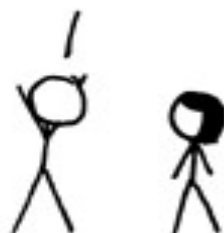


TERMS I HAVE USED OR HEARD USED
TO MAKE FUN OF THE INTERNET.

	NET	WEB	SPHERE	TUBES	BLAG
WORLD WIDE		X			
INTER-	X	X	X	X	X
BLOGO-			X		
BLAGO-	X		X		X
WEB-	X	X			

I HEARD ABOUT IT
ON THE INTERBLAG!



Ontologies and Object Models for Biological Experiments

Allyson Lister
Centre for Integrated Systems Biology of
Ageing and Nutrition (CISBAN)
Newcastle University
<http://www.cisban.ac.uk>

2nd BioModels.net Training Camp, 15 January 2007

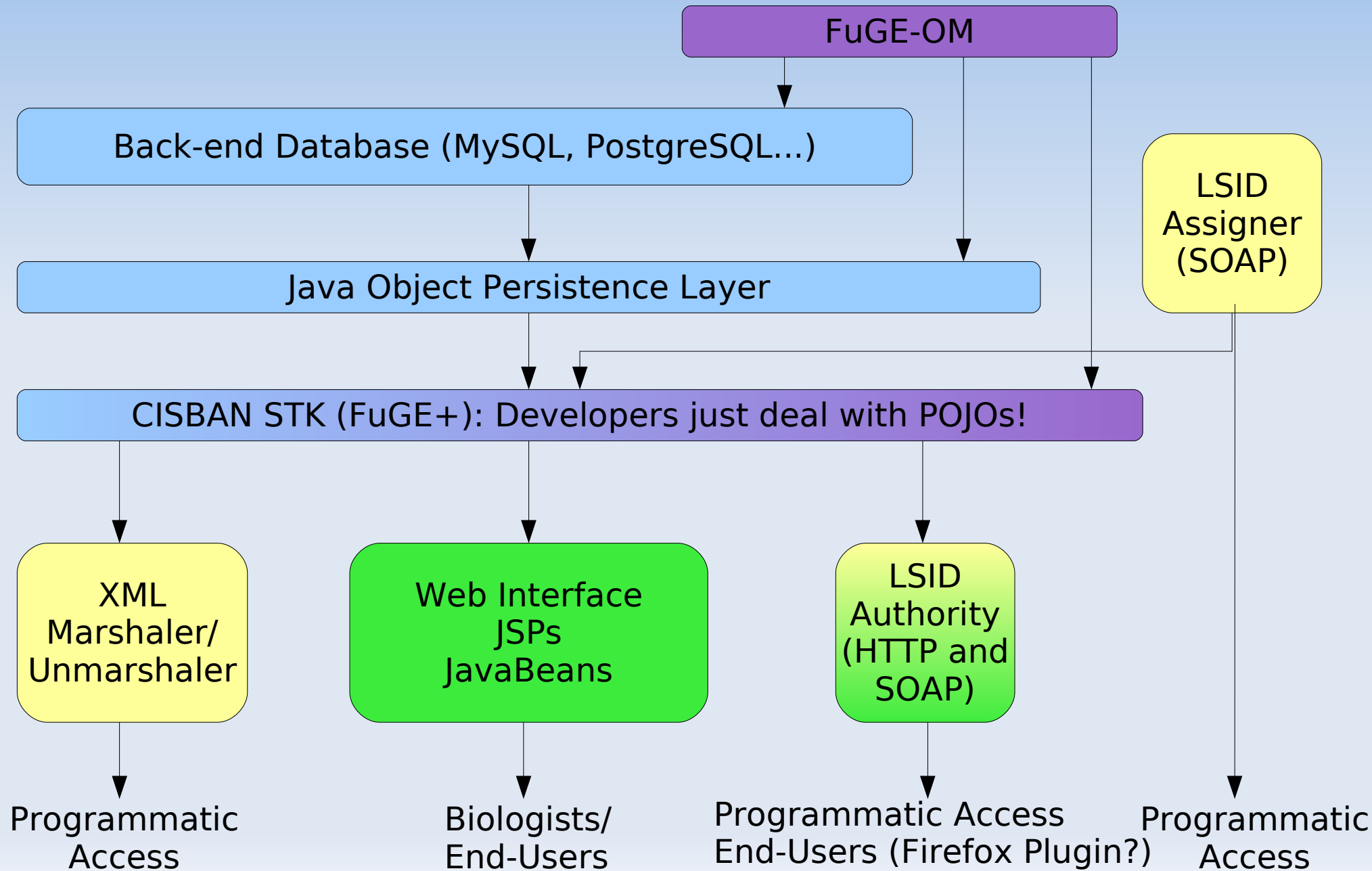
Overview

- Functional Genomics Experiment Object Model / Markup Language (FuGE-OM, FuGE-ML)
 - already described by Andy Jones
- Implementing FuGE for the CISBAN Data Portal and Integrator: a real-life example
- The Ontology for Biomedical Investigations (OBI)

CISBAN Data Portal: FuGE+ (Milestone 3)

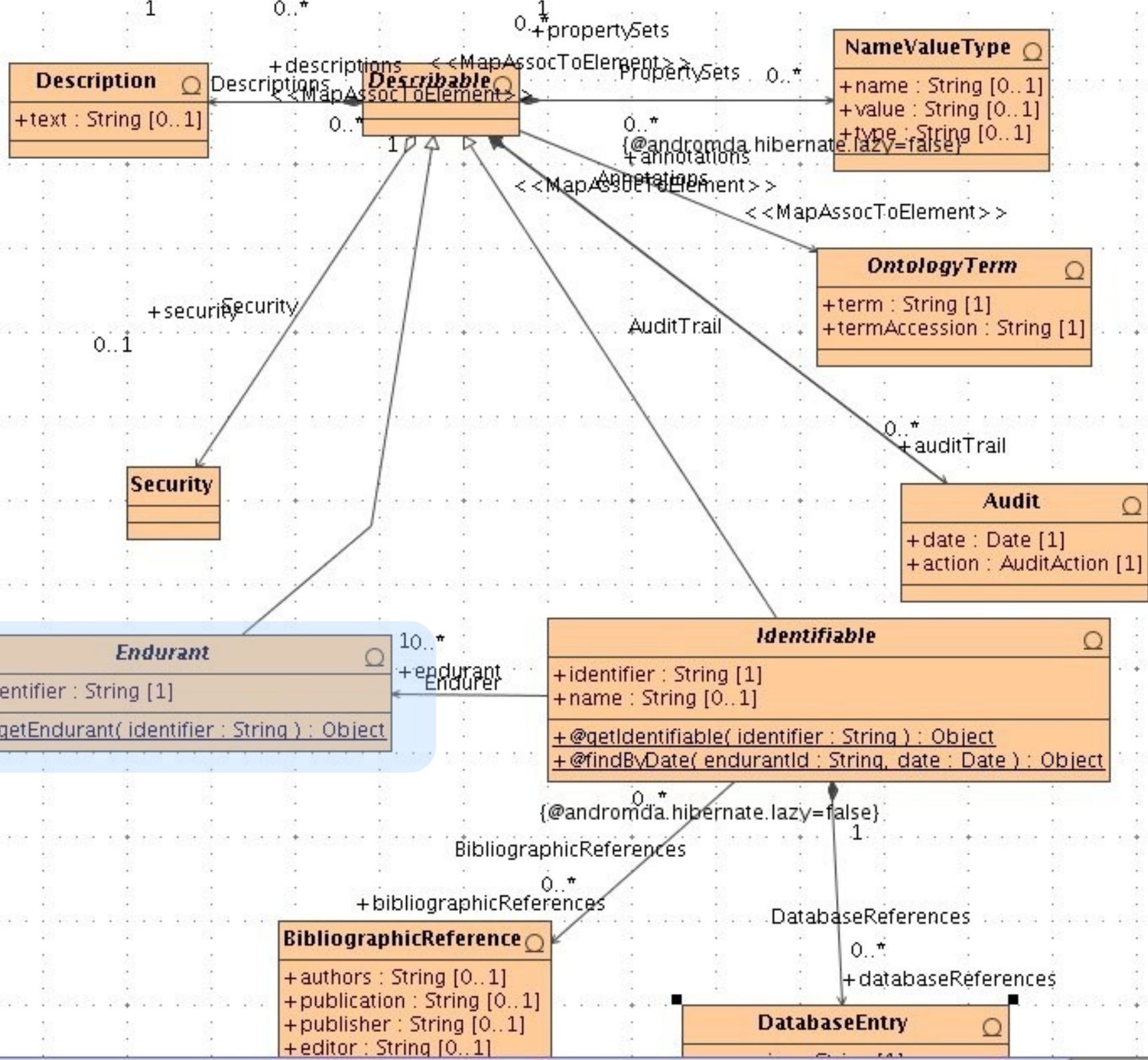
- FuGE Identifiers are LSIDs
- A more complex versioning system has been created
 - *ONLY* additions to the UML: no deletions or modifications of existing entities/relationships/attributes
- *Simple* front-end web view for experimentalists.

Graphical Overview



LSID authority

- Retrieving objects programmatically from the database in FuGE-ML is straightforward: just contact the LSID authority.
 - Programmers may construct LSIDs based on timestamps, to retrieve the data as it was at that time
 - Programmers may use existing LSIDs to retrieve specific versions



Versioning

- “Endurant” objects endure over space and time
 - Identified with an LSID
 - Because they do not change their intrinsic value over time
 - These LSIDs are unresolvable.
- All other objects identified with an LSID are just the *current state* of the endurant and are resolvable.
- Error-correction or adding more information (such as fixing a typo in the name of a robot) are new versions of the same endurant, and as such will only add a new object attached to the same endurant.
- A fundamental change in an object (such as the addition of a new camera on a robot) means that the robot is a new enduring object, and will be assigned a new endurant LSID

Versioning

Endurants

1234

Simple "Identifiabiles"

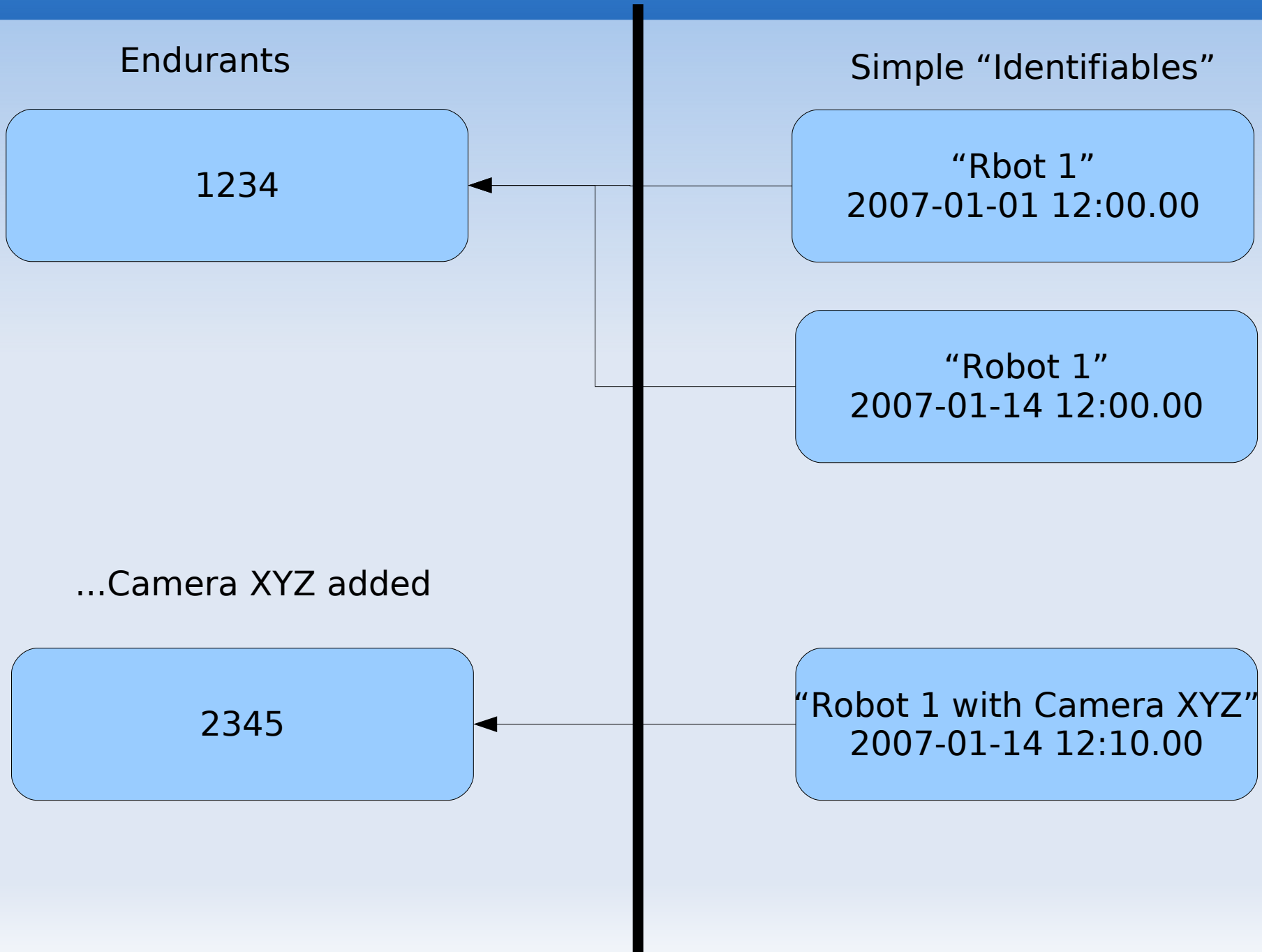
"Rbot 1"
2007-01-01 12:00.00

"Robot 1"
2007-01-14 12:00.00

...Camera XYZ added

2345

"Robot 1 with Camera XYZ"
2007-01-14 12:10.00



LOGIN SCREEN

User
(pre-assigned)

Login

Is this the
first login?

Yes

PERSONAL DATA
ENTRY SCREEN

Fill in Personal
Details

Decision

Screen

Green text
symbolizes
tasks to be
added at a
later date

SEARCH SCREEN

Various
Searches

Search

HOME SCREEN

Search, Tutorial,
Update Person,
Contact Us / Help

No

Enter New Data

View My Exps

View All

VIEW SCREEN

View Only
Click-through

link back to

RAW DATA ENTRY SCREEN

Request New
Type of Data

Add Performers?

Select Type & Upload

email helpdesk

PERFORMER
SCREEN

Contact(s)
Pull-down

new people?

email helpdesk

METADATA SCREEN

Confirmation of
Instrument(s)
and Protocol(s)

INSTR. SCREEN

Add new
instrument

PROTOCOL SCREEN

Add new
Protocol

New Exp

Existing Exp

EXP NAME SCREEN

Name of Exp,
Provider
Optional
Hypothesis a/o
Comment

CONFIRMATION
SCREEN

Automatic
Download
of Data

- Purpose
 - Provide a resource for the unambiguous description of the **components** of **biomedical investigations** such as the design, protocols and instrumentation, material, data and types of analysis on the data
 - NOT designed to model biology
 - NOT dependent on any given Object Model
 - NOT limited to only functional genomics
- Enables
 - Allow consistent annotation of data across different **technological** and **biological** domains
 - Enable powerful concept-driven queries
 - Facilitate semantically-driven data integration

Motivation for OBI

- Standardization efforts in **biological** and **technological** domains
 - Standard **syntax** - Data exchange formats
 - To provide a mechanism for software interoperability, e.g. FuGE Object Model
 - Standard **semantics** - Controlled vocabularies or ontology
 - Centralize commonalities for annotation term needs across domains to describe an investigation/study/experiment, e.g. OBI

OBI Top-Level Classes

- **Continuant:** *an entity that endure/remains the same through time*
 - **Dependent Continuant:** *depend on another entity*
 - E.g. Environment (depend on the set of ranges of conditions, e.g. geographic location)
 - E.g. Characteristics (entity that can be measured, e.g. temperature, unit)
 - **Realizable:** an entity that is realizable through a process (executed/run)
 - E.g. Software (a set of machine instructions)
 - E.g. Design (the plan that can be realized in a process)
 - E.g. Role (the part played by an entity within the context of a process)
 - **Independent Continuant:** *stands on its own*
 - E.g. All physical entity (instrument, technology platform, document etc.)
 - E.g. Biological material (organism, population etc.)
- **Occurrent:** *an entity that occurs/unfold in time*
 - E.g. Temporal Regions, Spatio-Temporal Regions (single actions or Event)
 - **Process**
 - E.g. Investigation (the entire 'experimental' process)
 - E.g. Study (process of acquiring and treating the biological material)
 - E.g. Assay (process of performing some tests and recording the results)

OBI Design Principles

- **OBO Foundry ontology, utilize ontology best practices**
 - Inherit top level classes from an Upper Level ontology
 - Use of the Relation Ontology
 - Follow additional OBO Foundry principles
 - Facilitates interoperability with other OBO Foundry ontologies
- **Develop recommendations for naming conventions and metadata**
 - Format for term names, e.g. underscore vs. camel case, no plurals
 - Use of Alphanumeric identifier for terms, i.e. something that does not have semantic meaning
 - Mechanisms for adding synonyms, etc.
- **Open source approach**
 - Protégé/OWL
 - Weekly conference calls
 - Shared environment using Sourceforge (SF) and SF mailing lists

OBI Collaborating Communities

- Crop sciences Generation Challenge Programme (GCP), www.generationcp.org
- Environmental genomics MGED RSBI Group, www.mged.org/Workgroups/rsbi
- Genomic Standards Consortium (GSC), www.genomics.ceh.ac.uk/genomecatalogue
- HUPO Proteomics Standards Initiative (PSI), psidev.sourceforge.net
- Immunology Database and Analysis Portal, www.immport.org
- Immune Epitope Database and Analysis Resource (IEDB), <http://www.immuneepitope.org/home.do>
- International Society for Analytical Cytology, <http://www.isac-net.org/>
- Metabolomics Standards Initiative (MSI), msi.workgroups.sourceforge.net
- Neurogenetics, Biomedical Informatics Research Network (BIRN), www.nbirn.net
- Nutrigenomics MGED RSBI Group, www.mged.org/Workgroups/rsbi
- Polymorphism
- Toxicogenomics MGED RSBI Group, www.mged.org/Workgroups/rsbi
- Transcriptomics MGED Ontology Group, mged.sourceforge.net/ontologies

Thank you! Questions?

- Tom Kirkwood, Anil Wipat
- Developers: Allyson Lister, Olly Shaw
- People who provided LOTS of help and advice: Andy Jones, Matt Pocock
- <http://fuge.sourceforge.net>, FuGE: Functional Genomics Experiment Object Model, Andrew R. Jones, Angel Pizarro, Paul Spellman, Michael Miller, Fuge Working Group, OMICS: A Journal of Integrative Biology Jun 2006, Vol. 10, No. 2: 179-184.
- <http://obi.sourceforge.net>. Next workshop Jan 29 - Feb 2 in San Diego. The 29th is an open day, please contact obi-devel@lists.sourceforge.net if you wish to attend that day.
- Initial documentation, including the UML diagram, for the CISBAN Data Portal, will be linked from <http://www.cisban.ac.uk> by the end of January. A few weeks after that, an alpha version of the data portal will be available for those who wish to play with it.

CISBAN is supported jointly by the Biotechnology and Biological Sciences Research Council (BBSRC) and the Engineering and Physical Sciences Research Council (EPSRC).

