

# BioPAX

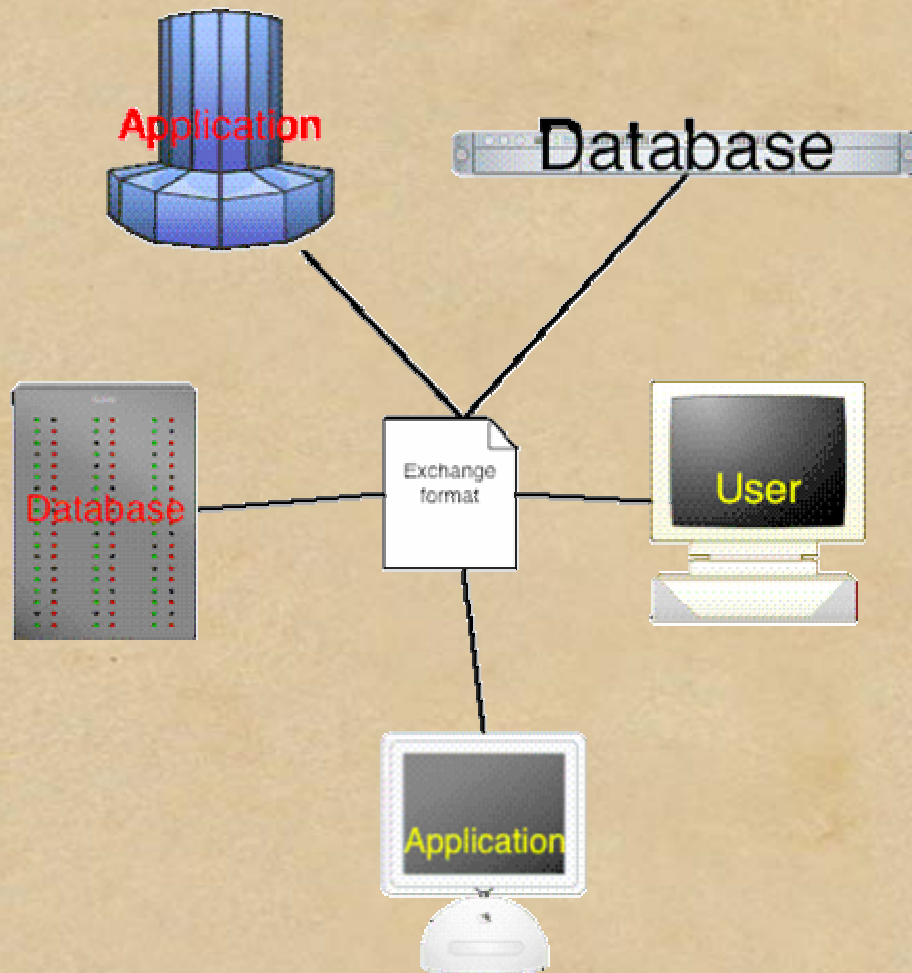
(and a "practical" introduction to OWL ontologies)

[andrea@pasteur.fr](mailto:andrea@pasteur.fr)

# PLR: Pathway Resource List

<http://www.cbio.mskcc.org/prl/index.php>  
>180 "pathway db" listed

# Exchanging information



- ♦ definition of a common exchange language...  
BioPAX

# BioPAX

(<http://www.biopax.org>)

- BioPAX = Biological Pathway Exchange (a data exchange format for pathway data)
- Include support for these pathway types:
  - Metabolic pathways
  - Signaling pathways
  - Protein-protein interactions
  - Genetic regulatory pathways
- Ideally all resources in the the Pathway Resource List

# BioPAX

- Conceptual framework based upon existing DB schemas (aMAZE, BIND, BioCyc, WIT, KEGG, Reactome, PATIKA, iNOH, TRANSPATH etc.)
- Allows wide range of detail, multiple levels of abstraction
- BioPAX is defined as an OWL Ontology

# BioPAX

- Development of BioPAX is an opensource project (GNU LGPL)
- To get involved:  
<http://www.biopax.org/mailman/listinfo/biopax-discuss> (biopax-discuss mailing list)
- Development of BioPAX follows a series of levels, each level extends and includes the previous one.

# BioPAX level 1

- It is targeted at representing **Metabolic Networks**

- Concepts:

- Entities: Small molecules, Proteins, RNA, Complexes

- Interactions: Biochemical Reactions, Enzyme Catalysis, Complex Assembly, Transport...

- Data from BioCyc, WIT, Amaze, KEGG...

- Released April 27, 2004 (Current version is 1.4)

- Data availability: BioCyc (under testing), KEGG, WIT coming soon

# BioPAX level 2

- It is targeted (also) at representing generic molecular interaction data.
- Adds DNA, physical-interaction (abstraction of control and conversion)
- Adds Evidence description
- Data from BIND, DIP, HPRD, MINT, IntAct, PSI-format
- Beta (version 0.9) available for testing and feedback (release date: April 13, 2005)

# BioPAX level 3

- Targeted at describing signaling pathways, regulation of gene expression, binding interaction involving DNA
- Will include concepts as: Gene, Genetic interactions, Gene regulation
- Will describe complex processes (eg. transcription, translation), nested networks.
- Data from TRANSPATH, TRANSFAC,...

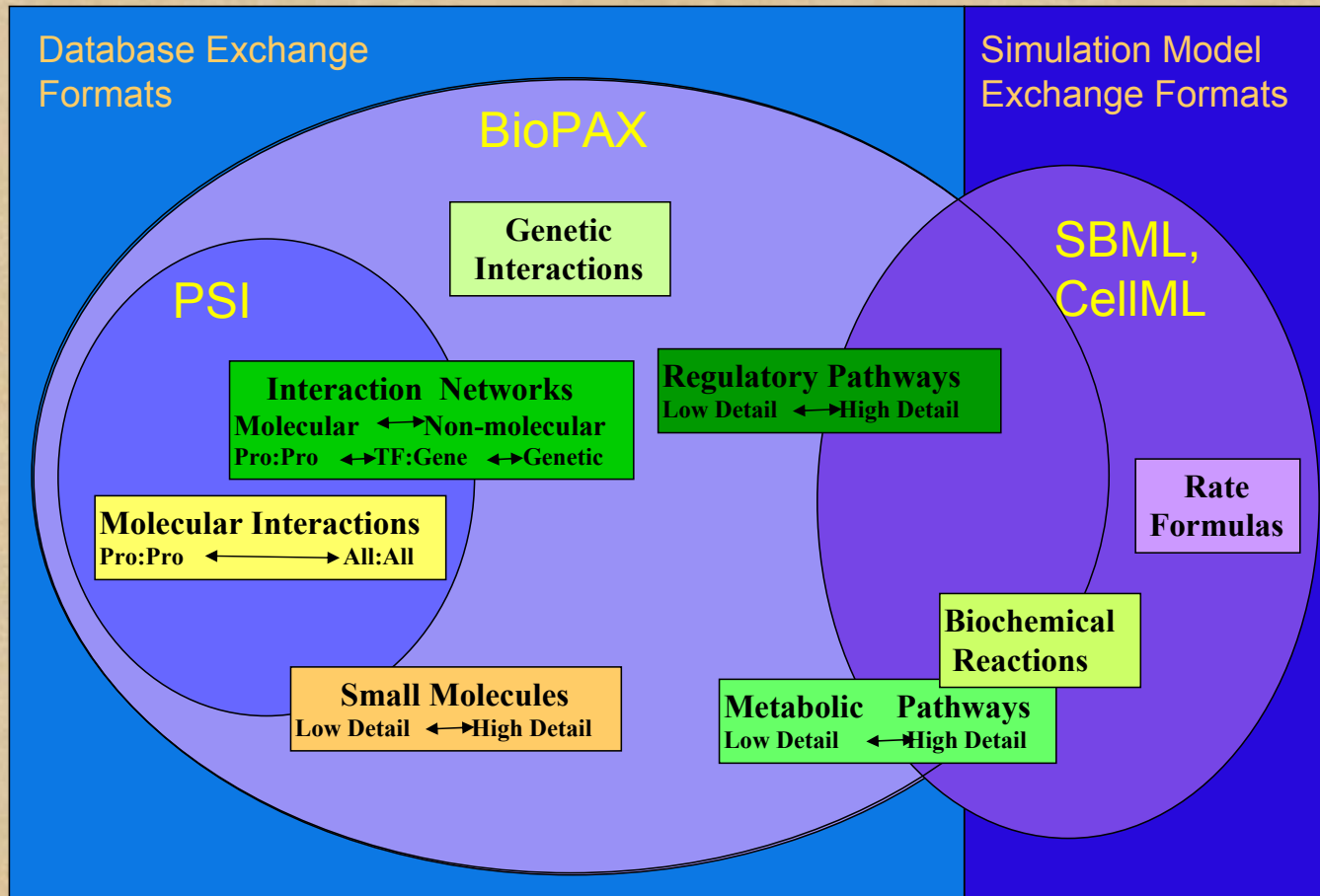
Future levels of BioPAX will include support for generic networks, abstract relationships (ex. from literature-mining), cell-compartments...

# SBML

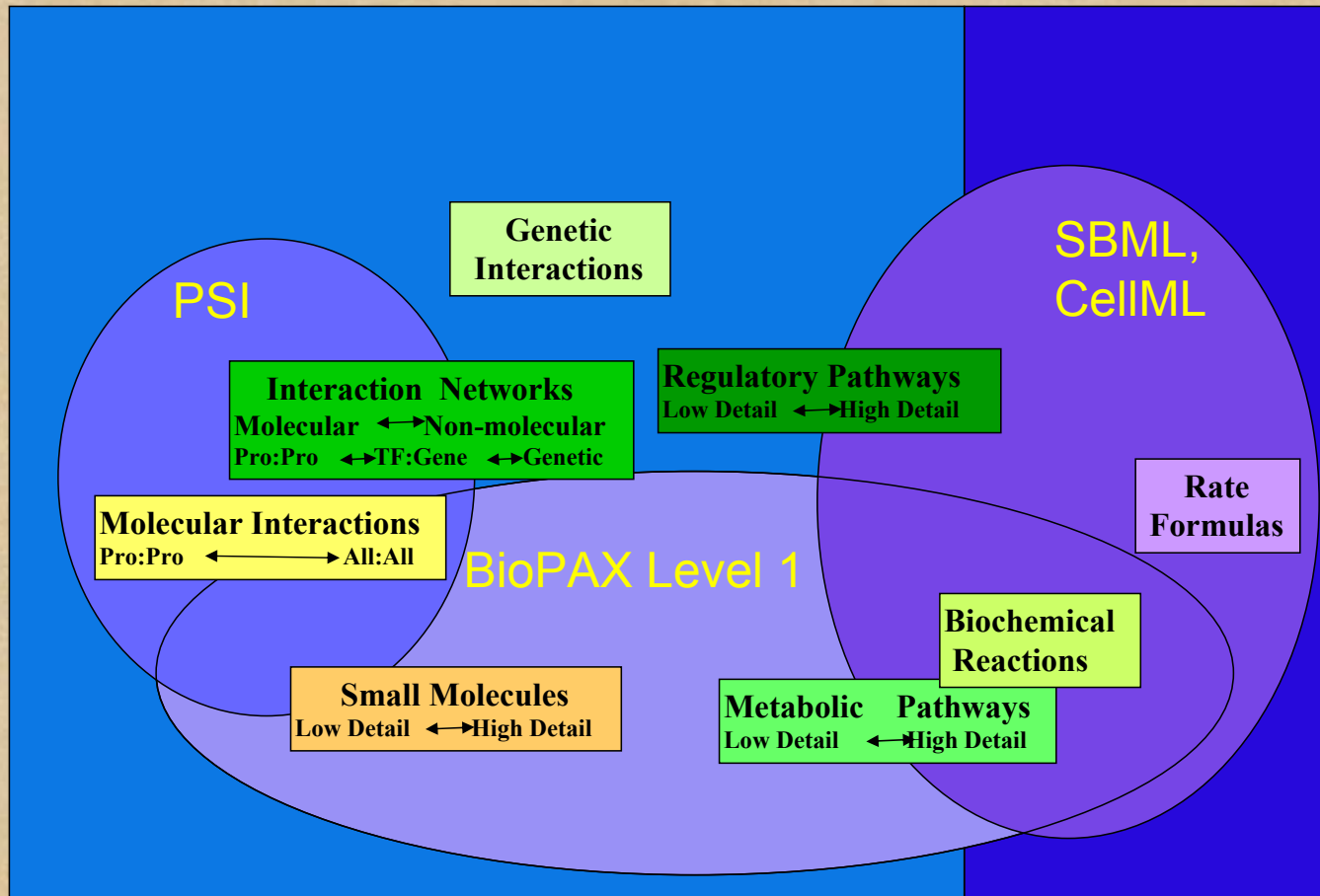
(<http://www.sbml.org>)

- The Systems Biology Markup Language is a language for representing models of biochemical reaction networks
- SBML is currently supported by more than 80 software systems
- It is defined in XML
- Its development follows an open source model (GNU LGPL) and is organized in levels (levels 1 and 2 are defined)

# BioPAX & SBML



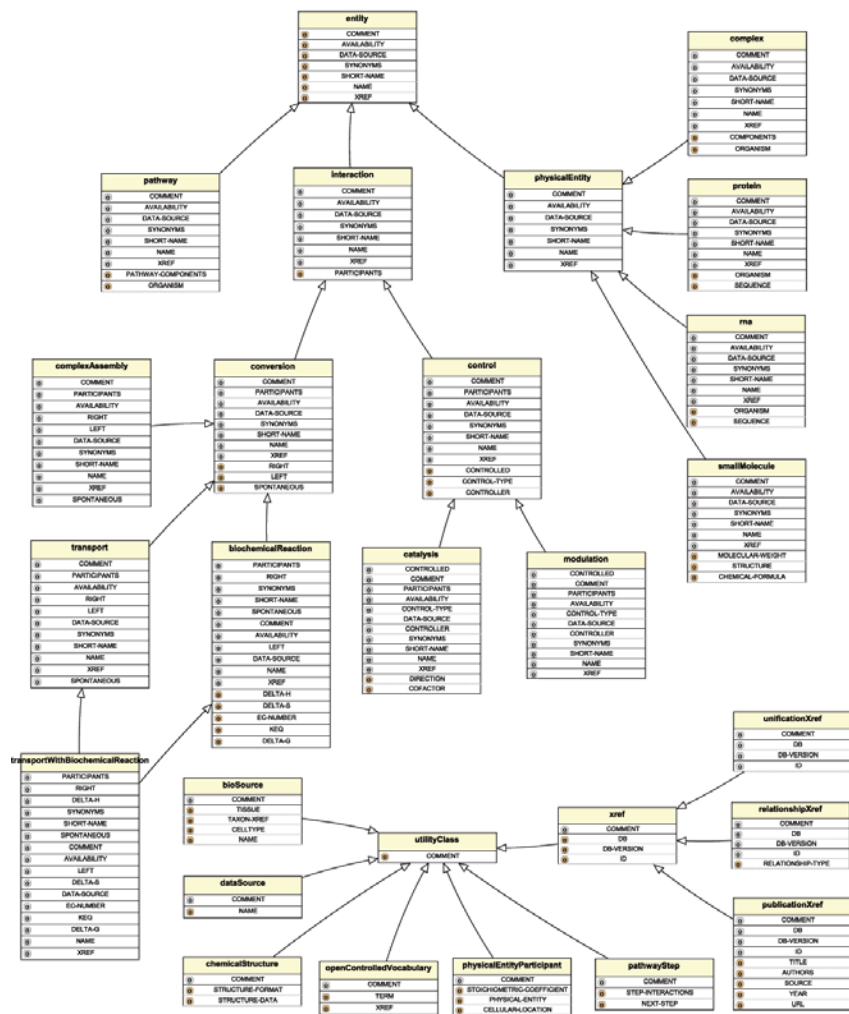
# BioPAX (1.1) & SBML

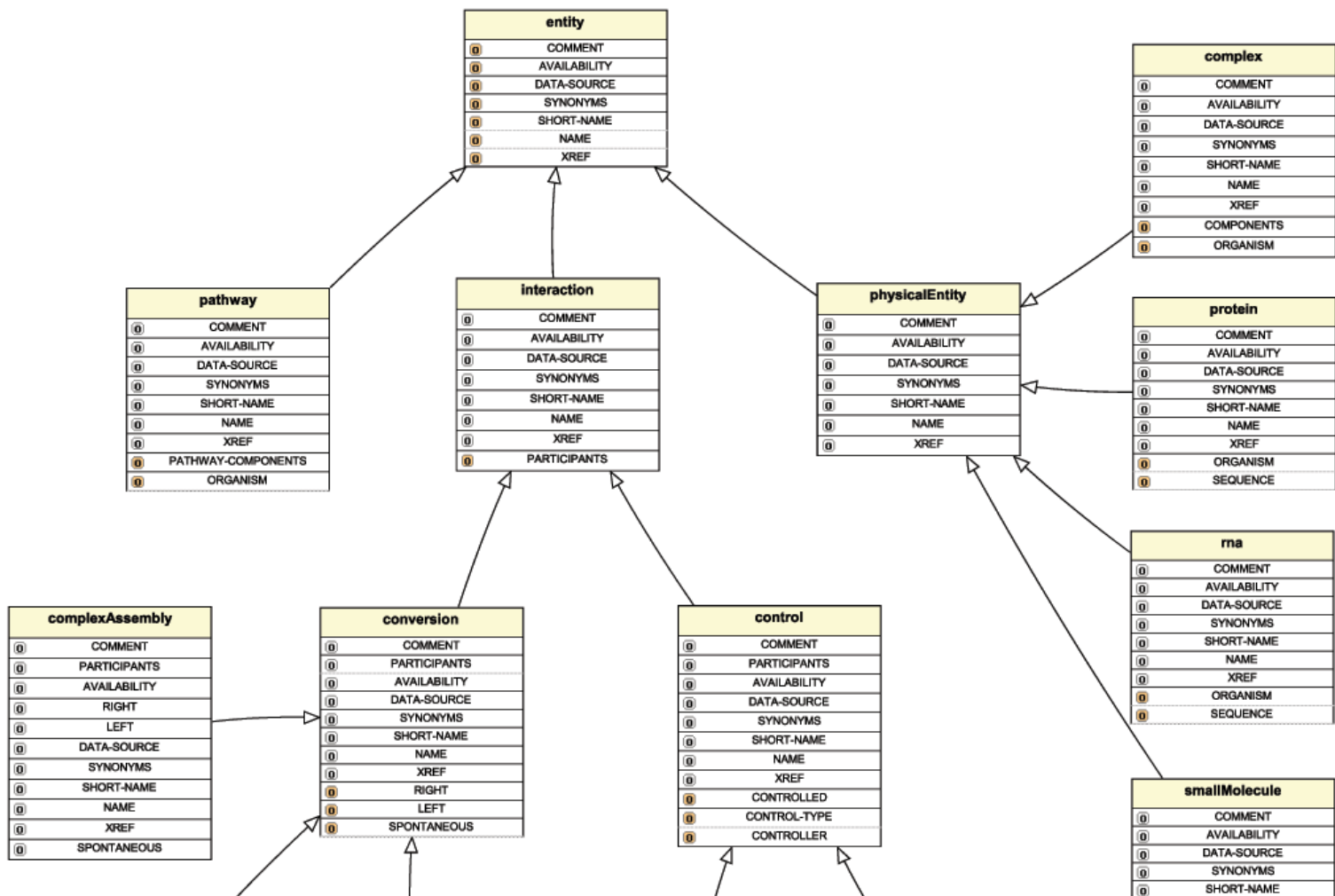


# BioPAX level 1 concepts

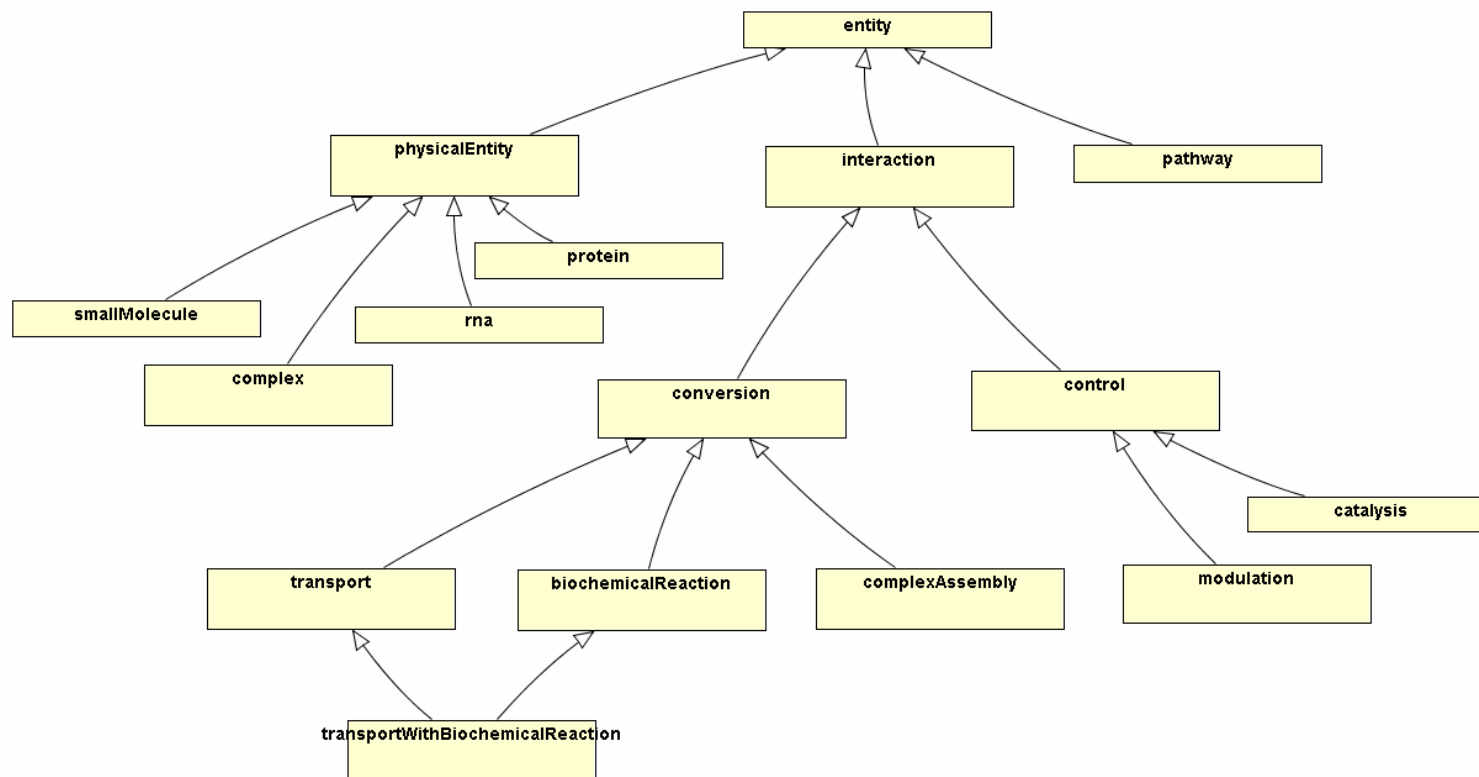
(slides from BioPAX consortium)

= agreement on how to define a pathway



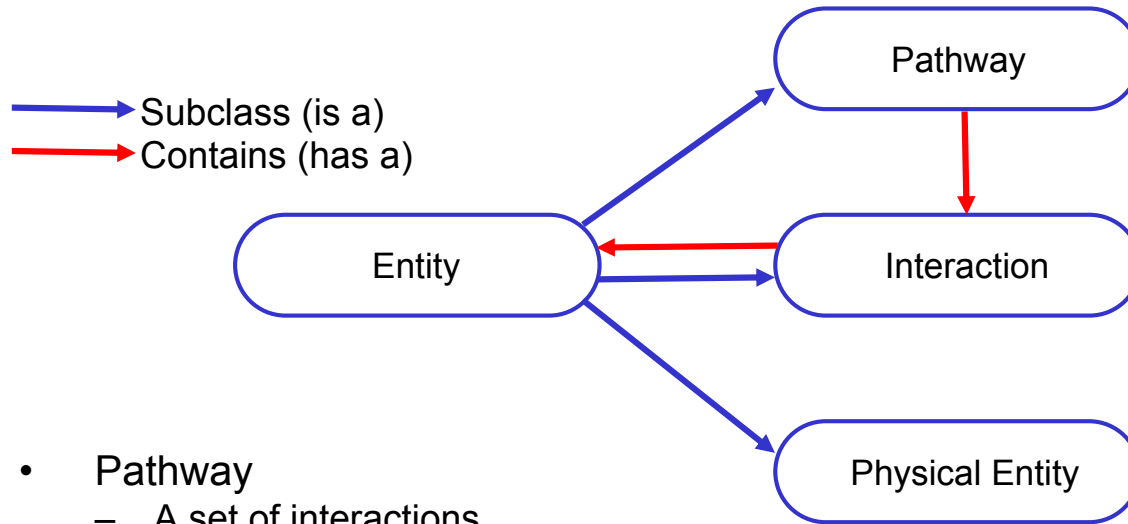


# Biological Ontology: Overview



Level 1 v1.0 (July 7th, 2004)

# BioPAX Ontology: Top Level



- Pathway
  - A set of interactions
  - E.g. Glycolysis, MAPK, Apoptosis
- Interaction
  - A set of entities and some relationship between them
  - E.g. Reaction, Molecular Association, Catalysis
- Physical Entity
  - A building block of simple interactions
  - E.g. Small molecule, Protein, DNA, RNA

# BiOFAAX Ontology: Root

- Root class: Entity
  - Any concept referred to as a discrete biological unit when describing pathways. This is the root class for all biological concepts in the ontology, which include pathways, interactions and physical entities

| entity                                                                              |              |
|-------------------------------------------------------------------------------------|--------------|
|    | SYNONYMS     |
|    | COMMENT      |
|    | DATA-SOURCE  |
|    | SHORT-NAME   |
|    | AVAILABILITY |
|  | NAME         |
|  | XREF         |

# BioPAX Ontology: Pathway

- Pathway
  - An entity that consists of a set of interactions. A pathway is a series of molecular interactions and reactions, often forming a network, which biologists have found useful to group together for organizational, historic, biophysical or other reasons.

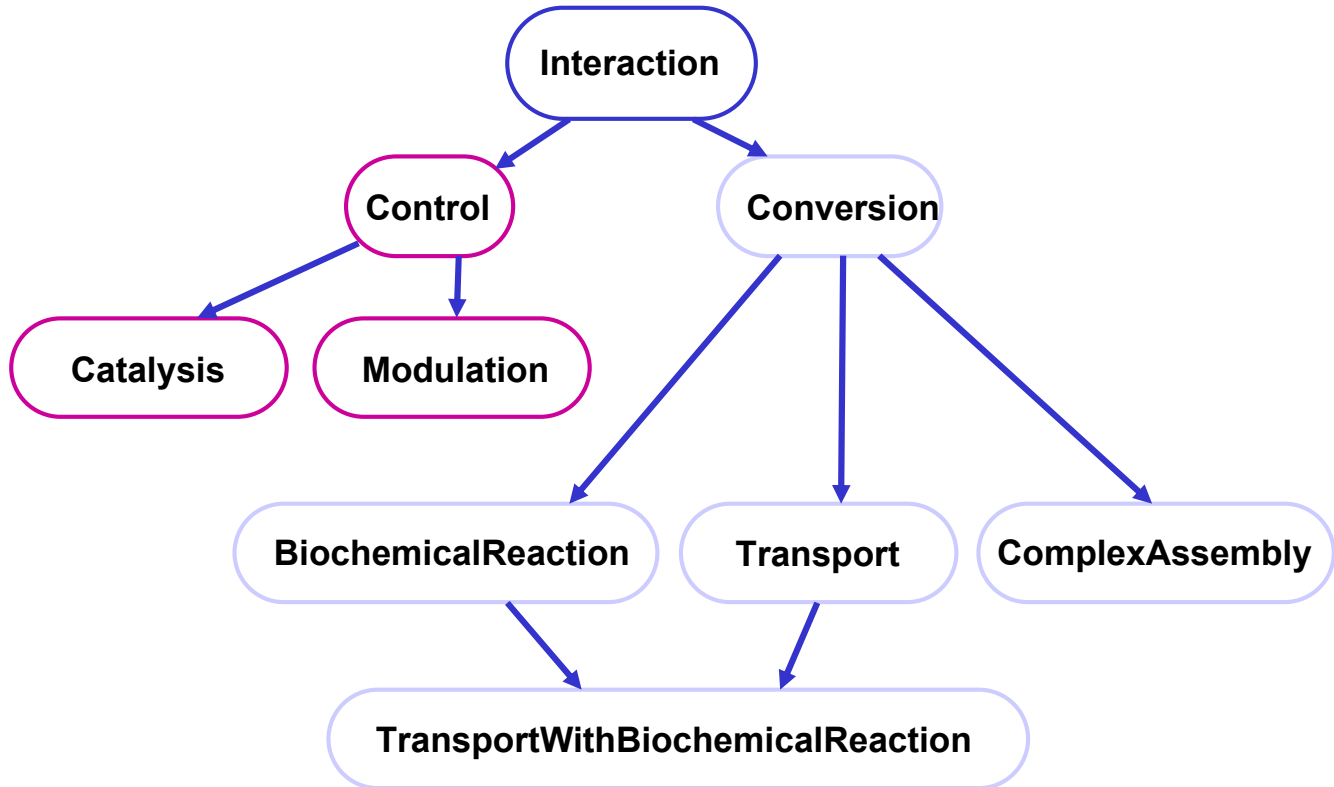
| pathway |                    |
|---------|--------------------|
| 0       | COMMENT            |
| 0       | AVAILABILITY       |
| 0       | DATA-SOURCE        |
| 0       | SYNONYMS           |
| 0       | SHORT-NAME         |
| 0       | NAME               |
| 0       | XREF               |
| 0       | PATHWAY-COMPONENTS |
| 0       | ORGANISM           |

# BioPAX Ontology: Interaction

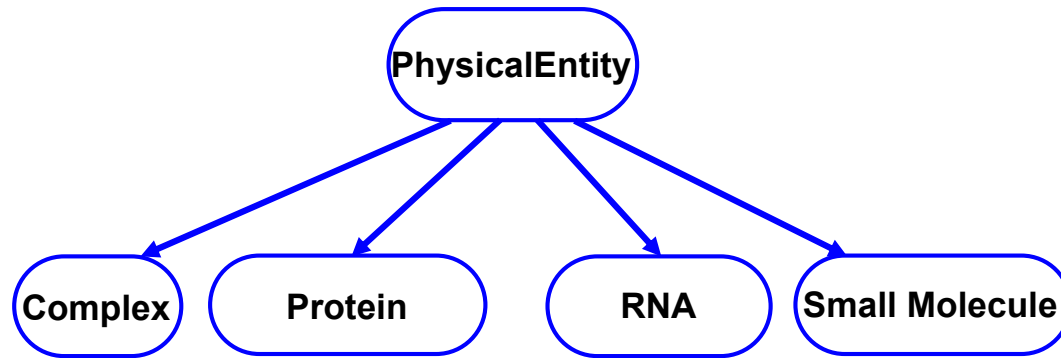
- Interaction
  - An entity that defines a single biochemical interaction between **two or more** entities. An interaction cannot be defined without the entities it relates.

| interaction |              |
|-------------|--------------|
| 0           | SYNONYMS     |
| 0           | COMMENT      |
| 0           | DATA-SOURCE  |
| 0           | SHORT-NAME   |
| 0           | AVAILABILITY |
| 0           | NAME         |
| 0           | XREF         |
| 0           | PARTICIPANTS |

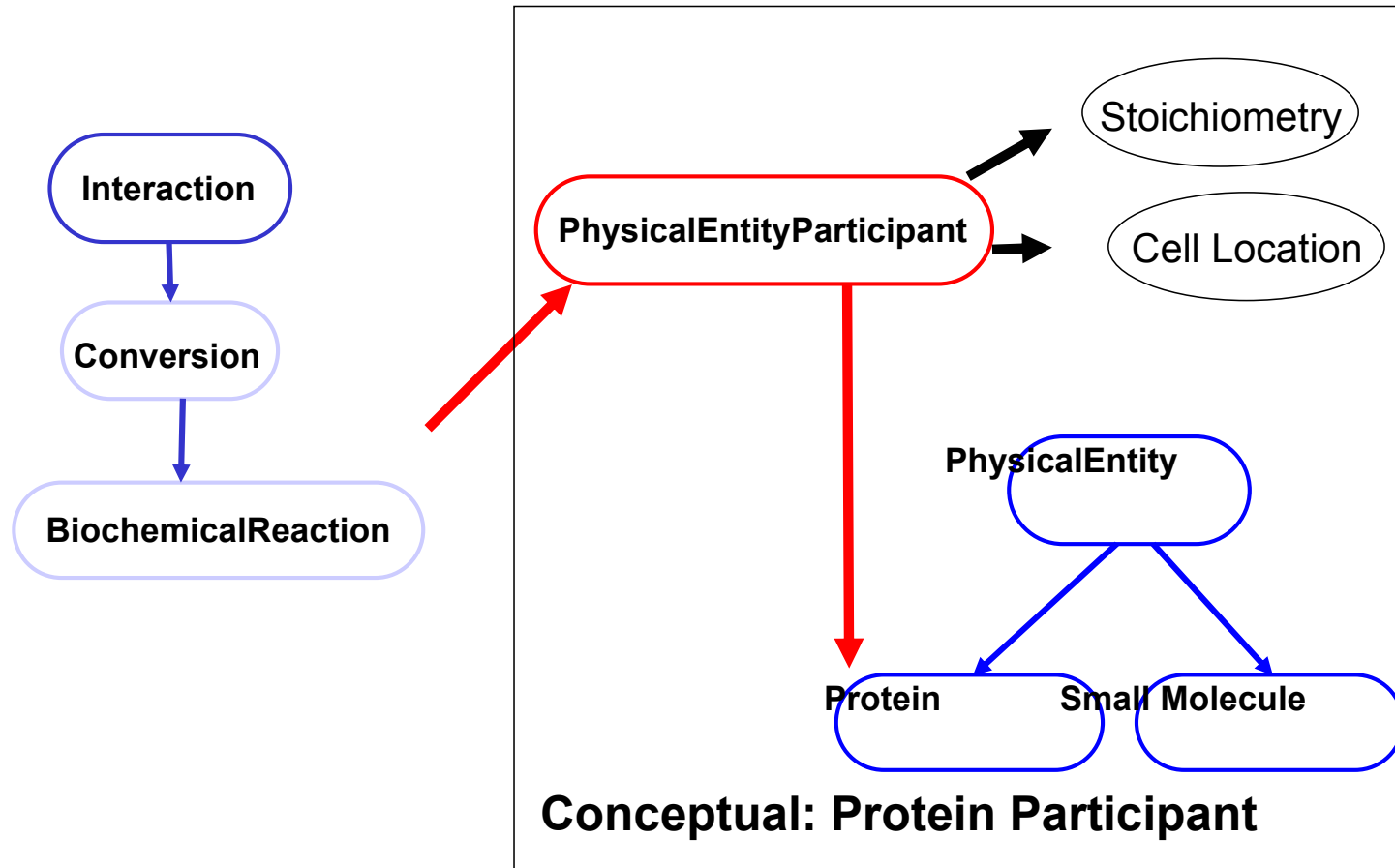
# BioPAX Ontology: Interactions



# BioPAX Ontology: Physical Entities



# Use of Physical Entities



“Practical” BioPAX

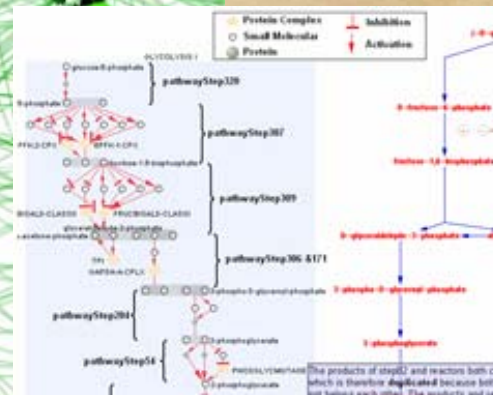
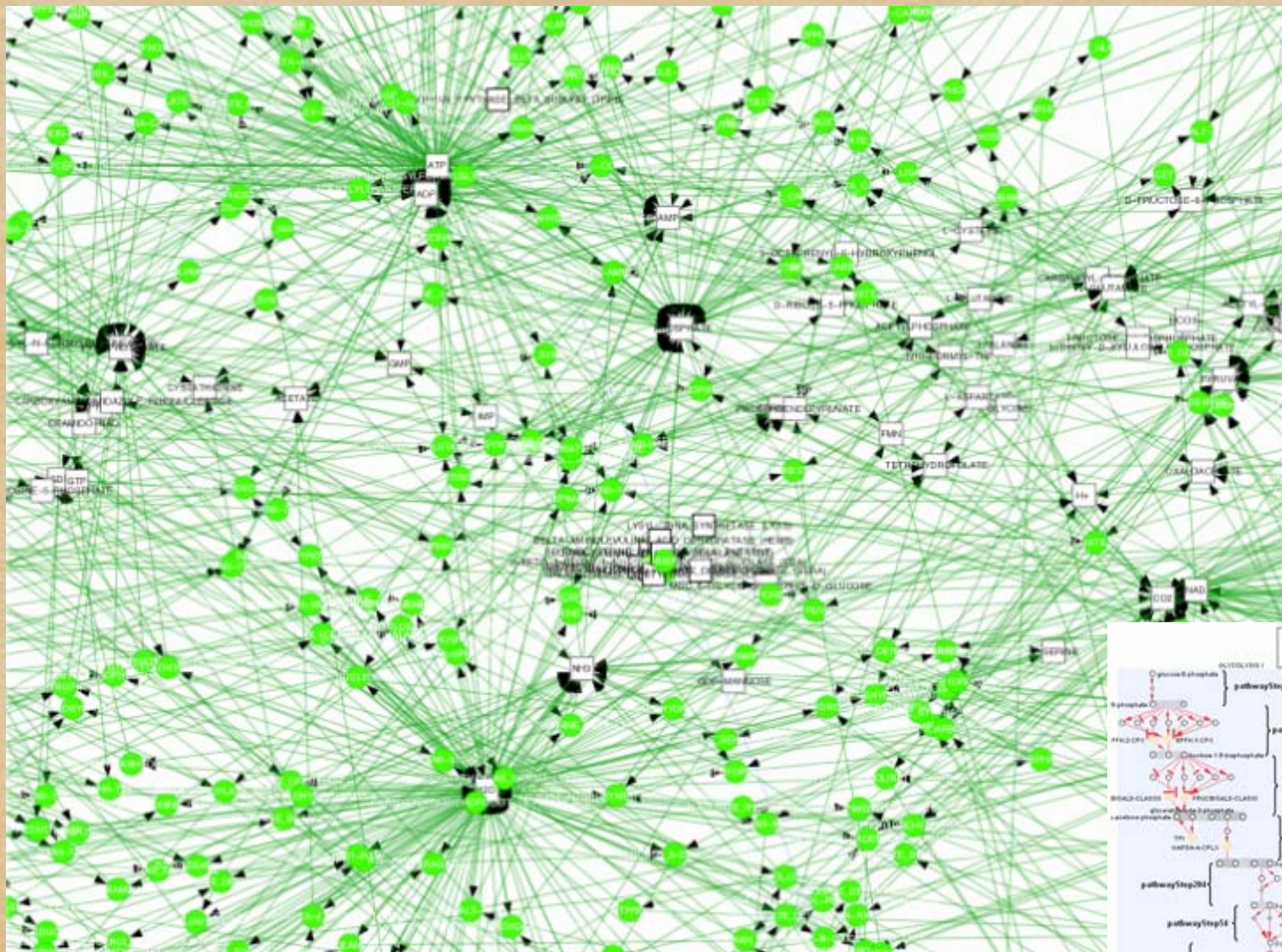
# from a BioPAX files...

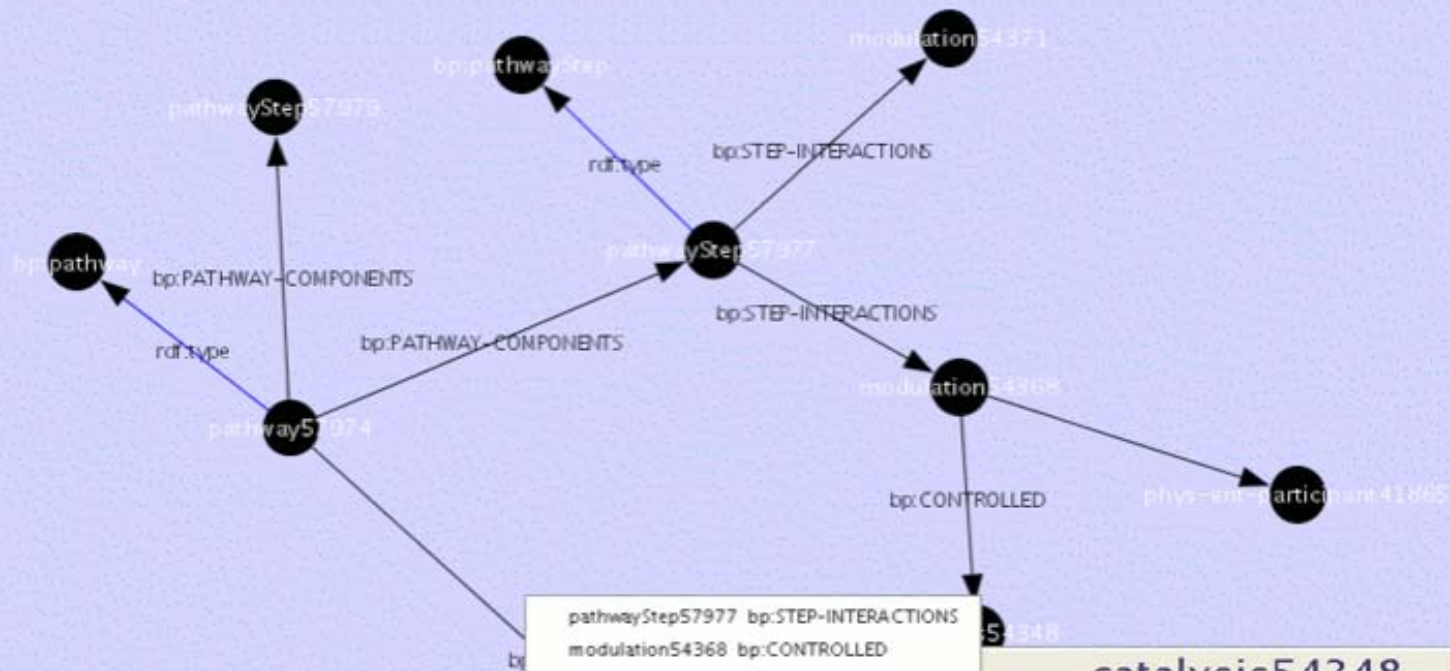
```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
  xml:base="http://biocyc.org/biopax/hpy-biopax#"
  xmlns="http://biocyc.org/biopax/biopax#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:xs="http://www.w3.org/2001/XMLSchema#"
  xmlns:bp="http://www.biopax.org/release/biopax-level1.owl#"
>
  <owl:Ontology rdf:about="">
    <owl:imports rdf:resource="http://www.biopax.org/release/biopax-level1.owl"/>
    <rdfs:comment>Copyright &#x00A9; 2001, SRI International. All Rights Reserved.</rdfs:comment>
  </owl:Ontology>
  <bp:unificationXref rdf:ID="unificationXref16778">
    <bp:ID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">ATPSYN-RXN</bp:ID>
    <bp:DB rdf:datatype="http://www.w3.org/2001/XMLSchema#string">HpyCyc</bp:DB>
  </bp:unificationXref>
  <bp:unificationXref rdf:ID="unificationXref16780">
    <bp:ID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">GO:0005737</bp:ID>
    <bp:DB rdf:datatype="http://www.w3.org/2001/XMLSchema#string">GO</bp:DB>
  </bp:unificationXref>
  <bp:unificationXref rdf:ID="unificationXref16782">
    <bp:ID rdf:datatype="http://www.w3.org/2001/XMLSchema#string">PROTON</bp:ID>
    <bp:DB rdf:datatype="http://www.w3.org/2001/XMLSchema#string">HpyCyc</bp:DB>
  </bp:unificationXref>
  <bp:unificationXref rdf:ID="unificationXref16786">
```

...

...

to...





pathwayStep57977 bp:STEP-INTERACTIONS  
 modulation54368 bp:CONTROLLED  
 modulation54367 bp:CONTROLLED  
 modulation54374 bp:CONTROLLED  
 modulation54371 bp:CONTROLLED  
 modulation54372 bp:CONTROLLED  
 modulation54365 bp:CONTROLLED  
 modulation54366 bp:CONTROLLED  
 pathwayStep58049 bp:STEP-INTERACTIONS

**catalysis54348**  
 Web Info  
 Attribute Browser  
 Node Editing (*short-term*)  
 Set First Neighbors (*short-term*)  
 Subject of  
 Object of

- Find all proteins in pathway  $X$
- Find all elements interacting with  $X$
- Find all elements controlling  $X$
- ...

- ◆ Useful information for analysis:

- ◆ Evidence

- ◆ “connectivity”

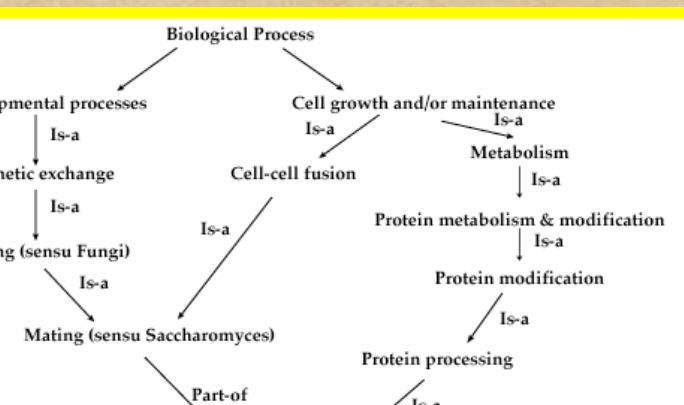
- ◆ Typed relations or sets

- ◆ ...

What is an OWL  
ontology ??

# What an ontology is...

An ontology is an explicit definition of terms and relations describing an area of knowledge (domain). The term ontology is used to refer many “artifacts”, from simple taxonomies as well as logical theories



| NECESSARY & SUFFICIENT                                |                                                     |
|-------------------------------------------------------|-----------------------------------------------------|
| NECESSARY                                             |                                                     |
| ● control                                             | <input type="checkbox"/>                            |
| ⌞ ∇ CONTROLLED catalysis                              | <input type="checkbox"/>                            |
| ⌞ CONTROLLER ≤ 1                                      | <input type="checkbox"/>                            |
| INHERITED                                             |                                                     |
| ⌞ CONTROL-TYPE ≤ 1                                    | [from control] <input type="checkbox"/>             |
| ⌞ NAME ≤ 1                                            | [from entity] <input type="checkbox"/>              |
| ⌞ ∃ PARTICIPANTS physicalEntityParticipant            | [from physicalInteraction] <input type="checkbox"/> |
| ⌞ ∃ PARTICIPANTS (entity ⊔ physicalEntityParticipant) | [from interaction] <input type="checkbox"/>         |
| ⌞ SHORT-NAME ≤ 1                                      | [from entity] <input type="checkbox"/>              |

Suppose we want to define a pathway exchange standard... we would need:

- A common syntax (a standard way of writing things)
- A common semantics (a standard representation of our domain)

# Syntax

- There is a common meta-language used to define the syntax of languages, XML
- XML files can be read (and written) with standard tools (DOM, SAX parsers...)

Some example will intuitively show how this meta-language works...

# SBML example (an XML language for systems biology)

More later!

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml xmlns="http://www.sbml.org/sbml/level1" level="1" version="2">
  <model name="Branch">
    <notes>
      <body xmlns="http://www.w3.org/1999/xhtml">
        <p>Simple branch system.</p>
        <p>The reaction looks like this:</p>
        <p>reaction-1:  X0 -> S1; k1*X0;</p>
        <p>reaction-2:  S1 -> X1; k2*S1;</p>
        <p>reaction-3:  S1 -> X2; k3*S1;</p>
      </body>
    </notes>
    <listOfCompartments>
      <compartment name="compartmentOne" volume="1"/>
    </listOfCompartments>
    <listOfSpecies>
      <speciesname="S1" initialAmount="0" compartment="compartmentOne"
        boundaryCondition="false"/>
      <speciesname="X0" initialAmount="0" compartment="compartmentOne"
        boundaryCondition="true"/>
      <speciesname="X1" initialAmount="0" compartment="compartmentOne"
        boundaryCondition="true"/>
      <speciesname="X2" initialAmount="0" compartment="compartmentOne"
        boundaryCondition="true"/>
    </listOfSpecies>
    <listOfReactions>
      <reaction name="reaction_1" reversible="false">
        <listOfReactants>
          <speciesReferencespecies="X0" stoichiometry="1"/>
        </listOfReactants>
        <listOfProducts>
          <speciesReferencespecies="S1" stoichiometry="1"/>
        </listOfProducts>
        <kineticLaw formula="k1 * X0">
          <listOfParameters>
            <parameter name="k1" value="1"/>
          </listOfParameters>
        </kineticLaw>
      </reaction>
    </listOfReactions>
  </model>
</sbml>
```

# KGML example (KEGG xml export)

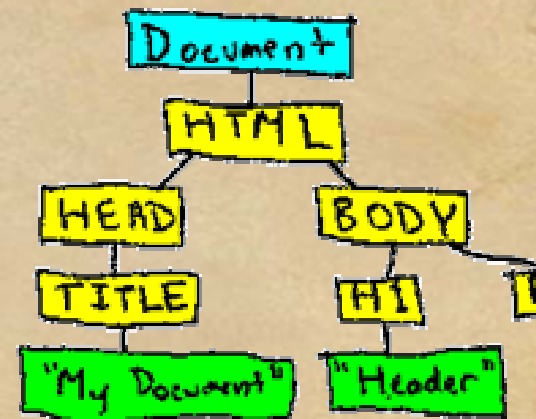
```
?xml version="1.0"?>
DOCTYPE pathway SYSTEM "http://www.genome.jp/kegg/xml/KGML_v0.6_.dtd">
-- Creation date: Jul 14 2004 02:10:15 +0900 (JST) -->
pathway name="path:map00730" org="map" number="00730"
  title="Thiamine metabolism"
  image="http://www.genome.jp/kegg/pathway/map/map00730.gif"
  link="http://www.genome.jp/dbget-bin/show_pathway?map00730">
<entry id="1" name="path:map00730" type="map"
  link="http://www.genome.jp/dbget-bin/get_linkdb?pathway+map00730">
  <graphics name="TITLE:Thiamine metabolism" fgcolor="#000000" bgcolor="#FFFFFF"
    type="roundrectangle" x="429" y="60" width="187" height="27"/>
</entry>
<entry id="2" name="path:map00010" type="map"
  link="http://www.genome.jp/kegg/pathway/map/map00010.html">
  <graphics name="Glycolysis / Gluconeogenesis" fgcolor="#000000" bgcolor="#FFFFFF"
    type="roundrectangle" x="77" y="261" width="72" height="26"/>
</entry>
<entry id="3" name="path:map00230" type="map"
  link="http://www.genome.jp/kegg/pathway/map/map00230.html">
  <graphics name="Purine metabolism" fgcolor="#000000" bgcolor="#FFFFFF"
    type="roundrectangle" x="95" y="157" width="116" height="27"/>
</entry>
<entry id="4" name="ec:2.5.1.2" type="enzyme" reaction="m:R02863"
  link="http://www.genome.jp/dbget-bin/www_bget?enzyme+2.5.1.2">
  <graphics name="2.5.1.2" fgcolor="#000000" bgcolor="#FFFFFF"
    type="rectangle" x="576" y="372" width="45" height="17"/>
```

# xhtml (almost html)

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en" lang="eng">
<head>
  <meta name="Description" content="An example XHTML page" />
  <meta name="Keywords" content="Some, Keywords, Here">
  <link rel="stylesheet" type="text/css" href="http://www.warwick.ac.uk/www/webstyle.css" title="webstyle" />
  <title>XHTML Example</title>
</head>
<body>
<h1>XHTML Example</h1>
<p>Here is an example of a page written in XHTML.</p>
<p>If you view the source for this page, you may spot a few differences e.g. </p>
<ul><li>The DTD definition on the top of the page (see <a href="http://www.w3.org/TR/xhtml1/">XHTML 1.0</a> for an explanation)</li>
<li>All tags should be in lower case for example: <lt;b></b>; (this is a <a href="/www/standards/">University Web Standard</a></li>
<li>All tags should be closed for example: <lt;p></p>; (this is a <a href="/www/standards/">University Web Standard</a></li>
<li>Attribute values should be in quotes for example: src="logo.gif" (this is best practice anyway)</li>
<li>Tags are correctly nested (this is best practice anyway)</li>
</ul>
<br />
<table bgcolor="#CCCCFF" width="600" cellspacing="4" cellpadding="4" border="0">
<tr>
<td width="300">1. The correct DTD must be used.</td>
<td width="300">2. Scripts and styles require extra syntax.</td>
</tr>
<tr>
```

note:

XML is a  
free language



# XML

- Each language differs from the specific entities (and their “syntactic” relations). These properties are defined in a language-specific file (DTD or XML-schema).
- An xml file is “correct” if it complies to the XML syntax
- An xml file is “valid” if it complies to the XML syntax and to the specific XML-Schema or DTD file

# XML

not correct



not valid



From xhtml dtd...

```
....  
<!ELEMENT textarea (#PCDATA)>  
<!-- multi-line text field -->  
<!ATTLIST textarea  
  %attrs;  
  %focus;  
  name CDATA #IMPLIED  
  rows %Number; #REQUIRED  
  cols %Number; #REQUIRED  
  disabled (disabled) #IMPLIED  
  readonly (readonly) #IMPLIED  
  onselect %Script; #IMPLIED  
  onchange %Script; #IMPLIED  
>  
....
```

# XML

- There is not an explicit definition of relations among data in the XML file.
- “find all control interaction” how can I find it if in the file the words “modulation” or “catalysis” are used?

# Describing semantics

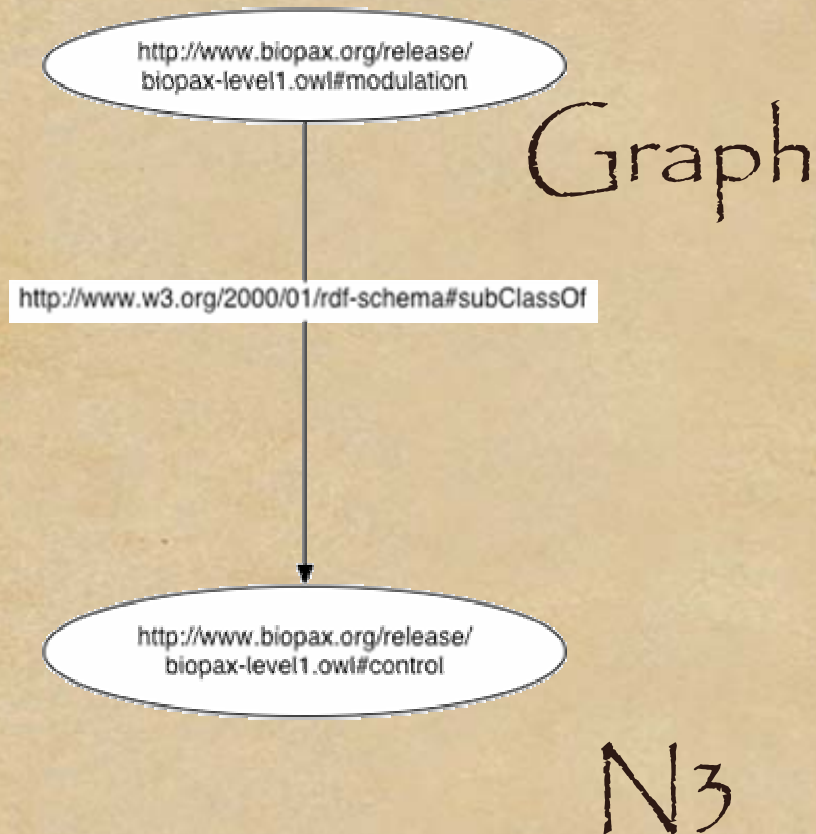
A simple way of describing relations among concepts is to assert a set of predicates of the form: <subject property object>

This is the basis of the RDF (Resource Description Format) specification.

RDF defines a data-model, that can be represented with several syntaxes  
(Graphs, N3, XML)

Note: in RDF subjects, properties and objects (not always) are univocally identified (// IRI)

# RDF-example



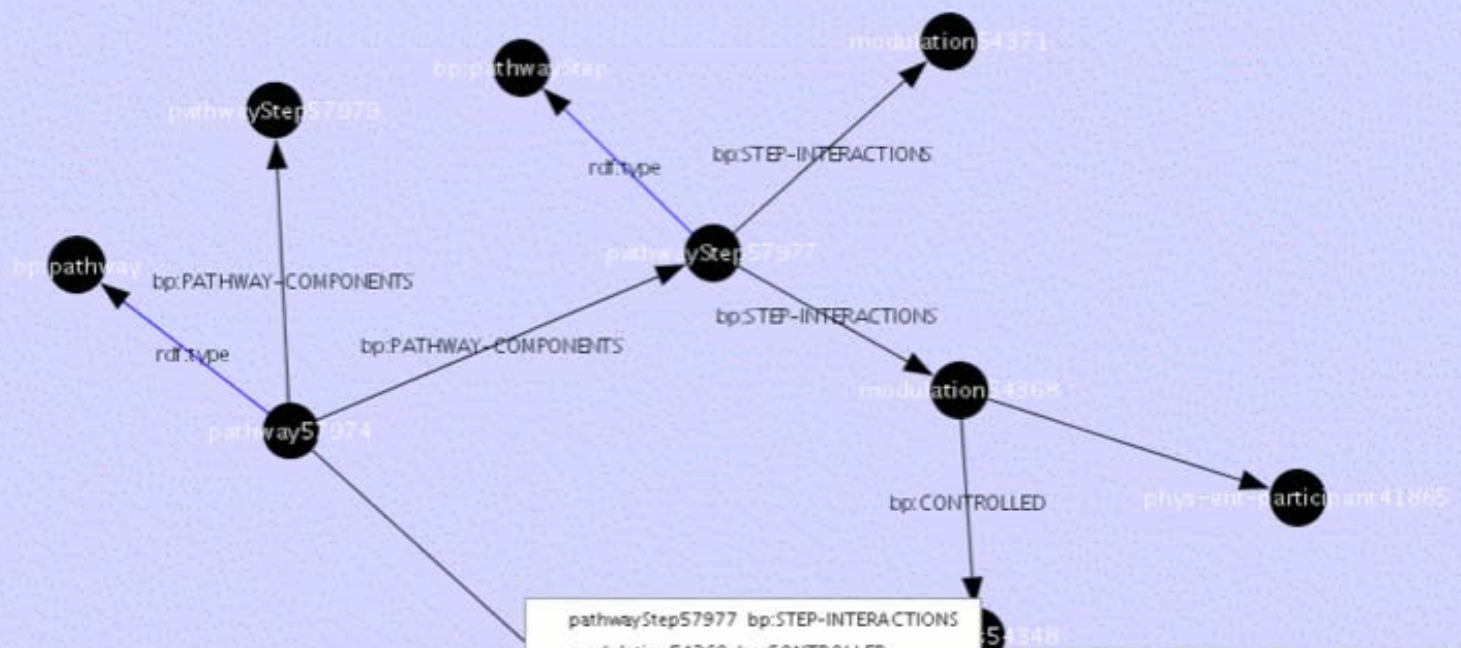
XML

```
<?xml version="1.0" ?>
<rdf:RDF
  xmlns:rdf="..."
  xmlns:rdfs="..."
  xmlns:biopax="..."
  <rdf:Description rdf:about="biopax:modulation">
    <rdfs:subClassOf>
      <biopax:control />
    </rdfs:subClassOf>
  </rdf:Description>
</rdf:RDF>
```

```
<http://www.biopax.org/release/biopax-level1.owl#modulation> <http://www.w3.org/2001/01/rdf-schema#subClassOf>
  <http://www.biopax.org/release/biopax-level1.owl#control>
```

Nodes: 11(0)    Edges: 9(0)

RDFScape



- pathwayStep57977 bp:STEP-INTERACTIONS
- modulation54368 bp:CONTROLLED
- modulation54367 bp:CONTROLLED
- modulation54374 bp:CONTROLLED
- modulation54371 bp:CONTROLLED
- modulation54372 bp:CONTROLLED
- modulation54365 bp:CONTROLLED
- modulation54366 bp:CONTROLLED
- pathwayStep58049 bp:STEP-INTERACTIONS

catalysis54348

Web Info ▶  
 Attribute Browser  
 Node Editing (short-term) ▶  
 Set First Neighbors (short-term) ▶  
 Subject of ▶  
 Object of ▶

# RDF

- Tools exists to read/write and “aggregate” different files ([jena.sourceforge.net](http://jena.sourceforge.net))
- Languages such as RDQL can be used to “query the graph”

RDQL query

```
SELECT ?modulation  
WHERE (?modulation biopax:CONTROLLER ?x)  
(?x rdfs:subClassOf biopax:protein)  
USING rdfs for...  
biopax for...
```

Don't try this!

# Semantics

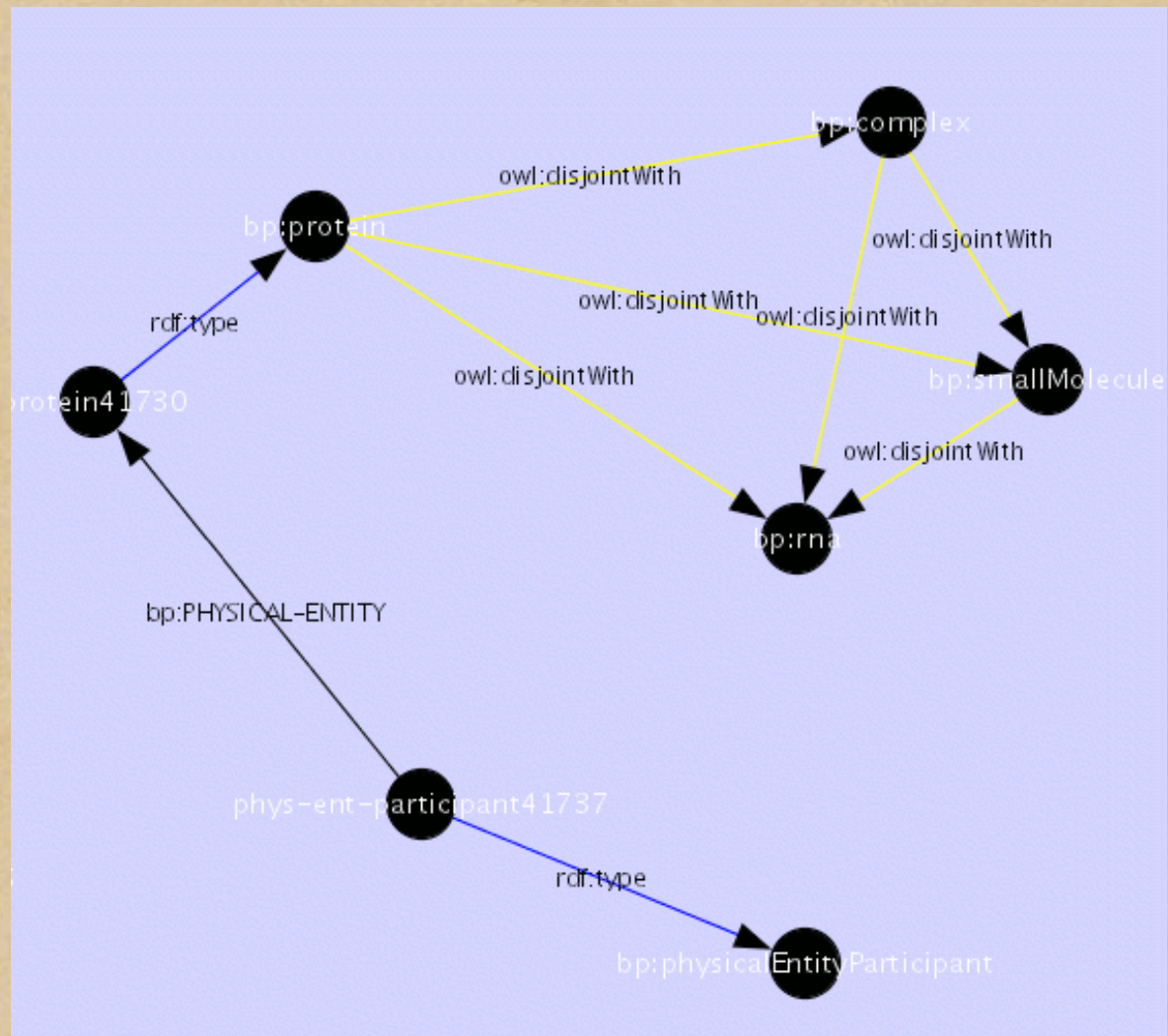
- ♦ RDF-schema adds specific semantics to represent properties and classes of RDF resources (ex. isSubclassOf, property restrictions)
- ♦ OWL (Web ontology language) extends RDF and RDF-schema to gain more expressiveness.

# OWL

- OWL provides three increasingly expressive sublanguages:
- OWL-Lite
- OWL-DL
- OWL-Full
- Distinction between classes and instances

# OWL-DL

- There is a trade-off between expressiveness of an ontology and its “tractability”: if and how much it takes to answer a query (a first-order logic is undecidable!).
- OWL-DL extends RDF with a set of concepts expressive enough yet allowing the ontology to be tractable (any query can be answered in a finite time).



# OWL

- Tools like Protégé can be used to edit and browse an OWL ontology
- Racer can be used to perform reasoning over an ontology

<http://protege.stanford.edu>

Protege

biopax-level1 Protégé 3.1 beta (file:/Users/andreasp Splendiani/Desktop/biopax-level1.pprj, OWL Files (.owl or .rdf))

File Edit Project OWL Code Window Tools Help

owl:Thing

- entity
  - interaction
    - control
      - catalysis
      - modulation
    - conversion
      - biochemicalReaction
        - transportWithBiochemical
      - complexAssembly
      - transport
        - transportWithBiochemical
  - pathway
  - physicalEntity
    - complex
    - protein
    - rna
    - smallMolecule
  - utilityClass
    - bioSource
    - chemicalStructure
    - dataSource
    - openControlledVocabulary
    - pathwayStep
    - physicalEntityParticipant
    - xref
      - publicationXref

Subclass Relationship

CLASS EDITOR

For Class: modulation (instance of owl:Class)

Name

SameAs

DifferentFrom

modulation

rdfs:comment

catalysis instance. A typical modulation instance has a small molecule as the controller entity and a catalysis instance as the controlled entity. Examples of instances of this class include allosteric activation and competitive inhibition of an enzyme's ability to catalyze a specific reaction.

Annotations

| Property     | Value                             | Lang |
|--------------|-----------------------------------|------|
| rdfs:comment | A control interaction in which... |      |

Asserted

Inferred

Asserted Conditions

NECESSARY & SUFFICIENT

control

∀ CONTROLLED catalysis

CONTROLLER ≤ 1

NECESSARY

CONTROLLED ≤ 1

NAME ≤ 1

∃ PARTICIPANTS (entity ∪ physicalEntityParticipant)

SHORT-NAME ≤ 1

INHERITED

(from control)

(from entity)

(from inter...)

Properties

CONTROL-TYPE (multiple owl:oneOf)

CONTROLLED (multiple entity ∪ physicalEntity)

CONTROLLER (multiple physicalEntity)

NAME (multiple xsd:string)

PARTICIPANTS (multiple entity ∪ physicalEntity)

SHORT-NAME (multiple xsd:string)

AVAILABILITY (multiple xsd:string)

COMMENT (multiple xsd:string)

DATA-SOURCE (multiple dataSource)

Disjoints

catalysis

Logic View

Properties View

# Back to BioPAX

BioPAX is defined as an OWL(DL) ontology,  
hence...

- It is syntactically written in XML (DOM or SAX parser could be used to read it).
- It is an RDF file. Tools like Jena can be used to read/write it. Queries based on graphs and using RDF-Schema properties are possible (all controls in a pathway... knowing a control is a modulation or catalysis )
- It is an OWL file. Could be tested for semantic coherence, aggregated with other ontologies. Reasoning (subsumption) could be done on it.

# Gene Ontology and OWL

- Other ontologies (like GO) could be represented in OWL (integration!)
- The information structure in GO is anyway low...
- Reasoning on it yields one redundant node!