

BioPAX

- Community update and BioPAX overview - Nadia Anwar
- BioPAX Roadmap - Level 4 Proposals - Gary Bader
- Workgroups - Emek Demir, Marijn van Iersel, Andrea Splendiani
- Discussion outline
 - BioPAX core/BioPAX addons
 - Generics
 - Polymerisation
 - Roadmap-Survey

November NYC meeting - Summary

- BP –L3 ratified
- Pathway layout
- Governance – editors – subgroups
- Next meeting - COMBINE

November NYC meeting - Discussion items

- Black box pathways
- Pathways with input and output (NCI-PID and Syngenta)
- Also perturbation/phenotype
- Production without consumption (non stoichiometric)
- Control logic (mim)
- Polymerisation
- Andrea's Semantic Integration wish list Controlled vocabularies
- Documentation
- What biopax does not do
- Granularity/Scope
- Reaction Rates/rate laws – link to SBML
- Outreach
- Paper

Governance

Role and process of the SAB and BioPAX Organisational Structure

BioPAX core group - Individuals funded to work on BioPAX

This is the administrative structure co-ordinated between MSKCC and University of Toronto.

Remit

- Workgroups/Editorial/SAB co-ordination
- BioPAX ontology dissemination, documentation and workshops, editorial mediation
- Organising submissions/proposals for workgroups
- Calls for workgroup participation and workshop organisation
- Align community requirements and SAB recommendations with community and workgroup efforts
- BioPAX outreach (e.g. tutorials, talks, posters at conferences)
- Annual workshop organisation

Editorial Board - Community vote

Technical experts to co-ordinate submissions (new classes/properties, refactorings, controlled vocabularies) into the BioPAX ontology and ensure compatibility with the ontology design.

Remit

- Technical implementation, Architecture, Modeling, Use Cases (Data providers)
- Stewardship of BioPAX Teams/Workgroups
- Propose BioPAX ontology recommendations to community at workshops and deliver recommendations for future ontology levels
- Align community requirements and SAB recommendations with community and workgroup efforts

Workgroups - Open participation

Remit

- outline and discuss recommendations for future ontology levels
- submit recommendations to Editorial Committee

Scientific Advisory Board - Stakeholders

Governance

Scientific Advisory Board/Steering Committee

- Gary Bader, University of Toronto
- Peter D'Eustachio, New York University
- Henning Hermjakob, European Bioinformatics Institute
- Peter Karp, SRI International
- Nicolas Le Novère, European Bioinformatics Institute
- Chris Sander, Memorial Sloan-Kettering Cancer Center
- Carl Schaefer, National Cancer Institute

Governance

Core Group - funded (project deliverables)

- Emek Demir - PaxTools, Technical Implementation, Validator
- Igor Rodchencov - PaxTools, Validator
- Nadia Anwar - Community Support, Technical Implementation, OWL support
- Ozgun Babur - graph querying, paxtools, pathway merging
- Benjamin Gross - cpath2 development
- Ethan Cerami - cpath2 development

Governance

Editorial Committee - elected (workgroup deliverables, specifications and proposals)

- Andrea Splendiani - Rothamstead Research, Harpenden, UK
- Peter D'Eustachio - NYU, New York, US
- Oliver Ruebenacker - University of Connecticut Health Center, Connecticut, US
- addition of 2-3 more editors in 12-18 months

2010 Milestone

- Website/Wiki/Mailing list updates
 - N.B. new mailing list biopax-discuss@googlegroups.com



2010 Milestone

- Level 3v1.0 Released July
- BioPAX Paper published!

The screenshot shows the Nature Biotechnology website interface. At the top left is the logo 'nature biotechnology'. On the right, it says 'Welcome back: Nadia Anwar' with a 'Logout' button. Below the logo is a search bar with a 'Go' button and a link to 'Advanced search'. A breadcrumb trail reads 'nature.com > journal home > current issue > research > perspective > full text'. Navigation links for 'previous article' and 'next article' are also present. The main article title is 'The BioPAX community standard for pathway data sharing'. The authors listed are Emek Demir, Michael P Cary, Suzanne Paley, Ken Fukuda, Christian Lemer, Imre Vastrik, Guanming Wu, Peter D'Eustachio, Carl Schaefer, Joanne Luciano, Frank Schacherer, Irma Martinez-Flores, Zhenjun Hu, Veronica Jimenez-Jacinto, Geeta Joshi-Tope, Kumaran Kandasamy, Alejandra C Lopez-Fuentes, Huaiyu Mi, Elgar Pichler, Igor Rodchenkov, Andrea Splendiani, Sasha Tkachev, Jeremy Zucker, Gopal Gopinath, Harsha Rajasimha, Ranjani Ramakrishnan, Imran Shah, Mustafa Syed, Nadia Anwar, Özgün Babur, Michael Blinov, Erik Brauner, Dan Corwin, Sylva Donaldson, Frank Gibbons, Robert Goldberg, Peter Hornbeck, Augustin Luna, Peter Murray-Rust, Eric Neumann, Oliver Reubenacker, Matthias Samwald. On the right side, there are links for 'Journal home', 'Current issue', 'For authors', 'Subscribe', 'E-alert sign up', and 'RSS feed'. At the bottom right, there is a 'nature' logo with the text 'Enjoy the world of science with a 30% discount to Nature' and a 'Science jobs from naturejobs' link.

nature
biotechnology

Welcome back: Nadia Anwar
Logout

Search Go Advanced search

nature.com > journal home > current issue > research > perspective > full text

< previous article next article >

NATURE BIOTECHNOLOGY | RESEARCH | PERSPECTIVE

The BioPAX community standard for pathway data sharing

Emek Demir, Michael P Cary, Suzanne Paley, Ken Fukuda, Christian Lemer, Imre Vastrik, Guanming Wu, Peter D'Eustachio, Carl Schaefer, Joanne Luciano, Frank Schacherer, Irma Martinez-Flores, Zhenjun Hu, Veronica Jimenez-Jacinto, Geeta Joshi-Tope, Kumaran Kandasamy, Alejandra C Lopez-Fuentes, Huaiyu Mi, Elgar Pichler, Igor Rodchenkov, Andrea Splendiani, Sasha Tkachev, Jeremy Zucker, Gopal Gopinath, Harsha Rajasimha, Ranjani Ramakrishnan, Imran Shah, Mustafa Syed, Nadia Anwar, Özgün Babur, Michael Blinov, Erik Brauner, Dan Corwin, Sylva Donaldson, Frank Gibbons, Robert Goldberg, Peter Hornbeck, Augustin Luna, Peter Murray-Rust, Eric Neumann, Oliver Reubenacker, Matthias Samwald,

Journal home
Current issue
For authors

Subscribe
E-alert sign up
RSS feed

nature
biotechnology

Enjoy the world of science with
a **30% discount** to Nature

Science jobs from naturejobs

2010 Milestone

- BioPAX Validator v2.0a

BioPAX Validator (v2.0a-20101006)

[home](#)

[webservice](#)

[settings](#)

[report issue](#)

Actions:

[Upload and Check Files](#)
[Check URL Resource](#)

Links

- [Validator's Wiki Page](#)
- [Rules's Wiki Page](#)
- [Docs \(auto-generated site\)](#)
- [Pathway Commons](#)
- [MIRIAM](#)
- [OBO](#)

Welcome!

BioPAX has become an important standard for communicating the knowledge about biochemical processes. But errors that arise from data transformation, OWL "Open World" semantics, and the extensive use of external references can be a real obstacle and a pain. To address this problem, the Validator has both syntactic and semantic rules together with the cross-cutting error reporting framework, and it also makes use of such magic components as: Paxtools (BioPAX API), Ontology Lookup Service (helps with controlled vocabularies), and MIRIAM database (to check external references).

So, the BioPAX rules were derived both from the OWL specification and the community best practices. They are generic Java classes based on the Paxtools in-memory BioPAX model, and more rules can be created and tuned into the application later. They can check across several BioPAX entities and can be nested or overlap in their subjects, which might take more care to implement. There are both "fail-fast" and "post-model" validation modes. However, in most cases (e.g., when one checks an OWL file), the former is not required, so the fail-fast mode will come to the scene in the future software that will allow interactive model assembling and merging and use the BioPAX Validator API.

Please feel free to post your comments, suggestions, and issues at the [Validator Wiki](#) and [BioPAX issue tracker](#).

2010 Milestone

- PaxTools
 - Converters L2toL3, PSI-MItoLevel3
 - Pathway alignment
 - DML functions
- Data Sources (see poster outside)

Data Source	Format	Size	Focus	Reference
BioGrid	PSI-MI 2.5	177804 Interactions	Model organisms	Breitkreutz et al. The BioGRID Interaction Database: 2008 update. <i>Nucleic Acids Res</i> , 36 , D637-640
Cancer Cell Map	BioPAX L2 10 Pathways	2104 Interactions	Human	http://cancer.cellmap.org
HPRD	PSI-MI 2.5	40618	Human	Keshava Prasad et al. Human Protein Reference Database :2009 update. <i>Nucleic Acids Res</i> , 37 , D767-772
HumanCyc	BioPAX L2	266 Pathways, 4879 Interactions	Human	Romero, et al. Computational prediction of human metabolic pathways from the complete human genome. <i>Genome Biol</i> , 6 , R2
IMID	BioPAX L2	1729 Interactions	Human	http://www.sbcny.org/
IntAct	PSI-MI 2.5	154567 Interactions	All	Aranda et al. The IntAct molecular interaction database in 2010. <i>Nucleic Acids Res</i> , 38 , D525-531

BioPax Overview

- Molecular Interactions
- Metabolic Networks
- Signaling Pathways
- Gene Regulation
- Genetic Interactions

What can be described in BioPAX

- entities

- Name(s), Description, Sequence, Chemical Structure
- Type
- Molecular complexes and binding
- Modifications to the sequence
- Subcellular location

What can be described in BioPAX

- interactions

- Biochemical Reactions, Complex (Dis)Assembly, Transport, Template Reactions
- Direction, stoichiometry
- Type and direction of the catalytic activity,
- Direct and indirect control
- Degradation
- Genetic and molecular interactions

Pathways

- Set (or optionally list) of interactions
- Controllers and controlled
- Blackbox pathway

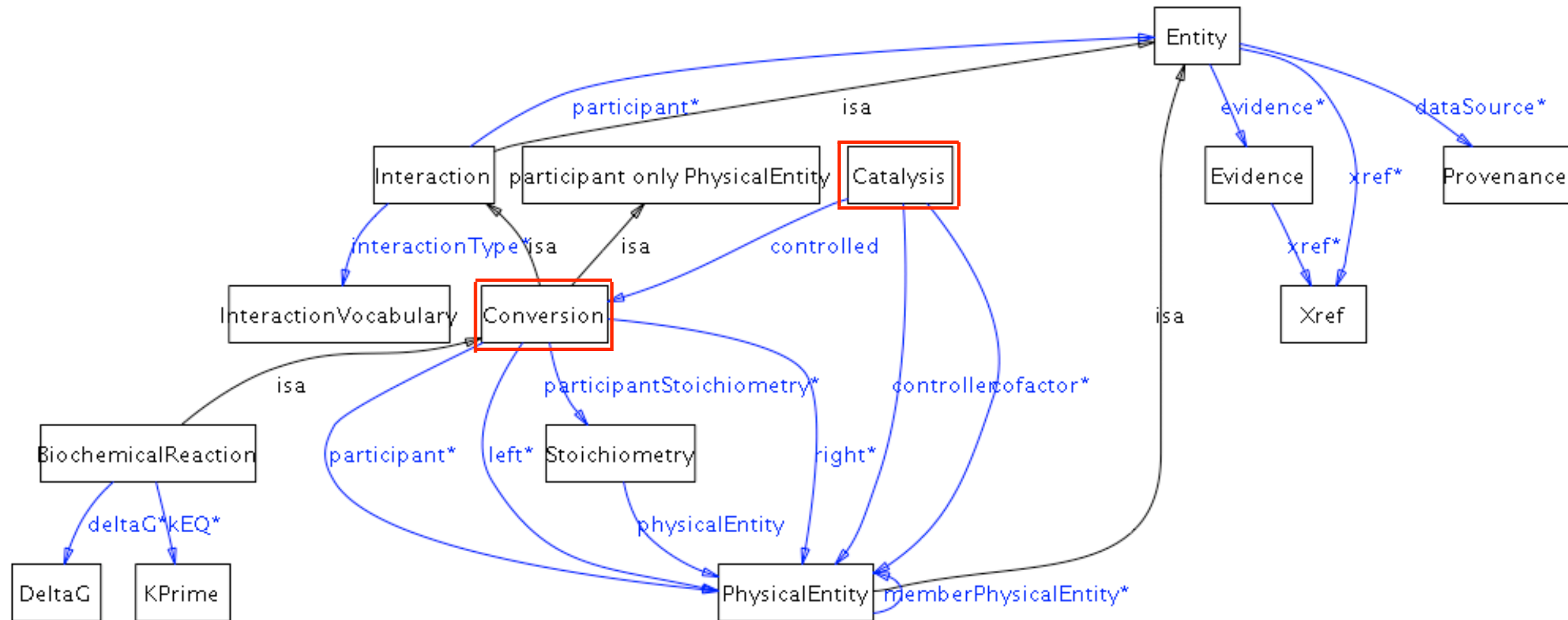
Annotation/Utility Classes/Meta Data

- Evidence
- Providence
- Organism
- Cell type, tissue
- External Links
- CVs

BioPAX - RDF/XML

- XML - syntax
- RDF/XML - syntax and semantics
- BioPAX uses OWL as a schema or validation scheme for RDF

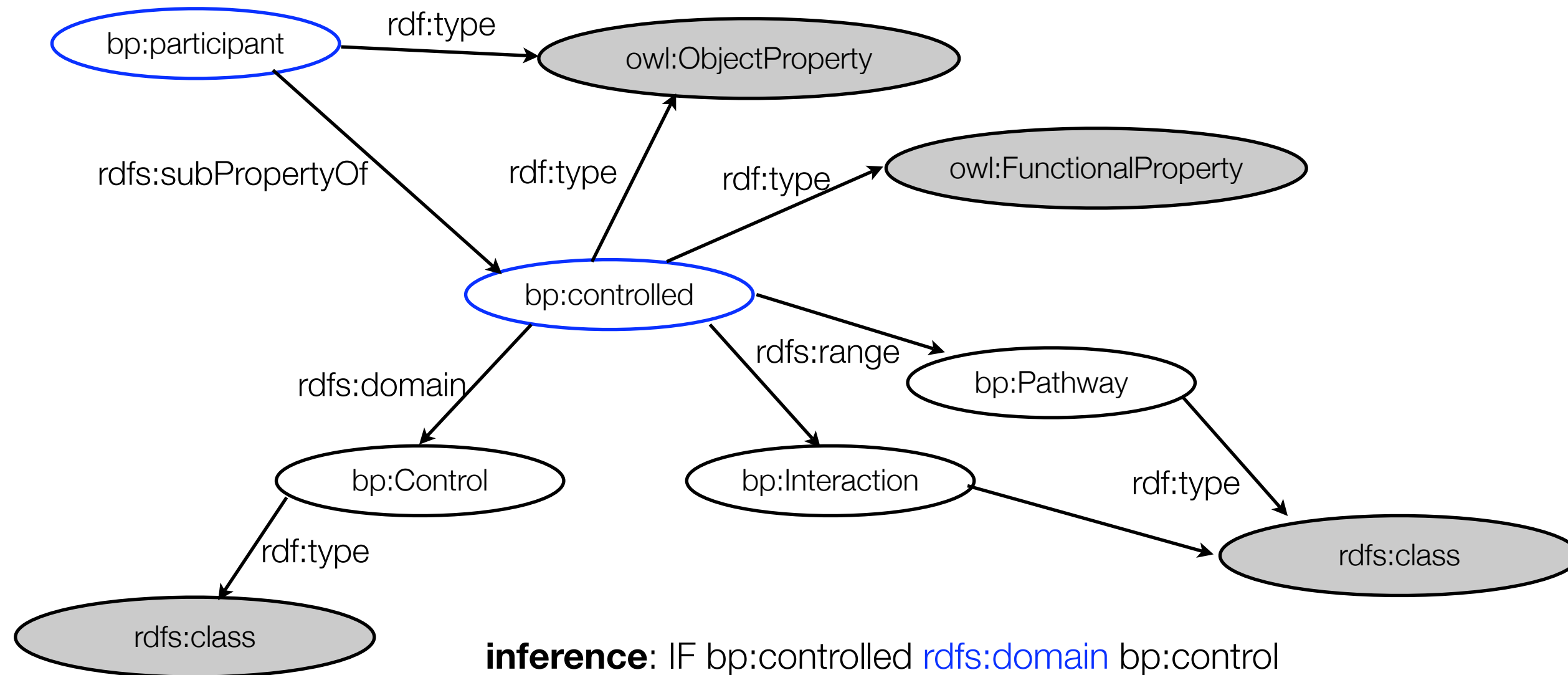
BioPAX - Relationship Propagation



domain $\xrightarrow{\text{object property}}$ range

catalysis $\xrightarrow{\text{controlled}}$ conversion

Semantics - Relationship Propagation

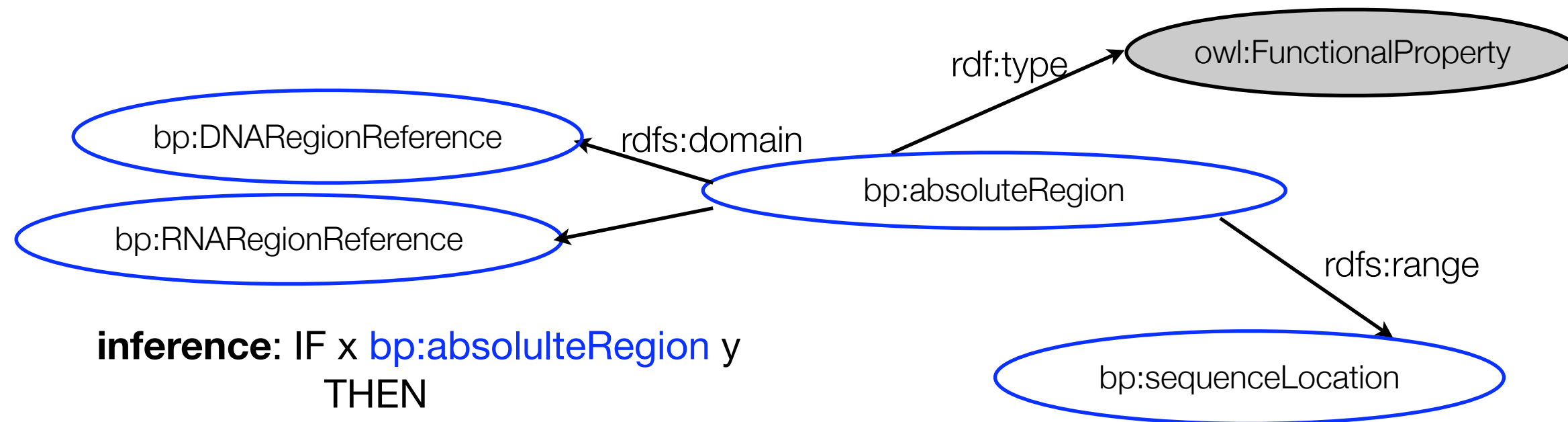


inference: IF bp:controlled **rdfs:domain** bp:control
and x **bp:controlled** y
THEN
x **rdfs:type** bp:control

NB: inference not inheritance
inferences are conjunctive - everything applies

RDFS - The semantics of domain and range

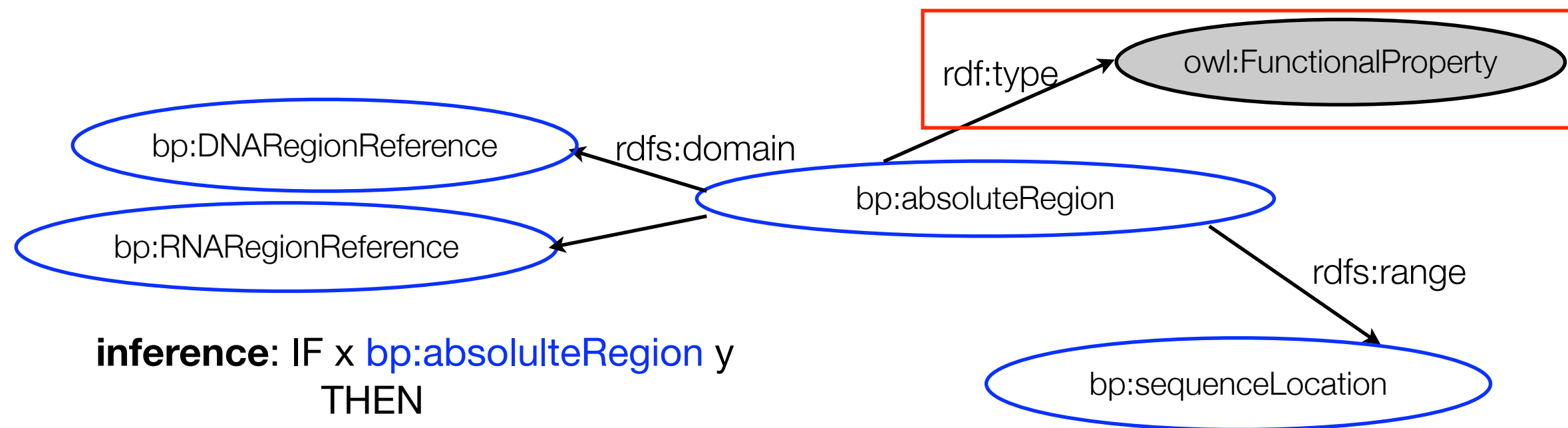
- N.B. inference (not restrictions)
- domain and range functions return sets which are to be interpreted as intersections
- Conjunctive:
- Properties can have any number of domains and ranges - **They all apply**



inference: IF x `bp:absoluteRegion` y
THEN
 x `rdf:type` `bp:DNARegionReference`
AND
 x `rdf:type` `bp:RNARegionReference`

RDFS - The semantics of domain and range

- N.B. inference (not restrictions)
- domain and range functions return sets which are to be interpreted as intersections
- Conjunctive:
- Properties can have any number of domains and ranges - **They all apply**



inference: IF x `bp:absoluteRegion` y
THEN
 x `rdf:type` `bp:DNARegionReference`
AND
 x `rdf:type` `bp:RNARegionReference`

BioPAX - ontology?

- BioPAX level 3 does not conform to OWL semantics, BioPAX assumes (sometimes, not all the time) a closed world

BioPAX assumes that there are no user defined classes

BioPAX uses domain and range as restrictions (in db speak - allowed values)

BioPAX objectProperties are usage (sort of treated like foreign key constraints)

BioPAX assumes that an interaction is complete (mostly, but not always) - but never explicitly stated.

BioPAX Roadmap - Level 4 Proposals - Gary Bader

Moving forward

- Roadmap

Level	Scope of Ontology				Data Source Compatibility
	Physical Entities	Interactions	Pathways	Metadata / Utility Classes	
Level 1	Small molecules Proteins RNA Complexes	Biochemical Reactions Enzyme Catalyses Transport Catalyses Assembly of Complexes	Metabolic pathways	X-refs Participants	KEGG, BioCyc, WIT/PUMA2, aMAZE
	Biological Rationale: Capture knowledge about simple metabolic pathways.				
Level 2	DNA	Binding Interactions	Molecular interaction networks	Evidence Confidence	BIND, IntAct, HPRD, MINT, DIP, PSI format
	Biological Rationale: Add support for molecular binding interactions.				
Level 3	Genes	Molecular states Gene regulation	Signal transduction networks	External controlled Vocabularies States	Transpath, PATIKA, CSNDB, Reactome, INOH
	Biological Rationale: Add support for signaling pathways and regulation of gene expression.				
Level 4	Generic physical entities	Genetic interactions Generic interactions	Genetic networks Generic pathways		FlyBase MIPS
	Biological Rationale: Add support for genetic interactions, generic entities and processes.				
Future Levels	Environmental effects Cells Cell compartments Photons	Abstract associations (e.g. co-occurrence in: pathways, literature abstracts, cell compartments, etc.)	Networks of abstract relationships	Experimental descriptions Provenance	PubGene GeneWays
	Biological Rationale: Capture abstract relationships between biological entities, cell-level interactions.				

- Survey - Community Priorities

Moving forward

BioPAX Level 4 Workgroups

- **Visualization and exchange** Status: Started - Point of Contact [Martijn Van Iersel](#)
The option of an extension to BioPAX to deliver layout (x,y coordinates only). A complete visualization extension for BioPAX would be a duplication of work and eventually SBGN will provide visualization exchange. In the interim, a workgroup could formulate a simple format for pathway visualization. This group would coordinate with SBGN.
- **Semantics: Generics/polymers/logic/** Status: Point of Contact [Emek Demir](#)
Works toward extending BioPAX e.g. to enable the capture and exchange of generic reactions and generic pathways and extensions to the entityreference class.
- **Semantic web/linking/CVs** Status: Started - Point of Contact [Andrea Splendiani](#)
Semantic web community wish list including Linked Data Project, Linking to other ontologies and ontology mapping, additional controlled vocabularies and architecture of controlled vocabularies are all areas where BioPAX needed improvement.
- **Validation/best practices** Status: Initiated - Point of Contact [Igor Rodchenkov](#)
Develop rules and best practices for data and documenting and deliver these into the BioPAX validator.
- **Quantitative modeling vs. static relationship/SBML/CellML/VirtualCell** Status: Started - Point of Contact [Oliver Ruebenacker](#)
With respect to reaction rates and rate laws, some BioPAX users would like to have a stronger link between BioPAX and SBML. In order to provide a technical link between BioPAX and SBML, this group will create a proposal and help to define a better bridge between the two languages and ultimately delegate to SBML.

L4 Proposals

Structural/New Classes and New Properties

- 1.1 Entity BioSource: propose subclasses organism, cell, tissue (rational tighter link to CVs)
- 1.2 Should Evidence have the object property BioSource?
- 1.3 Entity EntityFeature and EntityReference: propose to make these children of OWL:Thing (rational is that these, strictly speaking, are not meta data)
- 1.4 Entity Pathway: Create a Property "phenotype" (rational "black box" pathways)
- 1.5 Revisit the overall design of level of granularity for biological processes. We need some way of relating pathway to interaction (both are processes) and deals with input-output of processes, black box pathways, and subprocesses (part-of for processes).
- 1.6 Do we want to subclass pathway? – metabolism, signaling, MI, regulatory (regulation of one operon) – request from Peter Karp (long ago) and also RegulonDB
- 1.7 Extend physicalEntity to other molecules. You could conceive of an entire subclassing structure for all molecules.
- 1.8 Do we need an entity reference for complexes to tie together complex states (e.g. complexA and complexA-phosphorylated)?
- 1.9 Modulation Class, Controlled property can only be a Catalysis. Ashok Reddy, molecule pages example of Transport/Modulation
- Add cellularLocation to EntityReference class. This would be a list of all possible locations, in the same way that entityFeatures are listed in EntityReferences.

L4 Proposals

Controlled vocabularies/Semantics

- 2.1 ControlType Controlled Vocabulary
- 2.2 Update Small Molecule EntityReference with an object Property to SmallMoleculeClass (single value from SmallMoleculeVocabulary) And add subclass SmallMoleculeVocabulary(Chebi) to ControlledVocabulary.
- 2.3 Use of dublin core to store attribution? (suggestion by Augustin, MIM) Do we support all of dublin core, or only a subset? PaxTools and validator will have to be updated. Best practices for use?
- 2.4 Use of miriam CV for relationshipXref (suggestion by Augustin, MIM) <http://www.ebi.ac.uk/miriam/main/mdb?section=qualifiers>
- 2.5 Entity ExperimentalFormVocabulary: add property "Participant Identification Method"
- 2.6 Entity SequenceRegionVocabulary: points to disjoint classes Protein Domains, Promoters, UTRs
-

L4 Proposals

Others

- Negative Observations (Feature/NotFeature)
- Complex IA and IB have 3 different forms which are phosphorylated in 3 different places, but you don't know where. You need to create 3 different As, but they would look the same. There is no way to specify 2 unknown PTMs at different sites. (from Peter Karp)
- Review NIST PSL process specification language to see if we can use aspects of it.
- Model composition and layering, how do you refer to one BioPAX model from another model (for example a pathway not containing its participants that is described elsewhere).
-

An overview of changes from Level 2 to Level 3

Emek Demir

Proposals - Emek Demir, Marijn van Iersel

- Discussion outline
 - BioPAX core/BioPAX addons
 - Generics
 - Polymerisation