

BioPAX Work Group
October 22, 2003 Conference Call Minutes

Participants: Gary Bader, Michael Cary, Joanne Luciano, Suzanne Paley, Imran Shah

Overview:

1. Ontology work/status
 - a. 10/3/03 meeting issues logged into SF
 - b. Creating examples
 - i. Will be in SourceForge
 - ii. Glycolysis, MAPK, TGF-Beta, WNT
 - c. Examples are generating new issues
 - d. XML Schema
 - e. State document
2. Recent meetings
 - a. Conf call with Natalia
 - i. Main concern: resolving small molecule identities across different DBs
 1. Igor Goryanin (at GlaxoWellcome) may release S.M. dataset
 - ii. We will translate some WIT data into BioPAX
3. Next conference call date
 - a. November 5, 2003?
4. Next F2F meeting
 - a. Video conference?
 - i. Facilities, costs
 - b. Denver: December 8, 2003
5. Ontology issue(s) discussion
 - a. Molecular association solution: dual parentage?
 - i. http://sourceforge.net/tracker/index.php?func=detail&aid=781732&group_id=85345&atid=575904

Notes:

Suzanne launched a WebEx Now meeting, but Joanne could not view it via her Macintosh so we did not use it.

Ontology work/status, Subgroups

We are using SourceForge more heavily for issue and progress tracking – check the SourceForge bug-tracker for all current issues. We (primarily Joanne and Mike) are also creating more examples, using Protégé. We are uploading these into the Examples sub-directory on the SourceForge CVS. Creating these examples is generating new issues, which are making their way to SourceForge and to the BioPAX-discuss email list.

No news on the XML Schema version. The State subgroup document has been sent out and we have received some feedback. This feedback will likely result in us altering the document slightly, primarily to improve clarity.

Recent Meetings

Mike and Joanne called Natalia to update her on recent progress with BioPAX. The result of the call was a plan for Joanne and Mike to translate some WIT data into BioPAX, then show it to Natalia for her to comment on. One of Natalia's main concerns was resolving small molecule identities. She did not think CML sounded like the solution, she said we need a centralized database. We told her a little about ChemBank. She said that Igor Goriannon (sp?) from GlaxoWelcom plans to release their database of small molecules to the public (purportedly very large). Mike suggested Eric Brauner, Imran, and Igor should talk.

Upcoming Calls, Meetings

We set the next conference call for November 5 (tentative). Our next face-to-face meeting is still set for December 8 in Denver. Mike looked into video conferencing. MSKCC has a video conferencing facility, but in order to set up the call (and determine cost) we need to find out more about the facilities at each external site (Boston, U.C., SRI). Suzanne did not know how SRI conducted video conferencing.

Ontology Issue Discussion

Gary explained the issue with MolecularAssociation – some people use it as a complex, some people use it to represent protein-protein interactions. Gary said the proposal currently on the table was to use dual parentage (both PhysicalEntity and Interaction).

Suzanne asked what the advantage was of having it as an interaction. Mike said that conceptually, a complex is a relationship between instances of other classes in the ontology. The other physical entities also have parts, but these are not defined in our ontology. Suzanne said she did not have a major problem with using dual parentage.

Imran asked about the community's feedback on the issue. Gary said we have not had much of a response. Imran said his gut feeling was that it was dangerous to use dual parentage.

Suzanne asked if there were attributes on MolecularAssociation that mandated it to be an interaction. Gary said that all other physical entities were treated as atomic parts, MolecularAssociations can't be treated like this. Mike said that if it is only a physical entity, we break the idea of a separation between relationships and atomic parts.

Imran asked for examples. Gary said many protein-protein interactions (e.g. in DIP, BIND, etc.) that stably bind to each other were good examples (e.g. EFG-receptors that dimerize). Imran said the difference between complexes and interactions had to do with transience, and whether the two things that bound together then needed to be treated as a unit.

Gary spoke to Fabian Compagn from the SigPath database. They represent complexes as physical entities, which are created through a process called complex assembly. Gary said this didn't capture the transient interactions very well. Aviv (during a previous

conversation with Gary) said there was a duality (transient vs. strong) in molecular associations so dual parentage was OK.

Imran asked if perhaps transience should be dealt with via states. In other words, specify transience via an attribute (like half-life) of the interaction. He also said that dual parentage could create inference problems.

Gary said that a lot of data being called interactions (high throughput proteomics data, protein-DNA interactions, etc.) does not look like it should be captured as instances of physical entities. He also said that half-life is hardly ever known, and transience is not often known either. He suggested we hit the problem with more examples.

Suzanne asked if we could just keep calling it MolecularAssociation, but have it as a subclass of physicalEntity. Gary said that would still be calling protein-protein interactions physical entities. Perhaps we should allow all physical entities to have parts? This would require re-defining interactions a little. We could also just store PPI data in the complex assembly class, but this doesn't seem right either.

Imran agreed with the idea of allowing physical entities to have parts. Suzanne said that transience wasn't quite as important as the fact that the thing participates as a unit in some interaction. Even a very transient association could be called a complex if it does something.

Another possible solution would be to have both complex and molecular association. This creates a problem of allowing two possible classes for the same thing.

Suzanne asked about operons – they have parts (genes) but they are physical entities. Gary said that all of the physical entities really do have components, but up until now we have only thought about atomic physical entities. Perhaps we could distinguish between the compositional type of relationship of physical entities and the other set relationships.

Imran and Gary suggested that we write up a document on this issue, similar to the state document. This idea was agreed to.

We didn't have time to discuss the other issues (confidence, LSID) on the agenda.