The BioMANTA Ontology

Integrating Protein-Protein Interaction Data

InterOntology08 – Tokyo, Japan

Andrew Newman
ITEE, University of Queensland

Project Overview

- Development of Semantic Interactome Ontology
- Infrastructure for Integration, Processing and Querying
- RDF and Inferencing on Scale-Out Architecture
- Output for Visualisation and Improve Data Quality through Network Analysis

The Domain is a Problem

- Genes. What are genes anyway? Do they even exist?
  - Unit of hereditary (1860s-1900s)
  - Locus (1910s)
  - Coding for a Protein (1940s)
  - Molecule (1950s)
  - Transcribed Code (1960s)
  - Open Reading Frame (1970s-1980s)
  - Annotated Entity (1990s-2000s)

“…the gene is […] neither discrete […] nor continuous […], nor does it have a location […], nor a clearcut function […], not even a constant sequence […] nor definite borderlines” – Falk (1986)

Other Problems

- Properties tied to Local or Global IDs
- Data formats
- Quality – missing, incomplete, wrong
- Project size, scope, scale
- Differing semantics and specialities
- Poor understanding or usage of model semantics

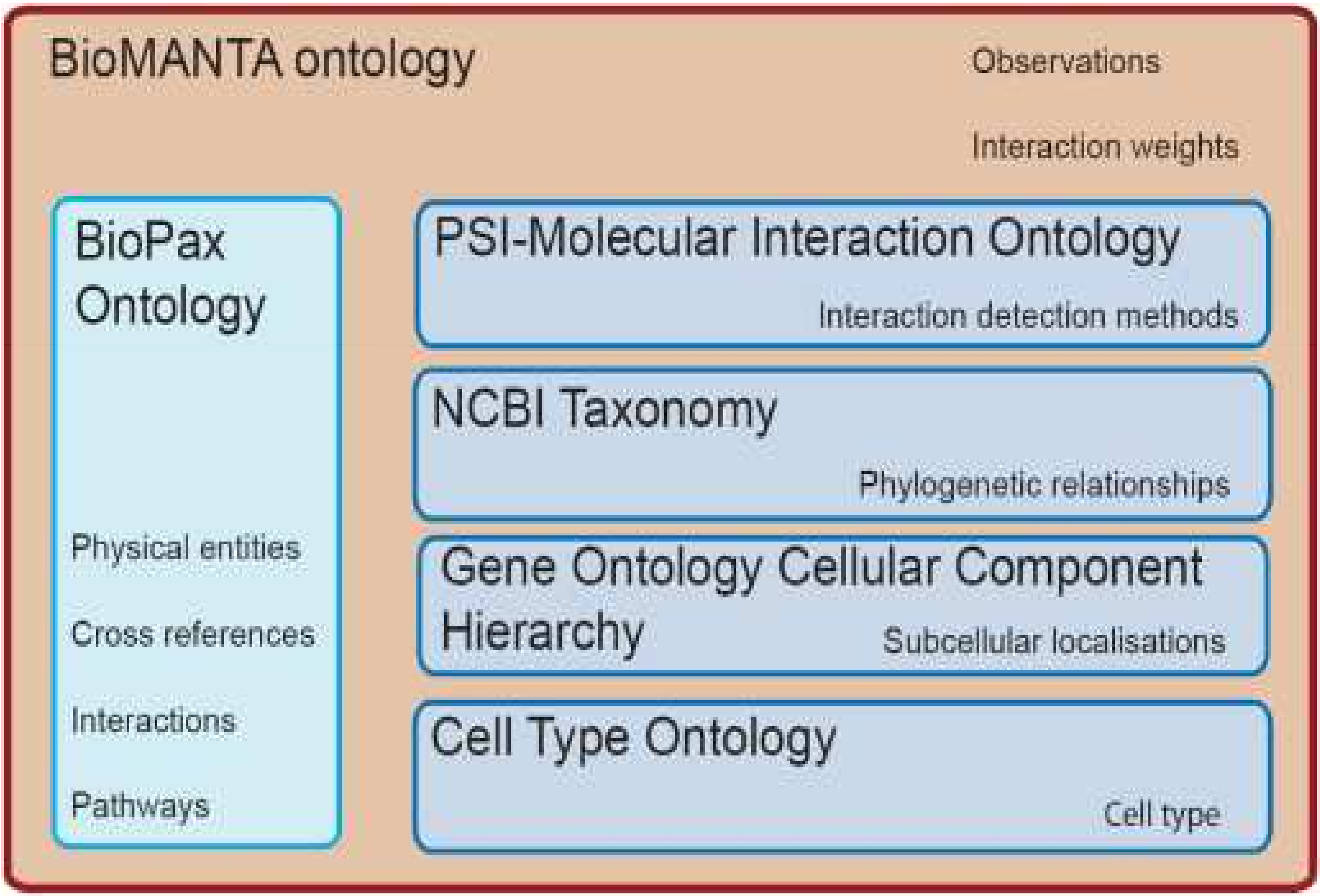
Some Questions

- What are we modelling? There is a record of an observation of an interaction in a given context about a protein.
- Do we need a new ontology for the domain? Not really.
- Why? We are integrating existing databases.
- We need to take the best bits from what we need to support our own in silico experiments.

Our Design

- Create or reuse dereferenceable URIs for names.
- Do not create another ID.
- The “thing” (protein, experiment, etc.) is a blank node.
- Global, local names and other properties are related to the blank node.
- Reusing existing terms where appropriate.
- It is a snapshot of the current data at a given time.
- Multiple versions of data and ontologies is expected.
- Instances of classes – for example Cell Type instances from Cell Type ontology (punning).

Ontology



BioMANTA ontology

Observations

Interaction weights

BioPax Ontology

Physical entities

Cross references

Interactions

Pathways

PSI-Molecular Interaction Ontology

Interaction detection methods

NCBI Taxonomy

Phylogenetic relationships

Gene Ontology Cellular Component Hierarchy

Subcellular localisations

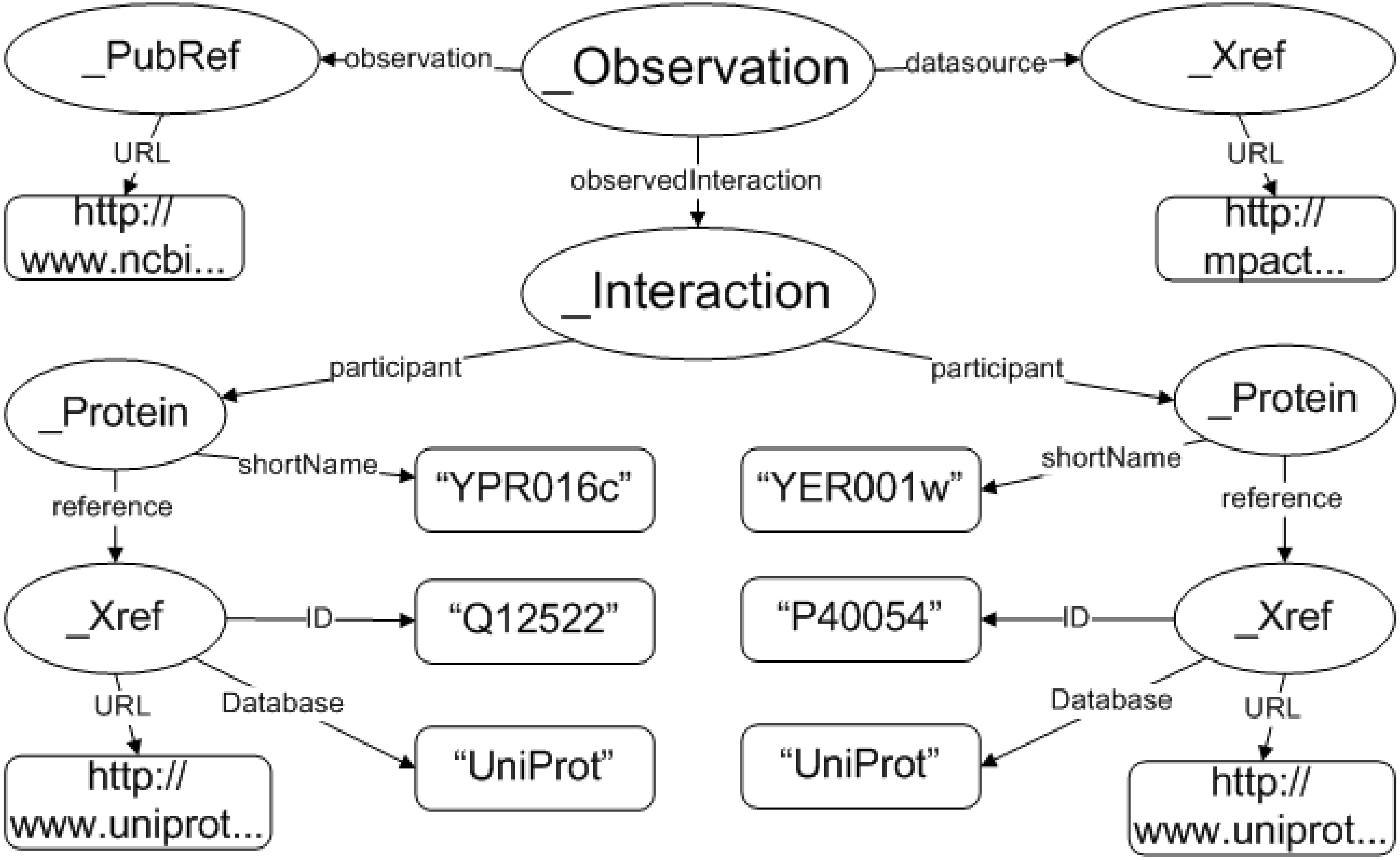
Cell Type Ontology

Cell type

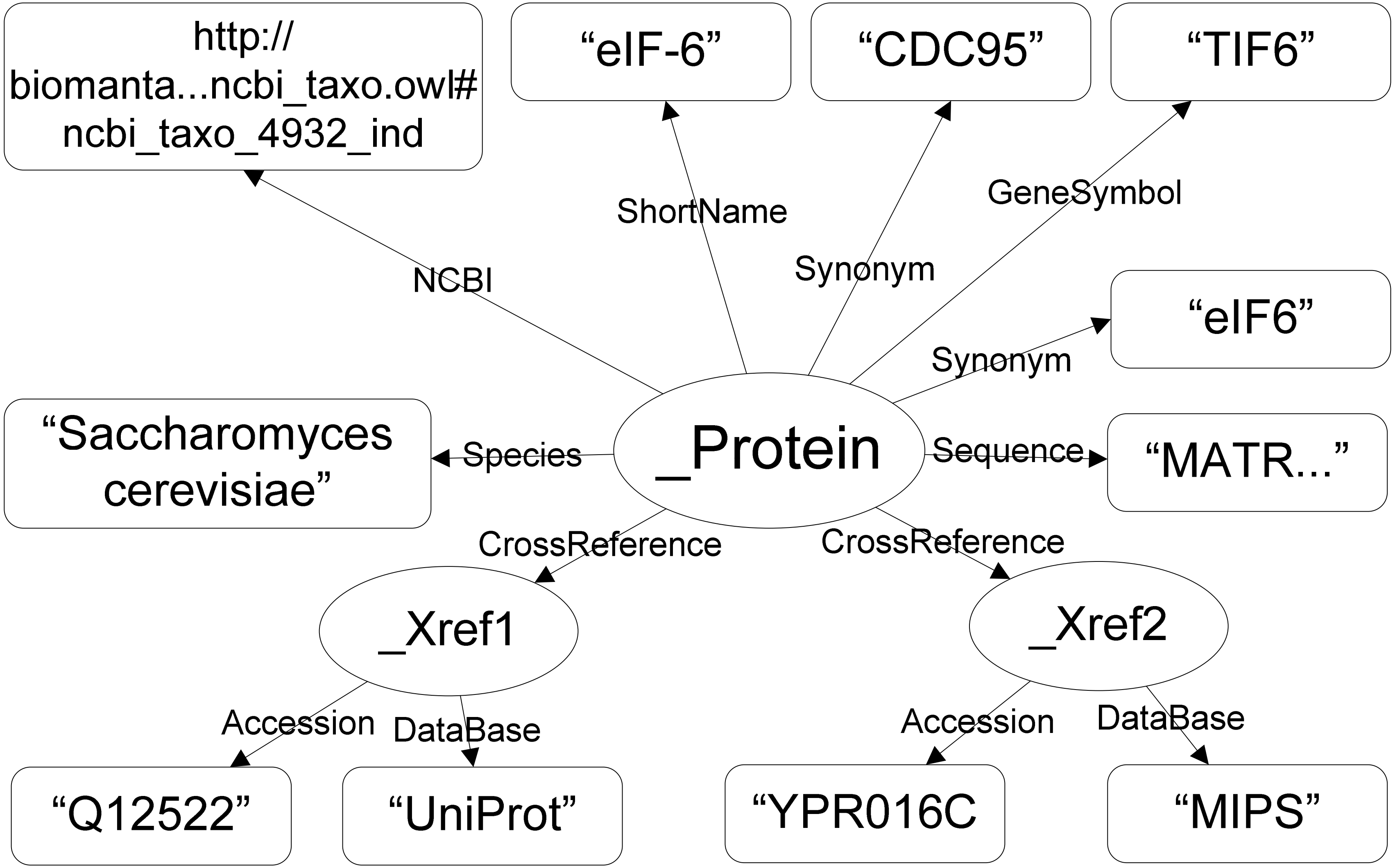
Examples of BioMANTA Extensions

- Experimental Observation ∩ Unspecified Observation ∩ Predicted Observation ∩ Inferred Observation = Ø
- sourceOfData – identity of 3rd party resource
- observedCellType – the cell type in which the experimental observation occurred.
- experimentalMethod – one type of evidence for a particular interaction.

An Overview of an Instance



A Protein


http://
biomanta...ncbi_taxo.owl#
ncbi_taxo_4932_ind
“eIF-6”
“CDC95”
“TIF6”
ShortName
GeneSymbol
NCBI
Synonym
“eIF6”
Synonym
“Saccharomyces cerevisiae”
Species
_Protein
Sequence
“MATR...”
CrossReference
CrossReference
_Xref1
_Xref2
Accession
DataBase
Accession
DataBase
“Q12522”
“UniProt”
“YPR016C”
“MIPS”

An Agile Process (Rip, Mix, Verify)

- People and interactions over process.
- Working ontologies over well documented ones.
- Collaboration over negotiation.
- Respond to change rather than follow a plan.
- This means:
  - Verify the output (interaction, PubMedID, Experimental method, etc).
  - Everything in RDF and OWL.
  - Feedback - find missing data or mistakes in the model.

Acknowledgements

Chris Bouton
Victor Farutin
Mike Schaffer
Fred Jerva

Computation Sciences Center of Emphasis,
Pfizer Global Research and Development,
Pfizer Inc.

Kevin Burrage
Jane Hunter
Mark Ragan
Melissa Davis
Yuan-Fang Li
Shoaib Sehgal

School of ITEE and
Institute of Molecule Bioscience,
The University of Queensland.

Links

Web Site

- http://biomanta.org/

Ontologies and other Results

- http://biomanta.org/downloads/

Sample Output

- http://biomanta.org/downloads/2008/yeast.zip

RDF Library for Scale-out (in development)

- http://jrdf.sf.net/