BIOPAX Work Group
5/14/03 Conference Call Minutes

Participants: Gary Bader, Michael Cary, Joanne Luciano, Eric Neumann, Suzanne Paley.

Purpose of call: Review and plan ontology work, discuss subgroup progress and upcoming meetings.

Overview:

1. Ontology work
   a. Root ontology looks good
   b. Interactions ontology
      i. People want to see where leaf classes would go
      ii. Should be able to represent indirect causality
   c. Other issues
      i. We will create a mirror ontology in Protégé
2. Subgroup status reports
   a. State subgroup – held conference call, working on summary document
3. Report of talk at I3C meeting Wed.May.7.2003
   a. Made clear the distinction between BioPAX and SBML
   b. Worked on use case that included both, to illustrate differences
   c. LSIDs
      i. Unique IDs for biological entities
      ii. Spec is done, gaining acceptance
      iii. We will learn more
4. Web-based project management announcement
   a. Site is www.biopax.org/pm
   b. Content must be added by Gary or Mike
5. Next meeting: Friday May 23 in NYC
   a. Tom Plasterer might attend (from Eric's group)
6. Next conference call: Wednesday June 7 (tentative)

Summary:

We began by talking about the root ontology structure. Overall the group liked the root structure, and no one raised any specific issues regarding it. There was a request for a picture file to supplement the documentation, similar to the one sent with the first interactions ontology.

Eric said that he thought the root ontology appeared to be in-line with what other groups were doing in this area. He especially liked the fact that we were thinking in terms of layers of abstraction.

We next discussed the interactions tree of the ontology. Mike sent out one possible

version, Gary plans to send out another version soon. There was a desire from the group to see the interactions ontology more fully developed (adding the leaves to the tree); people felt it was difficult to evaluate without the leaf classes.

Eric asked if we were making a distinction between a protein-protein interaction and the process of two proteins binding. Gary explained that a multi-protein complex would be represented as an undirected interaction between the proteins that comprised the complex. The process of assembly, however, would be a directed interaction going from a set of individual proteins to a multi-protein complex.

Eric said that he would like to see BioPAX be able to represent "proposed causality", i.e. events that appear to have some downstream consequence, but the intermediate steps or specific mechanisms are unknown.

Joanne pointed out that the ontology article that Suzanne sent to the group addressed this issue, and it was agreed that all of us should read that article.

In addition to the ability to represent indirect causal relationships, Eric said we should be able to distinguish between transient associations and stable complexes. If we use the same interaction class for protein-protein interactions and for two things that bind together covalently, we'll risk losing important information.

He also said that there is a trade-off between using abstract interaction classes and very detailed ones. With abstract classes, your tools are better able to treat the interaction network as graphs, and shortest-path algorithms (for example) can be applied to these graphs. With detailed classes, the interaction network is much less like a simple graph. While you are better able to represent human knowledge, it is more difficult to work with the data computationally.

Mike asked Eric if he would send the group a set of representational challenges. Eric agreed to send some examples from signal transduction databases.

We next talked about creating a version of the ontology in Protégé. There have been several requests for this, and Joanne volunteered to take on this task. The means for automatic exchange between the two (via OWL) should be available in July.

The only subgroup that had any news to report was the state subgroup. They conducted a conference call and made further progress on their documentation. They hope to have the document done by the next face-to-face meeting.

Joanne talked about her presentation at the I3C meeting. She said that people were not clear about the differences between BioPAX and SBML, and that following her talk and one by Michael Hucka they understood the difference better. Steve Burbeck suggested that a use case be developed that included both SBML and BioPAX, to illustrate the differences between the two.

Rainer Fuchs said that since many biological databases represent the same information in different ways, it would be useful to be able to see the differences and similarities of the different models. This should be easier with a common format. Joanne said that presumably this would be handled by application software but we need to be sure not to lose this information.

Joanne also mentioned that people at the I3C meeting were talking a lot about LSIDs (Life Science IDs). Eric explained that LSIDs were a means to give a unique identifier to an entity in a biological database. LSIDs are URNs (sets of alpha-numeric characters separated by colons) that point to a specific entry in a specific version of a specific database.

One problem with LSIDs is that they do not reveal the nature of the object being identified. To do this, metadata about the LSIDs must be provided by the biological databases. Eric also mentioned that there was some resistance to LSIDs by the W3C because they use URNs instead of URIs.

We agreed that we should learn more about the LSID project, as it may be something we should support in BioPAX, and Eric volunteered to send more information to the group.

Eric asked how we planned to use external ontologies in BioPAX. We said that this was something we planned to implement (e.g. GO), and Eric explained how it was very easy to do this in OWL and RDF.

Gary mentioned that the project management site was up and located at biopax.org/pm. At this time, all content must be manually added by Gary or Mike, so send one of them an email if you need something on one of the pm pages. Suzanne mentioned that she could not read the rtf files (upcoming agendas) from a Sun workstation; Mike agreed to make them available in another format.

We reviewed who planned to attend the next face-to-face meeting. Eric said that he could not make it, but may send Tom Plasterer in his stead. Tom has been working with Eric on developing an ontology in RDF.

We tentatively set the next conference call for the Wednesday two weeks after the May 23 meeting (June 4).