of Serotype	2005-05-11 08:22	5	nobody	parkinso
1151895	units in MAGE-OM and MGED	* 2005-02-25 08:03	5	nobody	jhinshel
1143662	tiling_path_design	* 2005-02-18 01:57	5	nobody	parkinso
1123262	platform_variation_design	* 2005-02-15 09:04	5	whetzel	eleh
1106824	moving average normalization	* 2005-01-21 08:47	5	nobody	nobody
1089227	new term for DataTransformationProtocolType	* 2004-12-21 10:08	7	nobody	manduchi
1088533	biofilm description	* 2004-12-20 09:26	5	nobody	nobody
1088517	class ConcentrationUnitOther	* 2004-12-20 08:51	1	nobody	susanna
1088438	has_image to point to an image	* 2004-12-20 06:34	5	nobody	susanna
1008762	has_uri contrainst on OntologyEntry Class	* 2004-08-13 08:36	5	nobody	procca
984123	Purity	* 2004-07-02 08:03	9	parkinso	parkinso
968992	Inheritance in Deprecated	* 2004-06-08 09:00	5	whetzel	parkinso
947591	duplicate term ?	* 2004-05-04 03:34	5	whetzel	parkinso
937876	BioSequence OE	* 2004-04-19 06:12	5	whetzel	parkinso

* Denotes Requests 30 Days Old

Priority Colors:
1 2 3 4 5 6 7 8 9

Fig. 4. Views of the MO. Panel (**a**) shows an html version of the MO available at http://mged.sourceforge.net/ontologies/MGEDontology.php along with links to files, notes and other views. Panel (**b**) The MO tracker at Sourceforge is used to coordinate development.

(http://www.biomoby.org/) and through annotation forms provided as part of microarray data management applications. Thus, anyone requiring a term from the most recent version of MO would be able to use the web service from their application to view the available data for classes, properties and instances and the relationships between them.

The MO has been implemented in annotation tools such as MIAMExpress, the RAD Study Annotator, SMD, MiMiR and others (Table 1). The groups managing and populating these resources collectively generate large amounts of data that present a rich source of information annotated with a common terminology. The use of common annotation among laboratories and experiments is expected to enhance the utility of all the data and to facilitate queries and data mining and thousands of experiments have been annotated using the MO to date.

The MO was originally developed to support the annotation of microarray experiments, however, many of the MO classes describing biomaterials, protocols and experimental design are independent of the technology used and applies to other functional genomics technologies (such as mass spectrometry, *in situ* hybridization, etc.). It is hoped that initiatives to provide standards in these other domains will leverage the terms and relationships contained in the MO. Work towards the development of a Functional Genomics Experiment Ontology (FuGO, http://fugo.sourceforge.net has already begun as part of a collaboration between the MO Working Group, the MGED Reporting Structure for Biological Investigations (RSBI, http://www.mged.org/Workgroups/rsbi/rsbi.html), the HUPO Proteomics Standards Initiative (http://psidev.sourceforge.net/) and the Metabolomic Society (http://www.metabolomicssociety.org/mstandards.html, Lindon *et al*., 2005) working groups. The resulting ontology will provide a consistent mechanism for annotating functional genomics experiments that encompass different technological and biological domains and assist in comparison of data across modalities. In the same way that the MO was developed in parallel with the MAGE-OM, FuGO will be developed in parallel with a Functional Genomics Object Model (FuGE; http://fuge.sourceforge.net/). The problems of representing complex semantics in an XML format, and the need to permit evolution of the ontology which have been problematic for the MO will inform such developments. In particular the difficulties in modelling a complex domain and developing an ontology simultaneously have resulted in a product that is MAGE-OM centric and therefore of limited use with other object models. We hope to avoid this in future by providing mapping to relevant object models rather than encoding these in the ontology. With this in mind we are currently reviewing the MO, with a view to participate in the development of FuGO. While FuGO is being developed the MO will continue to be maintained and extended for use in microarray-specific applications.

ACKNOWLEDGEMENTS

The authors would like to thank the members of the Ontology Working Group who have contributed to the MO especially Catherine Ball, Paul Spellman, John Matese and Angel Pizarro. The authors would also like to thank Robert Stevens for his help and guidance and the reviewers who provided a number of constructive comments that significantly improved this manuscript. This work was supported in part by NIH grant 1P41HG003619-01 and NIH-NIEHS contract 273-02-C-0027. Funding to pay the Open Access publication charges for this article was provided by NIH P41HG003619.

Conflict of Interest: none declared.

REFERENCES

Ball,C.A. *et al*. (2005) The Stanford Microarray Database accommodates additional microarray platforms and data formats. *Nucleic Acids Res*., **33**, D580–D582.

Barrett,T. *et al*. (2005) NCBI GEO: mining millions of expression profiles—database and tools. *Nucleic Acids Res*., **33**, D562–D566.

Bechhofer,S., Horrocks,I., Goble,C. and Stevens,R. (2001) OilEd: a Reason-able Ontology Editor for the Semantic Web. *Proc. KI2001*, **2174**, 396–408.

Berman,H.M. *et al*. (2000) The Protein Data Bank. *Nucleic Acids Res*., **28**, 235–242.

Brazma,A. *et al*. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. *Nat. Genet*., **29**, 365–371.

Covitz,P.A. *et al*. (2003) caCORE: a common infrastructure for cancer informatics. *Bioinformatics*, **19**, 2404–2412.

Eilbeck,K. (2005) The Sequence Ontology. *Comp. Funct. Genomics*, **5**, 642–647.

Gardiner-Garden,M. and Littlejohn,T.G. (2001) A comparison of microarray databases [Erratum (2001) *Brief. Bioinform*., **2**, 220.]. *Brief. Bioinform*., **2**, 143–158.

Gene Ontology Consortium. (2001) Creating the gene ontology resource: design and implementation. *Genome Res*., **11**, 1425–1433.

Ikeo,K. *et al*. (2003) CIBEX: center for information biology gene expression database. *C. R. Biol*., **326**, 1079–1082.

Kemp,T.J. *et al*. (2003) Changes in gene expression induced by H(2)O(2) in cardiac myocytes. *Biochem. Biophys. Res. Commun*., **307**, 416–421.

Lindon,J.C. *et al*. (2005) Summary recommendations for standardization and reporting of metabolic analyses. *Nat. Biotechnol*., **23**, 833–838.

Manduchi,E. *et al*. (2004) RAD and the RAD Study-Annotator: an approach to collection, organization and exchange of all relevant information for high-throughput gene expression studies. *Bioinformatics*, **20**, 452–459.

Mattes,W.B. *et al*. (2004) Database development in toxicogenomics: issues and efforts. *Environ Health Perspect*., **112**, 495–505.

Navarange,M. *et al*. (2005) MiMiR: a comprehensive solution for storage, annotation and exchange of microarray data. *BMC Bioinformatics*, **6**, 268.

Parkinson,H. *et al*. (2005) ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res*., **33**, D553–D555.

Rogers,J. *et al*. (2001) GALEN ten years on: tasks and supporting tools. *Medinfo*., **10**, 256–260.

Spellman,P.T. *et al*. (2002) Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biol*., **3**, RESEARCH0046.

Tomasulo,P. (2002) ChemIDplus-super source for chemical and drug information. *Med. RefServ. Q*, **21**, 53–59.

Wheeler,D.L. *et al*. (2005) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*., **33**, D39–D45.