

BIOPAX Work Group
8/20/03 Conference Call Minutes

Participants: Gary Bader, Michael Cary, Joanne Luciano, Peter Karp, Suzanne Paley, Aviv Regev, Imran Shah. (Minutes by Joanne Luciano)

Overview:

1. Ontology work
 - a. Progress on ontology ToDo list:
 - i. Add examples
 - ii. Protégé version → OWL proof of concept
 - iii. XML Schema version
 - b. Release of version 0.5
 - i. In two weeks? (Sept. 3 after next call)
 1. So that we could get feedback in time for Sept. 19 F2F meeting
 2. Pre-release to pax list by Sept. 1
 - ii. What will the release consist of?
 1. Ontology (GKB, Protégé) only or DEF (OWL, XML Schema) also?
 2. Documentation
 3. Publicity
2. Subgroup status reports – any news?
 - a. State subgroup
 - b. Small molecule subgroup
 - i. ChemBank status of metabolites?
 - c. Examples subgroup
 - d. XML Schema subgroup
3. Upcoming external meetings
 - a. OMG in Boston Sept. 9 – Sept. 11
 - i. Send an observer?
4. Next conference call date? Sept. 3?
5. Next F2F meeting: Sept. 19 in Denver?
6. Ontology issues of the day:
 - a. Should there be a class called *Polymer*?
 - b. Should we rename the *Control* class?
 - c. If there's time: Should *Molecular Association* be a *Physical Entity*?
 - i. Currently a subclass of *Interaction*

Summary:

The first item discussed was the ontology work. Glycolysis and MAPK would be added to the released version as instances illustrating the classes and how they would be represented. Joanne had started this work for Protégé, Mike will add to it; Gary has the XML-schema version as his next BioPAX to do. Targeted release date for Version 0.5 for review is scheduled for to precede the next face-to-face meeting with time for feedback to be received so that it could be reviewed.

The group agreed that it would document what it meant by version and release.

The discussion about OMG centered on the concern about getting bogged down by their process, who uses OMG standards and what our interaction and involvement should be. Ultimately, it was agreed that Joanne would attend the OMG meeting, which is in Boston, and that while we want to keep communication open, exchange information, and be happy to be adopted as a standard; we needed to remain independent and protect our process.

The exact logistics for the next face-to-face meeting has not yet been decided. There was some discussion, but things had to be checked and it was taken off line. The tentative preference was for October 3 in Denver. The plan is to review the ontology and resolve all outstanding issues, i.e. those that were open when the ontology was released and any new issues that are raised after the release of 0.5.

Ontology issues:

Polymer Class: It was decided not to have a class called polymer. It will be reconsidered for Level 2 because it may be useful for maintaining flexibility when describing certain interactions, such as gene-gene or gene-protein.

Control Class: The concern was that control might not be the best name (not general enough) for the class that is the super class of Enzyme catalysis, gene regulation, and transport facilitation. It was decided not to change the name of the control class, as there was not a clear consensus about what to change the name to.

There was discussion about who would be available to review version 0.5. It will be announced on BioPAX announce, Discuss, and the examples mailing list.

Meeting Details:

The meeting began with Mike asking if there was anything to add to the agenda. There wasn't. He started off with the ontology work. He reminded us that we came up with a short to-do list, of three items we wanted to have done which would allow us to release version 0.5, the trial release for feedback from the community.

Peter had suggested to Mike that to add examples to the ontology as instances of classes. Joanne had already begun this when she was in Manchester at the Protégé workshop.

Gary hadn't done this in the GKB version yet. Joanne mentioned problems with the two versions of Protégé. Protégé 2000, and the new version of Protégé 2, which contains the OWL constructs were not compatible. She started to implement BioPAX in Progege2000 and then learned that about Protégé 2 had OWL support (alpha release). She lost data (the entity class structure) and was not able to convert the ontology from Protégé2000 to Protégé2, so she reentered it into Protégé 2.

We agreed to stick with the simple examples pathways Glycolysis and MAPK. Joanne stated they are included in the example subgroup list and all had agreed that these pathways were a good place to start.

Gary asked if we had a to-do list on our project management web page yet. Mike said that we didn't. Adding examples needs to be added to the to-do list.

Next item: Creating a protégé version of the ontology. Prior to the conference call, Joanne sent the Protégé 2 files to the group. She explained that the file included the complete BioPAX class definition as of the end of June, i.e. it is what was presented at ISMB and any new changes since then, such as 'parts' being renamed to 'physical entities' is not included in this version. She had entered the slot definitions, but they were lost in the conversion and need to be re-entered. The utility classes were complete except for a few items— there were a few items in question, but after looking at it, one would get a good sense of the representation. Joanne wants to contact Alan Rector to obtain his suggested work-around for the enumerated type (symbol class) that has not yet been implemented. Joanne said it is on going.

Mike said he'd take a look at it and add to it based on what is discussed today and recent ontology development. Joanne cautioned that we would need to be careful about having multiple versions and versioning in general. Mike said he and Joanne should pass it between each other, when Gary suggested using CVS. They all agreed that it is a good idea.

Joanne then mentioned that the new version supported multiple users and Mike said he would look into that as well.

Peter mentioned that GKB editor also supports multiple users.

Joanne asked if GKB supports OWL yet and Peter said No, and asked if that is a rate limiting step? Gary explained that the current work in Protégé is to get some OWL proof of concept underway so that we could

wrap our heads around it. That we would probably make the Protégé version also available as well as GKB and support both versions if that were possible to satisfy people who have requested both.

XML-Schema: No progress, but it is next on Gary's BioPAX to do list.

Release of version 0.5: Mike reported that he Gary and Chris Sander discussed it and would like to release it in the next two weeks if it were possible. This way they can get some feed back to guide us during our next face-to-face meeting, which is tentatively set for Sept 19 in Denver (we still need to talk to Imran about that). So it can go out by Sept 3, which would have a goal to pre-release it to the PAX list by Monday and we could talk about it during the September 3rd conference call. And then if there were no objections we could then release it to the biopax-discuss list. Mike asked if that plan is agreeable to everyone; everyone acknowledged agreement.

Peter mentioned that we need a mechanism in place to have updated that people want to make to have happen in a coordinated way and added that it sounds like source forge can help us do that.

Mike went on to say that this release would consist of he, Chris, and Gary talked about yesterday. This would be included on the BioPAX web site. Mike said they would also like if it could also be included on the BioPathways Consortium web site and Joanne said sure. We would have documentation, not only the html generated version from GKB, but also more thorough documentation in paragraph form that we have an older version of that we need to update to the recent changes, then GKB files, Protégé files, and also the OWL output from that. The XML Schema, if we can get it done in time. Gary agreed that it should be included if it can get done in time and that's why it's next on his list.

[Imran joins the meeting and explained that he was writing up the CellML SBML BioPAX interface summary from ISMB. Gary suggested giving Imran a 2-minute update and Mike did that.]

When he got to the part about the next face to face meeting being in Denver, Imran appreciated the thought but had to inform us that he would be out of town and suggested that we move it earlier or later or somewhere else. He said he'd be closer to NY. Some discussion ensued and Gary explained that we did it to keep the rotation going. Discussion taken off-line, unless Chris joins at end of call when we might have enough to make the decision. That would give a little more time for version 0.5, but we still would like to get it out in 2 weeks, but 3 would still be enough time to get some feedback before our face-to-face. Play it by ear. Mike asked if there was anything else on 0.5? Gary asked if we finished discussing what the release would consist of.

OWL, XML. GKB Protégé well documented.

Joanne brought up a technical detail we need to be mindful of conventions, naming conventions including issues of case sensitivity, underscores versus dashes in names. Imran expressed a concern for dashes or other separators rather than upper and lower case letters in that one might lose meaning if case sensitivity were lost. Gary asked if it was an issue in Lisp and Imran said no. He hadn't upgraded to the current version that supports case sensitivity, but he could. He said Have to put things in vertical bars. Allegro Common Lisp – vertical bars is fine and that what was tentatively agreed upon. All agreed that consistency was the main issue. We could use dashes too.

Subgroups Status Reports

State Subgroup

Aviv – was swamped and was very apologetic. Gary said they were swamped too. Aviv said she'd do it by then of this week. Mike suggested she reply to the emails that were sent out a few weeks ago or maybe late July; follow the thread and reply.

Small Molecules Subgroup

Gary had a question: Have we contacted ChemBank; no, Are they going to put metabolites in ChemBank and if so, what is their time line?

Gary suggested we could help that along, recalling a previous conversation, that if we donated something from a database he (Eric Brauner) could use it, which would probably speed things up.

Imran commented that they have several bioactive compounds in the database already, and referred everyone to their demo page. Several people indicated that there wasn't much there. Gary said that if there was enough to do Glycolysis then we might want to use it for our example. Imran then asked Peter if he is interested in donating some molecules. Peter said yes, he is interested. Peter then reminded us that there are a couple of questions that have been raised that he has not heard answers to, like "How is he going to give credit to contributors?" and "How do updates of the data get coordinated?" For example, if he fixes a bug how do they get that update back? And, for that matter, who controls the data? Imran said he would coordinate a conference call with Eric and Peter acknowledged with interest. The rest was taken off line and Imran said he'd cc the BioPAX group so others interested could join in if they want to, otherwise it would be Imran, Peter and Eric.

Gary mentioned he just noticed something on the OBO web page, i.e. the GO extension. Gary explained they are trying to extend the GO type of ontologies to lots of different things and one of the things is biochemical substance, which Michael Ashburner is created. There's also a demo ontology, biochemical.ontology that is in the GO format. Gary said he didn't know where the names are from, but there are thousands of names each with an identifier and they are cross-reference to a KEGG ligands database. The sources indicated Ligand, Enzyme, UMB and MetaCYC. Compound Ontology, i.e. CO instead of GO (Gene Ontology). Gary said he would send the link.

Examples Subgroup

Not much has happened – those that offered to help were in transition during August and Joanne didn't expect much activity until after Labor Day.

OMG Meeting in Boston September 9 – 11.

Joanne asked if we should contact the OMG Group. Mike said the SBML group contacted us. Gary explained that the interest of OMG (the Object Management Group) is that they want to standardize data models. They have a request out for a pathway standard and they have asked SBML to be that standard initially. SBML sent a note to Gary (us) asking what were planning to do with the OMG, if they have contacted BioPAX, if we think it's a good idea, a good use of resources to get involved in it.

Joanne filled in that Scott Markel from OMG was at the BioPathways Meeting at the ISMB in Brisbane, and attended both Birds of a Feather Sessions. Scott gave a pitch during the BioPathways SIG meeting for participation. Joanne recounted what she had said at the meeting in response to Scott. She explained that we are a small group with very limited resources and that we didn't feel it was wise for us to get involved now. She explained that we were making good progress and invited OMG to adopt BioPAX when it is completed. Joanne said she was surprised to see the message and wanted to follow up with him to find out what happening. Joanne doesn't think that we should allocate resources to OMG efforts at this time; however, she stated that the more OMG knows about what we're doing the more likely they are to adopt it. Joanne thinks it is important that we keep the communication open with OMG and that since the meeting is local (for her, i.e. is in Boston) that it makes sense to go. By going we would learn about the OMG process, which we don't know about now. The purpose would be to keep the lines of communication open, but not to be regulated or delayed by the OMG process. Gary agreed, that it's a matter of resources. We do not want OMG to control any part of our process, but if we can help each other out, that would be great.

Gary asked if anyone know anybody who is using OMG data models? Aviv said that all she understood from them is that it's a pretty elaborate process. The SBML people were very concerned about the amount of work that goes into being compliant with OMG. Imran said that industry groups like OMG because they can point to OMG in companies and say they are using standards that have followed through with certain protocols. He was not sure if any company had adopted OMG standards and whether they are being used. Then MIAMI came up, which was OMG approved and from what we understand, it was a grueling process. Imran asked if anyone have any concerns about SBML being adopted as a default? Imran said he just sent around a quick email about SBML and BioPAX to give an intuitive idea about why it will be easy for us to use SBML and vice versa. Imran said Mike Hucka is very open to talking to us about this because they

don't necessarily want to get into the details that we do and we don't necessarily want to get into the details of describing the dynamics of pathways mathematically, so it's a good combinations, the came with CellML. If OMG wants to adopt SBML then that really cant be a deficiency for them [couldn't make out what was said].

The only company Gary knew that used OMG was Netgenics, which was bought out by LION and LION is currently involved in OMG. Both Scott Markel and Eric Minch work for Lion

Imran said he wanted to talk with Alan Robinson from EBI. Imran asked if anyone knew him, none of us do. Imran said he works a lot of EBI stands with CORBA and web services and the like, who may be an interesting person to ask if they are using or any of the OMG standardization processes. But if they start off an initiative, they're in England; they're not near by. Gary said we agreed earlier that we weren't going to go for OMG, but that he didn't say this in response to what Imran had said, just as a separate comment. Gary said that what he remembered about OMG was that Eric Minch / Lion wanted to provide help to people to get through the process and would be willing to provide support, like for instance, SBML with resources, in that they would actually spend time writing up documentation. So, he guesses they may have some vested interest in getting those things involved. Joanne then added that help and resources would be welcome if they wanted to adopt us, however, the help would have to be help and not something that would hinder our process. Everyone agreed on the above. Mike then asked if BioPAX is going to send Joanne to the meeting and then asked if Joanne wanted to go. Everyone agreed that it would be a good idea for her to go. Joanne said she'd contact Scott to get more details and to keep the communication lines open.

Mike asked everyone if Wednesday September 3 would be OK with everyone for the next conference call. It was fine with everyone present, however Gary said that he knew Chris would be away that whole week. Gary said he'd check with Chris as to whether to hold the meeting on the 3rd or reschedule for another date.

The next Face-to-Face meeting was discussed next. The plan for meeting is to review the ontology and resolve all outstanding issues, i.e. those that were open when the ontology was released and any new issues that are raised after the release of 0.5. The group discussed the 1st week of October because Imran could not make it in September except for the 1st week. Peter expressed a preference for October 3rd Denver, and then the issue of location, i.e. at the University or at the Airport. The last meeting in Denver was at the Airport (hotel). Imran offered the University because there of the excellent meeting facilities (over head projectors, connectivity, etc.) and food is nearby. Peter expressed preference for the airport to cut down on travel, which was fine with Imran, when Suzanne reminded us that we didn't have connectivity last time. Peter said it could be arranged. All agreed to think about it. There are tradeoffs – the university is nicer but you lose about an hour. Taking advantage of the Airlines (e.g. United Red Carpet) was brought up, but it seems you can only invite one guest. That may be for using the lounge, but not for setting up a meeting – which may be possible for a price. Mike will look into it and said that we should have more info and decide by the next conference call. Imran will follow-up with an email response as the next step.

Ontology issues

Mike reminded us that at our last conference call the idea was raised as to whether it would be fruitful during our conference calls to try to resolve some of the issues that we have listed open on the source forge site. We agreed that it would as long as we limit the discussion to those issues that could be resolved.

The issues:

Whether we should have a class called polymer as a super class for protein DNA and RNA.

The advantage: You save time defining slots because a lot of those things have common slots.

The disadvantage: It adds and extra class. we'd like to keep things simple and adding a class complicates things.

Mike asked if any one think of a case where an instance of the polymer class would be useful. Peter said you wouldn't use it to create instances. Then consider it as a generic non-used super class. What are the common slots: Answer:

Common slots: sequence name, database ID, common, then realized that it was probably just sequence, and if that was the case, then it probably didn't make sense to create a class just to save one slot. Then the

question was raised about whether it was macromolecule that was thought to be the proposed super class, because we have small molecule. Gary thought there are two separate things on the source forge site. HE would check. So then we talked about a macromolecule. Would it save slots? Peter took a look to see what slots if any they have defined in their databases. While he was doing this, Suzanne said polyvinyl something would be a polymer, but wouldn't have a sequence. Imran asked if we are thinking about a polymer as a super class for all DNA and RNA, answer was yes. Polymer class is useful if using it on sequences of monomers? Discerning some common set of features of Monomers – Gary said he probably wasn't thinking about it along those lines. Peter reported that they are under physical entities. A child of that being macromolecules, and there are three children of that. They are, and noted that they are not totally orthogonal, complexes, proteins, and polynucleotides. Under polynucleotides are DNA and RNA. He said there aren't many a lot of slots defined in these classes, but one nice reason to have this is to separate small molecules from macro molecules is for the purpose of querying. However, this may already be done because generally you want to query all the small molecules, you don't want to query all the macromolecules, so he said he didn't feel that strongly about it. He said he thinks it's nice to group things together in that way, but it's not a big deal. Gary said, the only argument he had for not putting it in is "Keep it Simple," when in doubt. Chris says that a lot and Mike added that it saves on documentation. Suzanne mentioned that the other reason is because if you have some slot in another class that takes it as a value then it's value can be a protein or a DNA, then you might want to have a super class for that, which is something we should be thinking about when deciding whether or not to define new classes. Gary then asked if we should resolve the issue by not putting it in now but keeping these things open so that if we need to do later than that's OK or if we need to do it later than we should do it now. Imran said that Suzanne raises an important issue. If you do want to maintain flexibility into gene-gene gene-protein interactions, then having a polymer-polymer interaction that is defined by having some region of the polymer, would be a generalized way of representing that, and that could be specialized to become a or a protein DNA molecule later, so there is an advantage to that – that you don't have to make a commitment right up front, Mike added yes, because it requires putting the equivalent interactions into our interaction tree. Imran said there does seem to be a representational advantage in that you don't have to make a commitment. The question is overhead. Gary said we could resolve these issues in two ways. One we can say the issue is closed and we are nor or we are going to do it in this specific way, or we don't do it for Level 1, but we have it on the list for Level 2. So, this might be a Level 2 issue. Peter thinks its worth reconsidering for the reasons Imran raises, and suggests we put it on the Level 2 list. Gary agreed, as did everyone else.

Conclusion: No, but possible feature for level 2.

Next issue: CONTROL class: Do we want to rename it?

Is there a better work than control because control suggests maybe a speed up or a slow down? For example an enzyme cannot slow down a reaction. So could it be called a modulating interaction?

Inhibition, activation, accumulation also under control. Yes if we had those classes. The names regulation and modulation were brought up and discussed: There were preferences for control and regulation, but then Aviv mentioned that regulation is often specific in other contexts like transcription regulation. But, it is a good word otherwise. Aviv asked for the specific example that is problematic. Gary offered that an enzyme doesn't control a reaction. Then she said what's wrong with catalysis, that's the exact word. The problem is that it isn't general enough for the other classes, it's not for just enzymes and control doesn't imply catalysis. The problem is that we have under control,

Control:

Enzyme catalysis

Gene regulation

Transport facilitation (like a transporter)

Mediator, mediation – discussed: And it doesn't have to be a protein? No.

Mediator doesn't have a connotation. Then we looked up mediation in the dictionary. Control is short. Imran looked at WordNet and reported Catalysis is a specific case of chemical process, which is a specific case of natural process, which is a specific case of process of phenomenon. We agreed they're missing a few steps in there. Peter reported the dictionary – mediation is intervention between conflicting parties to

promote reconciliation. That's not good (we agreed). Gary came up with exhibiting indirect causation or connection or relation.

Taking a step back, Gary explained that it an issue that was raised by one person.

Peter suggested that part of the problem with control is that the reasons we want to use control rather than catalysis; those examples aren't well flushed out. So, perhaps we should postpone this decision to Level 2 until we actually see how we're going to represent gene regulation or attenuation or something, it's hard to know what this all means. So, he wouldn't mind postponing this whole thing. Then Gary said that it might be difficult to postpone the renaming of a class because if you rename a class then you might make Level 2 incompatible with Level 1. If you add a new class, that's not the same. Peter said that we need to be clear about what we mean by Level 1 and Level 2. He was assuming that the first release that we're about to do, we're not encouraging anybody to use it, that it's more for comments. Gary explained that there's a difference between Level and Version and Release. Level is the way SBML is use it, and other groups like Adobe Postscript. Level 1 is a way to get things done quickly, deal with 80-90 percent of the cases. Level 2 you add more features, and hopefully they are compatible with each other, but they were going for compatibility. Versions: he explained that there could be multiple versions of a level. Level 1, version 1; Level 1 version 2. They can be bug fixes for example. Joanne said that Levels define the scope and version defines iterations within the scope. Imran cautioned that Level 1 SBML is what is used a lot and they're sort of stuck with that. Gary verified that this is Level 1 version 0.5, but this isn't a recommendation until we're satisfied with it. This would be a pre-alpha version. We could reconsider it later for Level 1, Peter suggested, which is the same as limiting out discussion, Gary replied.

Imran suggested we have a formal description of formal description of Level's and Joanne said the meanings are specific in the software industry, Imran agreed and Peter pointed out this isn't software, so it's a little different. We all agreed that we should have it defined and if there's an existing standard for representations or ontologies, we should adopt it. The group didn't know of any. The issues are major versus minor changes and backwards compatibility is separate but important and should be stated.

Summary of control issue discussion: We suggested a number of different words, regulation, modulation, and mediator. Some people liked control.

Conclusion: In the absence of a clear consensus for what we can change it to that we should leave it the same for now. All agreed, Peter suggested that Gary make some notes in the frame itself about some alternative possible wording. Gary said he has a slot in there for that and that we should consider it before the final release of Level 1.

An aside: SBML has been using a website called surveymonkey.com. You can create your own surveys and you can use it to vote. We can use it like they have for people voting on some of these things, if we need to. It might be too much trouble to do it for small things. If there are a number of people who want to vote on something and they are not on the conference call, then we can use it to track people.

Imran asked whom, in terms of numbers and members of the community will be commenting on the BioPAX release.

Gary named:

Andrey Rzhetsky would like to give feedback; said he'd be available for that
Stan Letovsky

Eric Neumann

Vincent Schacter has helped but Gary's not sure if he's on board for that

Gary said we have a list:

Internal list: 24 on the internal list

Public list: 77 people

Announce list: 105 people

Example list: 27 people

Joanne mentioned the example subgroup members at the same time Imran said that there were 20 or 30 people at ISMB that showed a keen interest being involved in this process. Joanne confirmed they are all on the example subgroup mailing list. Mike asked if they were also on the biopax-discuss list. Joanne said that they weren't automatically signed onto the discussion list, but she knew some joined on their own, some at the meeting, and some afterwards, but we didn't automatically sign them up when we created the list and couldn't say for sure.

Mike said we should definitely give it to them (the examples subgroup). Joanne added that Sylvia Nagl's group in London would definitely want to have it for comment and they have been waiting for it.

Joanne – in response to Gary's email from Genomeweb, was interested in learning more about the Jubliant database and wanted them to know about us. Imran said he looked into BioPath and that there are lots of diagrams that it is like biocarta. The entries appear to be collected 100 different resources. Joanne pointed out that someone bought it.

Joanne to follow-up Imran – follow with about NSF Grant

Suzanne asked if someone would update source forge to reflect the changes? Gary said he would.

Gary recommended that we should get an account on source forge, which is free; all you have to do is fill out a web-based form. Then he can put each of us on as a member of the group. This will allow members to make changes to the tracker and also check into CVS, which answers Peter's question about coordinating updates. Gary will administrate. When you get the source forge account, send Gary the name of the account and he will add it.