



MGED Standards usage for Toxicogenomics

Philippe Rocca-Serra, EMBL-EBI

Microarray Team



Application to Toxicogenomics - Overview

Toxicogenomics : *The study of Stressors (chemical, radiation, environmental factor) and their effects on biological systems.*

Evaluation and, potentially, Validation of high-throughput technologies (initially, microarray for gene expression) in classic toxicology and monitoring of genomic response to toxicant exposure.

Evaluation of Bioinformatics requirements for information sharing.

Definition of information sharing policies.

Definition of specification and requirement for data submissions [Applied for complex annotation representation in Tox-MIAMExpress]

How to use the Standards ?

Recurrent Questions when setting up pipelines

i. What Should I use in the model ?

ii. What can I leave out ?

i.e . What is really mandatory and what can be considered as optional !

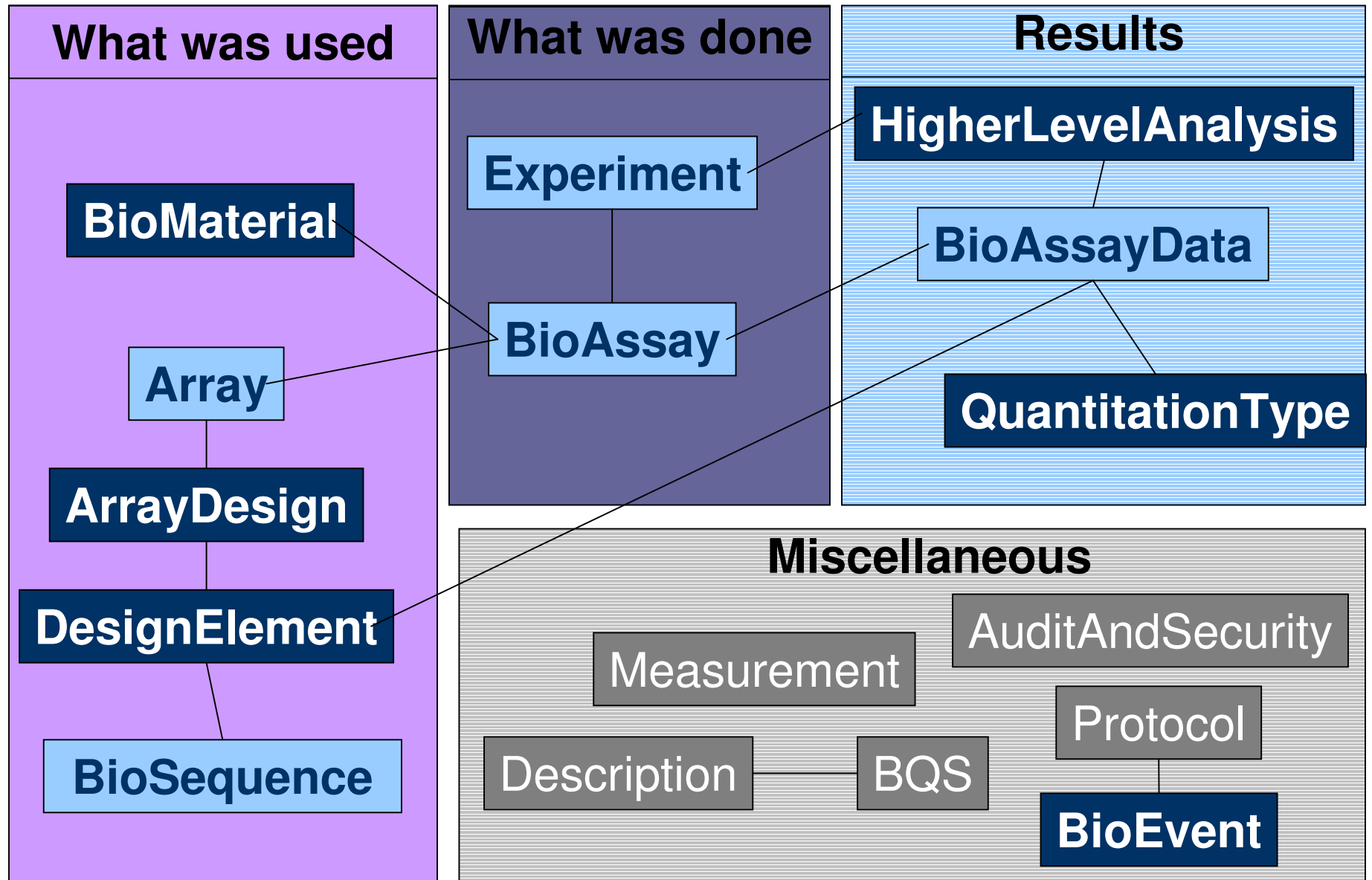
-> Need for guidelines on how to use the model

-> MAGE good practice: a set of rules and guidelines for MAGE-ML users :

- ensuring efficient data capture and formatting.

- ensuring efficient training (and therefore diffusion of Standards)

MAGE overview



(U. Sarkans)

Adapting MAGE to domain specific application ?

Recurrent Issue with first users MAGE

i. I can not find a class for “....” in MAGE? Should I dump MGED standards ?

- MAGE is OMG accepted standard for Microarray world.

- can not be altered before MAGE-Mk II.

ii. MAGE is array centric and leaves biological world out of the scope of the model

iii. How to reconcile Model and Annotation ?

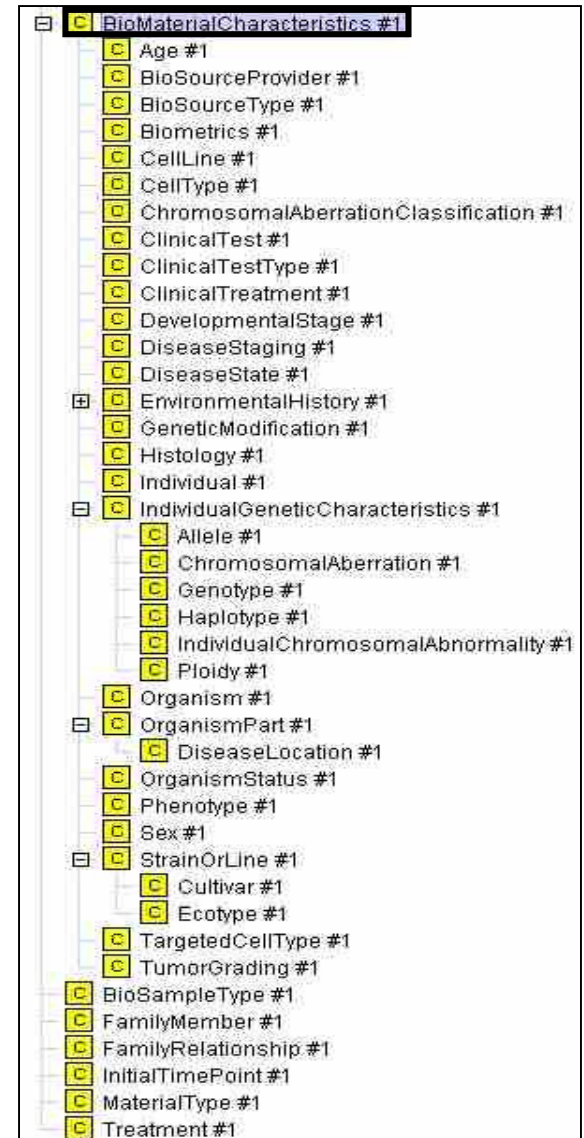
MGED Ontology (MO)

Ontology -
WG

C. Stoeckert
U.Penn

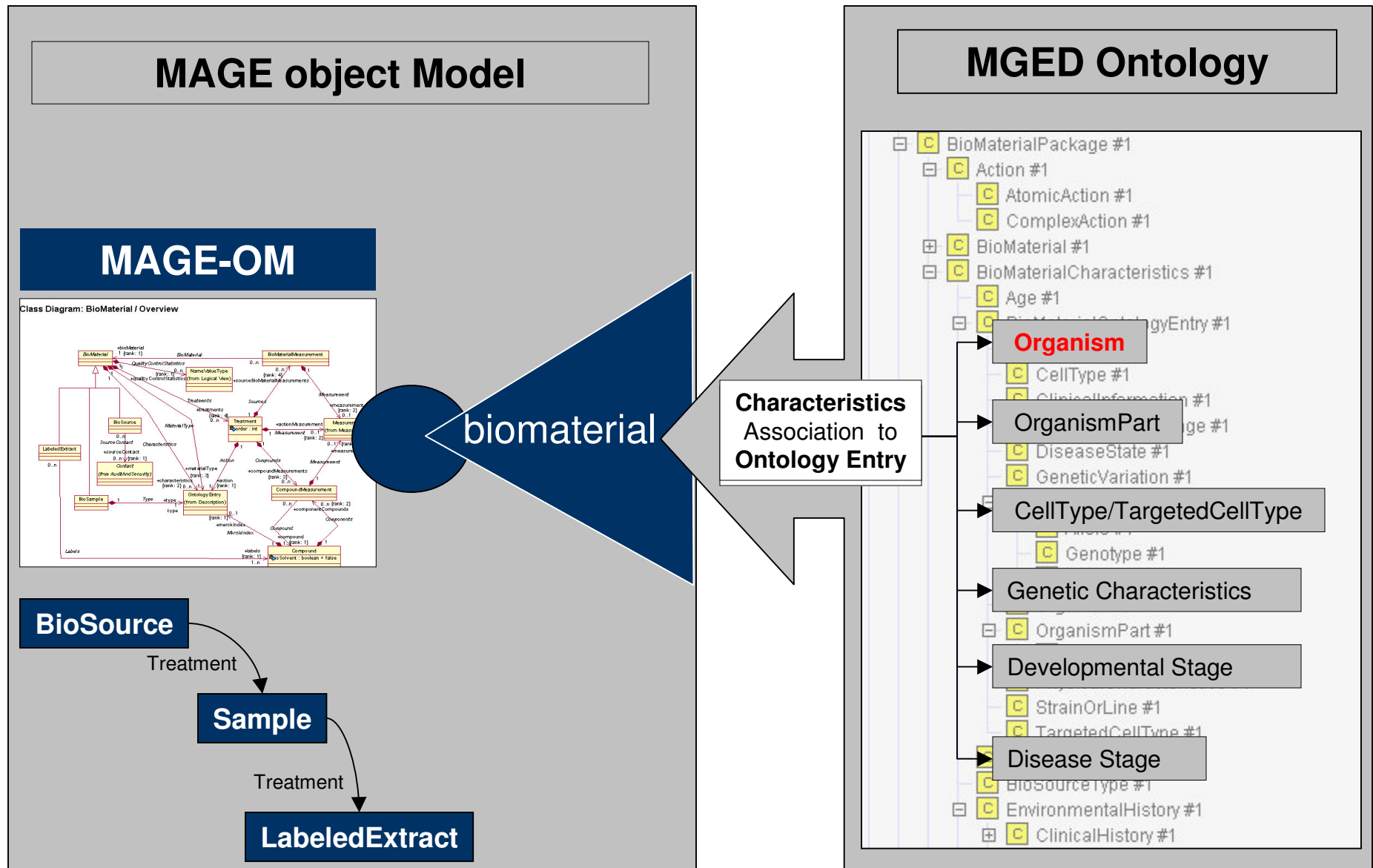
Standard for
experiment
annotation
MGED Ontology

- Formal Knowledge Representation
- Tested vs reasoner
- Standard terms
 - Unambiguous description
 - Structured queries
- To be used by
 - Investigators
 - Software
 - Databases

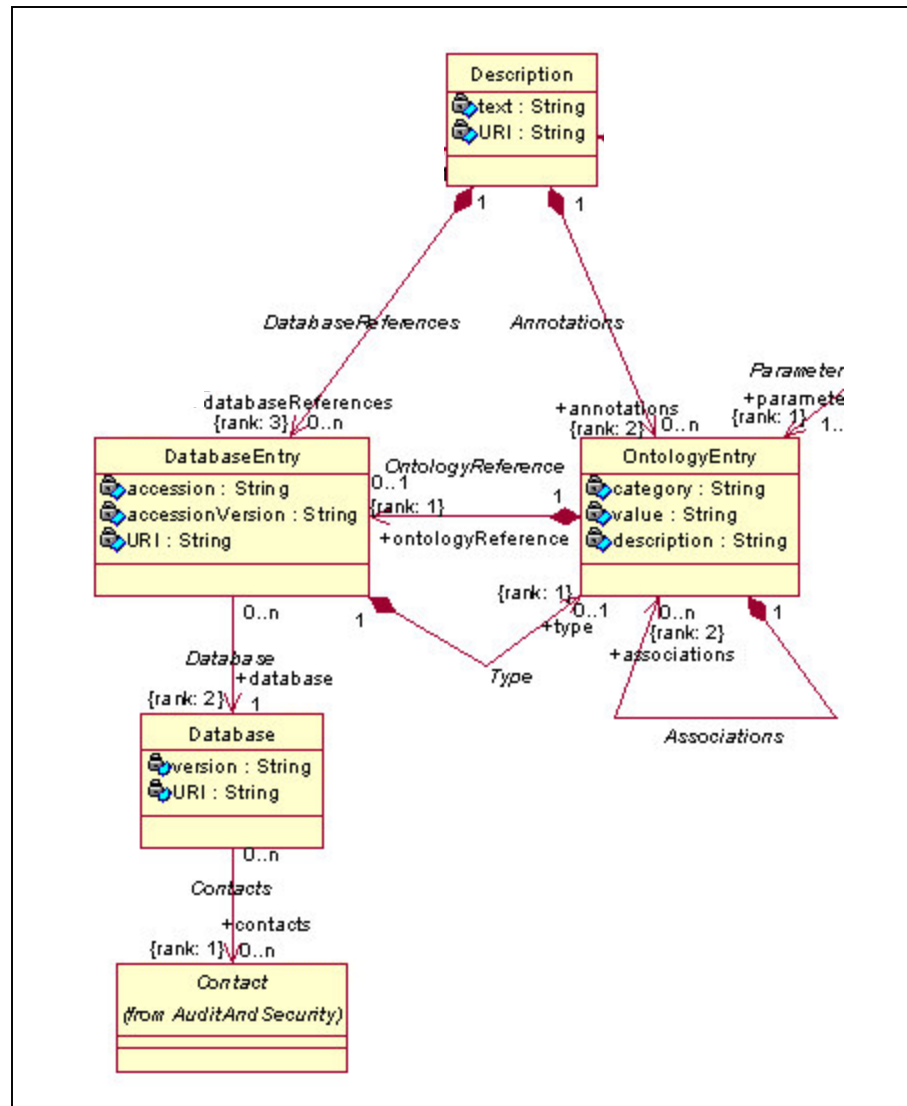


(Stoeckert et al. Nature Gen, 2002)

How can Standards work together?



How to Code MO in MAGE ?



Description Package:

Feature of the

Ontology Entry Class

- Association to itself
- Association to DatabaseEntry



How to Code MO in MAGE ?

PseudoCode for Coding MO in MAGE-ML

A set of rules has been devised

**First presented during MGED6
meeting in September 2003**

Formally posted March 2004

**Extensively Applied for complex
annotation representation in Tox-
MIAMExpress**

```
DatabaseEntry (OntologyEntry) {
    create MAGE DatabaseEntry that references whatever is in
    OntologyEntry.value, if applicable
}

#rule1:
-----
CreateOE(Term ) {
if (Term == Instance) => {
    OE.category = parent of Term ,
    OE.value = instance term
    OE.OntologyReference = DatabaseEntry(OE);
}

#rule2:
-----

elseif (Term == Class || Term == Property ) => {
    OE.category = term ,
    OE.value = term,
    OE.OntologyReference = DatabaseEntry(OE);
    OE.Association = [ foreach ChildTerm of Term
                        { CreateOE(ChildTerm) }
                      ]
}

#rule3:
-----

else (Term == user-defined ) => {
    OE.category = parent of Term || base type of Term || Term
    OE.value = Term
}
}
```

How to Code MO in MAGE ?

Simplest Case:

Class with instances

Example: >>>>>>>>>>

To be compared to
initial way of coding >>>

```
<OntologyEntry category="SomeMOClass" value="some_MO_instance">
  <OntologyReference_assn>
    <DatabaseEntry accession="# some_MO_instance "
URI="http://mged.sourceforge.net/ontologies/MGEDontology.php#
some_MO_instance ">
    <Database_assnref>
      <Database_ref identifier="MO"/>
    </Database_assnref>
  </DatabaseEntry>
</OntologyReference_assn>
</OntologyEntry>
...
<Database name="The MGED Ontology" identifier="MO"
URI="http://mged.sourceforge.net/ontologies/MGEDontology.php" />
```

```
<OntologyEntry category="MaterialType" value="cell_lystate">
  <OntologyReference_assn>
    <DatabaseEntry accession="#cell_lystate"
URI="http://mged.sourceforge.net/ontologies/MGEDontology.php#cell_lystate">
    <Database_assnref>
      <Database_ref identifier="MO"/>
    </Database_assnref>
  </DatabaseEntry>
</OntologyReference_assn>
</OntologyEntry>
```

```
<OntologyEntry category="MaterialType" value="cell_lystate">
</OntologyEntry>
```

How to Code MO in MAGE ?

Subclass/Class:

Rule of thumb:

**Go to root class to
provide higher
level container**

(bottom-up)



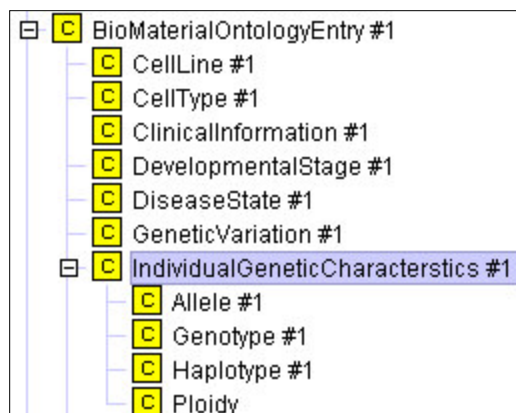
Subclass

(Same as previous)

```
<OntologyEntry category="Some MO super class ">
  <OntologyReference_assn>
    <DatabaseEntry accession="# Some MO super class "
URI="http://mged.sourceforge.net/ontologies/MGEDontology.php# Some MO super
class ">
      <Database_assnref>
        <Database_ref identifier="MO"/>
      </Database_assnref>
    </DatabaseEntry>
  </OntologyReference_assn>
  <Associations_assnlist>
    <OntologyEntry category=" SomeMOClass " value=" SomeMOClass ">
      <OntologyReference_assn>
        <DatabaseEntry accession="# SomeMOClass "
URI="http://mged.sourceforge.net/ontologies/MGEDontology.php# SomeMOClass ">
          <Database_assnref>
            <Database_ref identifier="MO"/>
          </Database_assnref>
        </DatabaseEntry>
      </OntologyReference_assn>
    </Associations_assnlist>
```

```
<OntologyEntry category=" SomeMOClass" value=" some_MO_instance">
  <OntologyReference_assn>
    <DatabaseEntry accession="#some_MO_instance"
URI="http://mged.sourceforge.net/ontologies/MGEDontology.php#some_MO_instance
">
      <Database_assnref>
        <Database_ref identifier="MO"/>
      </Database_assnref>
    </DatabaseEntry>
  </OntologyReference_assn>
</OntologyEntry>
</Associations_assnlist>
</OntologyEntry>
```

How to Code MO in MAGE ?



Example for previous slide

Coding Diploid status

```
<OntologyEntry category=" IndividualGenetic Characteristics " value="
IndividualGenetic Characteristics ">
  <OntologyReference_assn>
    <DatabaseEntry accession="# IndividualGenetic Characteristics "
URI="http://mged.sourceforge.net/ontologies/MGEDontology.php# IndividualGenetic
Characteristics ">
    <Database_assnref>
    <Database_ref identifier="MO"/>
    </Database_assnref>
  </DatabaseEntry>
</OntologyReference_assn>
<Associations_assnlist>
  <OntologyEntry category="Ploidy" value="Ploidy">
    <OntologyReference_assn>
      <DatabaseEntry accession="#Ploidy"
URI="http://mged.sourceforge.net/ontologies/MGEDontology.php#Ploidy">
        <Database_assnref>
          <Database_ref identifier="MO"/>
          </Database_assnref>
        </DatabaseEntry>
      </OntologyReference_assn>
```

```

<OntologyEntry category="Ploidy" value=" diploid">
  <OntologyReference_assn>
    <DatabaseEntry accession="#diploid"
URI="http://mged.sourceforge.net/ontologies/MGEDontology.php#diploid">
      <Database_assnref>
        <Database_ref identifier="MO"/>
        </Database_assnref>
      </DatabaseEntry>
    </OntologyReference_assn>
  </OntologyEntry>
</Associations_assnlist>
</OntologyEntry>
```

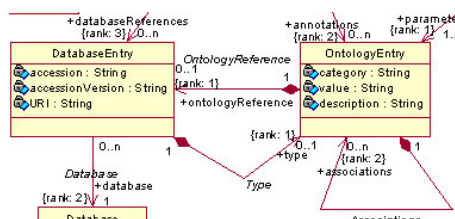
How to Code MO in MAGE ?

Class/Constraints:

Rule of thumb:

Create nested structures using mage

“association_assn”



Individual #1	has-class	ID #1	thing
---------------	-----------	-------	-------

```
<OntologyEntry category="someMOClass" value="someMOClass">
```

```

<OntologyReference_assn>
  <DatabaseEntry accession="#someMOClass "
URI="http://mged.sourceforge.net/ontologies/MGEDontology.php# someMOClass">
    <Database_assnref>
      <Database_ref identifier="MO"/>
    </Database_assnref>
  </DatabaseEntry>
</OntologyReference_assn>

```

```
<Associations_assnlist>
```

```
<OntologyEntry category=" has_MO_restriction" value=" has_MO_restriction">
```

```

<OntologyReference_assn>
  <DatabaseEntry accession="# has_MO_restriction "
URI="http://mged.sourceforge.net/ontologies/MGEDontology.php#has_MO_restriction">
    <Database_assnref>
      <Database_ref identifier="MO"/>
    </Database_assnref>
  </DatabaseEntry>
</OntologyReference_assn>

```

```

<Associations_assnlist>
  <OntologyEntry category=" has_MO_restriction" value="some-value">
  </OntologyEntry>
</Associations_assnlist>

```

```

</OntologyEntry>
</Associations_assnlist>

```

How to Code MO in MAGE ?

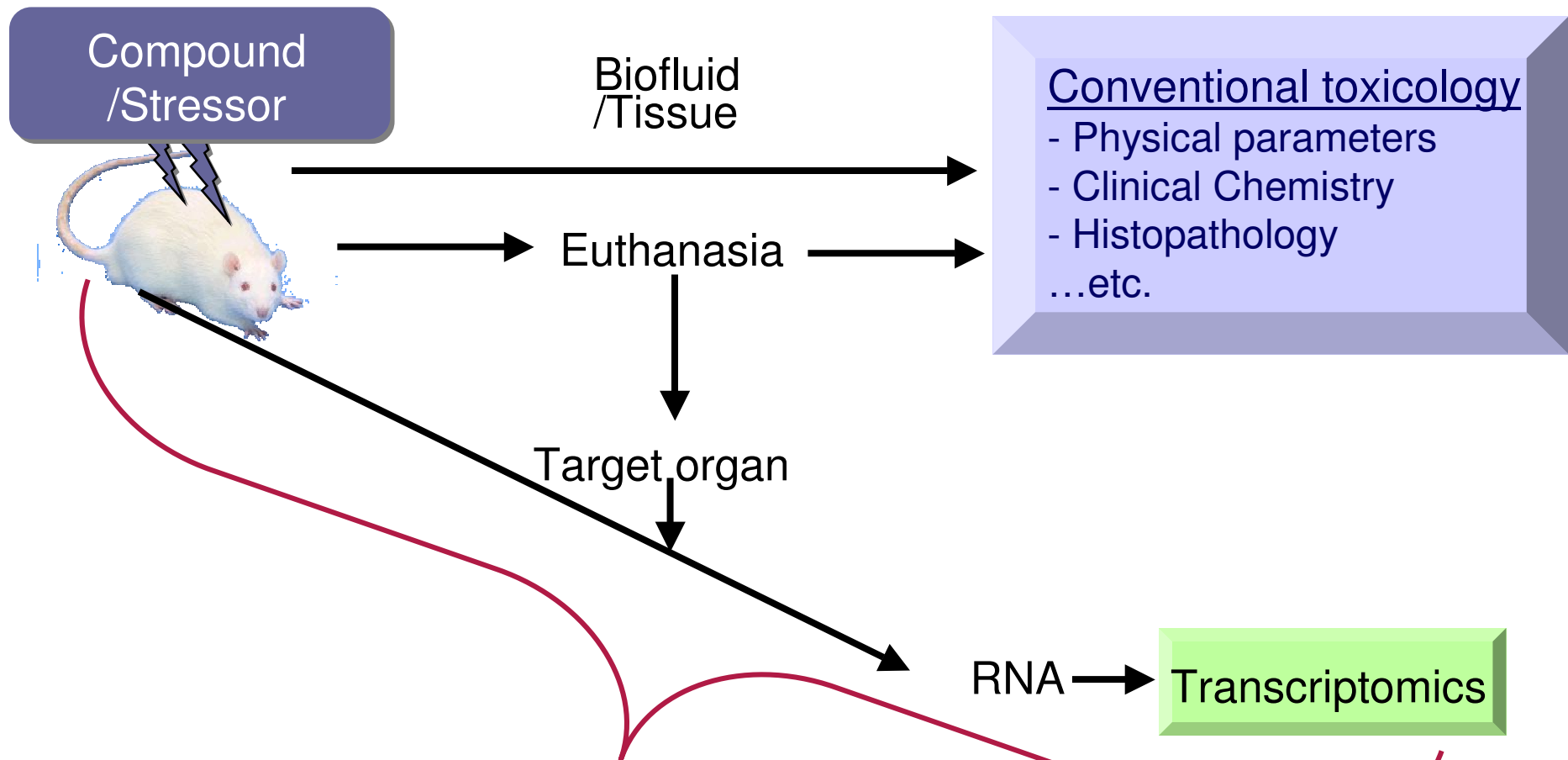
Example for previous slide

Coding Animal Study identifier

mager-ontmat-1	has-class	ID #1	thing
Individual #1			

```
<OntologyEntry category="Individual" value="Individual">
  <OntologyReference_assn>
    <DatabaseEntry accession="#Individual"
URI="http://mged.sourceforge.net/ontologies/MGEDontology.php# Individual">
      <Database_assnref>
        <Database_ref identifier="MO"/>
      </Database_assnref>
    </DatabaseEntry>
  </OntologyReference_assn>
  <Associations_assnlist>
    <OntologyEntry category="has_ID" value="has_ID">
      <OntologyReference_assn>
        <DatabaseEntry accession="#has_ID"
URI="http://mged.sourceforge.net/ontologies/MGEDontology.php# has_ID">
          <Database_assnref>
            <Database_ref identifier="MO"/>
          </Database_assnref>
        </DatabaseEntry>
      </OntologyReference_assn>
    </Associations_assnlist>
    <OntologyEntry category="has_ID" value="mouse123">
      <OntologyEntry>
    </Associations_assnlist>
  </OntologyEntry>
</Associations_assnlist>
```

A case study: Toxicogenomics



■ Bioinformatics

- Methods for analysis
- Protocols for data management

ILSI

- Login
- Help page
- Submission Steps
- Contact Curation Team

Navigation: Submission types > Experiment submissions > Experiment: **ILSI-HESI hepatotoxicity study** > Sample: 1 of 48

First Sample Previous Sample New Sample

Paste existing sample description

Sample name: [GSK]-S10

Organism: Rattus norvegicus

Sex: male

Animal strain: SD/SD

Submit

Navigation: Submission types > Experiment submissions > Experiment: **ILSI-HESI hepatotoxicity study** > Sample: [GSK]-S1-Rat clofibrate-0-day1* [1 of 48]

Navigation: Submission types > Experiment submissions > Experiment: **ILSI-HESI hepatotoxicity study** > Sample: [GSK]-S1-Rat clofibrate-0-day1* [1 of 48]

Clinical observations

General: DIARRHEA, EYE ABNORMALITY

Pathology Observations data sheet

	Sample identifier	Body weight (g)	Brain weight (mg)	Target Organ	Organ weight (mg)	Site	Morphology	qualifier	value	qualifier
1										
2										
3	S10-[GSK]	150	2500	Liver	7850	Left Lat Lobe	Necrosis	Distribution	-	Duration
4	S10-[GSK]	150	2500	Kidney	1950	-	-	-	-	-
5				Liver			-	-	-	-
6				Kidney			Accessory Spleen	-	-	-
7				Liver			Accumulation	-	-	-
8				Kidney			Adenocarcinoma	-	-	-
9				Liver			Adenolipoma	-	-	-
10				Kidney			Adenoma	-	-	-
11				Liver			Adhesion	-	-	-
12				Kidney			Alveolar/Bronchiolar Carcinoma	-	-	-
13				Liver			-	-	-	-
14				Kidney			-	-	-	-

Protocols

Sample protocols

Toxicological endpoints

Clinical Observations

Clinical Chemistry

Pathology Observations data sheet

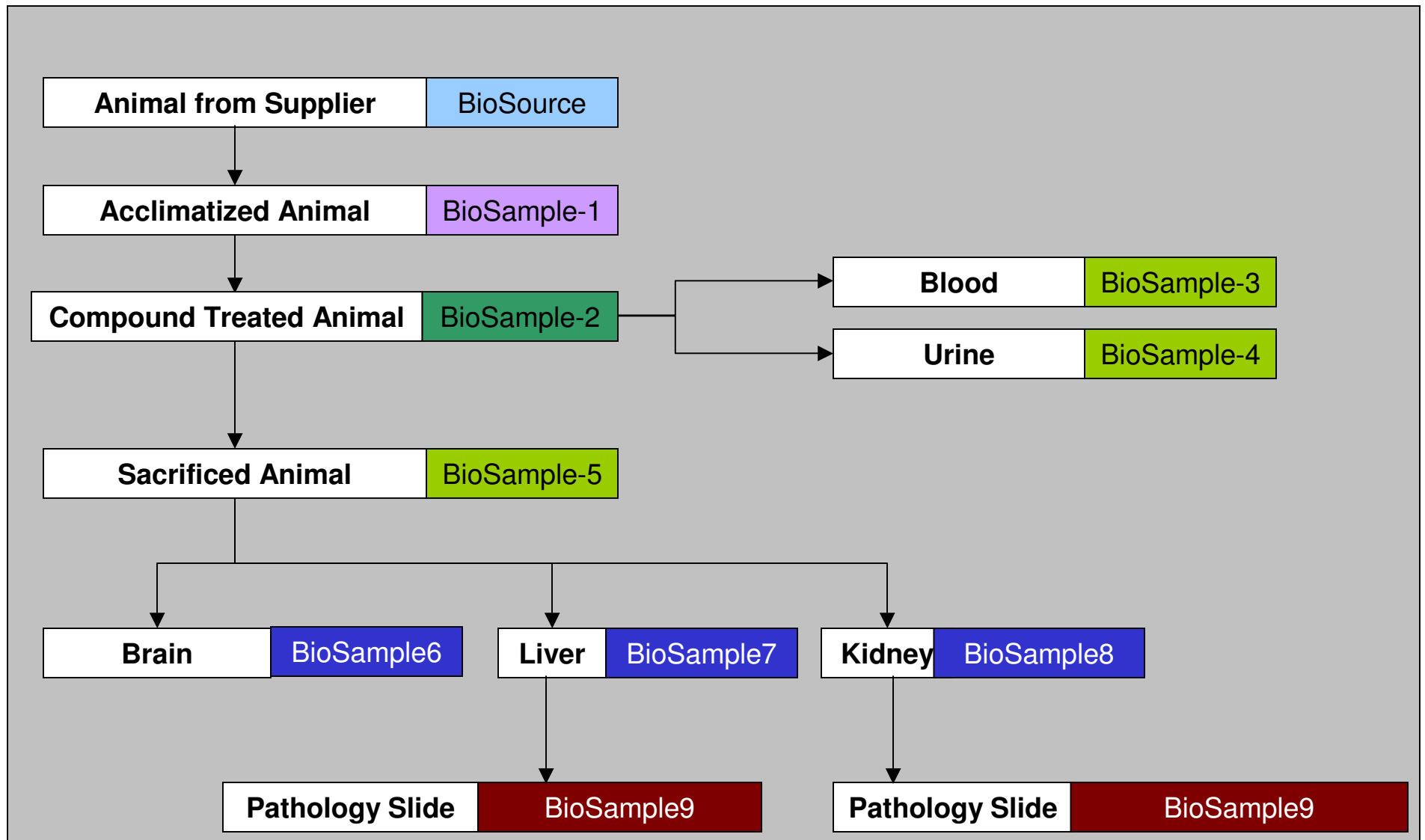
Extracts of sample

Eosinophilocytes

Controlled vocabulary from NIH/NIEHS-NTP Pathology Code Tables (PCT)

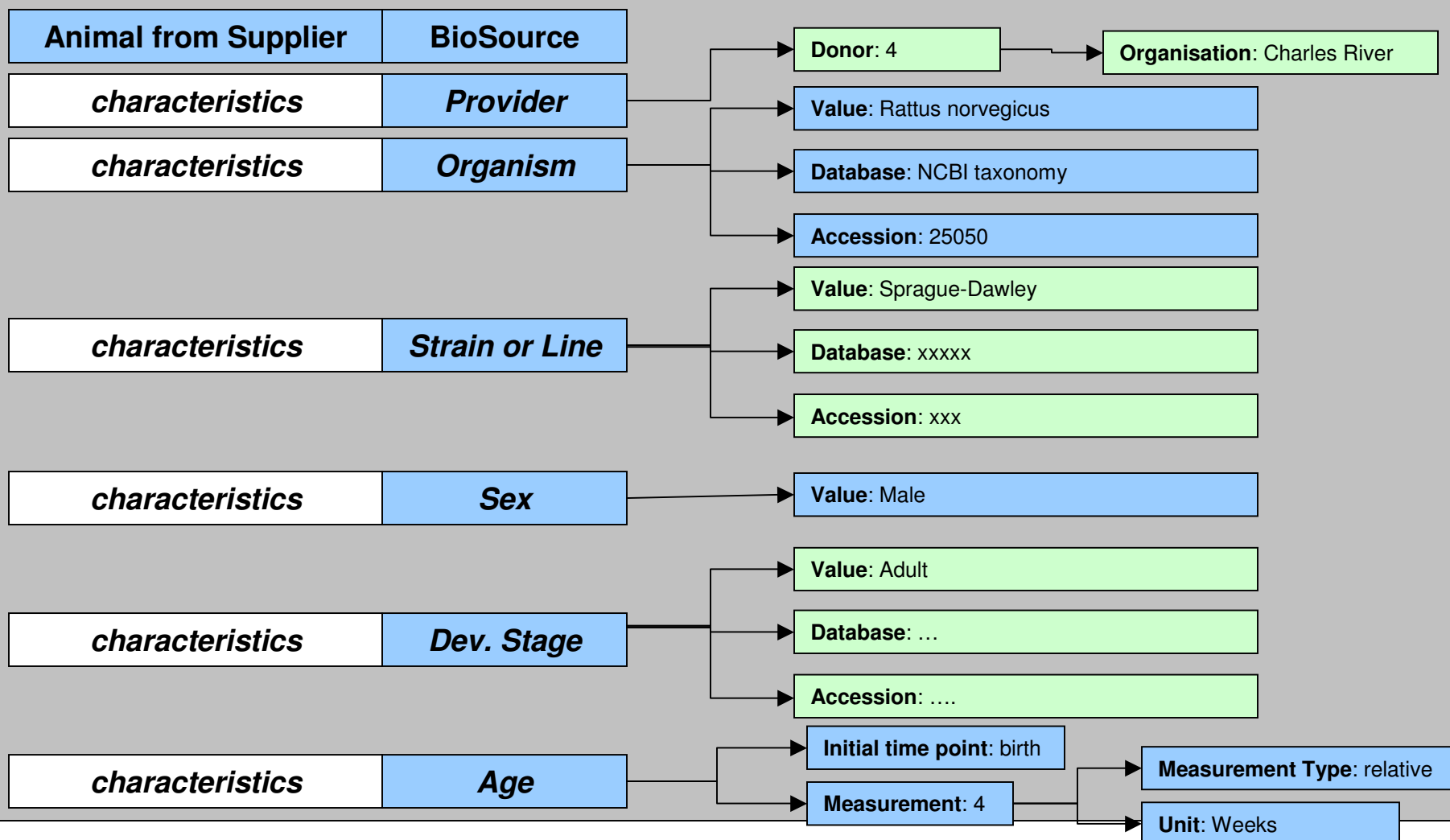
Toxicogenomics: ILSI case study

BioMaterial Overview



MIAMExpress-Tox: BioSource characteristics

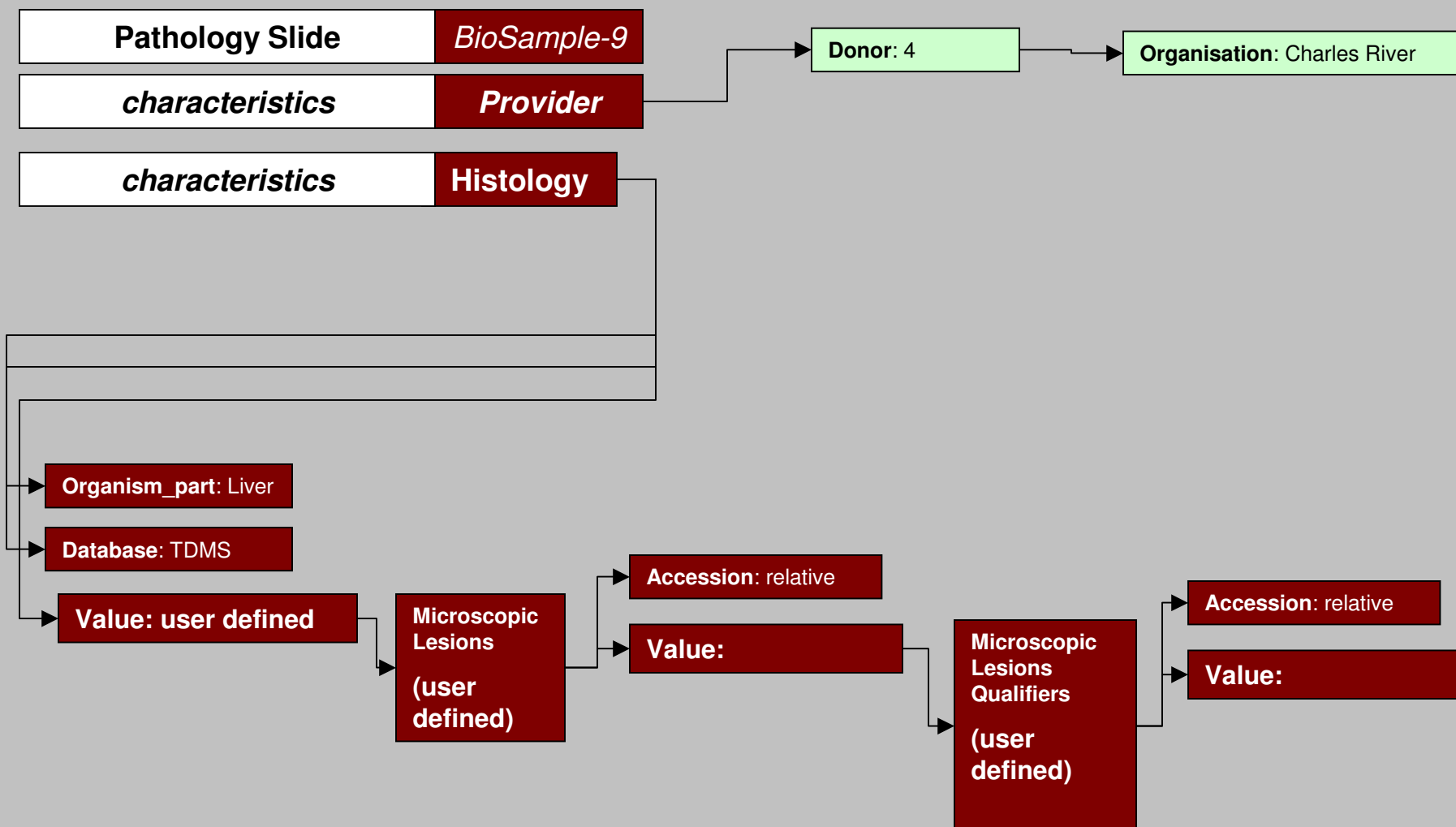
Attributes: Provider, Organism, Strain or Line, Age, Developmental Stage, Sex,



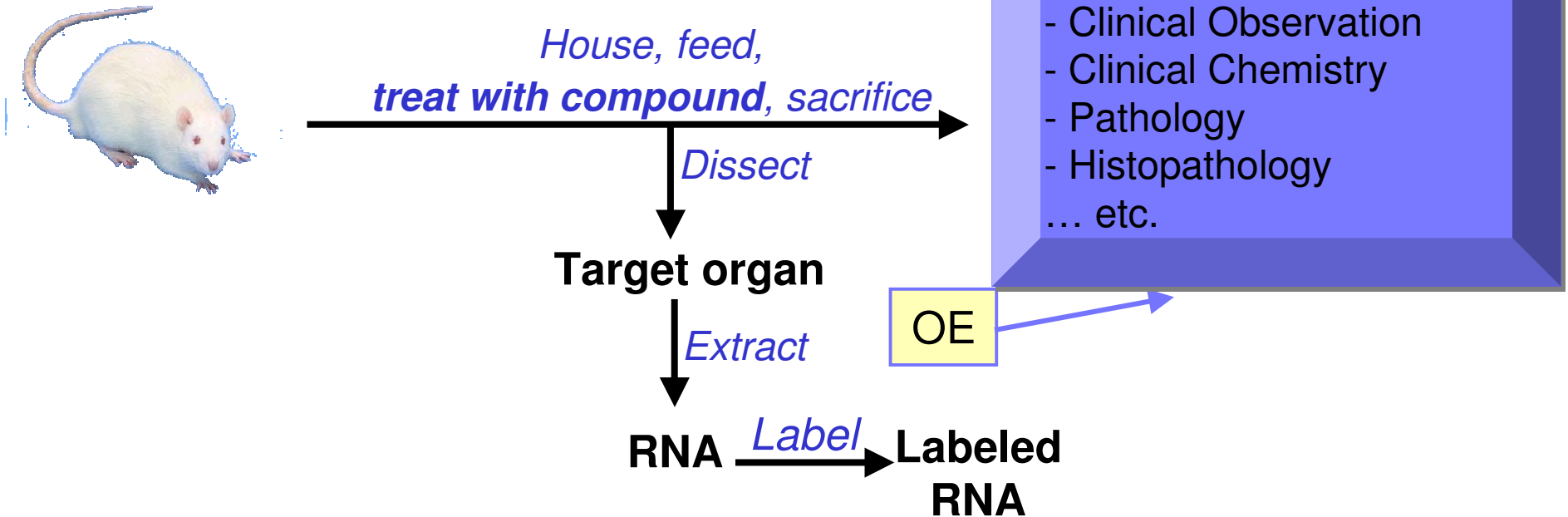
MIAMExpress-Tox: Biomaterial Details BioSource

Before viewing MAGE-ML documents

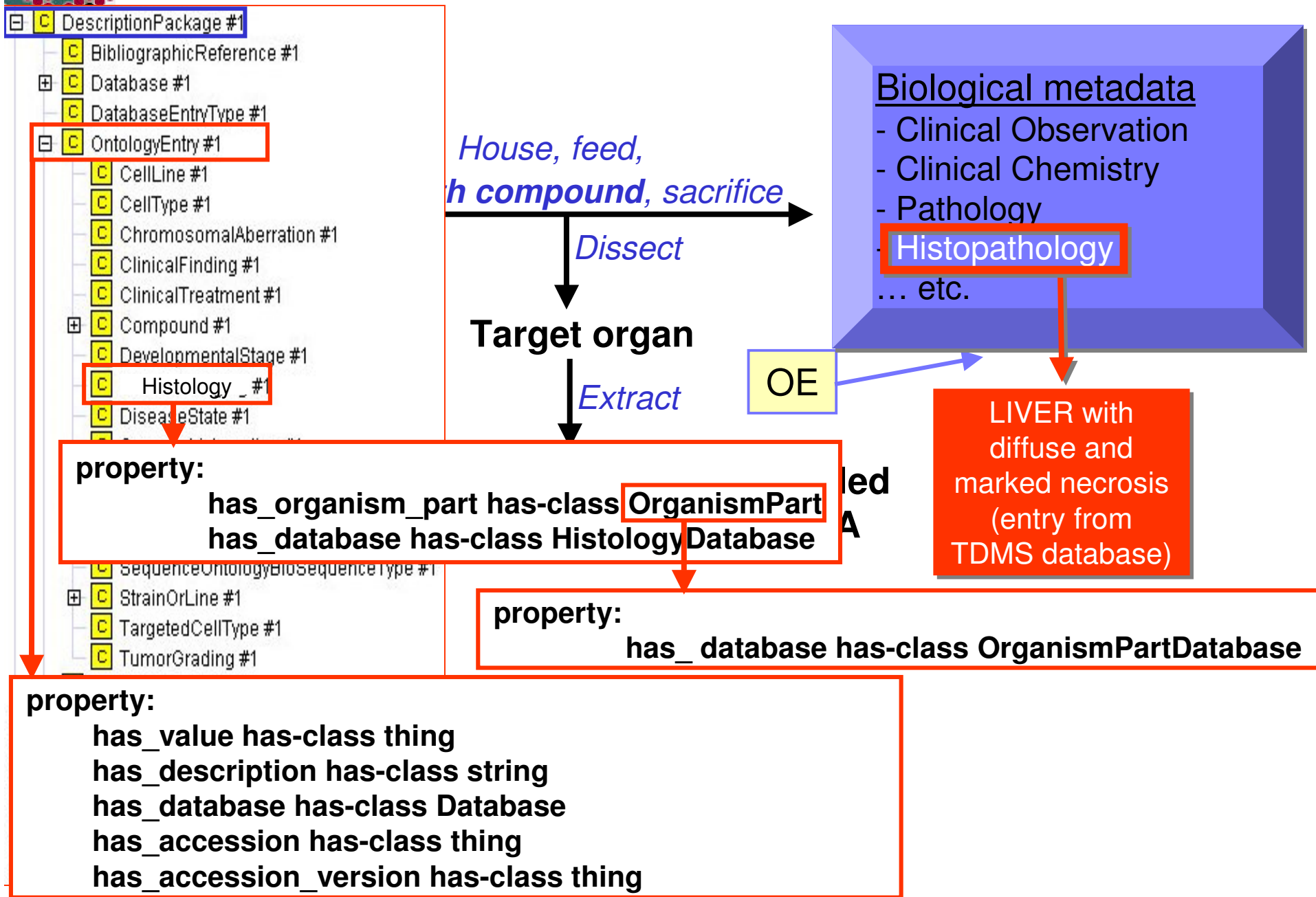
Attribute: Histology



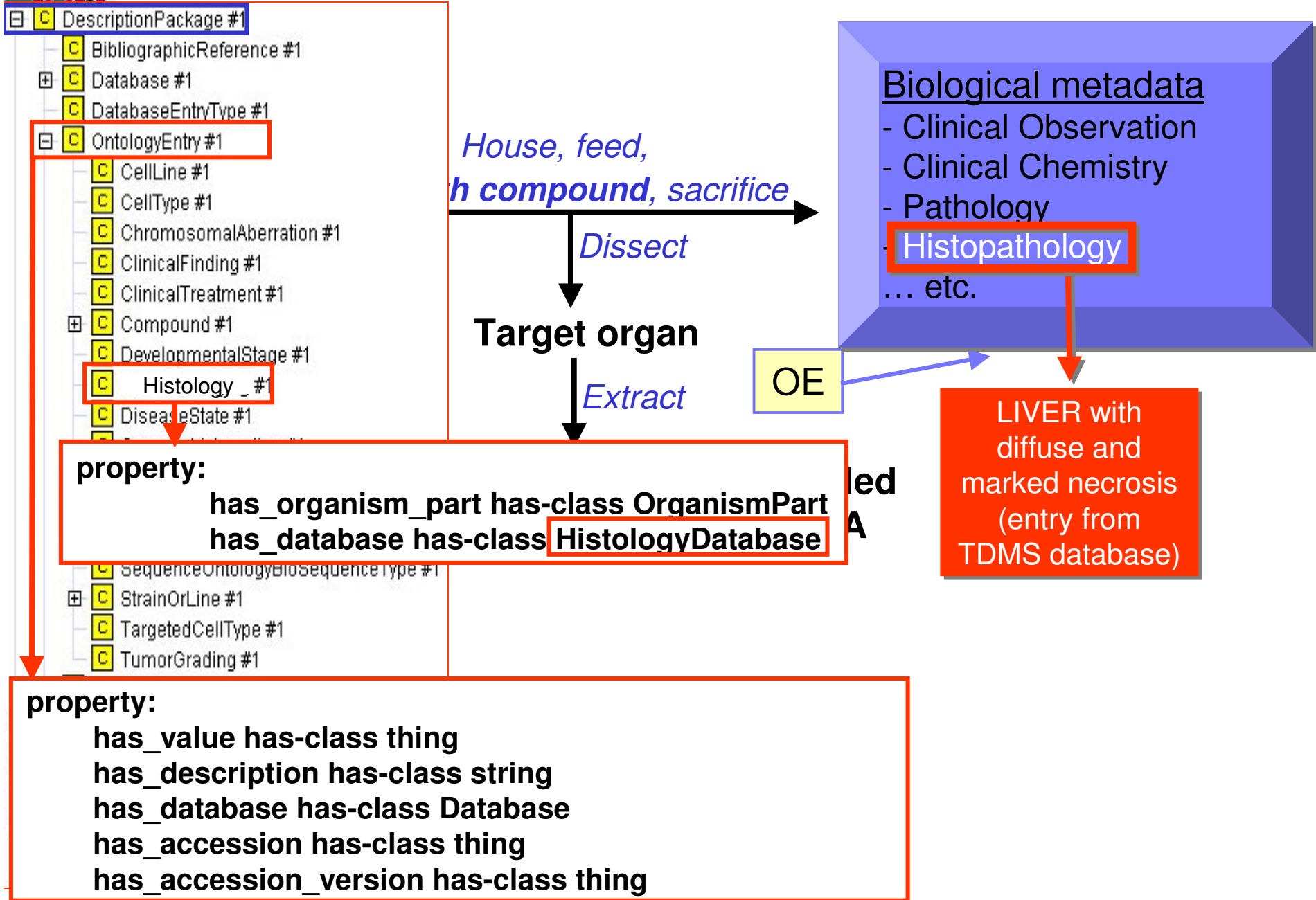
“Liver has a marked necrosis”



“Liver has a marked necrosis”



“Liver has a marked necrosis”

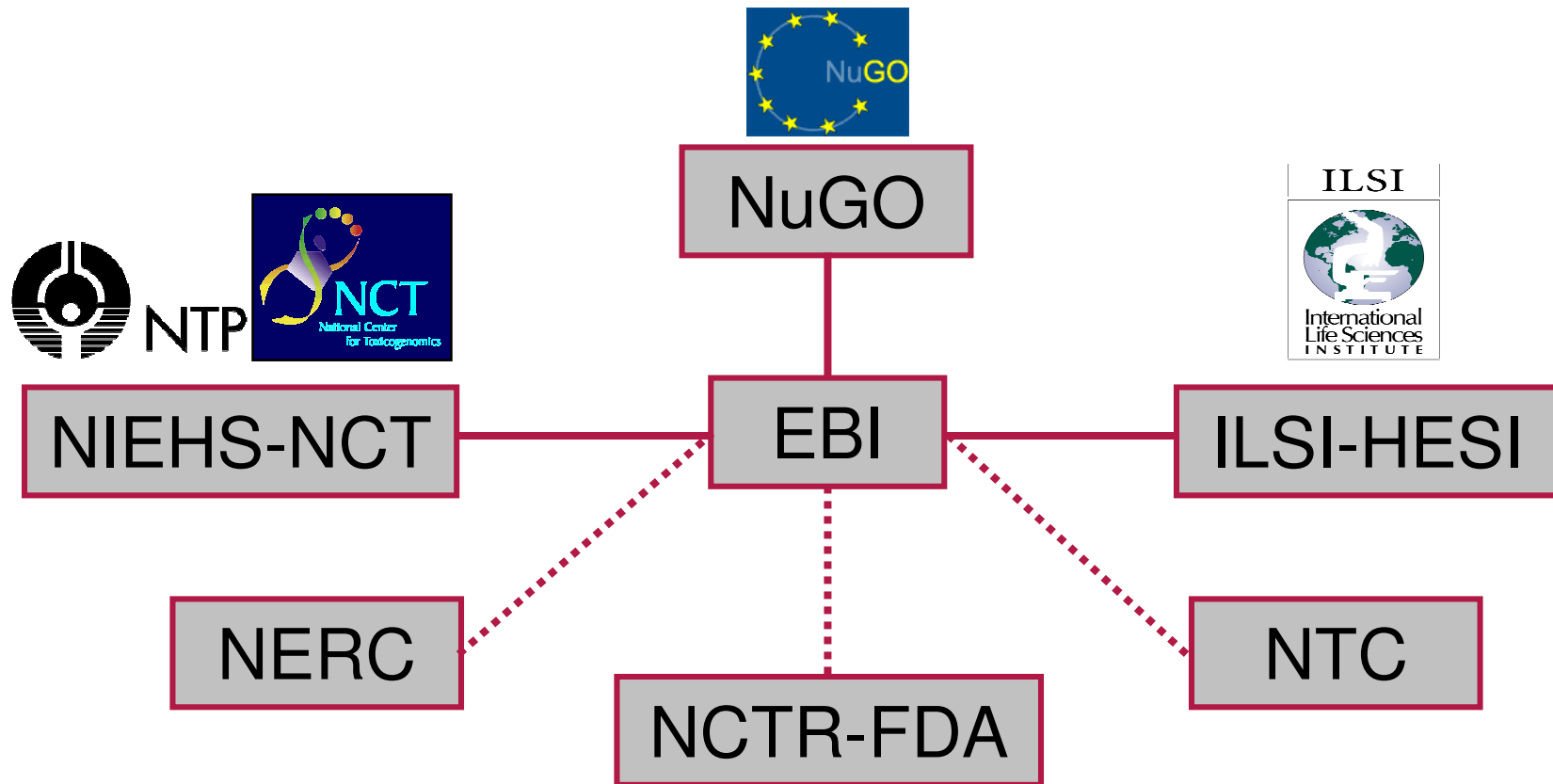


Information to be recorded – Obstacles?

- **Regulatory bodies focus**
 - Tox metadata-centric vision
 - Data submission models
 - Minimal burden on sponsors
 - Transport file
 - Flexibility with the terms
 - Data storage
 - Visualization
 - Verification
 - Emphasis on data review
 - IP issue
- **Research community focus**
 - Genomic data-centric vision
 - Data submission tools
 - Users support
 - Mandatory data fields
 - Controlled vocabularies
 - Structured databases
 - Query-driven
 - Interoperability
 - Emphasis on data comparability
 - Open source approach

Looking for an optimal level of interaction !

Toxicogenomics WG

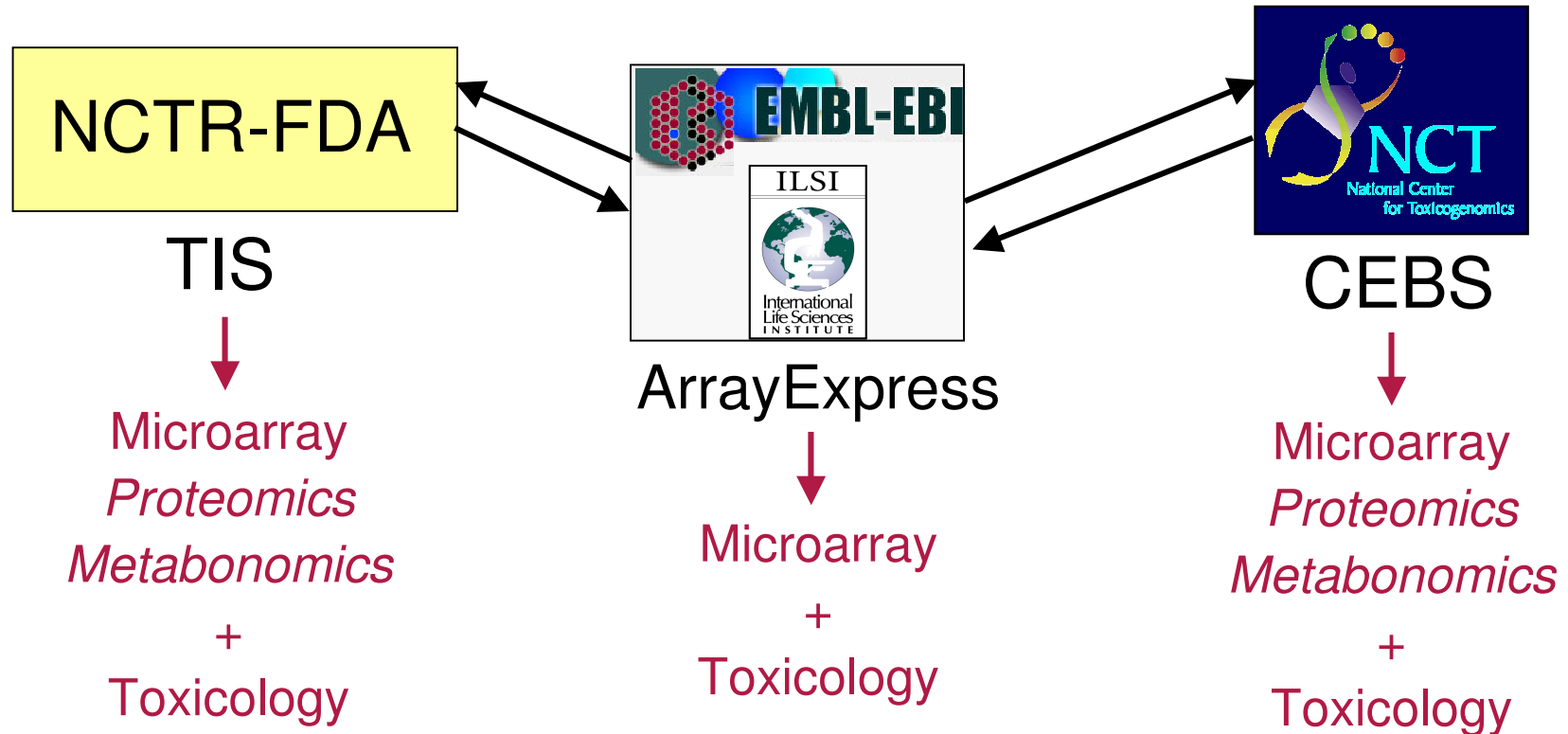


- International partners
- Discussion forum within MGED

— = *Funding*
... = *Collaboration*



Compatible Informatics Platform





Compatible Informatics Platform

NCTR-FDA

TIS



ArrayExpress



CEBS

SEND guidelines

version 1.4 3/2/04

SEND+CDISC=
Data Tabulation Model
(DTM)



TDMS

ECVAM

Pharmacogenomic
Data Standards

I3C/HL7/CDISC

Standards - State of the Art

- Public, community vetted efforts
 - **Content**
 - Minimal descriptors
 - **Terminology**
 - Controlled terms, ontology
 - **Storage**
 - Database models
 - **Exchange**
 - Interoperability format

Conventional toxicology

- Physical parameters
- Clinical Chemistry
- Histopathology
- ...etc.

Transcriptomics

Proteomics

Metabonomics

Standards - State of the Art

- Required
- Regulatory arena
 - SEND and CDISC models (DTM)
- A wealth of terminologies
 - Public, e.g.:
 - NIEHS-NTP, LOINC, RENI
 - Proprietary, e.g.:
 - UMLS, SNOMED
- Problems
 - Species-specific terms hamper cross species comparisons

Conventional toxicology

- Physical parameters
- Clinical Chemistry
- Histopathology
- ...etc.

Transcriptomics

Proteomics

Metabonomics



Standards - State of the Art

- Ongoing...
- HUPO-PSI
(Proteomics Standards Initiative)
 - Data model and exchange (XML)
 - IntAct, MINT, BIND, DIP and other databases
 - Industry
 - Academia
 - Mass spectrometry Group
 - Weimin Zhu (EBI)
 - Randy Julian (Ely Lilly)
 - Molecular integration Group
 - Henning Hermjakob (EBI)
 - Proteomics integration Group
 - Chris Taylor (EBI)

(Hermjakob *et al. Nature Biotech*, 2004)

Conventional toxicology

- Physical parameters
- Clinical Chemistry
- Histopathology
- ...etc.

Transcriptomics

Proteomics

Metabonomics

Standards - State of the Art

- Starting...
- Standardized Metabonomics Reporting Structures (SMRS)
 - J. Nicholson, J. Lindon, Imperial College, UK
 - Industry
 - FDA (NCTR)
 - Academia

Conventional toxicology

- Physical parameters
- Clinical Chemistry
- Histopathology
- ...etc.

Transcriptomics

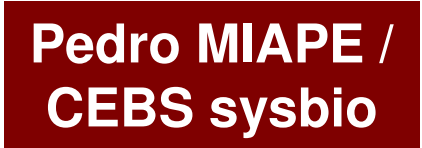
Proteomics

Metabonomics

Synergy – Opportunities

Name	Scope	Exchange format	Database model	CVs	Minimal descriptors	Sample	Tox	uA	Prot	Met
SEND	Model of non-clinical data	Flat file (XML later)	✗	✗	✓	✓	✓	✗	✗	✗
CDISC	Model of clinical data	Flat file (XML later)	✗	✗	✓	✓	✓	✗	✗	✗
(MGED) MAGE	General model of microarray	XML	✓	✓	✓	✓	✗	✓	✗	✗
(PSI) MIAPE	General model of proteomics	XML	✓	✓	✓	✓	✗	✗	✓	✗
(NIEHS -NCT) SysBio -OM	MGED+PSI model of proteomics and metabonomics for CEBS	XML	✓	✓	✓	✓	✓	✓	✓	✓

Xirasagar S *et al.* CEBS object model for systems biology data, SysBio-OM. *Bioinformatics* (2004) Mar 25.



Toxicogenomics WG

- **Consensus on minimal descriptors**
 - Extend MIAME to toxicogenomics experiments
 - > **MIAME/Tox**
- **Toxicology terms harmonization**
 - Promote development of ontologies
 - > **Standard Ontology for Functional Genomics (SOFG) conference**
- **Storage and data sharing**
 - Agree on standard format for toxicological domain
 - > **Use MAGE and the MGED Ontology?**
 - > **Develop a specific model for tox data?**
 - Adopt proteomics and metabolomics standards
 - > **Interaction with PSI and SMRS initiatives**



Acknowledgements

EMBL-EBI

Microarray Informatics

Alvis Brazma , Ugis Sarkans
Susanna Sansone
(Toxicogenomics Coordinator)

MGED Society

Angel Pizarro

NIH-NIEHS

National Center Toxicogenomics
Ray Tennant (Director)

Mike Waters
Jennifer Fostel

NCT Microarray Group
Richard Paules
Pierre Bushel

National Toxicology Programme
Skip Eastin

NCTR-FDA

Dan Casciano (Director)
Weida Tong

ILSI-HESI

Syril Pettit
Bill Mattes
Bill Pennie (Chair)