

Birds of a Feather #1 - Data Exchange Formats for Pathway Databases
ISMB 2003 Brisbane Australia
30 June 2003

This session was announced at BioPathways and BioOntologies SIG meetings on 28 June and posted on the announcement boards across from the conference registration. Monday was the first day BOF sessions were held was chosen for the first BioPAX BOF because it was expected that a follow-up BOF session would be desired. Scheduling early would give us time maximum time.

During the first BOF we had roughly 20 attendees. We were able to complete proper introductions and brief statements of interest during this first session and agreed to schedule another BOF for the next day to discuss these interests and decide how to move forward. We wish to thank Phil Lord for his suggestion to abandon the long table and move our chairs to a circle and for having the music turned off in the area where we met. Thanks go to Frank van Enckevort for arranging for a room for the second BOF session.

30-June - 1-July BOF attendees:

Joanne Luciano
Ken I Fukuda
Yu-Tai Wang
Ugur Dogrusoz
Emek Demir
Poul Nielsen
Mark Lacy
Phil Lord
Robert Stevens
Kam Dahlquist
Eunok (Eunice) Paek
Frank van Enckevort
Peter Hunter
Matt Halstead
Steffan Hennig
Steve Burbeck
Scott Markel
Michael Hucka
Barry Wanner
JunBai Wang
Jim Butler
Klaus Lehnert

Re: Use Cases Robert Stevens comments:

Develop and decide what is the core of what your producing
Mechanism for Use Case Submission
Purpose of the Ontology is to capture knowledge of the community

Interest of Community members:

- Fakuda – pathway database interest and interest in communicating with pathway groups
- PATIKA – exchange pathway data
- CellML integrate to be more specific about meanings and terms
- Physiome – Peter Hunter (Auckland) – (Physiome project is in collaboration with Oxford University, Denis Noble)
- NIZO, Nijmegen, The Netherlands Food Research (Frank V.E.); tools and recreation of pathways from multiple data sources
- USCB – Pathway graphs, comparison and analysis KEGG, WIT, Flux reactions

- U of Manchester – Developing and using bioinformatics tools. Hear what the community is interested in with respect to tools, make sure they are building tools that meet the communities needs. [Resident experts in Ontology use and building] “Biologists make the best ontologies, computer scientists are there to help them do this” Robert Stevens. Learn about OWL and applications of ontologies to better meet the needs and make sure they are hearing the users – Phil Lord
- LION-OMG Lion- user of pathway data in current project. OMG Representative, put out an RBP, would like to receive submission.
- ___ - Curates pathway information providing visualization tools, different levels of representation ensuring they agree and are interpreted in same way.
- GenMAPP (Kam D). Viewing Microarray data on pathways. (User). Draw pictures of pathways with gene IDs. Wants to make their system compatible with an exchange format (BioPAX).
- U of Seoul, Korea – Eunice Paek: Data modules for signaling pathways.
- Max Planck Institute, Berlin Steffan H. Uses GO, interested in pathway ontology (user).
- IBM, I3C Steve Burbeck, Interested in interoperability, that’s the goal, the question is “How to achieve it?” What are the gaps?
- SBML – Mike Hucka – Interoperability

Summary of attendees’ interests:

Standards/Interoperability
 Curators of data / data base developers
 Modelers
 Users
 Ontologists
 Tool developers

Groups for use cases – formal representations such as aMAZE, representation is hypergraph; been asked to export in DAML+OIL.

Discussion:

All agreed to have a 2nd BOF the next day; proposed agenda: discuss use cases; Robert Stevens, Natalia Maltsev and Phil Lord regrets – they have organized a BOF on GRID technologies at that time.

Birds of a Feather #2

Agenda for BOF 2 – BioPAX (drafted by Joanne and Imran, input from attendees of 1st BOF).

Handouts:

1. Poster copies
2. Use case diagram of reverse engineering of gene network made by BioPAX, SBML and CellML

Use Cases:

Prioritize

Automatic submission (suggestion made by Robert Stevens at previous BOF. Need to follow-up with an email to Robert, he offered to look up the screens for auto-submission)

Examples

Subgroups - Group philosophies should be defined by each subgroup

Existing:

Small Molecule
 State

Examples/Use Cases

New:

Tools – what tools will be needed; what are we going to represent?

Ontology (formalism)

Validation & Quality Control

(Paper/Grantwriting? – jsl)

Questions/Main Issues

What do you want to see?

What's missing?

Next Steps?

With the introductions out of the way we could talk about the needs of the community. We came up with about 5 pathways/parts of pathways that constituted a satisfactory proof of concept for the user community. [I'll have to fill this in later – I don't have the big flip chart sheets with me at the moment.] The idea would start with Kam (GenMAPP) who would generate the pathway diagrams/maps which would be converted into BioPAX. Members of the BOF would write the tools and pass the exchanged little data base from one group to another. These details would be worked out in a subsequent follow-up email to the group (which is forthcoming). This was thought to be a natural extension to the test the Small Molecule Subgroup ran on CML. All liked the idea and agreed to contribute.

Layout example of data loss.

Legacy systems (P&G)

-annotations – only the tool that puts it in knows how to use it.

-choosing tags to... (name space issue)

Fields that are required in the exchange format

?few common pathways

Models for glycolysis and MAPK exist

Complex – oxidated phosphorolase KEGG cytochrome(?)

Phosphotransphorase (transport system)

Create a benchmark for exchange from several sources and minimal set of pathways that are representative of the community (this was what we all agreed to do).

Tags – what follows is CellML, SBML – allow for representations

Out of Scope of BioPAX

- Geometry
- Junctions
- Cell-level events

Overlap with CellML/SBML and BioPAX: Interaction/Transformation

Source info (authorship, reliability)

Time Line (for getting the benchmark done? For BioPAX?)

Also was discussed the need for the creation of additional subgroups:

1. Tools
2. Validation

Follow-up

1. Start the Examples sub group mailing list
2. Get the examples benchmark underway. The following people agreed to contribute to creating a set of pathways for a benchmark for the BioPAX data exchange format. This will involve creating the pathway representations and tools to exchange the pathway data. We will also need to develop

a mechanism to verify the integrity of the data and feed back to the biopax-discussion group problems that arise so they can be addressed. The following agreed to contribute to this process, now we have to work out a plan.

- a. Kam Dahlquist (GenMAPP) will write up how GenMAPP will use BioPAX and send the example pathways
- b. Eunok (Eunice) Paek (University of Seoul) in studying heat shock signaling pathway
- c. Steffen Hennig (Max Planck Inst, Berlin) is interested in structural views of pathways by Gene Ontology sequence mapping and clustering by GOs
- d. Ken Ichiro Fukuda (Computational Biology Research Center, AIST, Tokyo Japan) is interested in signaling transduction pathways:
 - Reference pathway
 - Weekly updated pathway (journals)
 - Literature based curation
 - Logical inference of pathways e.g. cross talk
- e. Jim Butler (GSK) is interested in representations that scientists can actually look at in order to understand data (as opposed to pathway modeling). Interested in ways to categorize pathways to improve intuitive understanding
- f. Klaus Lehnert (ViaLactia Biosciences, Auckland NZ)
- g. Matt Halstead (BioEngineering Institute, U of Auckland, NZ) CellML
- h. Ugur Dogrusoz and Emek Demir (Belkint University, Turkey) PATIKA – interested in modeling, integration, visualizing and analyzing pathways.