



Biological Data Migration in Pathway Simulation

Biological Data Migration
in Pathway Simulation

***Abel Gómez**, José Á. Carsí, Artur Boronat, Isidro Ramos*
agomez@dsic.upv.es

Technical University of Valencia.

Departamento de Sistemas Informáticos y Computación.

Software Engineering and Information Systems Research Group



Table of contents

- Motivation.
- Case study.
- Model-Driven Engineering: standards and tools.
- Proposal.
- Application of the proposal to the case study.
- Conclusions and future work.



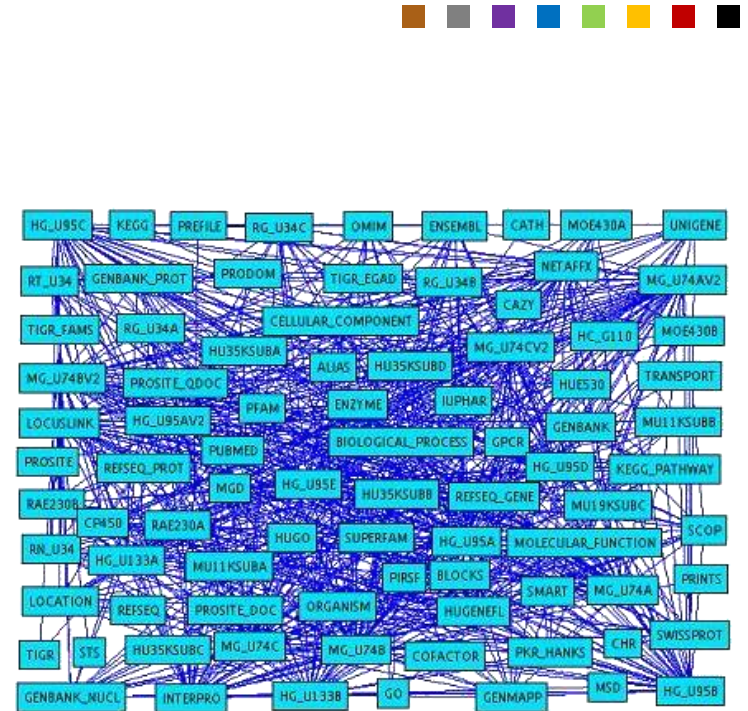
Motivation

MOTIVATION



Motivation

- Nowadays there are approximately 1078 databases that store biological data¹.
- Some of them may contain repeated or even contradictory information.

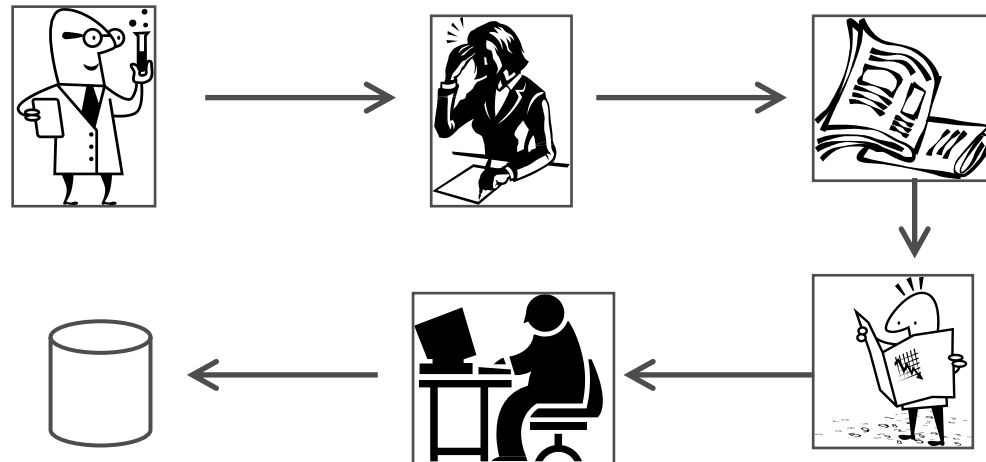


¹ Nucleic Acids Research, Volume 36, Database issue.
http://nar.oxfordjournals.org/content/vol36/suppl_1/index.dtl

Motivation



- The typical information retrieval process is prone to introduce errors in the information extracted.



- It is necessary to provide interoperable simulation tools in order to validate (and to be able to work with) the great amount of data obtained from the scientists.

Studying the signaling pathways

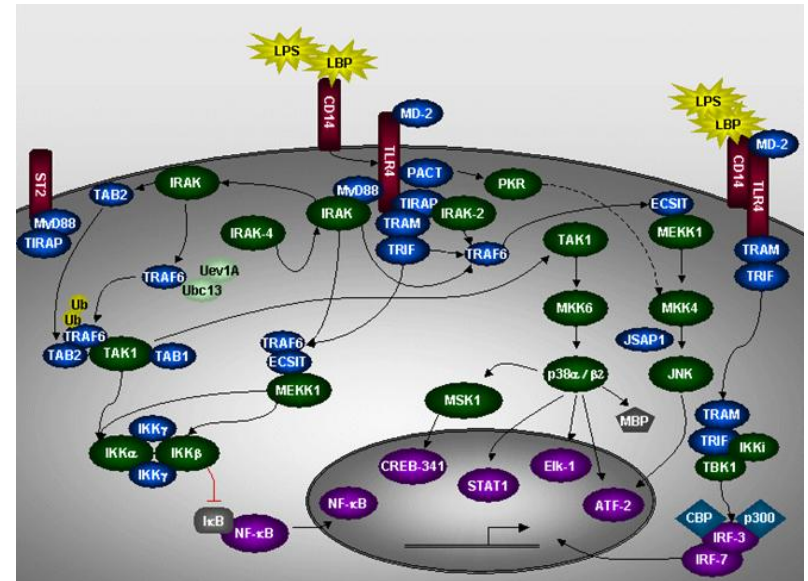
21001100 100 21001100 10000000



The Toll-Like Receptor 4 pathway.



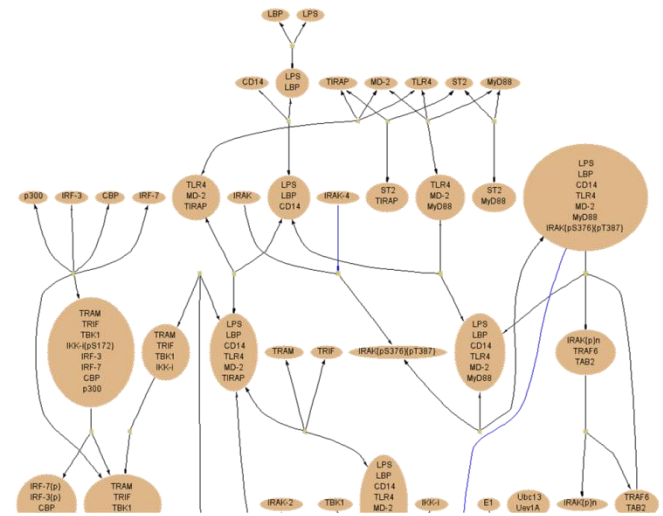
- Databases store information about proteins and the interactions between them.
- Interactions between proteins are described by means of signalign pathways.



The Toll-Like Receptor 4 pathway.

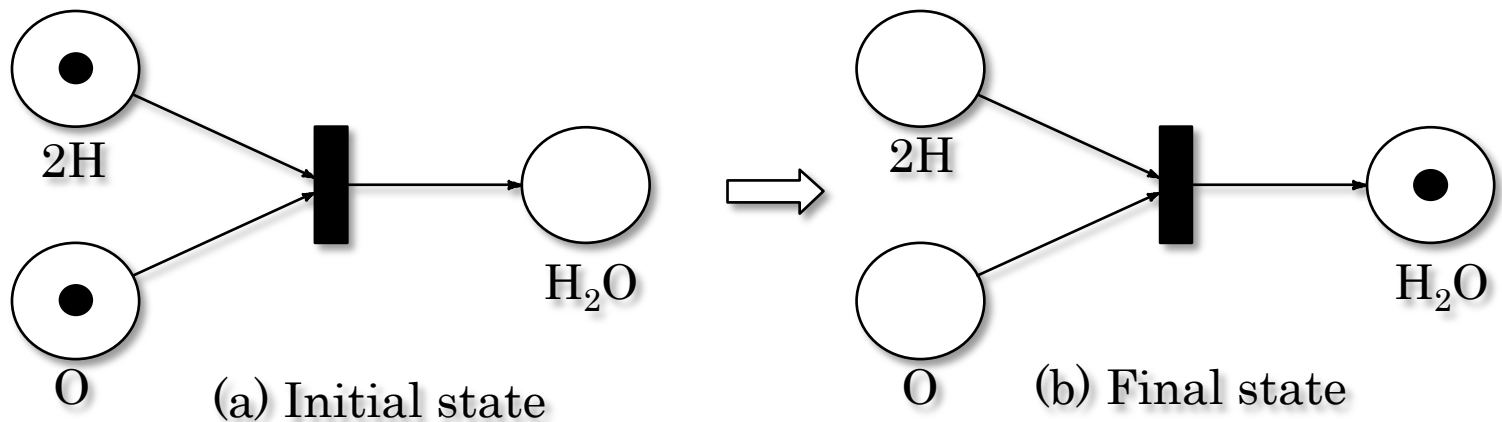


- Mechanistic view of the TLR4 signalign pathway.
- A set of chemical chain reactions occur. These reactions can also be simultaneous.



Petri nets.

- Petri nets are a formal representation that is able to represent concurrent events. It is used to simulate discrete distributed systems.
- Petri nets can be used to represent chemical reactions.



Case Study.



- The case study has been extracted from the INTERGENOMICS project.
 - Data source: TRANSPATH[®] (BIOBASE).
 - Simulation Tool: CPN Tools.
- Traditional approach: Manual creation of the TLR4 pathway extracted from the TRANSPATH[®] database into CPN Tools
 - ≈ 75 Places.
 - ≈ 47 Transitions.
 - ≈ 100 Colors.





Model-Driven Engineering

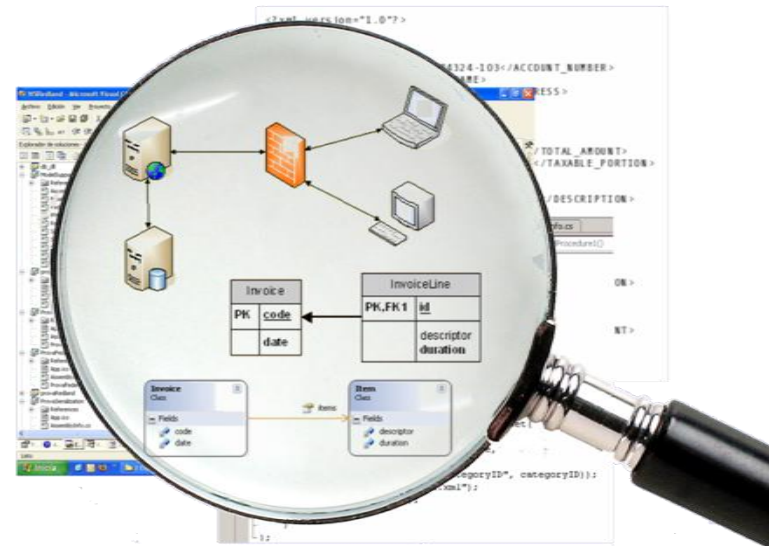
MODEL-DRIVEN ENGINEERING



Model-Driven Architecture



- Model-Driven Architecture.
 - Models as the main assets in the software development process.
 - More productivity, more portability, more interoperability.
- MDA standards.
 - MOF.
 - UML+OCL.
 - Query/Views/Transformations.
- Advantages.
 - Widely used modeling frameworks.
 - Automatic code generation.
 - Standard interchange formats.



Technology in MDA

- Eclipse (and the Eclipse Modeling Framework) is becoming the standard de facto in MDA. It is strongly supported by IBM.
- Eclipse is a platform highly extensible all its functionality is provided by means of plug-ins.
- Ecore can be seen as a subset of the UML2 class diagram, and it allows to define new models/metamodels.
- EMF provides automatic code generation capabilities for:
 - The system structure in Java.
 - XML persistence for the data of the system.
 - Graphical Editors (Tree editors, visual editors).
 - Etc.

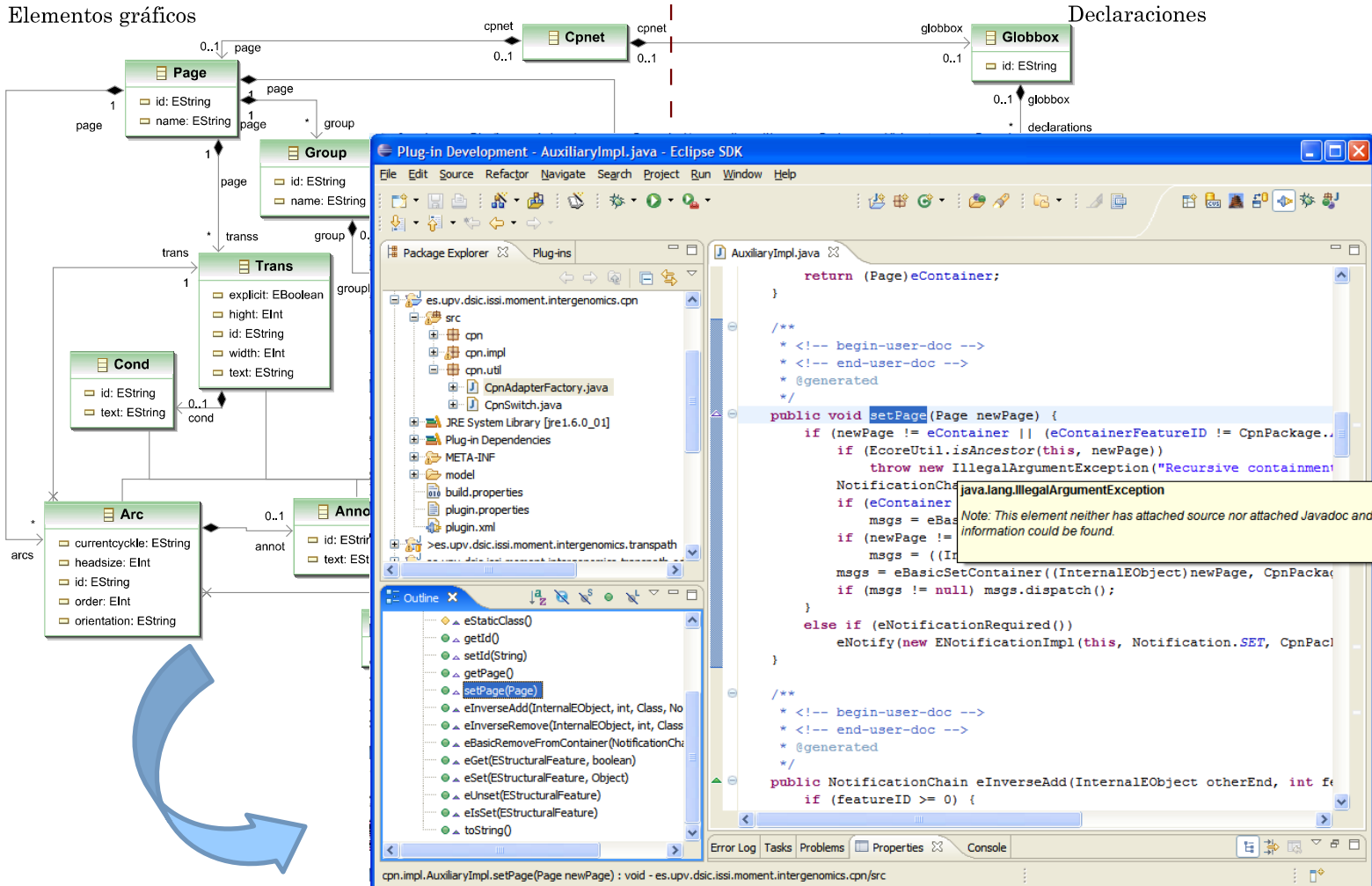


Technology in MDA

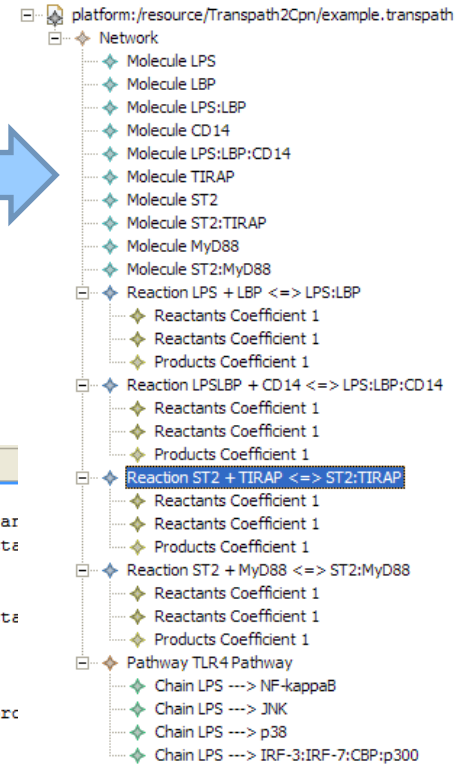
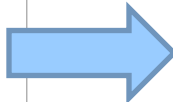
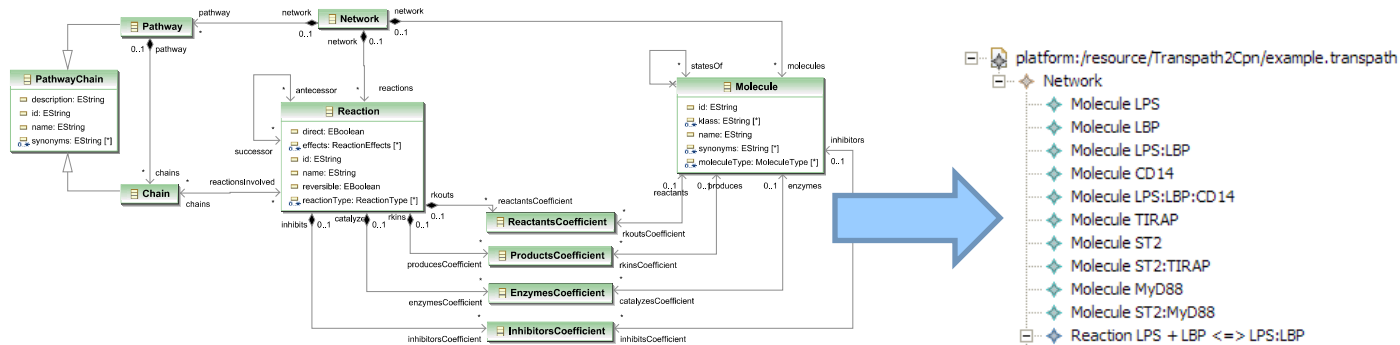


Elementos gráficos

Declaraciones



Technology in MDA



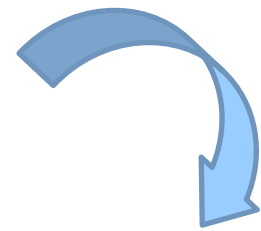
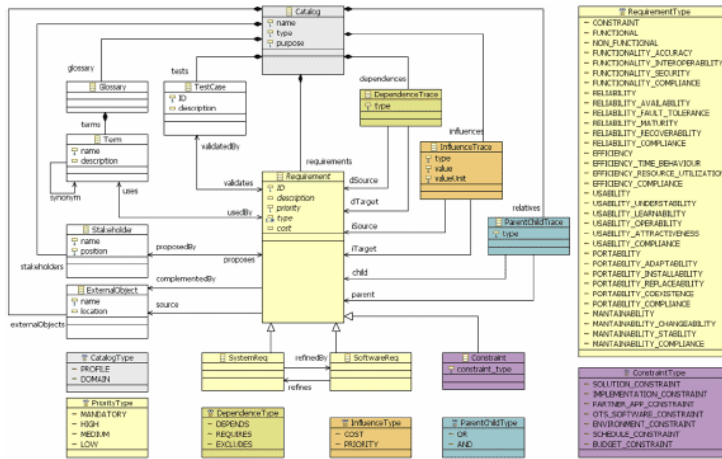
```

example.transpath
<?xml version="1.0" encoding="UTF-8" standalone="no"
es.upv.dsic.issi.moment.intergenomics.transpath/model/example.transpath
>
<transpath:Network xmi:version="2.0" xmlns:xmi="http://www.omg.org/XMI"
xmlns:trar="http://www.omg.org/spec/Transpath/2.0/Model/Transpath"
>
  <molecules id="MO000016882" name="LPS" rkoutsCoefficient="//@reactions.0/@reactantsCoefficient"
  <moleculeType>other</moleculeType>
  </molecules>
  <molecules id="MO000019420" name="LBP" rkoutsCoefficient="//@reactions.0/@reactantsCoefficient"
  <moleculeType>orthogroup</moleculeType>
  <class>adaptor proteins</class>
  </molecules>
  <molecules id="MO000021928" name="LPS:LBP" rkinsCoefficient="//@reactions.0/@productsCoefficient"
  <moleculeType>orthocomplex</moleculeType>
  </molecules>
  <molecules id="MO000018132" name="CD14" rkoutsCoefficient="//@reactions.1/@reactantsCoefficient"
  <moleculeType>isogroup</moleculeType>
  <class>membrane-transducting</class>
  <class>components</class>
  <class>receptors</class>
  </molecules>
  <molecules id="MO000021929" name="LPS:LBP:CD14" rkinsCoefficient="//@reactions.1/@productsCoefficient.0"
  <moleculeType>orthocomplex</moleculeType>
  </molecules>
</transpath:Network>

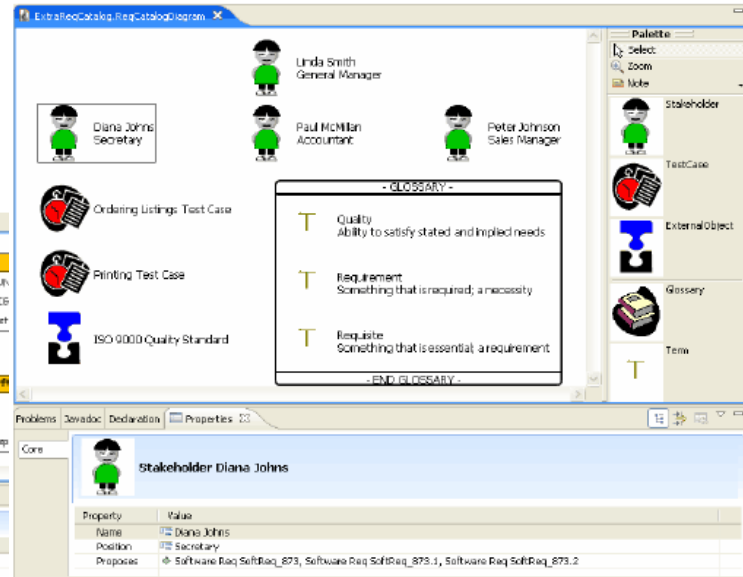
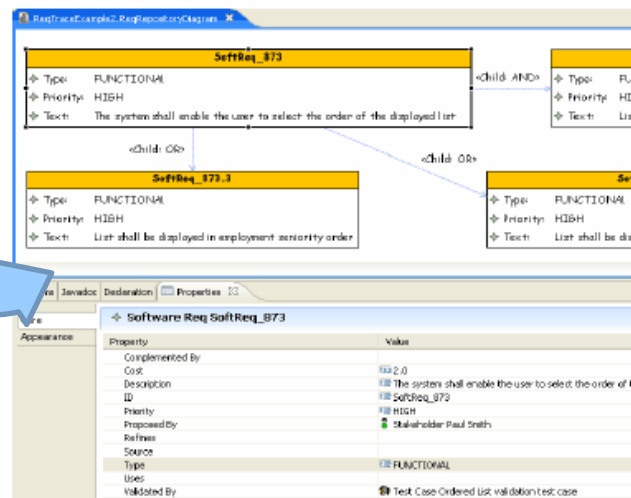
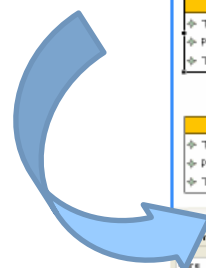
```



Technology in MDA



Credits: Vicente-Chicote, Moros, Toval.
REMM-Studio: an Integrated Model-Driven Environment for Requirements Specification, Validation and Formatting.





Proposal

Propobots



Goals



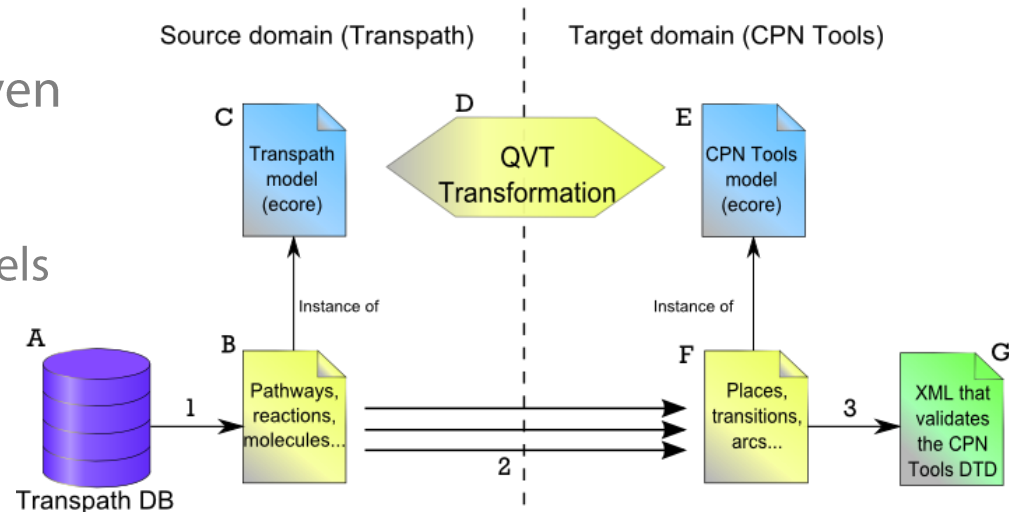
- Not to reinvent the wheel. We want to share with biologists how we manage, organize and share the information in the software engineering field.
- To transfer the widely used open standards that are in use in software engineering (if no other previous and specific standard yet exists).
- Fill the gap between computer scientist and biologists by using high level concepts to manipulate the data and by using declarative languages to deal with the information.
- To apply all the previous technology developed by third parties, in order to decrease the cost of software development.



Architecture of the tool



- The data migration process is done in 3 steps:
 - Preprocessing.
 - Transformation.
 - Postprocessing.
- When using a Model-Driven approach...
 - Biological data are represented using models very easily.
 - We can deal with data independently of the persistence format.
 - Higher level of abstraction.
- Technologies:
 - EMF.
 - MOMENT-QVT.





Application of the proposal to the case study

to the case study
with a focus on the bioinformatics



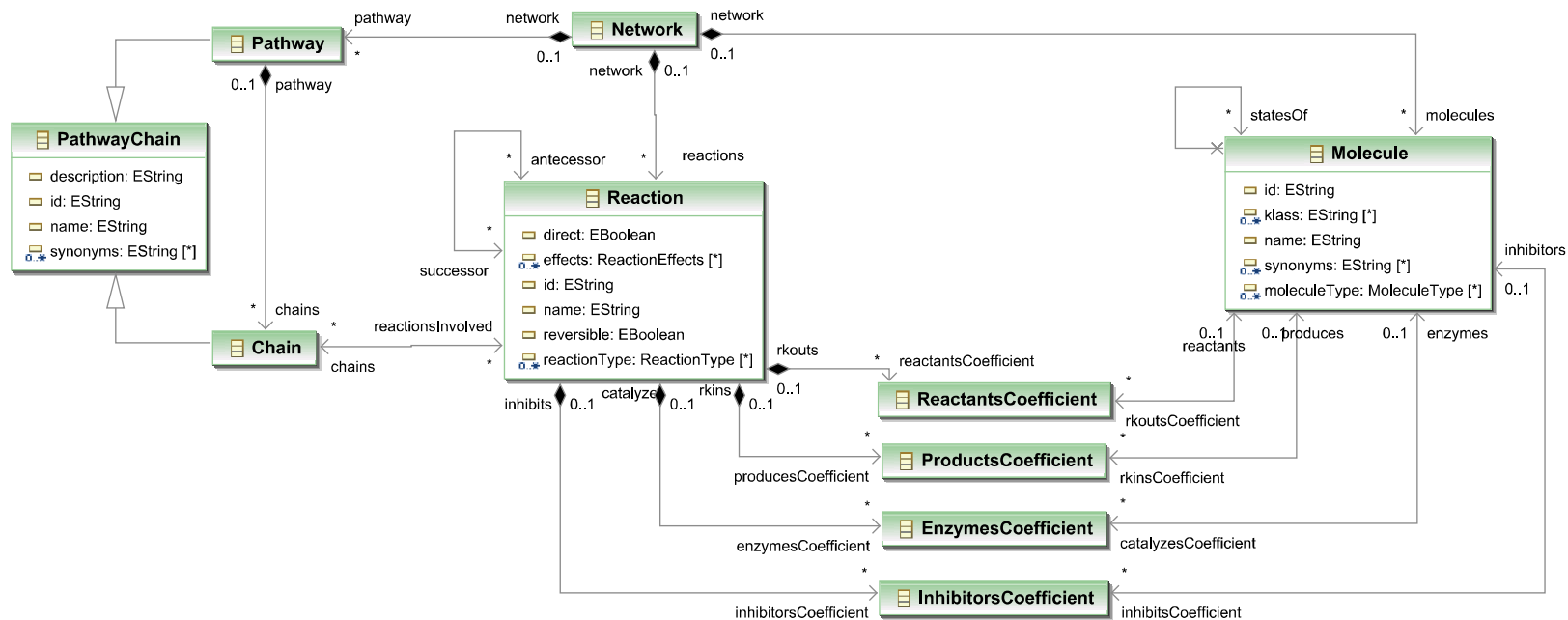
The TRANSPATH[®] database.



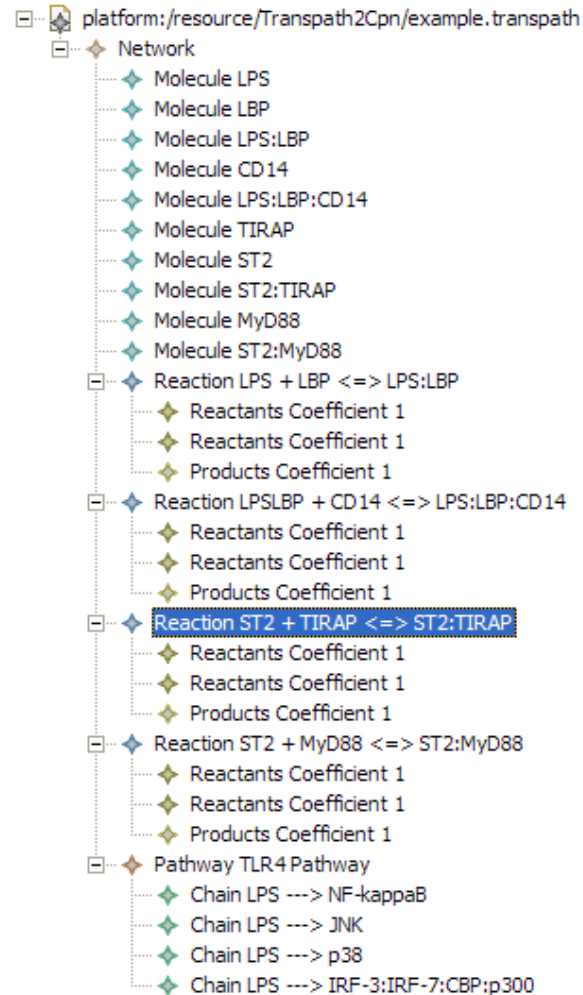
ID	Accession number	MO000022132	
	Created 2001-11-05 by	Claudia Choi (click for feedback)	
	Updated 2006-03-29 by	Sriranjini (click for feedback)	
Modification information	Copyright	Copyright (c) Biobase GmbH.	
	Name	Molecule name	TLR4(m)
		Synonyms	TLR-4; toll-like receptor 4.
Species	Encoding gene	G006820; TLR4.	
	Species	mouse, Mus musculus.	
	Classification	membrane-transducing components; receptors; cytokine receptor family; TIR superfamily; TLRs; TLR4 . membrane-transducing components; receptors; pathogen-recognition receptors; TLRs; TLR4.	
External links	Type	isogroup.	
	Supertamilies	MO000019394; TLR4.	
	Subfamilies	MO000079660; TLR4(m).	
	External database hyperlinks	MGD: MGI:96824; . SWISSPROT: Q9QZF5; . SWISSPROT: Q9QUK6; . BKL: HumanPSD:Tlr4. INTERPRO: IPR000157; TIR. [4]. INTERPRO: IPR000887; KDPG and KHG aldolase. [4]. INTERPRO: IPR000483; Cysteine-rich flanking region, C-terminal. [4]. INTERPRO: IPR001611; Leucine-rich repeat. [4].	
	External database hyperlinks (of encoding gene)	show	
	GO: biological process, molecular function	show	
	Location positive and experiment(s) More info...	GO: cellular_component: integral to membrane; GO:0016021. GO: cellular_component: lipopolysaccharide receptor complex; ISS; GO:0046696. GO: cellular_component: membrane; IEA; GO:0016020.	
Reactions where this molecule appears	Complexes	show	
	Reaction	XN000005346; TLR4(m) + MD-2(m) <=> TLR4(m):MD-2(m) (binding) [2]. XN000005380; TLR4(m) + MyD88(m) <=> TLR4(m):MyD88(m) (binding) [3]. XN000101120; Bcl-10(m) + TLR4(m) <=> Bcl-10(m):TLR4(m) (binding) [1].	
	Reaction upstream	XN000020629; TLR4(m) -> TLR4(m) (expression).	
Information about the source publication	Visualization	Show graphics for the signaling network In this window In a new window	
	Hand-drawn map(s)	TLR4 pathway	
	Reference number	[1] PUBMED:15213237.	
Author(s), Title, Journal	Liu Y., Dong W., Chen L., Xiang R., Xiao H., De G., Wang Z., Qi Y. BCL10 mediates lipopolysaccharide/toll-like receptor-4 signaling through interaction with Pellino2		



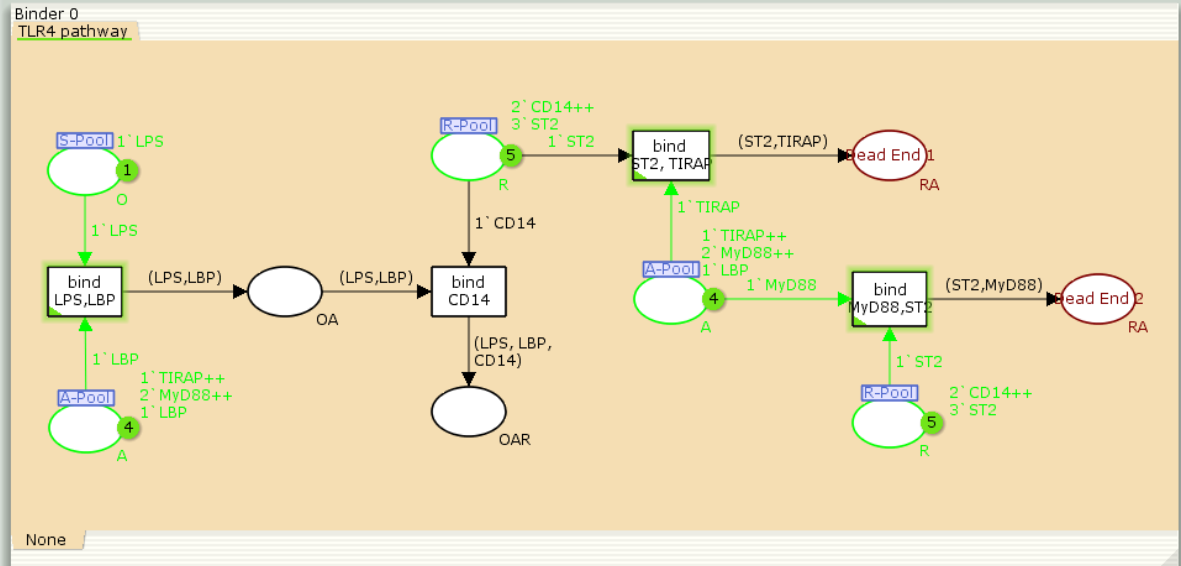
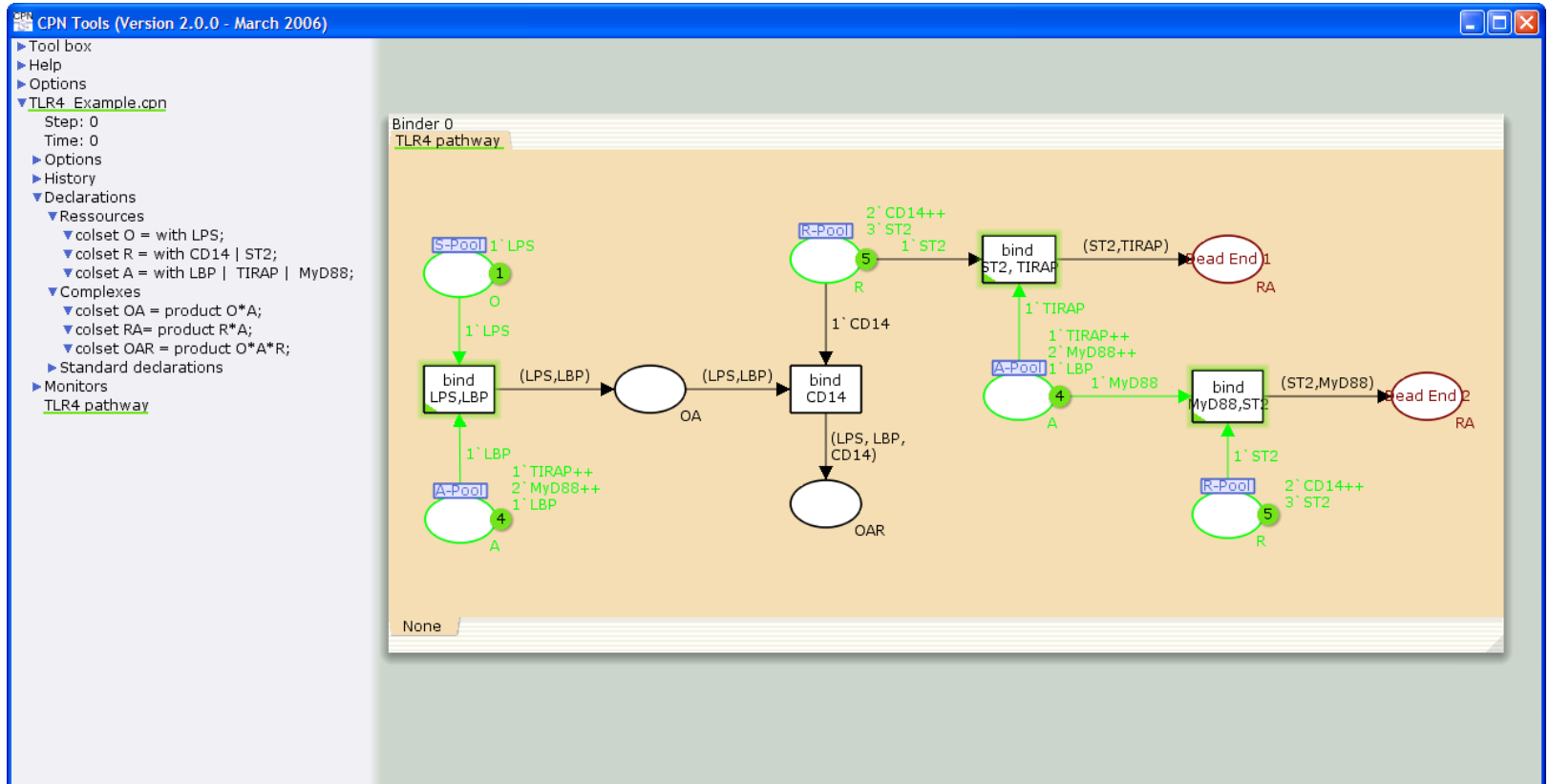
The TRANSPATH[®] database model.



Instance of the TRANSPATH© model.



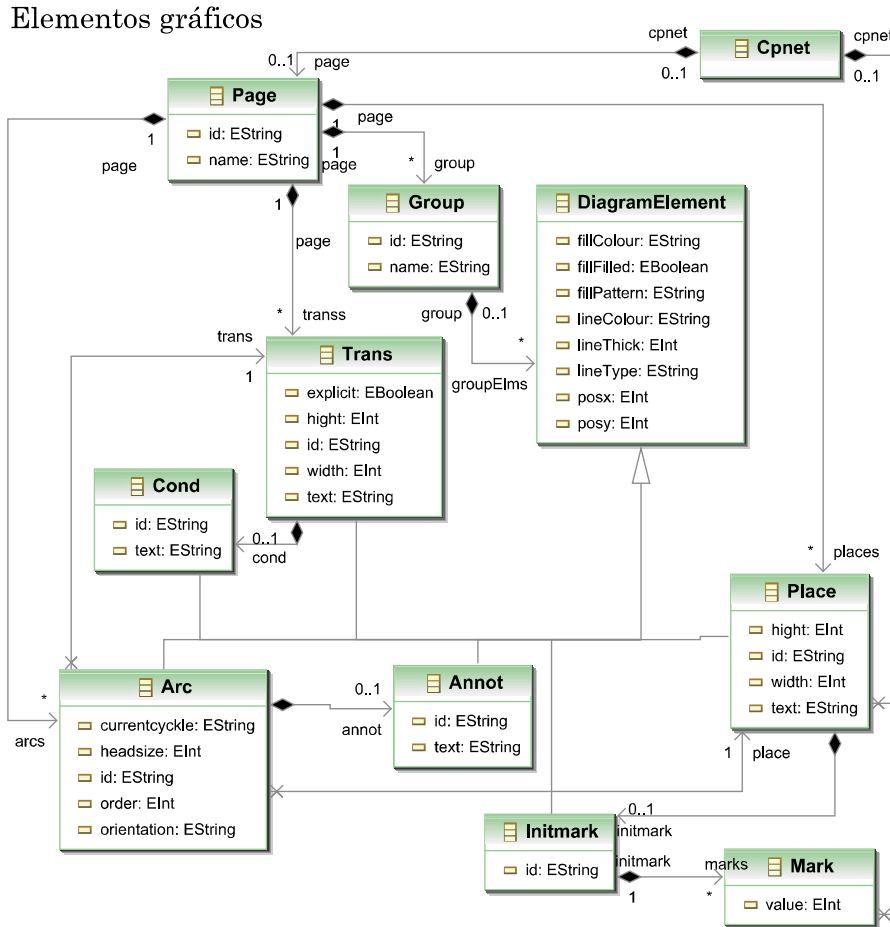
CPN Tools interface.



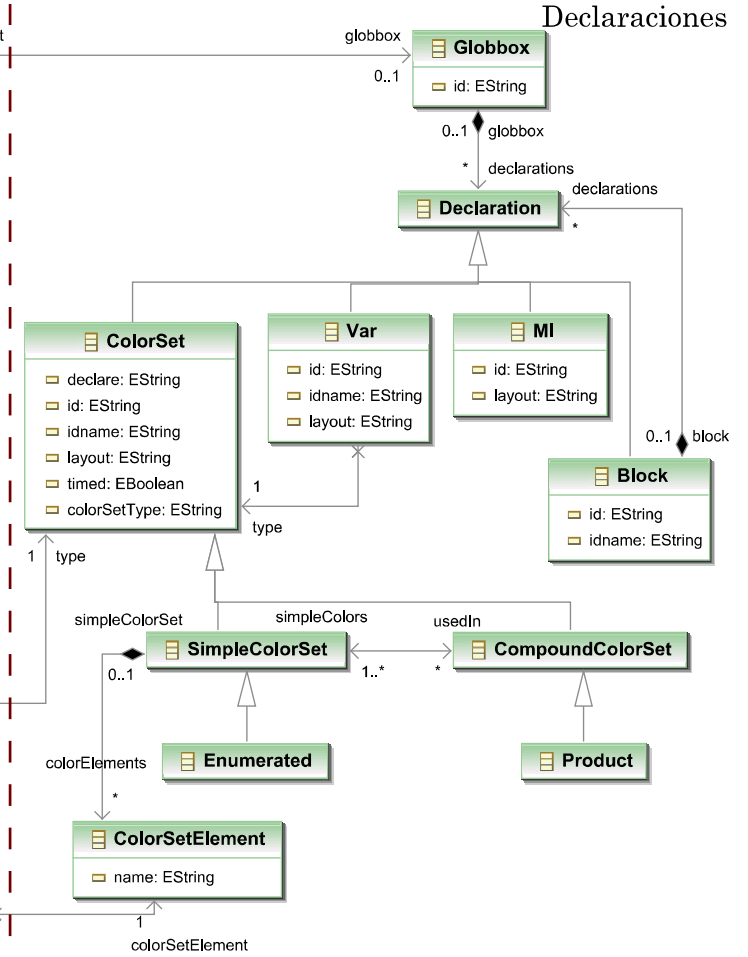
CPN Tools model.



Elementos gráficos



Declaraciones



Transpath2CPNTools Transformation.

top relation NetworkToCpnet {

```

checkonly domain tpDomain nt1:Network {
  molecules = molec1 : Molecule {},
  reactions = react1 : Reaction {}
};
    
```

enforce domain cpnDomain cn1:Cpnet{

```

page = page1 : Page {
  id = GetPathwayName(nt1),
  name = GetPathwayName(nt1)
},
    
```

```

globbox = gb1 : Globbox {
  id = 'Declarations'
}
    
```

};

where {

```

  ComplexMoleculeToComplexesBlock(mole
  c1,gb1);
    
```

```

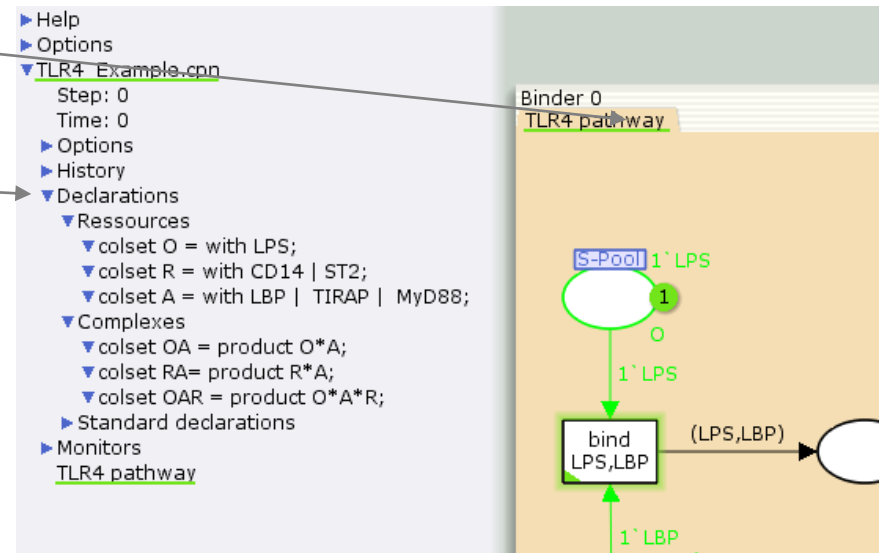
  ReactionToGUIElements(react1,cn1,page1);
    
```

}

}

- **TLR4 pathway:**

- $LPS + LBP \rightleftharpoons LPS:LBP$
- $LPS:LBP + CD14 \rightleftharpoons LPS:LBP:CD14$
- $ST2 + TIRAP \rightleftharpoons ST2:TIRAP$
- $ST2 + MyD88 \rightleftharpoons ST2:MyD88$



Transpath2CPNTools Transformation.

relation ReactantsToPlaces {

molecName1 : String;
transId1 : String;

```

checkonly domain tpDomain
  reactantMolecule1 : Molecule {
    name = molecName1
  };
    
```

//...

```

enforce domain cpnDomain place1 : Place {
  id = GetMoleculeType(reactantMolecule1),
  page = page1
};
    
```

where {

//...

```

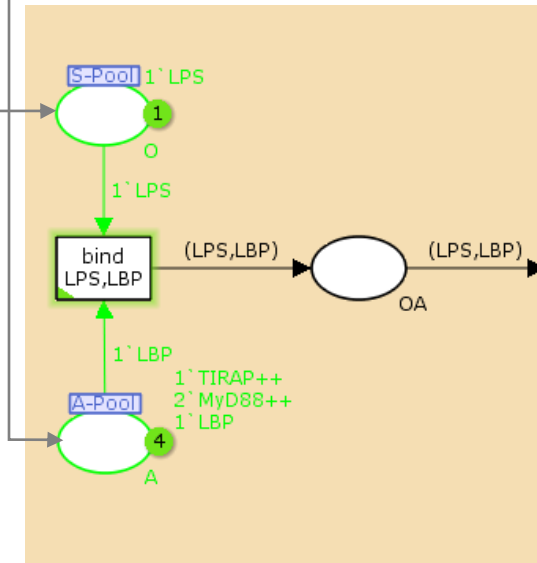
  ReactantsToArcs(reactantMolecule1,page1
  ,trans1,place1...);
    
```

}

}

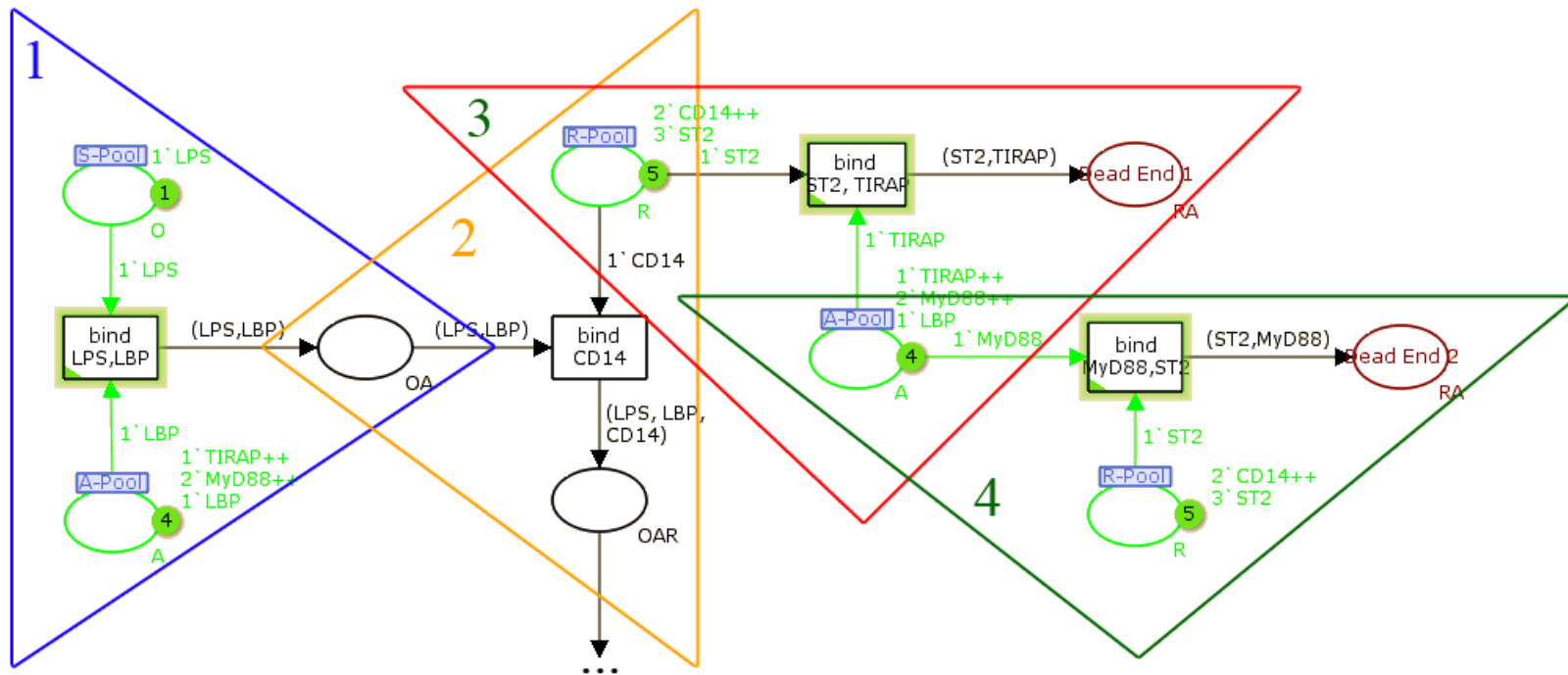
• TLR4 pathway:

- LPS + LBP \rightleftharpoons LPS:LBP
- LPS:LBP + CD14 \rightleftharpoons LPS:LBP:CD14
- ST2 + TIRAP \rightleftharpoons ST2:TIRAP
- ST2 + MyD88 \rightleftharpoons ST2:MyD88

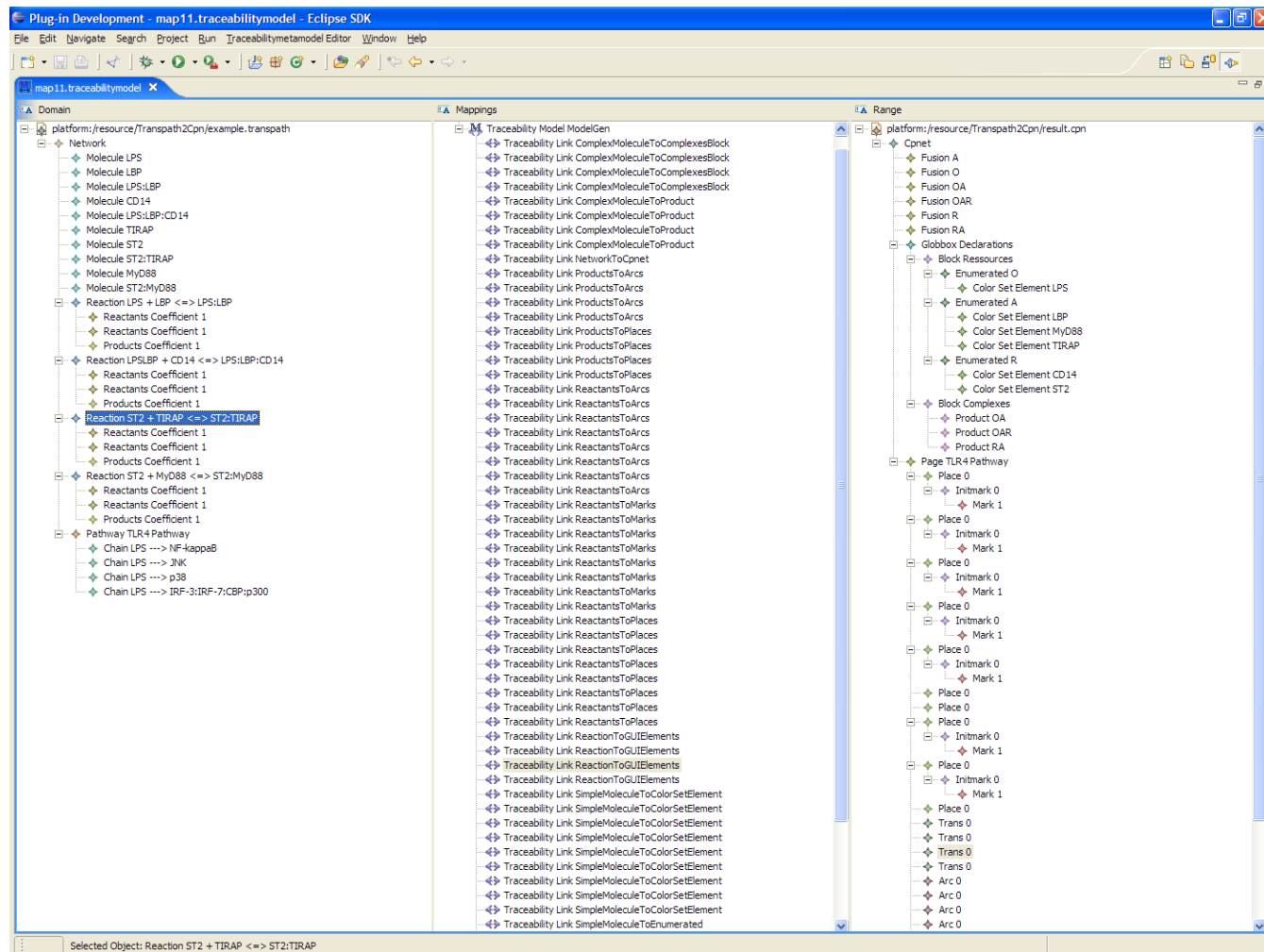


Representation of the case study in CPN Tools.

- TLR4 pathway:
 - $LPS + LBP \rightleftharpoons LPS:LBP$
 - $LPS:LBP + CD14 \rightleftharpoons LPS:LBP:CD14$
 - $ST2 + TIRAP \rightleftharpoons ST2:TIRAP$
 - $ST2 + MyD88 \rightleftharpoons ST2:MyD88$



Execution in MOMENT-QVT.



Conclusions

CONCIOUS



Conclusions and future work



- Conclusions obtained for this small case study:
 - It allows the automation of manual tasks.
 - We can develop more modular tools and these are independent of the data persistence format.
 - It takes advantage of model transformations technologies, allowing us to deal with high level concepts.
 - It provides traceability capabilities implicitly. This permits to correct the detected errors in the source databases.
 - QVT-Relations offers several advantages against imperative approaches given its declarative nature. This language is very expressive when describing the mappings between the source and the target domain.
- In the future we would like to validate this approach by using some other case studies. The model Driven engineering techniques can help with:
 - Building interoperable tools.
 - The implementation of GUIs to represent and manipulate data.
 - Providing standard file formats based in XML.
 - Querying and filtering data, by using model transformations (OCL and pattern matching mechanisms).





Biological Data Migration in Pathway Simulation

Biological Data Migration
in Pathway Simulation

***Abel Gómez**, José Á. Carsí, Artur Boronat, Isidro Ramos*
agomez@dsic.upv.es

Technical University of Valencia.

Departamento de Sistemas Informáticos y Computación.

Software Engineering and Information Systems Research Group

