

Working With Different Pathway Standards

Dagmar Köhn

University of Rostock
Graduate School diEM oSIRIS

13. Januar 2007

University of Rostock
Graduate School diEM oSIRIS

Content

- 1 Motivation
- 2 Existing technologies
 - Data storage
 - Data transformation
 - Model matching
- 3 Solution
 - The architecture
 - Achievements
 - Problems
- 4 Future Work
 - Architecture
 - dIEM oSIRIS

Motivation

Standards

- Biological Pathway Data
- Different standards for storage and exchange of pathway data
 - SBML
 - PSI MI
 - bioPAX
 - CellML ...
- Different formats for definition of models
 - XML Schema
 - OWL model
 - And others, such as Java

Motivation

Standards

- Different standards make work harder for biologists.
 - Different databases for each standard
 - Different queries for each standard



Motivation

Standards

- Different standards make work harder for biologists.
 - Different databases for each standard
 - Different queries for each standard



Thesis

A comparison of different formats with each other is needed.

Existing Technologies

... for data storage

What do we already have?

Existing Technologies

... for data storage

What do we already have?

- Biological data from publications
- Stored in models/databases
- Data can be queried from databases

Existing Technologies

... for data storage

What do we already have?

- Biological data from publications
- Stored in models/databases
- Data can be queried from databases

But:

- Data is stored in **different formats**
SBML → XML Schema
PSI MI → XML Schema
bioPAX → **OWL model**
- Biologists are restricted to a limited number of software

Existing Technologies

... for data storage

What do we already have?

- Biological data from publications
- Stored in models/databases
- Data can be queried from databases

But:

- Data is stored in **different formats**
SBML → XML Schema
PSI MI → XML Schema
bioPAX → **OWL model**
- Biologists are restricted to a limited number of software

Goal

Work with pathway data should not be limited to one standard.

Existing Technologies

... for data and model comparison

How can one **compare** data?

Existing Technologies

... for data and model comparison

How can one **compare** data?

- 1 Comparison on XML Schema level
 - Tools: Clio (IBM), COMA++(University of Leipzig)
- 2 Comparison on OWL model level
 - Tools: SAMBO (University of Linköping), COMA++, Protégé
- 3 Comparison of different formats
 - SBML and bioPAX (XML Schema/OWL model)

Existing Technologies

... for data and model comparison

How can one compare data?

- 1 Comparison on **XML Schema** level
 - Tools: Clio (IBM), COMA++(University of Leipzig)
 - 2 Comparison on **OWL model** level
 - Tools: SAMBO (University of Linköping), COMA++, Protégé
 - 3 Comparison of different formats
 - SBML and bioPAX (XML Schema/OWL model)
- Most important standards use XML Schema and OWL model

Existing Technologies

... for data and model comparison

How can one compare data?

- 1 Comparison on **XML Schema** level
 - Tools: Clio (IBM), COMA++(University of Leipzig)
 - 2 Comparison on **OWL model** level
 - Tools: SAMBO (University of Linköping), COMA++, Protégé
 - 3 Comparison of different formats
 - SBML and bioPAX (XML Schema/OWL model)
- Most important standards use XML Schema and OWL model

Conclusion

A translation between XML Schema and OWL model helps to compare and query existing data available in different standards.

Existing Technologies

...for the transformation of XML Schema and OWL model

What does the translation look like?

Existing Technologies

...for the transformation of XML Schema and OWL model

What does the translation look like?

Transformation

A transformation is only possible in **one direction**, XML Schema to OWL model.

Existing approaches:

- Mapping XML to OWL (Auer/Bohring, University of Leipzig)
- PSI MI to bioPAX transformation (Frank Gibbons)
- Recall transformation (Köhn, University of Linköping)
 - XML Schema to OWL model transformation (XSLT)
 - Recreation of original XML Schema (XPath)

Existing Technologies

...for the comparison of XML Schema and OWL model

Example (Recall Transformation using XSLT)

```
<xsd:schema [...]>
  <xsd:simpleType name="SId">
    <xsd:restriction
      base="xsd:string">
      <xsd:pattern
        value="(_|[a-z]|[A-Z])
          (_|[a-z]|[A-Z]|[0-9])*"/>
      </xsd:restriction>
    </xsd:simpleType>

  <xsd:complexType name="SBase">
    <xsd:sequence>
      <xsd:element name="notes"
        type="SId"/>
    </xsd:sequence>
  </xsd:complexType>
</xsd:schema>
```

```
<rdf:RDF [...]>
  <owl:Class rdf:ID="SBase" /> [...]
  <owl:DatatypeProperty
    rdf:ID="notes">
    <rdfs:range
      rdf:resource="xsd:string" />
    <rdfs:comment>
      The following restriction [...]
    </rdfs:comment>
    <rdfs:domain
      rdf:resource="#SBase"/>
    <!--XPath information (@notes/
      xsd:element/xsd:sequence/
      xsd:complexType[@SBase]/
      xsd:schema)-->
  </owl:DatatypeProperty>
</rdf:RDF>
```


Existing Technologies

...for matching on OWL model level

What happens after the transformation?

Existing Technologies

...for matching on OWL model level

What happens after the transformation?

- Models are available in a common format
- Matching

Definition (Match)

“A **match** is a function which takes two schemas S1 and S2 as input and returns a mapping between those two schemas as output [...]” (Rahm)

Example (Match function)

schema 1

crocodile

rough

duck



schema 2

rough

crocodile

roughly

Existing Technologies

...for matching on OWL model level

What happens after the transformation?

- Models are available in a common format
- Matching

Definition (Match)

“A **match** is a function which takes two schemas S1 and S2 as input and returns a mapping between those two schemas as output [...]” (Rahm)

Example (Match function)

schema 1

crocodile

rough

duck



schema 2

rough

crocodile

roughly

Existing matching tools

...for matching on OWL model level

- SAMBO
 - OWL and DAML+OIL



Alignment Candidate Details

SBML	PSI MI
ListOfParameters definition: synonym: part-of:	parameter definition: synonym: part-of:
ListOfParameters definition: synonym: part-of:	parameterList definition: synonym: part-of:

ListOfParameters
parameter

ListOfParameters
parameterList

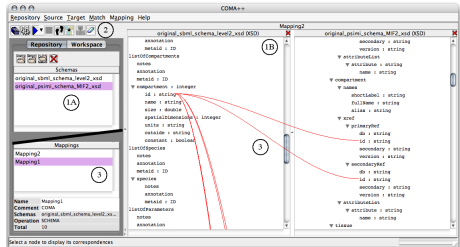
new name for the equivalent concepts:

5 Remaining Suggestions, Previously Aligned Concepts

Existing matching tools

...for matching on OWL model level

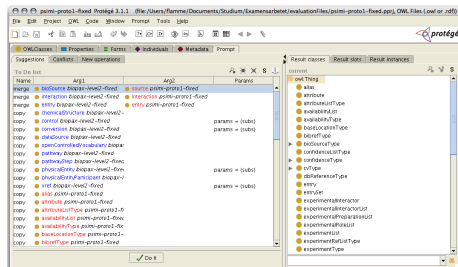
- SAMBO
 - OWL and DAML+OIL
- COMA++
 - OWL conversion to internal format



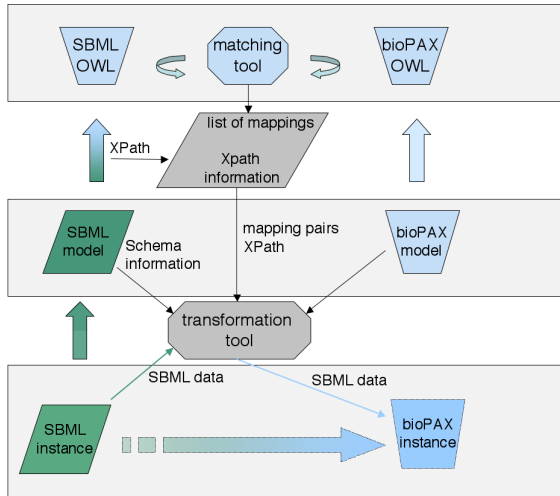
Existing matching tools

...for matching on OWL model level

- SAMBO
 - OWL and DAML+OIL
- COMA++
 - OWL conversion to internal format
- Protégé Prompt plugin
 - OWL version control



Solution! Architecture



Solution

Achievements

What can we do with this architecture?

Solution

Achievements

What can we do with this architecture?

- Comparison of different standards (models and data)
 - Transformation of SBML models/data into other formats
 - Retrieval of SBML models/data out of other formats
- Storage of data in different formats
 - Store SBML data in a PSI MI format
- Enhanced queries
 - Query SBML data using PSI MI, bioPAX... syntax
 - Query other formats using SBML syntax
- Version control

Solution!?

Problems

And what can we not do?

Solution!?

Problems

And what can we not do?

- XML/OWL translation only in one direction
- Matching only on OWL model level
 - Quality of transformation
 - Quality of OWL matching tool

Solution!?

Problems

And what can we not do?

- XML/OWL translation only in one direction
- Matching only on OWL model level
 - Quality of transformation
 - Quality of OWL matching tool

Evaluation

First evaluations showed that the idea is working, but both, matching algorithms and transformation, still have to be improved.

Solution!?

Problems

And what can we not do?

- XML/OWL translation only in one direction
- Matching only on OWL model level
 - Quality of transformation
 - Quality of OWL matching tool

Evaluation

First evaluations showed that the idea is working, but both, matching algorithms and transformation, still have to be improved.

- Standards aim at different targets
 - SBML → description of biochemical reactions
 - PSI MI → protein interaction
 - bioPAX → metabolic pathways, molecular interactions
- Are the different standards comparable at all?

Future Work

Proposed Architecture

The following main points will be considered in future work:

- Biologists and the different standards
- Differences/common attributes in different standards
- Implementation of the architecture
- XSL Transformations XML Schema → OWL model
- Evaluation of SBML OWL model

Future Work

dIEM oSIRIS project

How can this architecture contribute to the dIEM oSIRIS project?

Future Work

dIEM oSIRIS project

How can this architecture contribute to the dIEM oSIRIS project?

- Integrative Development of Modeling and Simulation methods for Regenerative Systems (University of Rostock)
- Biologists explore signaling pathways
- Computer Scientists store recorded data
 - Large scale model storage for simulation tools
- Computer Scientists simulate pathway models
 - Efficient storage of pathway data and request of model components (queries on structure)

Future Work

dIEM oSIRIS project

How can this architecture contribute to the dIEM oSIRIS project?

- Integrative Development of Modeling and Simulation methods for Regenerative Systems (University of Rostock)
- Biologists explore signaling pathways
- Computer Scientists store recorded data
 - **Large scale model storage** for simulation tools
- Computer Scientists simulate pathway models
 - Efficient **storage of pathway data** and request of model components (**queries on structure**)

Last words

When trying to get biologists to work with software, it has to be kept as simple as possible. Providing a tool to compare and query any data from any pathway database, can be of big help.

Last words

When trying to get biologists to work with software, it has to be kept as simple as possible. Providing a tool to compare and query any data from any pathway database, can be of big help.

Thank-you for your attention!

contact

<http://wwwmosi.informatik.uni-rostock.de/diemosiris>

dk103@informatik.uni-rostock.de