

# BioPAX Update of Activities

Report made to I3C/Pathways

Authors: BioPAX Group

Web: [www.biopax.org](http://www.biopax.org)

Correspondence: [biopax-discuss@biopax.org](mailto:biopax-discuss@biopax.org)

Presentation made by Joanne Luciano

Boston, Massachusetts

USA

7 May 2003

## Talk Script

### PART 1 – Introduction to BioPAX its Background, mission and philosophy

- Introduction
  - BioPAX name means BioPATHways data eXchange format
  - Brief history – interest in open source pathway database meeting in response to invitation by keynote speaker Chris Sander, ISMB 2001 Copenhagen, Denmark
  - Chris invited interest again at BPC meeting at ISMB 2002, Edmonton, Canada.
  - Exchange Format recognized as 1<sup>st</sup> step at ISMB 2002, Edmonton, Canada
  - Agreement from major DB sources EcoCYC and BIND to collaborate
  - Core group established, and 1<sup>st</sup> meeting November 2002, NYC
  - Publicly announced at Pacific Symposium on Biocomputing, Hawaii, USA
- Why BioPAX?
  - Growing need for unified framework for pathway representation
  - Scientists need data from multiple **pathway** sources – need to integrate these data
  - [more stuff like this (use existing slides)]
- Who?
  - DB sources (critical to get buy in from pathway data base sources)
  - Pathway experts
  - Tool developers
- How?
  - Logistics
    - Small & efficient
    - Very carefully pick core group members
      - Work cooperatively towards the common goal
      - Have complementary skill sets
    - Expand through formation of subgroups
    - Actively reach out to other DB and Community members for collaboration where areas overlap (and equally be receptive to interest from other groups)
    - Apply to DOE for funding, Initial Funding by each core group members and by Chris Sander.
  - Technical
    - Present templates/examples of how to represent from different dbs
- Where & When?

- Rotate location of face-to-face meetings to balance out travel expense and inconvenience
- Bi-weekly conference calls for decision making, next steps, strategy, raise suggestions, issues, etc.
- Face-to-face and subgroup/off line meetings work out technical details

## PART 2 – Examples

- Framework
  - For building record (as opposed to record)
    - *Record structure currently under development; not ready to announce it as such*
    - *No details present other than to describe the basic format of the record structure.*
  - **Red indicates small but extensible controlled vocabulary**
  - **Blue – need established, ideal representation not yet known** – areas for working groups to define further (example – how to handle timing, states) might want to add type “cell” to capture cell-to-cell interactions.
  - We think we can capture all representations (we could be wrong), tried not to paint ourselves into a corner. We invite your feedback (at end will tell you how to get in touch)
- Examples – Introduction – will present a number of examples:
  - Increasing complexity
  - Illustrate flexibility of the PAX record
  - Multiple representations of the same pathway are common, here we show how one might represent them as a PAX record
  - These examples will illustrate our thinking and that we can capture
    - Interactions and reactions
    - Multiple Sources of Data and therefore multiple representations
    - Arbitrary levels of detail
  - Will show examples of Signal Transduction and Metabolic Pathways of increasing complexity (simplistic and informal to show the kinds of things we’re thinking about)
  - **Example 1** – Glycolysis (1) (METABOLIC)
    - Representation – class notes found on web from 4<sup>th</sup> year course
      - Simple Relationship - Enzyme Substrate Product
      - Relationship are defined by Left and Right sides
      - Enzyme is part of both L & R, but **it** doesn’t change
      - Only substrate and product change
    - Simple, but doesn’t specify glucose →glucose-6 phosphate
  - **Example 2** – Glycolysis (2) (METABOLIC)
    - Same Pathway, represented with nested PAX records
    - Left and right parts of metabolic reaction
    - Catalysis
      - Represented as directed association
      - Similar to MAPK (3) – flat representation
    - Different people represent same thing in different ways, with different levels of detail – goal support all the different representations

- **Example 3** – Signaling MAP Kinase (1) (SIGNAL TRANSDUCTION)
  - Ref: Molecular Biology of the Cell
  - Same Pathway
  - Can represent the detail – capturing more detailed information in yet another way
    - Relationships among relationships
    - Different types of relationships
      - Biochemical reactions
      - Catalysis
        - left side and catalysis
        - Right side and catalysis
    - Need 2 relationships to define every step
  - Same Pathway can be represented in different ways
    - Signal transduction people understand state transition from inactive to active
    - (Point: We can describe it)
- **Example 4** – MAP Kinase (2) (SIGNAL TRANSDUCTION)
  - Ref: Molecular Biology of the Cell
  - Same Pathway, more detail, different representation
    - Relationship among relationships
      - Biochemical reactions
- **Example 5** – mRNA (1) (TRANSCRIPTIONAL REGULATION)
  - 2 mRNA transcription factors present in cell at same time (gene chip data)
- **Example 6** – Protein-Protein Interaction (MOLECULAR INTERACTIONS)
  - Pairs of proteins
  - Don't know much about them
  - Know there is some binary interaction

*This format as it stands is the PSI format – i.e. how PSI represents molecular interactions.*

- **Example 7** – Mass Spectroscopy and Y2H (PROTEIN COMPLEX)
  - Associations
  - From different experiments
  - Know 2 & 3 are interacting with each other
  - Same complex, but not just flat – it is a complex
  - Does not include the process of assembly – it does not comment about assembly, but rather just says that the complex exists (we don't know how it assembled)
  - Yeast Protein names – biology: polymerizes actin directs assembly of the actin cytoskeleton

This is an example of how general we can go. The types of pathway and pathway data we can encode

May be beneficial to text mining people (text mining slide)

### **PART 3 – Next Steps**

- Draft in XML Schema and OWL available “soon”
  - We will recommend specific ways of encoding specific data
  - Point – we can support the representations out there

**BioPAX – Take Home Messages**

- Made by people who are making the databases
- Main purpose is data exchange format between databases.
- SBML's main purpose is to share models between software simulation packages
- The two things are related and where they overlap we will try to work together to maximize compatibility