

Experiences from CellDesigner to BioPAX conversion

Anushya Muruganujan

University of Southern California

SBGN-10 08/17/2014

CellDesigner to BioPAX conversion

Two versions

1. Java CellDesigner Plugin version
2. Java application that uses the CellDesigner code.

Both output BioPAX document in owl format

CellDesigner Plugin version

- The CellDesigner plugin interface allows access to the Model.
- There are some model annotation objects that are not accessible via the plugin interface; For example, MIRIAM references for Cellular components, etc. This is a vital omission, since, this is precisely the kind of information that BioPAX requires to link to external data. MIRIAM references were added 'recently' to CellDesigner, they may become available via the plugin interface.

CellDesigner Plugin version cont...

- Uses webservices to query EBI's Ontology Lookup Service to create BioPAX xRef's, if possible.
- The plugin interface allows the current model to be output in BioPAX. It does not allow the plugin to open multiple models. Hence, batch conversion of multiple models is not possible.
- Relies on PaxTools Java library to output BioPAX file in owl format.

Java application version

- Uses the CellDesigner code to access the CellDesigner internal objects.
- Requires changes to some of the CellDesigner code to enable batch processing.
- Relies on PaxTools Java library to output BioPAX file in owl format.
- Uses webservices to query EBI's Ontology Lookup Service to create BioPAX xRef's if possible.
- Has feature to add's xRefs for PANTHER pathways through webservice calls to www.pantherdb.org. A standard protocol has not been established for CellDesigner diagrams from other sources.

Unmapped items

- Pathway order is not specified in CellDesigner although it can be calculated.
- Sequence association and Evidence is not stored in CellDesigner.
- Activation state is not available in BioPAX.
- ...

Mapping for main CellDesigner types

CellDesigner	BioPAX
Model	Pathway
Species	Physical Entity
Reaction	Interaction
Compartment	CellularLocationVocabulary

Mapping details for components

CellDesigner	BioPAX
Protein	Protein
RNA and AntiSenseRNA	RNA
Gene	DNA
Simple molecule	SmallMolecule
Ion	SmallMolecule
Unknown, Degraded, Drug	PhysicalEntity
Phenotype	Pathway
Heterodimer	Complex
Modification	ModificationFeature SequenceModificationVocabulary from ModiciationState SequenceSite from ModificationResidue also SequencePosition or SequenceRegionVocabulary
Binding region	BindingFeature

Details for Protein mapping

1. A protein can appear in the same cellular location in within a CellDesigner diagram. In BioPAX, a single Protein object will be created. This has implications when we try to go back from BioPAX to CD, SBGN, etc.
2. One instance of ProteinReference will also be created for the protein in each unique cellular location.
3. If there are protein references that can be mapped to the Protein, ProteinReferences will also be created for the protein. These will be added as memberEntityReferences to the ProteinReference from step 2.

Mapping details for reactions

CellDesigner	BioPAX
State transition, truncation,	Biochemical Reaction
Transcription, translation	TemplateReaction
UnknownTransition, KnownTransitionOmitted	Conversion
Association or disassociation	ComplexAssembly
Transport	Transport with biochemical reaction
Truncation	Degradation
Catalysis	Catalysis
Modifier, Physical stimulation, modulation, trigger, inhibition, unknown inhibition,	Control

However...

Creation of a valid BioPAX document does not mean that it can be linked with external data sets.

“An Inconvenient Model”



CellDesigner or BioPAX?

Answer: Both!

Why? They serve different purposes

BioPAX

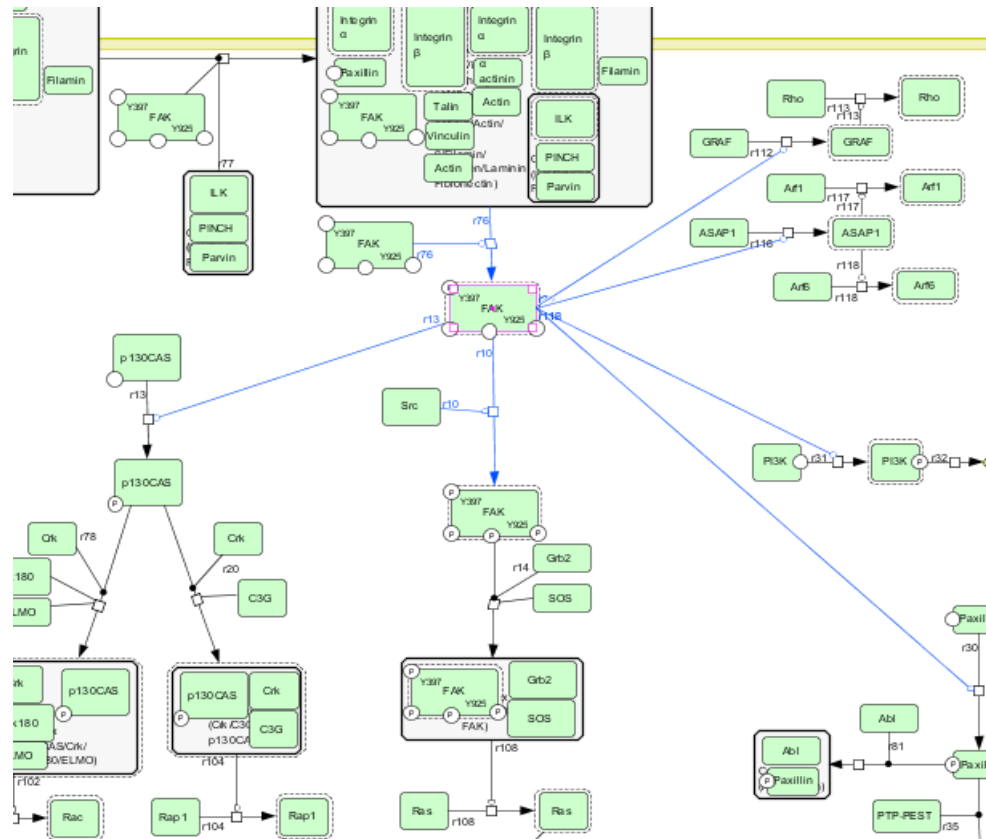
- The rationale for BioPAX is exchange of information.
- BioPAX is not meant to be 'human readable'.
- It has helped in cleaning up some of the CellDesigner diagrams.

```
<bp:ProteinReference rdf:about="http://identifiers.org/uniprot/P10894">
  <bp:name rdf:datatype = "http://www.w3.org/2001/XMLSchema#string">1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-1</bp:name>
  <bp:xref rdf:resource="http://identifiers.org/panther.family/PTHR10336:SF12" />
  <bp:xref rdf:resource="#uniprot_P10894_1094" />
  <bp:organism rdf:resource="http://identifiers.org/taxonomy/9913" />
</bp:ProteinReference>

<bp:ProteinReference rdf:ID="_Wnt_PROTEIN">
  <bp:xref rdf:resource="http://identifiers.org/panther.pthcmp/P01444" />
  <bp:displayName rdf:datatype = "http://www.w3.org/2001/XMLSchema#string">Wnt</bp:displayName>
  <bp:standardName rdf:datatype = "http://www.w3.org/2001/XMLSchema#string">Wingless-type MMTV integration site family member</bp:standardName>
  <bp:comment rdf:datatype = "http://www.w3.org/2001/XMLSchema#string">ENTITY_REFERENCE_PROTEIN_TYPE=GENERIC</bp:comment>
  <bp:comment rdf:datatype = "http://www.w3.org/2001/XMLSchema#string">ENTITY_REFERENCE_NOTES=
Long Name: Wingless-type MMTV integration site family member
Synonym: Wingless
Accession: P01444
</bp:comment>
  <bp:comment rdf:datatype = "http://www.w3.org/2001/XMLSchema#string">ENTITY_REFERENCE_ID_DESC=Wnt signaling pathway.PROTEIN.GENERIC.Wnt</bp:comment>
  <bp:memberEntityReference rdf:resource="http://identifiers.org/uniprot/F1P5C4" />
  <bp:memberEntityReference rdf:resource="http://identifiers.org/uniprot/F1N7G3" />
```

CellDesigner

- CellDesigner is useful for providing a visual representation of components, their interactions and locations.



CellDesigner cont...

- CellDesigner does not impose any restriction on creating elements that can only be cross-referenced from external sources.
- CellDesigner's popularity is due to its flexibility. The user has freedom when it comes to naming components and depicting the components, interactions and locations. Things are usually self-explanatory although they may not always follow standard conventions. In biology it is impossible to specify a set of rules to cover all cases. This is similar to computer science where a sorting algorithm can perform poorly for a given dataset.

Lessons learned/implications for other software

- Text Mining capability for linking external data. This can be a back end process that continuously mines for data as the user edits the document.
- The editor should prompt the user to enter cross reference data. The software can perform federated searches and present the results to the user. Similar to a personal assistant.
- Interface has to be established for searching and returning the data.