

SBML → SBGN → SBML

Just my 2 cents

Alice C. Villéger

COMBINE 2010

# Disclaimer

- Fuzzy talk
  - work in progress
  - last minute slides...
- Someone else has been working on *very* similar stuff
  - and should really have been talking today
  - YES, I'm talking about YOU Frank Bergmann

# Context

- University of Manchester, July 2007 – October 2010
  - Douglas Kell, Steve Pettifer, BBSRC funded project
- Let's have a **look** at SBML files...
  - e.g. yeast metabolic reconstruction
  - or BioModels.net Database
- Software applications to create/edit **layouts**
  1. **Arcadia** <http://arcadiapathways.sourceforge.net/>
  2. **“CellWhatever”** [WORKING TITLE]  
(Arcadia + Dunnart + CellDesigner)
- *Before* SBGN-ML and LibSBGN

# What do we mean by SBML files?

- “Raw” SBML
- **Semantic** annotations
  - SBO terms
  - MIRIAM
- **Layout & render extension**
  - “official” way to store visual information in SBML
  - growing tool interoperability (e.g. COPASI, Frank’s SBML Layout)
- **CellDesigner** annotations
  - “historical” file format: a lot of legacy models
  - most (?) popular SBML visual editor
- JDesigner annotations
  - not covered in this talk

# How to draw a SBML file? (1/2)

- **Rendering**

- partly cosmetic

- affects usability (font **size**, **color** contrast, etc.)
- but not a huge priority for me (at that stage)

- partly linked to **semantics**

- **SBGN** = new standard (esp. SBGN PD for SBML)
- older conventions (e.g. CellDesigner notation)
- can encode extra semantics (e.g. through color)

# How to draw a SBML file? (2/2)

- **Layout**

- **difficulty** explodes as networks grow

- manual layout takes time
- automatic layout is a bit rubbish...

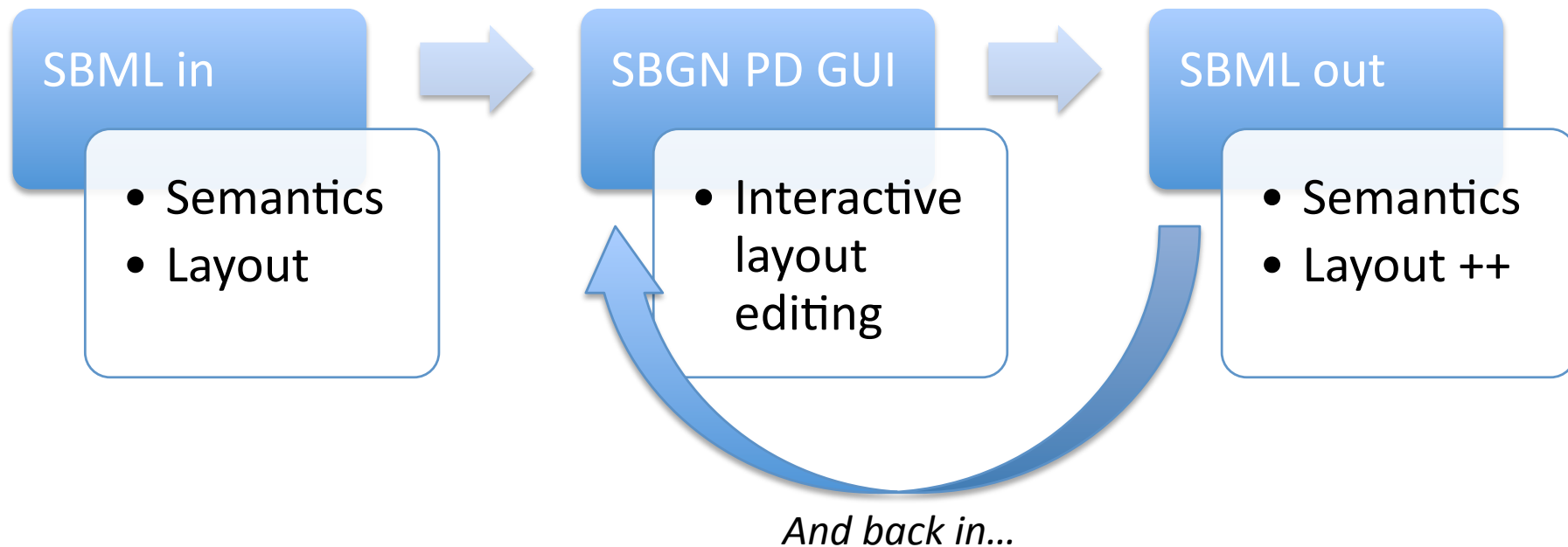
- can be described as **static coordinates**

- e.g. layout extension, CellDesigner annotations

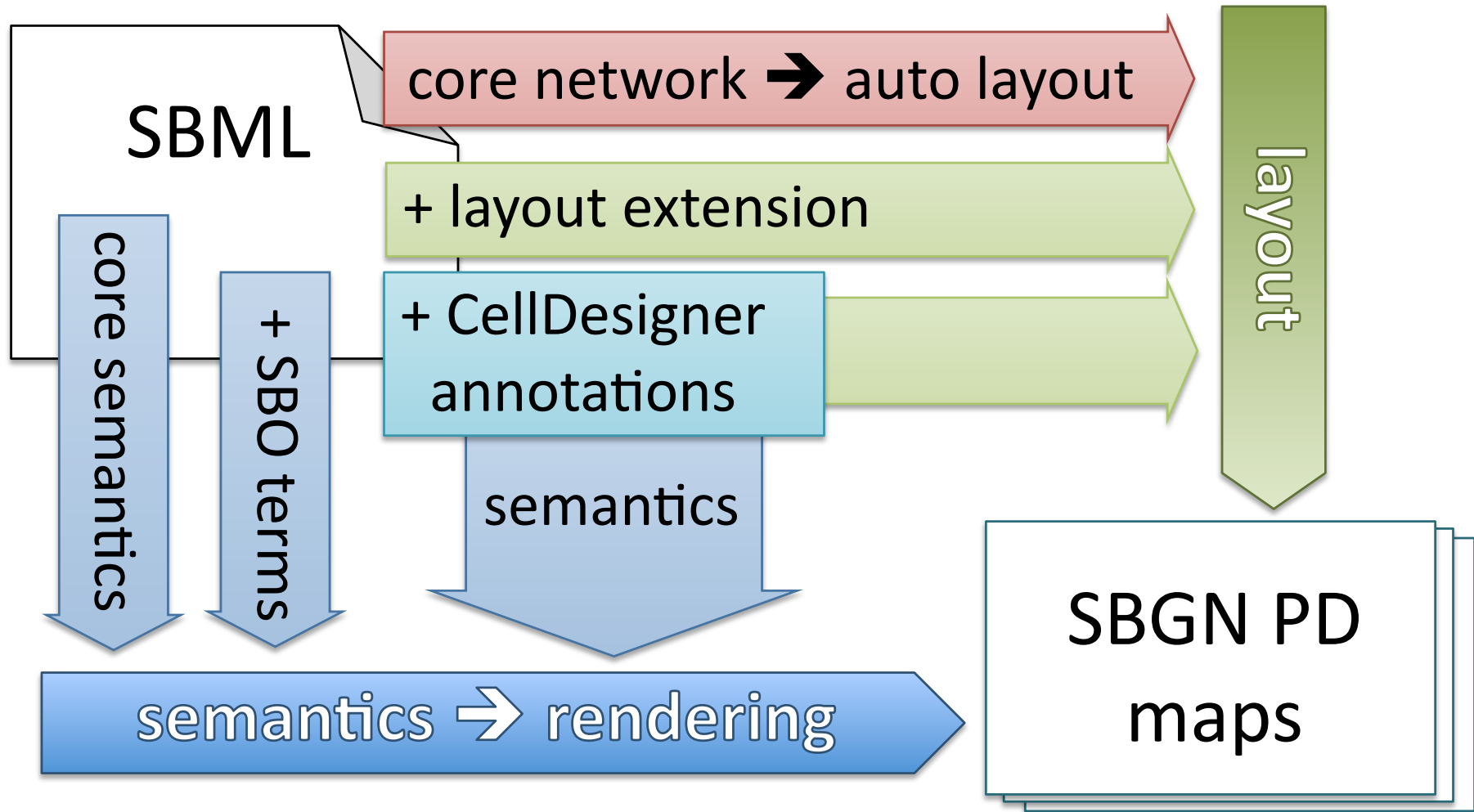
- can also be expressed as **dynamic rules**

- simple constraints = alignment, distribution, separation, etc.
- domain specific templates (e.g. cycle, cascade, etc.)
- currently no standard way to store that information in SBML

# SBML to SBGN PD to SBML...



# Input





# Output

- SBML
  - **Layout extension**  
(multiple layouts per file)
  - **CellDesigner extension**  
(only one layout per file)
  - **Dynamic layout information** \*new\*
    - Arcadia annotations → will be deprecated (or updated?)
    - CellWhatever annotations (from Dunnart) → work in progress
- Image export
  - SVG, PDF, PNG, ... (can improve rendering with SVG editor)  
e.g. Inkscape
- Other formats
  - e.g. SBGN-ML → soon...

# Semantics and Rendering (1/3)

- Raw SBML → SBGN PD
  - pretty straightforward?
    - Species → Unspecified Entity
    - no Reactants/Products on a Reaction → Source/Sink
    - Reaction → Transition
    - Compartment → Compartment
    - Modifier → Modification Arc
    - Product → Production Arc  
(but sometimes the rate equation actually goes the other way!)
    - Reactant → Consumption Arc, or Production Arc if Reaction is reversible  
(default case if attribute not set !)
  - semantically poor (uses only a small subset of SBGN PD)
  - missing bits? (could use a closer look at the math?)
    - events, assignment rules, etc. → [?] (e.g. perturbation or observable???)
    - impact of modifiers in rate equation → more specific modification arc?  
(also, logical operators???)
    - **real** stoichiometry for SBGN arcs? (cf. SBML Units)

# Semantics and Rendering (2/3)

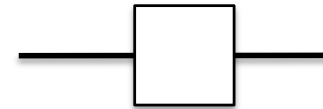
- SBML + SBO → SBGN PD
  - fairly straightforward? Cf. `arcadia/apps/pathways/pathwaystylesheet.cpp`
  - semantically richer: covers most of SBGN PD
    - some missing bits: state variables, units of information, “inside” of complex, submaps (modular SBML?)
  - not always obvious which SBML element would carry the SBO term (cf. SBML specs)
    - e.g. confusion for modifiers → required new SBO terms (e.g. “catalyzer” not “catalysis”)
    - can relevant SBO terms be on other things than reactions, species, species refs, or compartments?
    - are some SBO terms relevant to SBGN actually never used on any SBML element?
- Availability of this conversion information?
  - SBGN specs (PDF)
  - Some talks I gave in March/April 2009
  - Arcadia source code (1<sup>st</sup> version summer 2009, updated around Xmas 2009)
  - Frank has done something similar => SBML Layout code on SourceForge???
  - What’s the best way to share that information?

# Semantics and Rendering (3/3)

- Questions
  - SBML notes → SBGN info glyph?
  - Source and sink: implicit (zero reactant/product) or explicit? (species with corresponding SBO term)
- Todo
  - Use MIRIAM annotations
    - Turn instance (e.g. “is ATP”) into class (“simple molecule”)
    - Get info on inside of complex? (hasPart)
    - Any other use? (e.g. state variables?)
  - Not reinvent the wheel
    - Cf. SBML ↔ BioPAX conversion tools?
  - Rendering extension (I was waiting for the new LibSBML patch)
    - Useful for non-SBGN info (color, font, etc... shape of compartments)
    - But what about glyph shape?
      - Infer semantics? Difficult, and not reliable in general
      - May be OK if there was an official, identifiable SBGN stylesheet for the render extension

# Layout

- Raw SBML?
  - network connectivity → graph → automatic graph layout
    - EPN, PN = nodes
    - Arcs = directed (?) edges
    - Compartments = ? (clusters? Subgraphs?)
  - no cloning information → lots of edge crossings, and strange map shape
  - some networks are just too big to be shown on a single map in a way that makes sense
- Layout extension
  - ✓ Quite straightforward (cf. Layout extension specs, + LibSBML support for parsing)
  - ✓ Supports EPN cloning
  - X More generic than SBGN → can describe non SBGN compliant maps → validation? ignore some data? (e.g. ignore label bounding box position if not centered)
  - X Not sure if source/sink should be a SBML glyph, or a “other graphical object” (no SBML element)
  - X Reactions are hyperedges...  
Process Node: Bounding box + curve segment?
  - X Describe some semantics! (arcs role)  
... What if contradicts SBML + SBO?
  - Discuss and define a standardized way to store SBGN layout info in layout extension?



# CellDesigner Annotations

- Semantics, rendering AND layout [Work in progress...]
- Parser... from official xsd and specifications
  - Tried automatic binding <http://www.codesynthesis.com/products/xsd/>
  - But failed (→ feedback to improve CellDesigner XSD?)
    - no single root element (annotations for model, species, reactions, ...)
    - a few missing types
    - some name collisions between elements and attributes
  - Wrote parser by hand
    - Lots of work...
    - But good insight in the format data model
- Rich semantics... Almost a bit too rich, actually (from an SBGN perspective)
  - Done: most EPN and PN, compartments, inside of complex
  - Some glyphs require a bit of translation, not documented yet (Cf. SBGN viewer in CellDesigner)
  - Work in progress: state variables, multiple views (expand/collapse complex)
- Cosmetic rendering (color, font, etc.): ignore at the moment (but preserve info when saving)

# CellDesigner Annotations

- Layout
  - Species alias allows cloning
  - Compartment alias? Not an SBGN concept (?), not sure what to do about it
  - Some limitations, e.g. discrete port for connecting arcs on nodes
    - OK for CellDesigner → SBGN
    - approximation for SBGN → CellDesigner
  - Curious definition of arcs and reactions (solved, but should document more)
    - Sometimes not SBGN compliant (“branch” reactions)

# Layout rules

- Old system: Arcadia annotations
- Now using Dunnart annotations
  - Work in progress
  - References objects described in layout extension or CellDesigner map



# Open questions...?

- How to handle potential semantics conflict
  - E.g. SBO vs CD inconsistencies?
  - Or rendering extension vs SBO
  - Would be nice to insert missing SBO inferred from other sources of semantic info...?
- How well do various layout annotations coexist?
  - Could we say that the CD layout is the same as one of the Layout extension maps?
  - What about submaps? How to represent their relationship? (bundle some of the layouts extension maps together via an annotation?)

# Conclusion

- What is the best way to share the work I've done on this conversion problem?
  - Library? (cf. LibSBGN)
    - What language? (C++? Java?)
  - write some technical documentation?
  - some XML representation of some relationships? (e.g. SBO  $\Leftrightarrow$  SBGN mapping?)
- Who is interested in using these results?
- Who can help?