

**BioPAX Work Group**  
**November 12, 2003 Conference Call Minutes**

**Participants:** Gary Bader, Michael Cary, Peter Karp, Joanne Luciano, Suzanne Paley

**Overview:**

1. Ontology work/status
  - a. Several updates in to-do queue (see <http://cvs.sourceforge.net/cgi-bin/viewcvs.cgi/biopax/biopax/>, select “Level 1 To Do” group)
  - b. Next release: 0.6 by end of November
2. Project management calendar (<http://www.biopax.org/pm/calendar/November2003.html>)
3. External meetings
  - a. Report by Gary: NIST meeting (<http://www.nist.gov/director/NIH/workshop.htm>)
4. Subgroup status reports
  - a. State subgroup – Emek’s comments need to be integrated
  - b. Small molecule subgroup – PubChem project starting up
  - c. Examples subgroup – received community comments on data integration issue
5. Next conference call date: Nov. 26, 2003?
6. Next F2F meeting: Denver, 12/8/03
  - a. Location: Hotel near airport is first choice
  - b. Travel, hotel will be paid for with DOE grant
7. Ontology issue discussion
  - a. Cofactor slot
    - i. [http://sourceforge.net/tracker/index.php?func=detail&aid=828568&group\\_id=85345&atid=575904](http://sourceforge.net/tracker/index.php?func=detail&aid=828568&group_id=85345&atid=575904)
  - b. Confidence
    - i. [http://sourceforge.net/tracker/index.php?func=detail&aid=823093&group\\_id=85345&atid=575904](http://sourceforge.net/tracker/index.php?func=detail&aid=823093&group_id=85345&atid=575904)
  - c. LSID
    - i. [http://sourceforge.net/tracker/index.php?func=detail&aid=825160&group\\_id=85345&atid=575904](http://sourceforge.net/tracker/index.php?func=detail&aid=825160&group_id=85345&atid=575904)
    - ii. Presentation: [http://industry.ebi.ac.uk/~senger/talks/Life\\_Sciences\\_Identifiers.ppt](http://industry.ebi.ac.uk/~senger/talks/Life_Sciences_Identifiers.ppt)
    - iii. IBM LaunchPad: <http://www-124.ibm.com/developerworks/opensource/lsid/>

**Notes:**

**Ontology Development**

There has been some progress on a few conceptual issues via discussions on the biopax-discuss list, but many of the recently resolved issues still need to be added to the ontologies (GKB and Protégé versions). One major task that we still need to address, and

plan to do so before the next meeting, is allowing entities to acquire additional attributes. This is how we will implement states in BioPAX.

Mike suggested that we publish a new release before the next meeting that shows all of the changes we have agreed to since our last meeting. We agreed to do this – we will call it version 0.6 and it will be a silent release (i.e. we will not make a big fuss about it like we did with version 0.5).

### **Calendar**

Mike mentioned that he put up a calendar on the project management page to help us keep track of upcoming BioPAX events (<http://www.biopax.org/pm/calendar/November2003.html>).

### **Recent Meetings**

Gary gave a talk about BioPAX at a recent NIST meeting (<http://www.nist.gov/director/NIH/agenda.htm>). There was good, general interest in BioPAX. Gary said that it was the only proposed standard presented at the conference – the other presentations were on tools (like Protégé) or on implementations of existing standards. After his talk, someone suggested to Gary that we should look into the Process Specification Language (PSL) (<http://ats.nist.gov/psl/>). Gary sent this link to the pax list, but no one has had a chance to look into it yet.

Joanne mentioned that Eric Neumann also attended the conference, and that he suggested that she pursue further funding for BioPAX from Eric Jakobsson (Director, Center for Bioinformatics and Computational Biology, BISCTIC, NIH [jakobse@nigms.nih.gov](mailto:jakobse@nigms.nih.gov)) and Karen Skinner (Deputy Director, Science and Technology Development, Division of Neuroscience and Behavioral Research, NIDA, NIH, [kskinner@nida.nih.gov](mailto:kskinner@nida.nih.gov)). Joanne will email them this week.

### **Subgroups**

The state subgroup sent out its final draft to interested parties. Emek Demir made comments which still need to be integrated into the document. Peter asked Gary to send him the latest draft.

Joanne got some feedback from the examples list about an integration issue that she posted. She is working to integrate those responses into a proposal. She is also creating an example that uses LSIDs to integrate identical elements in separate BioPAX documents.

Gary mentioned that Steve Bryant (NCBI) is working to create a small molecule database called PubChem. The project has not been started yet but is on NCBI's roadmap.

### **Next F2F Meeting**

So far, it appears that the following people will attend the Denver meeting: Gary Bader, Eric Brauner, Mike Cary, Joanne Luciano, Peter Karp, Andrey Rzhetsky, and Chris Sander. Joanne asked Mike to keep Eric Brauner apprised of the details, as he is not a member of the pax list.

We voted to conduct the meeting at a hotel near the airport, as this will save on cab fare and travel time. Mike will try to find one with Internet access.

### **Integration Issue**

Joanne mentioned that in a recent discussion with Eric Brauner he mentioned his concern that various databases might have their data in BioPAX format but still not be able to easily integrate their data. Gary called this issue “semantic mapping”. The best solution currently appears to recommend “best practices” to our users, i.e. suggestions on how various types of data should be represented. If two users represent the same concepts identically, then data integration becomes much easier.

Mike said that data integration is a software issue. It involves identifying identical elements, incorporating any additional annotation they may provide, then deleting them and remapping any dependent pointers. Joanne said that she agreed, but that the issue was whether the ontology needed to be altered to better facilitate this.

### **Cofactor Slot**

The issue is whether or not we need a specific slot(s) for cofactors and coenzymes in the enzyme catalysis class. Without such a slot, cofactors could be specified by either a) including the cofactors along with the enzyme in the CONTROLLER slot, or b) first creating a complex between the enzyme and the cofactor, then using this complex as the CONTROLLER. If we decide to include a slot for cofactors (inorganic), should we also include a slot for coenzymes (organic)?

Since both EcoCyc and WIT have a slot (or field) for cofactors, we decided to do the same. We will not, however, distinguish between cofactors and coenzymes (as WIT does); both shall go in the COFACTOR slot. In cases where an enzyme may use a number of different cofactors to catalyze the same reaction, we decided that we will ask users to represent this using separate enzyme catalyses for each cofactor, rather than bundling the cofactors in a group in the COFACTOR slot. This solution is the same as the solution of the case of multiple enzymes catalyzing a single reaction.

### **Confidence**

During the last face-to-face meeting, Peter asked for clarification on the definition of the confidence slot. The slot was created to maintain compatibility with PSI, which uses it to specify the statistical confidence that a particular interaction occurs. Many high-

throughput proteomics methods produce a good deal of false-positive interactions; being able to specify confidence is a way to deal with this.

Gary mentioned the possibility that we could use an RDF-based approach, called Annotea (<http://www.w3.org/2001/Annotea/>), to provide supplemental annotation, such as confidence, to BioPAX instances. Annotea was described to Gary by Marja-Riitta Koivunen at the NIH BISTIC meeting. Using XPointer, it can point to any XML document on the web and provide annotation to it. Mike said that it sounded like using Annotea might be a valid option, but that perhaps we would want to use it for more than just confidence. He said we might want to think about restricting BioPAX to describing purely mechanistic biology, and “sub-contract” all metadata (like confidence) out to Annotea. Gary said this approach was problematic because it would be difficult to get people to agree on what is metadata and what is not.

Peter said he thought we should try to improve our efficiency during these ontology discussions. For example, we should have provided a more concrete definition of confidence first, and then discussed it if there were any issues surrounding it. This was agreed to; we will try to focus future ontology discussions better.

We decided that we need to consider which classes in the ontology the CONFIDENCE slot is applicable to, and what its precise definition is for each such class. We can then revisit the issue if anyone has any questions or comments to make about it.

## **LSID**

The issue is whether we need to modify anything in the ontology in order to deal with LSIDs. LSIDs are a way to point to specific entries in any biological database. It looks like we can use the XREF slot, without modification, to handle LSIDs. The XREF slot is currently in the Entity class, and therefore (by inheritance) in all other (non-utility) classes in the ontology.

## **Preparing for December 8, 2003 meeting**

Joanne said that we should try to have examples for each issue we wish to discuss. She also advocated stronger user involvement in picking which issues we should resolve first, and suggested that the community could supply or develop examples for the issues they felt most strongly about. We agreed that the best way to do this would be for Mike to re-prioritize the open issue list in SourceForge, then ask the users (via biopax-discuss and biopax-eg) for their comments on the issue priorities and to contribute examples that illustrate the issues that they feel are important. At the December meeting, we would give priority to those issues that have clear examples, as the examples would likely lead to faster issue resolution.