The Synthetic Biology Open Language (SBOL) Data Model

Chris J. Myers
on behalf of the SBOL community

University of Utah

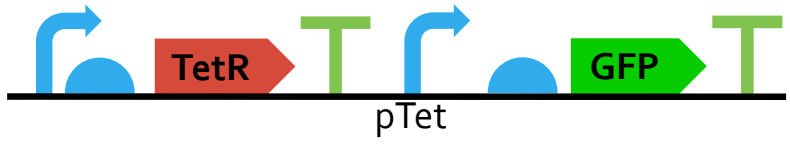
SBOL Workshop
August 15, 2016

Sequence Data Standards

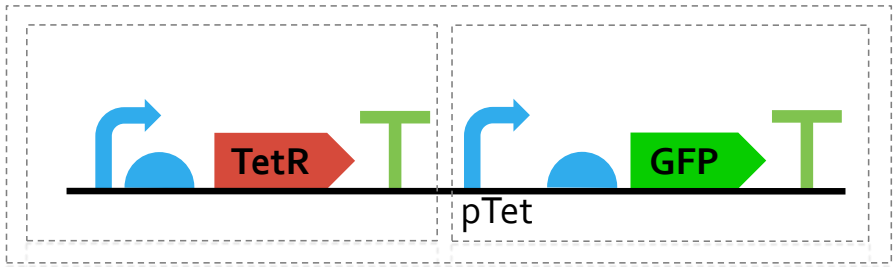
FASTA

ACTGTGCCGTTAAACGTGATTAAATCCGTACTGATAT...

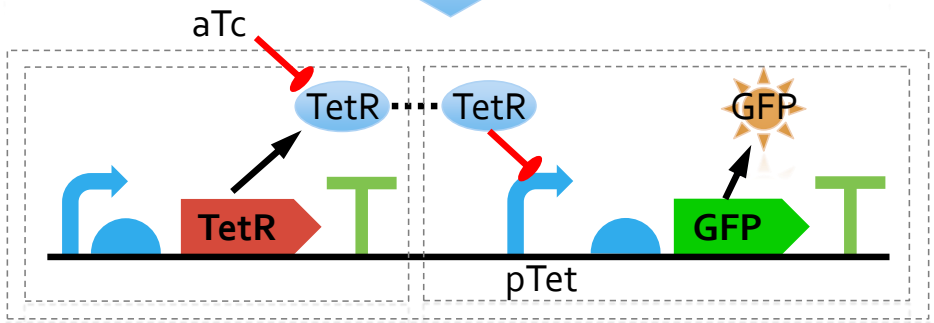
GenBank



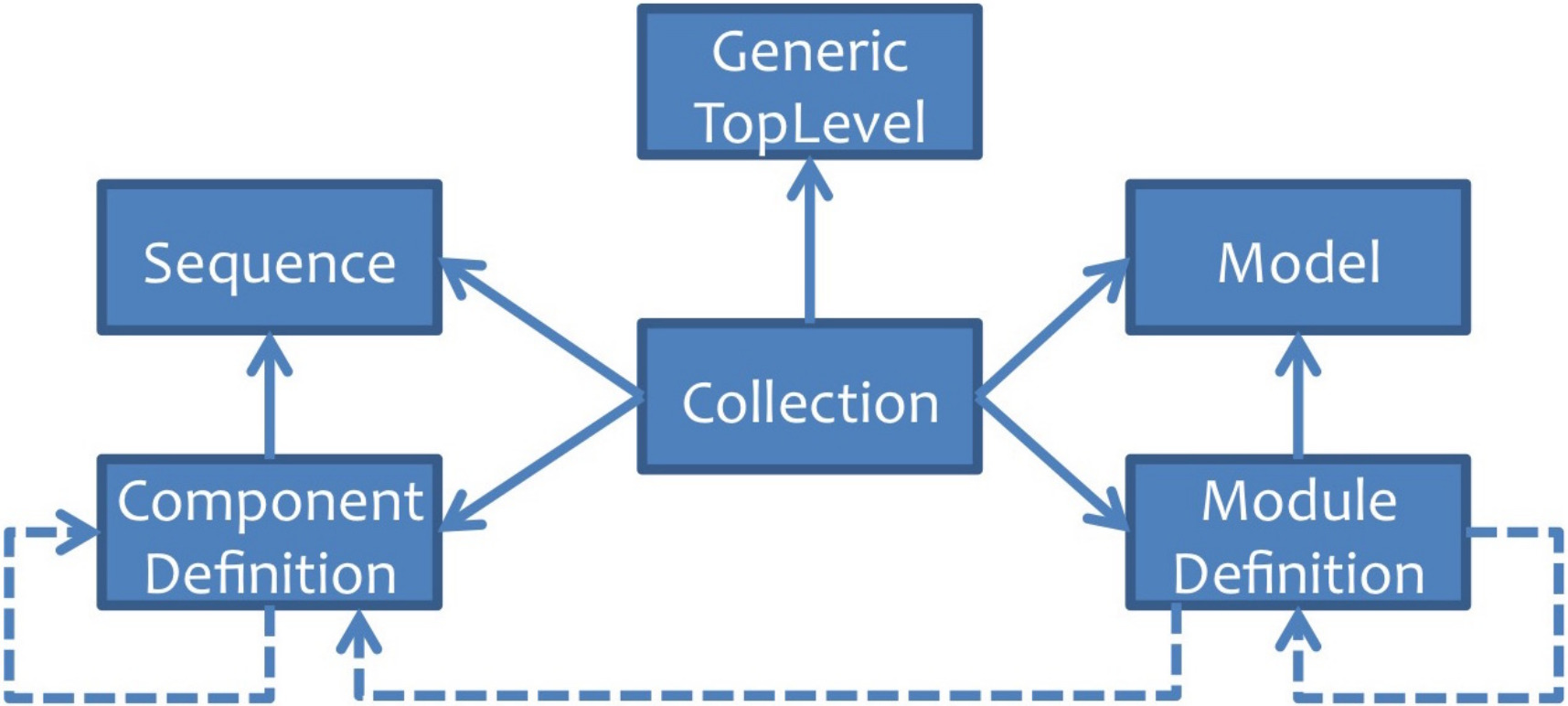
SBOL 1.1



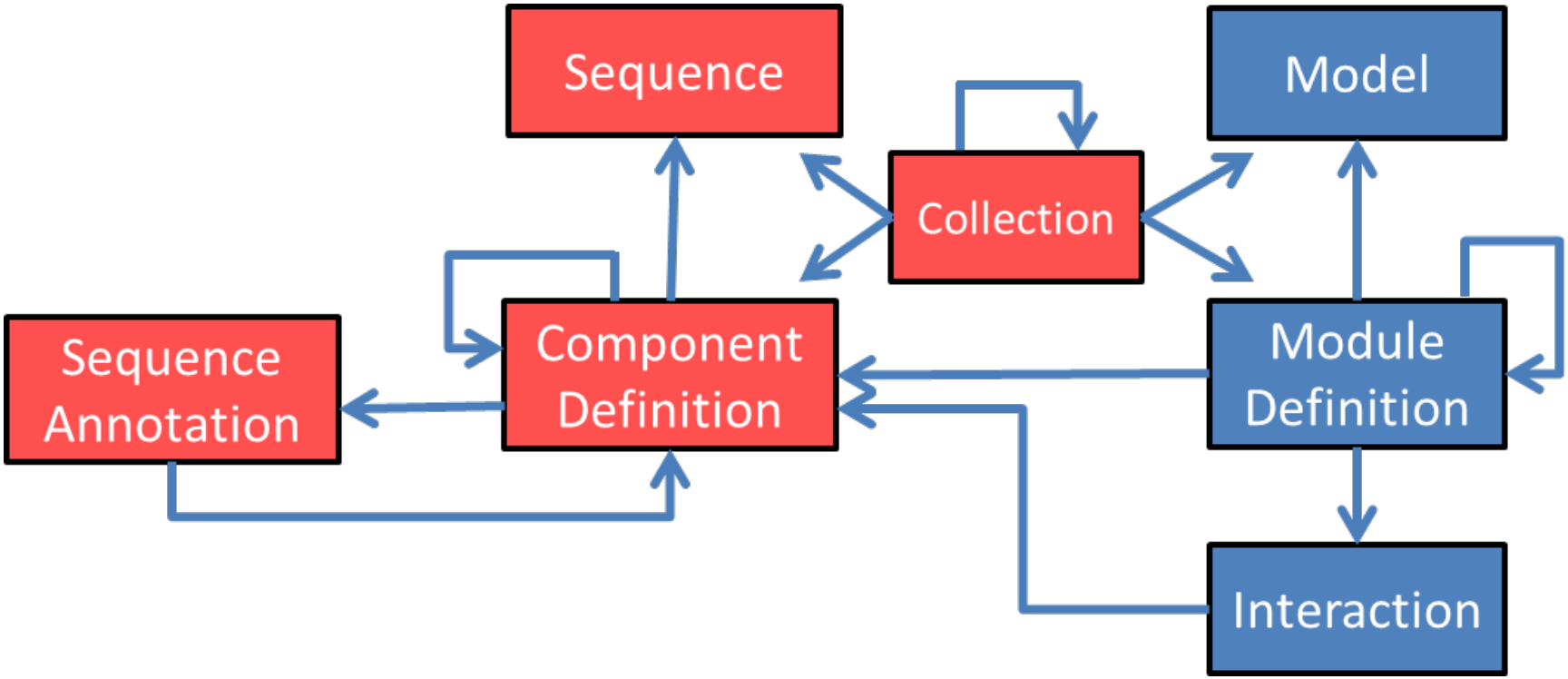
SBOL 2.0



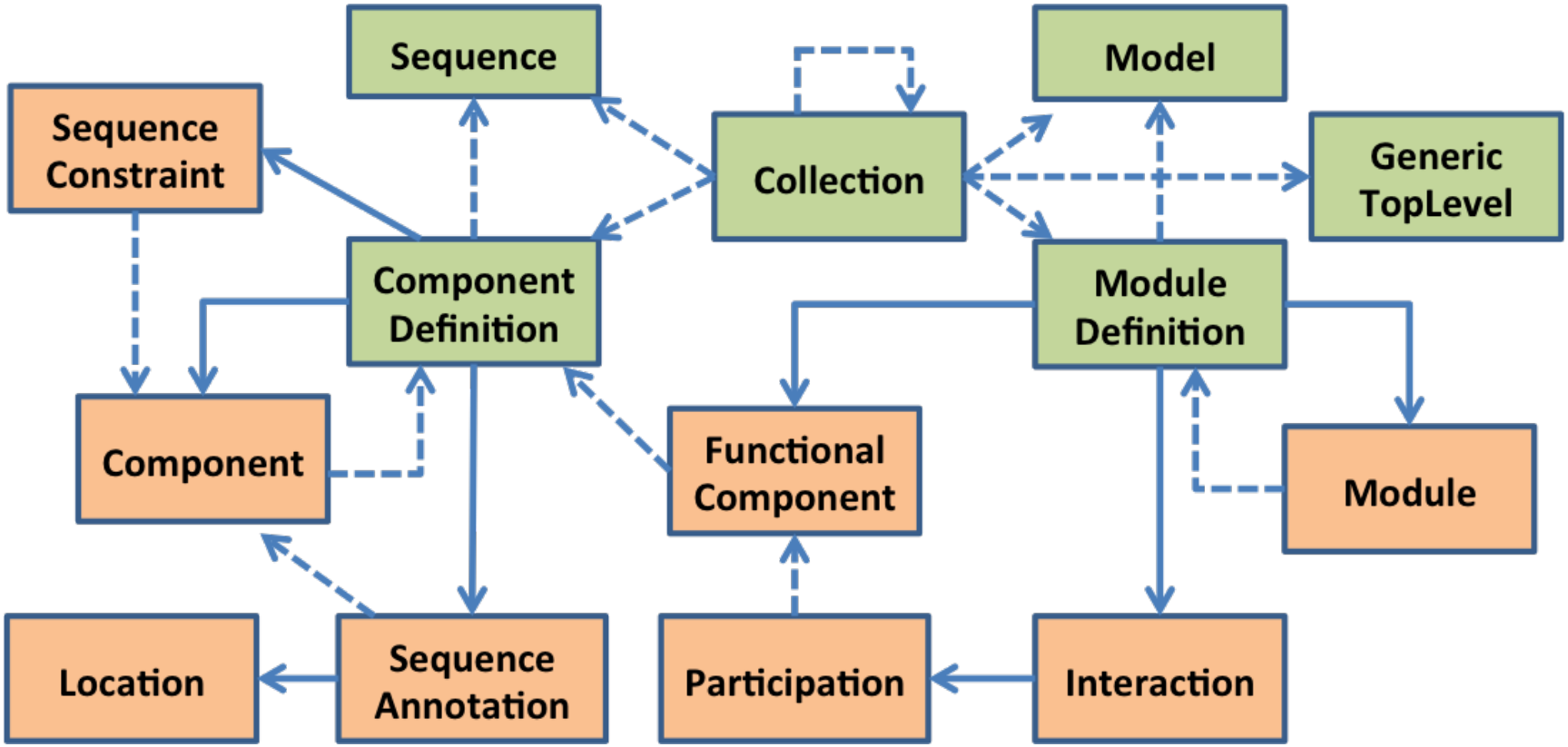
SBOL 2.0 Data Model



SBOL 2.0 Data Model



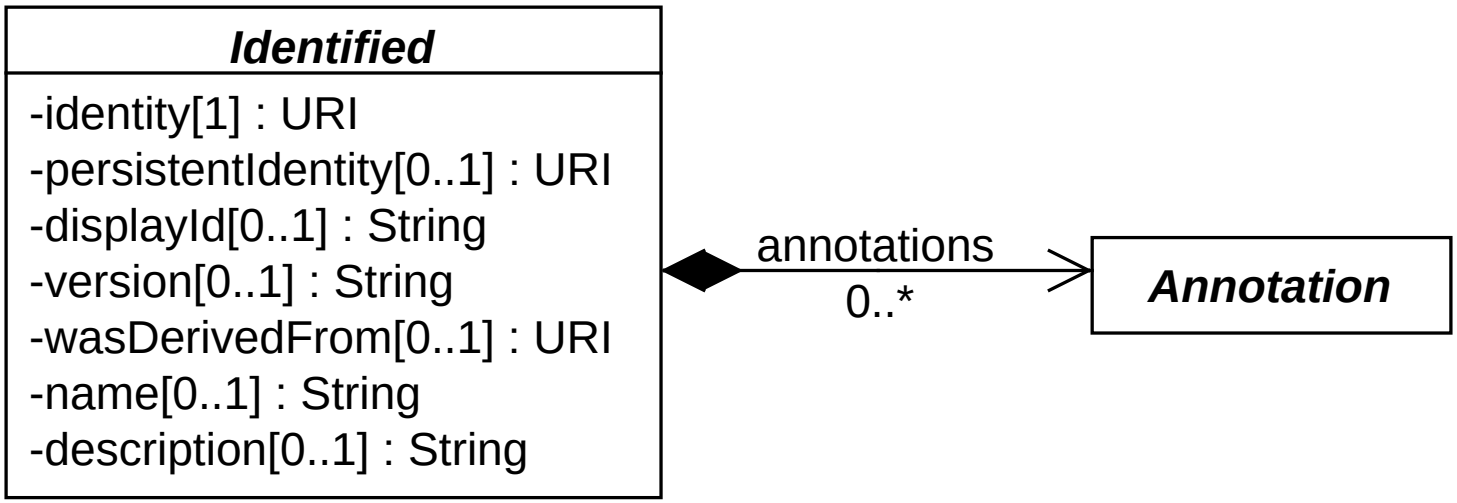
SBOL 2.0 Data Model



Understanding UML Diagrams

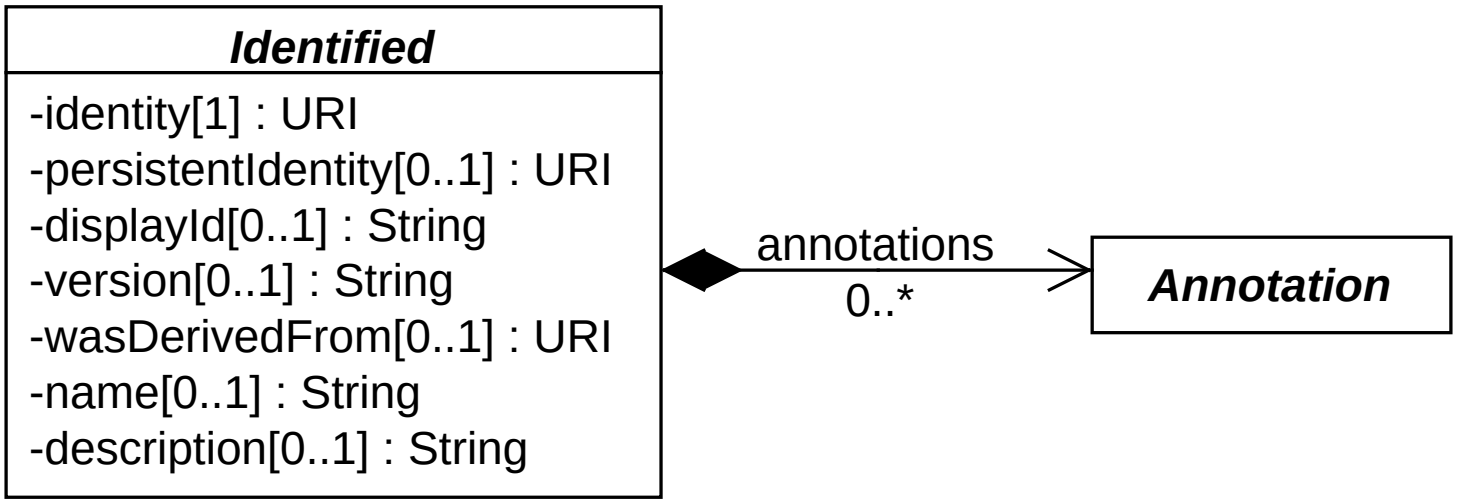
- Classes represented as boxes labeled with their member variables with types and cardinalities.
- Arrows with cardinality indicate associations between classes.
- A hollow diamond at the origin of an arrow represents shared aggregation (i.e., object is referenced and not owned).
- A solid diamond at the origin of an arrow represents composite aggregation (i.e., child object is owned by its parent object).
- Hollow arrows are used to represent inheritance.

Identified (Base Class for All SBOL Objects)



- *identity* - globally unique URI to identify this object (required).
- *persistentIdentity* - identity shared by multiple versions of the same object (optional).
- *displayId* - human-readable id composed of alphanumeric and underscore characters (optional).
- *version* - uses semantic versioning to identify multiple versions of the same object (optional).

Identified (Base Class for All SBOL Objects)



- *wasDerivedFrom* - identity of object that this is derived from (optional).
- *name* - human-readable String of arbitrary characters (optional).
- *description* - thorough text description of the object (optional).
- *annotations* - additional data about this object (more later).

Identified (Example)

```
<?xml version="1.0" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:dcterms="http://purl.org/dc/terms/"
 xmlns:prov="http://www.w3.org/ns/prov#"
 xmlns:sbol="http://sbols.org/v2#">
  <sbol:ComponentDefinition rdf:about="http://www.partsregistry.org/BBa_J23119/2">
    <sbol:persistentIdentity rdf:resource="http://www.partsregistry.org/BBa_J23119"/>
    <sbol:displayId> BBa_J23119 </sbol:displayId>
    <sbol:version> 2 </sbol:version>
    <prov:wasDerivedFrom rdf:resource="http://www.partsregistry.org/BBa_J23119/1"/>
    <dcterms:title> J23119 promoter </dcterms:title>
    <dcterms:description> Constitutive promoter </dcterms:description>
  </sbol:ComponentDefinition>
</rdf:RDF>
```

Compliant Top-Level URIs

$$\langle \textit{URI prefix} \rangle / \langle \textit{displayId} \rangle / \langle \textit{version} \rangle$$

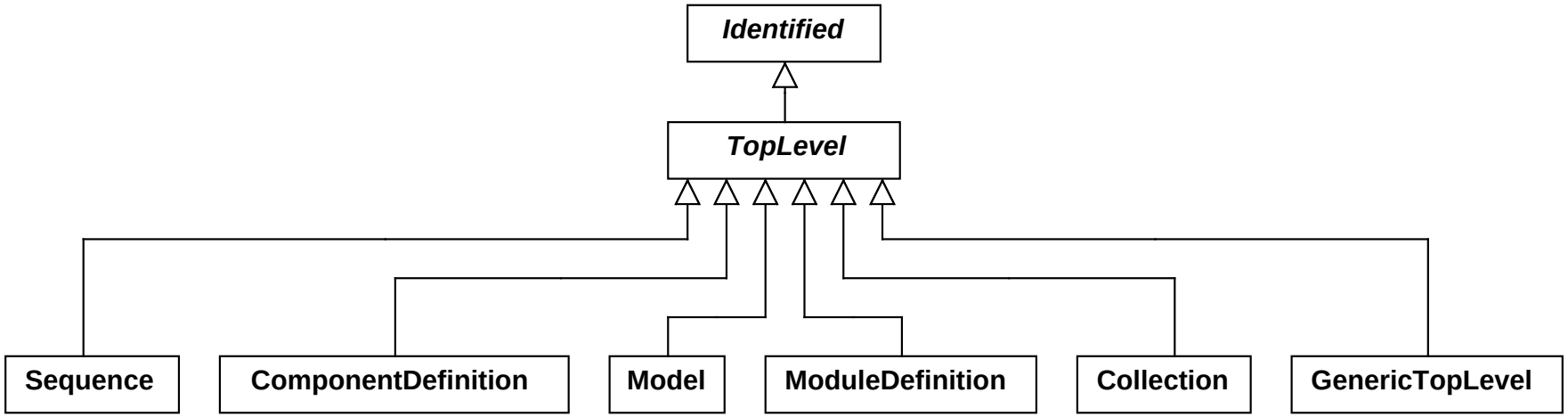
1. The *identity* MUST begin with a *URI prefix* that maps to a domain over which the user has control.
2. The *persistentIdentity* and *displayId* properties are REQUIRED.
3. The *persistentIdentity* MUST end with a delimiter (’/’, ’#’, or ’:’) followed by the *displayId* of the object.
4. If an object is not given a *version*, then its *identity* and *persistentIdentity* properties MUST contain the same *URI*.
5. If an object has a *version*, then its *identity* property MUST contain a *URI* of the form $\langle \textit{persistentIdentity} \rangle / \langle \textit{version} \rangle$".

Compliant Child URIs

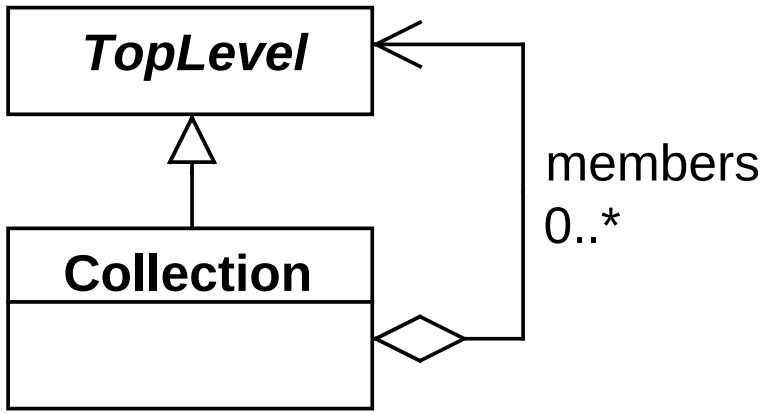
$$\langle parent\ persistentIdentity \rangle / \langle displayId \rangle / \langle parent\ version \rangle$$

1. The *persistentIdentity* MUST begin with the *persistentIdentity* of its parent object and be immediately followed by a delimiter (’/’, ’#’, or ’:’) and the *displayId* of the object.
2. The *version* MUST contain the same *String* as the *version* property of the object’s parent object.

TopLevel


Identified
TopLevel
Sequence
ComponentDefinition
Model
ModuleDefinition
Collection
GenericTopLevel

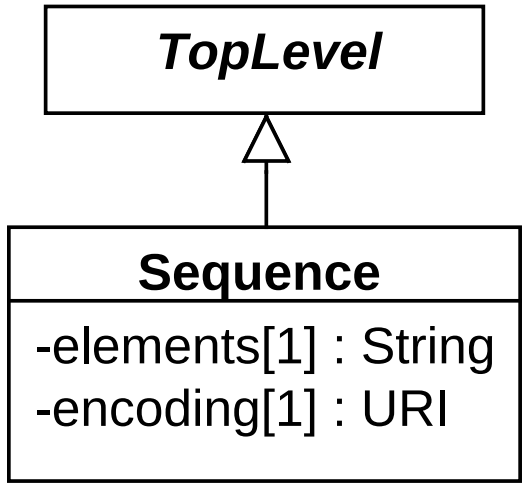
Collection



Collection (Example)

```
<?xml version="1.0" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:dcterms="http://purl.org/dc/terms/"
 xmlns:prov="http://www.w3.org/ns/prov#"
 xmlns:sbol="http://sbols.org/v2#">
  <sbol:Collection rdf:about="http://parts.igem.org/Promoters/Catalog/Anderson">
    <sbol:persistentIdentity rdf:resource="http://parts.igem.org/Promoters/Catalog/Anderson"/>
    <sbol:displayId> Anderson </sbol:displayId>
    <dcterms:title> Anderson promoters </dcterms:title>
    <dcterms:description> The Anderson promoter collection </dcterms:description>
    <sbol:member rdf:resource="http://partsregistry.org/Part:BBa_J23119"/>
    ...
    <sbol:member rdf:resource="http://partsregistry.org/Part:BBa_J23118"/>
  </sbol:Collection>
</rdf:RDF>
```

Sequence



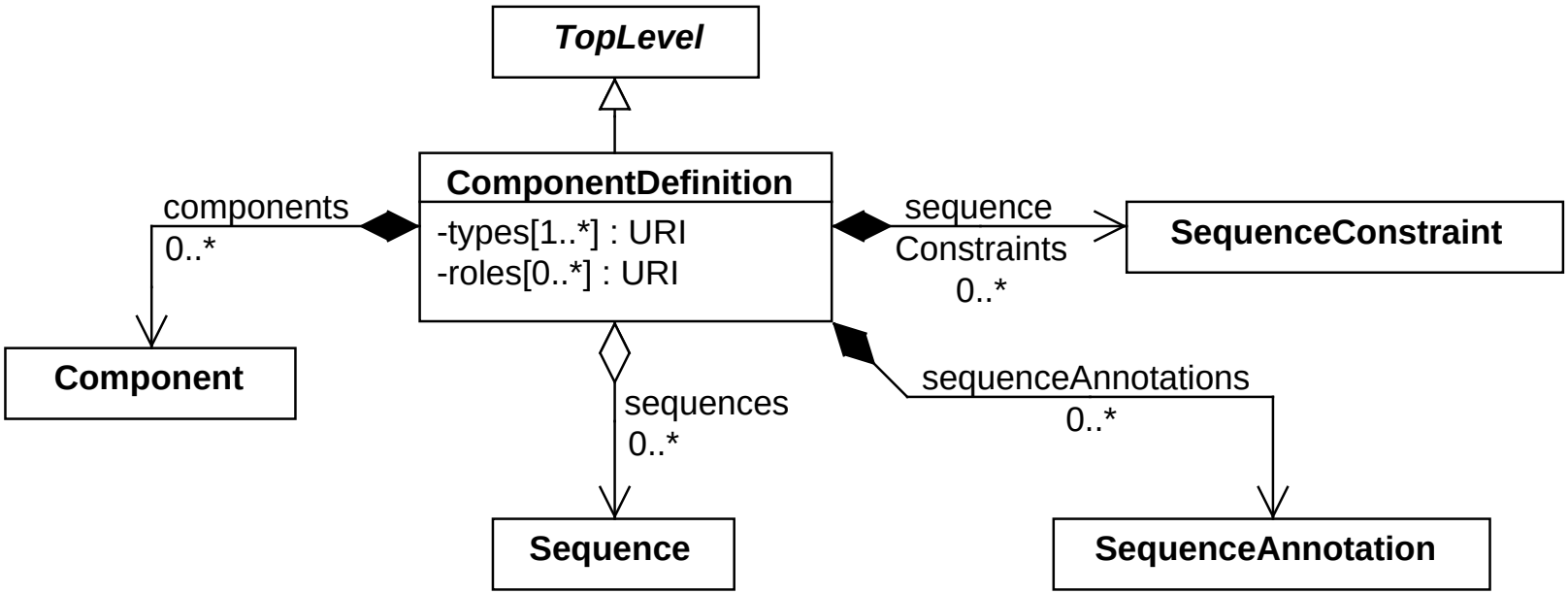
- *elements* - String of characters representing constituents of a biological or chemical molecule.
- *encoding* - URI indicating how elements are to be interpreted.

Encoding	**URI**	**CD Type**
IUPAC DNA, RNA	`http://www.chem.qmul.ac.uk/iubmb/misc/naseq.html`	DNA, RNA
IUPAC Protein	`http://www.chem.qmul.ac.uk/iupac/AminoAcid/`	Protein
SMILES	`http://www.opensmiles.org/opensmiles.html`	SmallMolecule

Sequence (Example)

```
<?xml version="1.0" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:dcterms="http://purl.org/dc/terms/"
 xmlns:prov="http://www.w3.org/ns/prov#"
 xmlns:sbol="http://sbols.org/v2#">
  <sbol:Sequence rdf:about="http://partsregistry.org/seq/BBa_J23119">
    <sbol:persistentIdentity rdf:resource="http://partsregistry.org/seq/BBa_J23119"/>
    <sbol:displayId> BBa_J23119 </sbol:displayId>
    <prov:wasDerivedFrom rdf:resource="http://parts.igem.org/Part:BBa_J23119:Design"/>
    <sbol:elements> ttgacagctagctcagtcctaggtataatgctagc </sbol:elements>
    <sbol:encoding rdf:resource="http://www.chem.qmul.ac.uk/iubmb/misc/naseq.html"/>
  </sbol:Sequence>
</rdf:RDF>
```

ComponentDefinition



ComponentDefinition (Types/Roles)

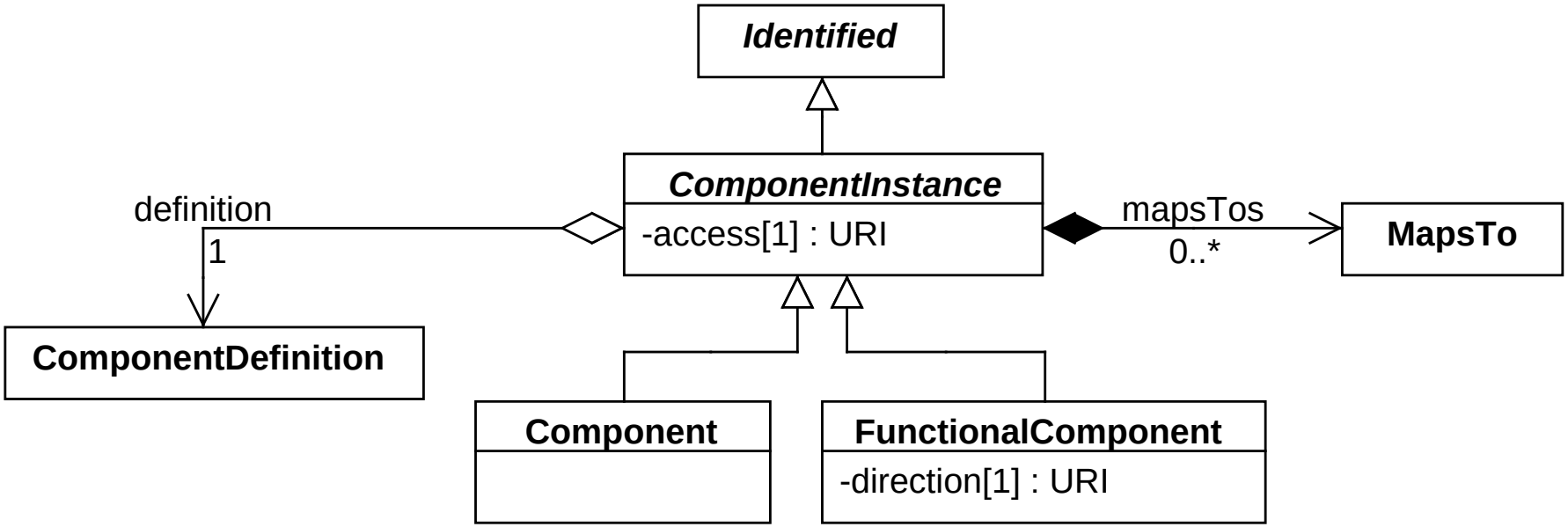
CD Type	URI for BioPAX Term
DNA	http://www.biopax.org/release/biopax-level3.owl#DnaRegion
RNA	http://www.biopax.org/release/biopax-level3.owl#RnaRegion
Protein	http://www.biopax.org/release/biopax-level3.owl#Protein
Small Molecule	http://www.biopax.org/release/biopax-level3.owl#SmallMolecule
Complex	http://www.biopax.org/release/biopax-level3.owl#Complex

CD Role	URI for SequenceOntology Term	CD Type
Promoter	http://identifiers.org/so/SO:0000167	DNA
RBS	http://identifiers.org/so/SO:0000139	DNA
CDS	http://identifiers.org/so/SO:0000316	DNA
Terminator	http://identifiers.org/so/SO:0000141	DNA
Gene	http://identifiers.org/so/SO:0000704	DNA
Operator	http://identifiers.org/so/SO:0000057	DNA
Engineered Gene	http://identifiers.org/so/SO:0000280	DNA
mRNA	http://identifiers.org/so/SO:0000234	RNA

ComponentDefinition (Example)

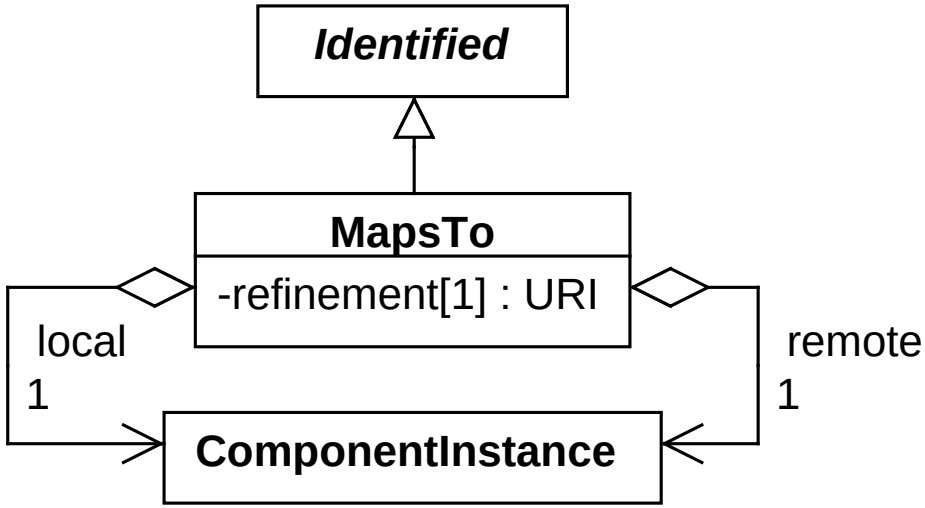
```
<?xml version="1.0" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:dcterms="http://purl.org/dc/terms/"
 xmlns:prov="http://www.w3.org/ns/prov#"
 xmlns:sbol="http://sbols.org/v2#">
  <sbol:ComponentDefinition rdf:about="http://partsregistry.org/cd/BBa_J23119">
    <sbol:persistentIdentity rdf:resource="http://partsregistry.org/cd/BBa_J23119"/>
    <sbol:displayId> BBa_J23119 </sbol:displayId>
    <dcterms:title> J23119 promoter </dcterms:title>
    <dcterms:description> Constitutive promoter </dcterms:description>
    <sbol:type rdf:resource="http://www.biopax.org/release/biopax-level3.owl#DnaRegion"/>
    <sbol:role rdf:resource="http://identifiers.org/so/SO:0000167"/>
    <sbol:sequence rdf:resource="http://partsregistry.org/seq/BBa_J23119"/>
  </sbol:ComponentDefinition>
</rdf:RDF>
```

Component (Child of ComponentDefinition)



Access URI	**Description**
`http://sbols.org/v2#public`	MAY be referred to by remote **MapsTo** objects.
`http://sbols.org/v2#private`	MUST NOT be referred to by remote **MapsTo** objects.

MapsTo (Child of ComponentInstance)



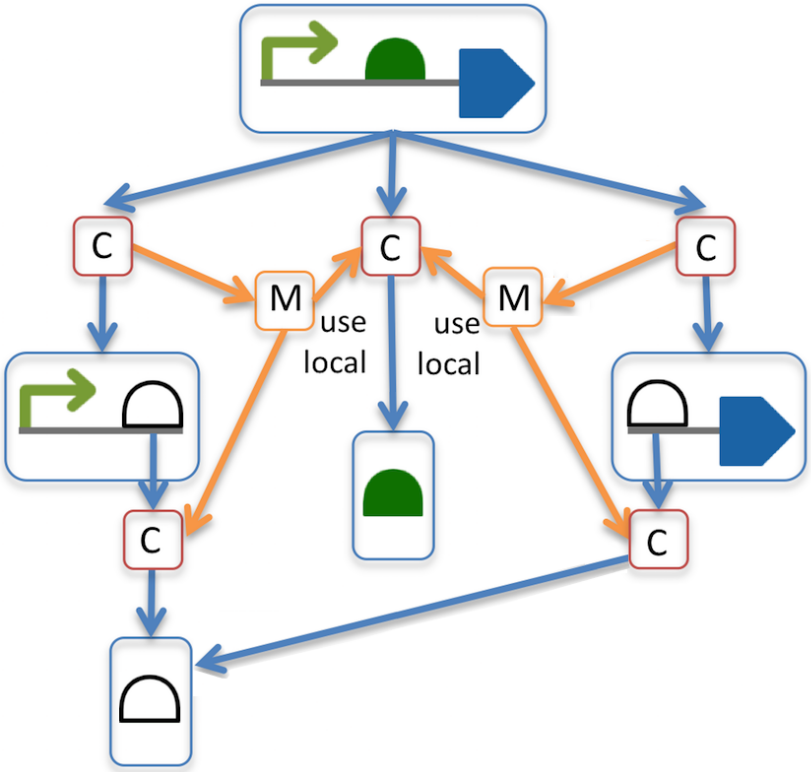
Refinement URI

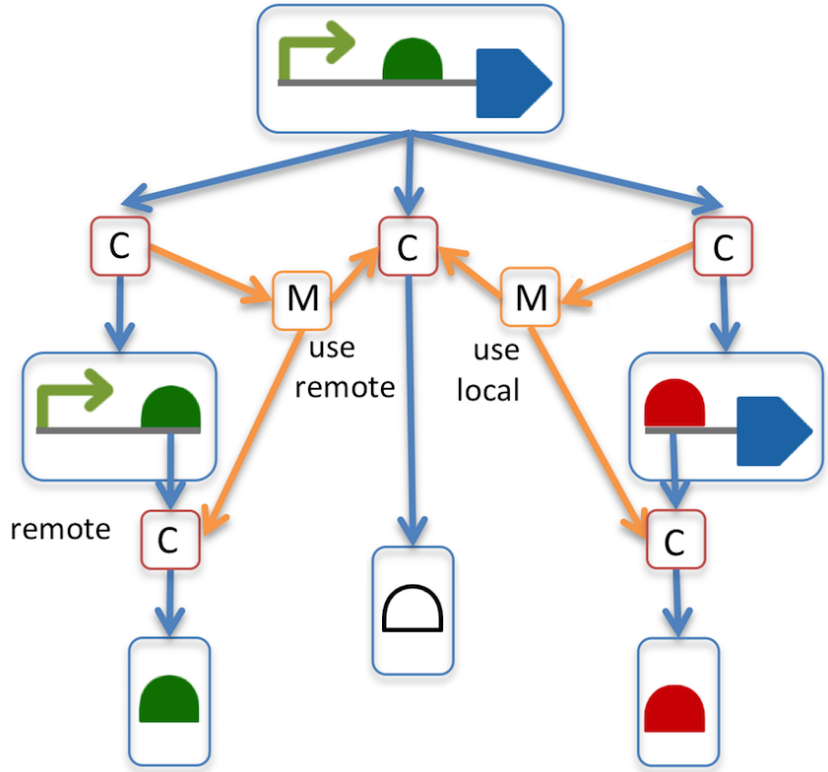
`http://sbols.org/v2#useRemote`
`http://sbols.org/v2#useLocal`
`http://sbols.org/v2#verifyIdentical`
`http://sbols.org/v2#merge`

Description

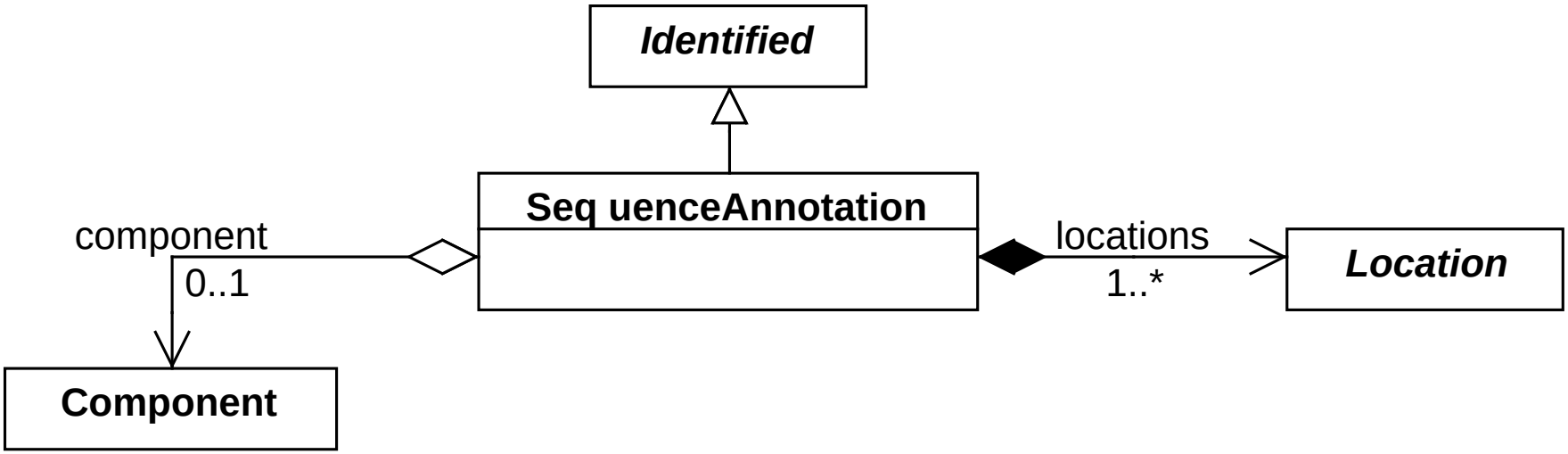
All references MUST dereference to the *remote* **CI**.
All references MUST dereference to the *local* **CI**.
The *definition* of the *local* and *remote* MUST be same **CD**.
All references MUST dereference to both objects.

MapsTo (Example)

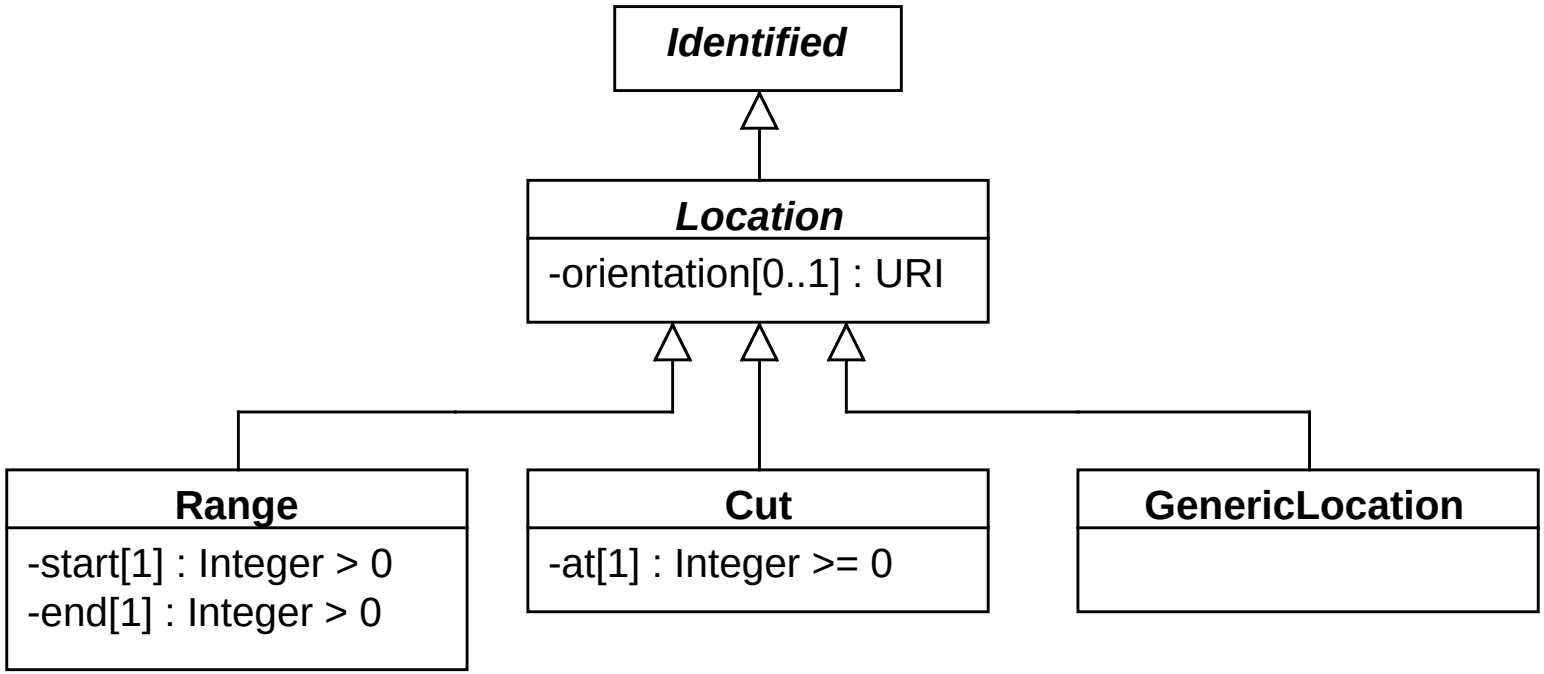




SequenceAnnotation (Child of ComponentDefinition)



Location (Child of SequenceAnnotation)



Orientation URI

`http://sbols.org/v2#inline`

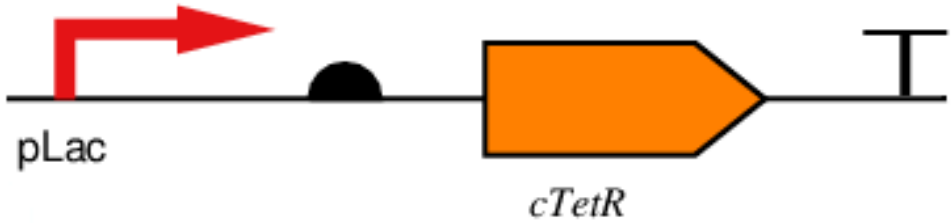
`http://sbols.org/v2#reverseComplement`

Description

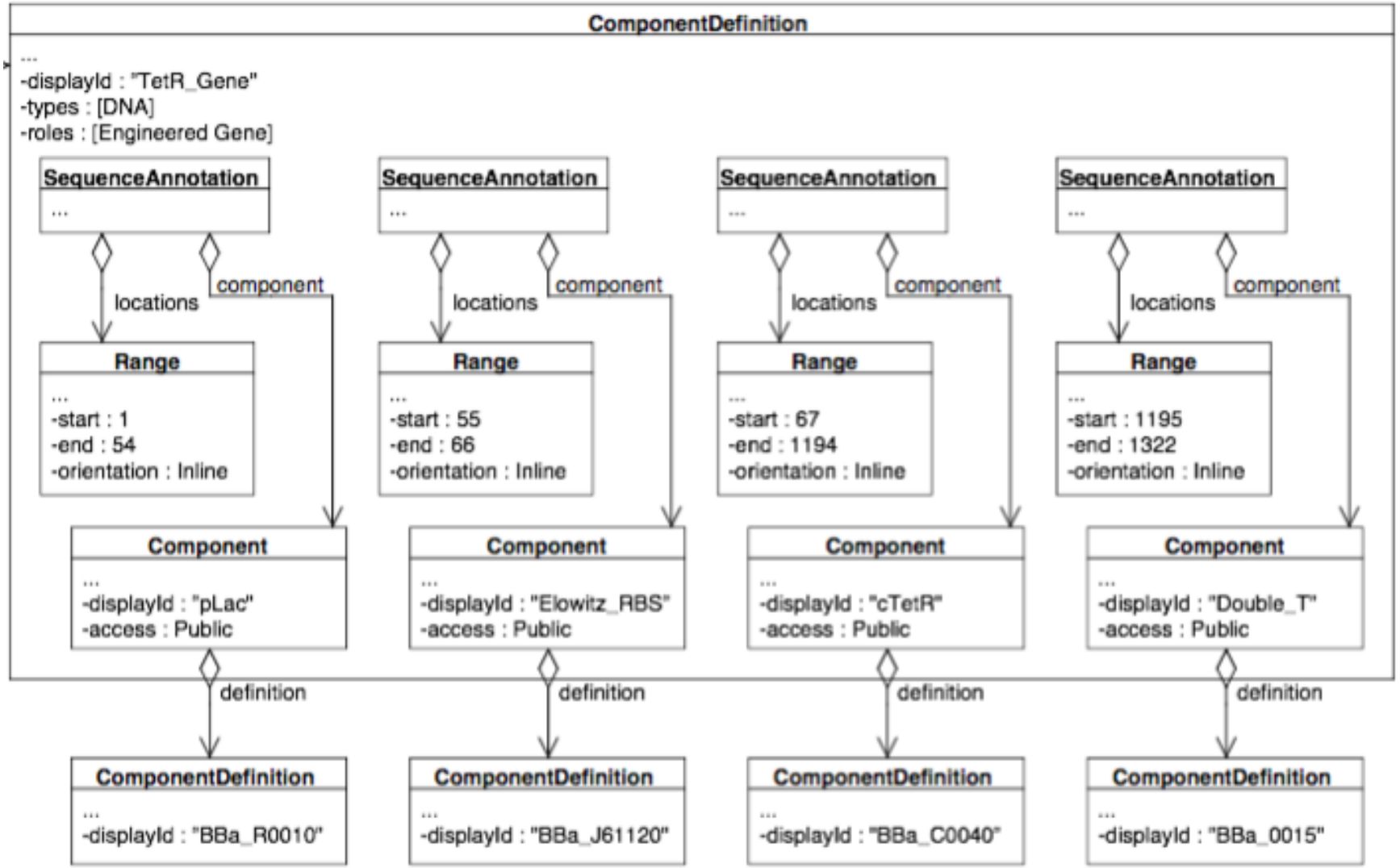
The region is inline with the sequence.

The region is on the reverse-complement translation.

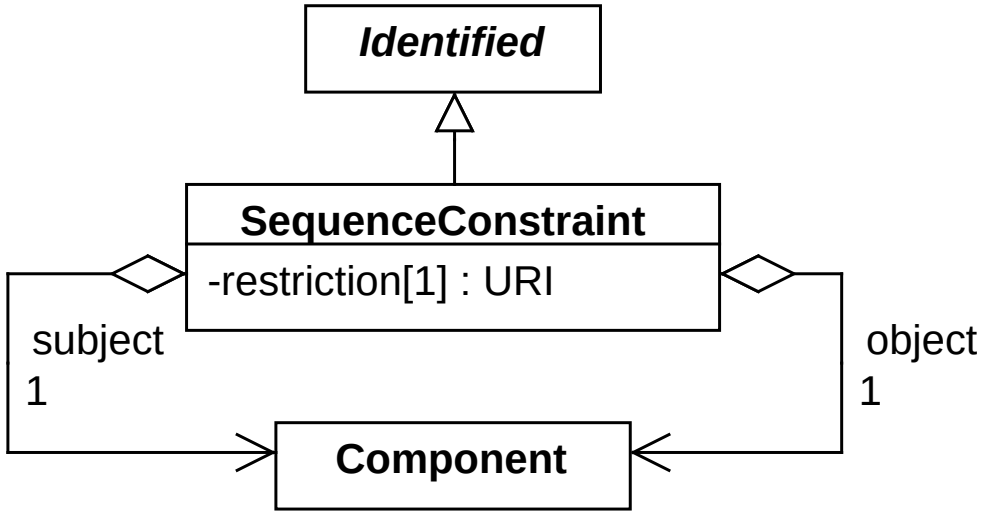
LacI Inverter



LacI Inverter (SBOL)



SequenceConstraint (Child of ComponentDefinition)



Restriction URI

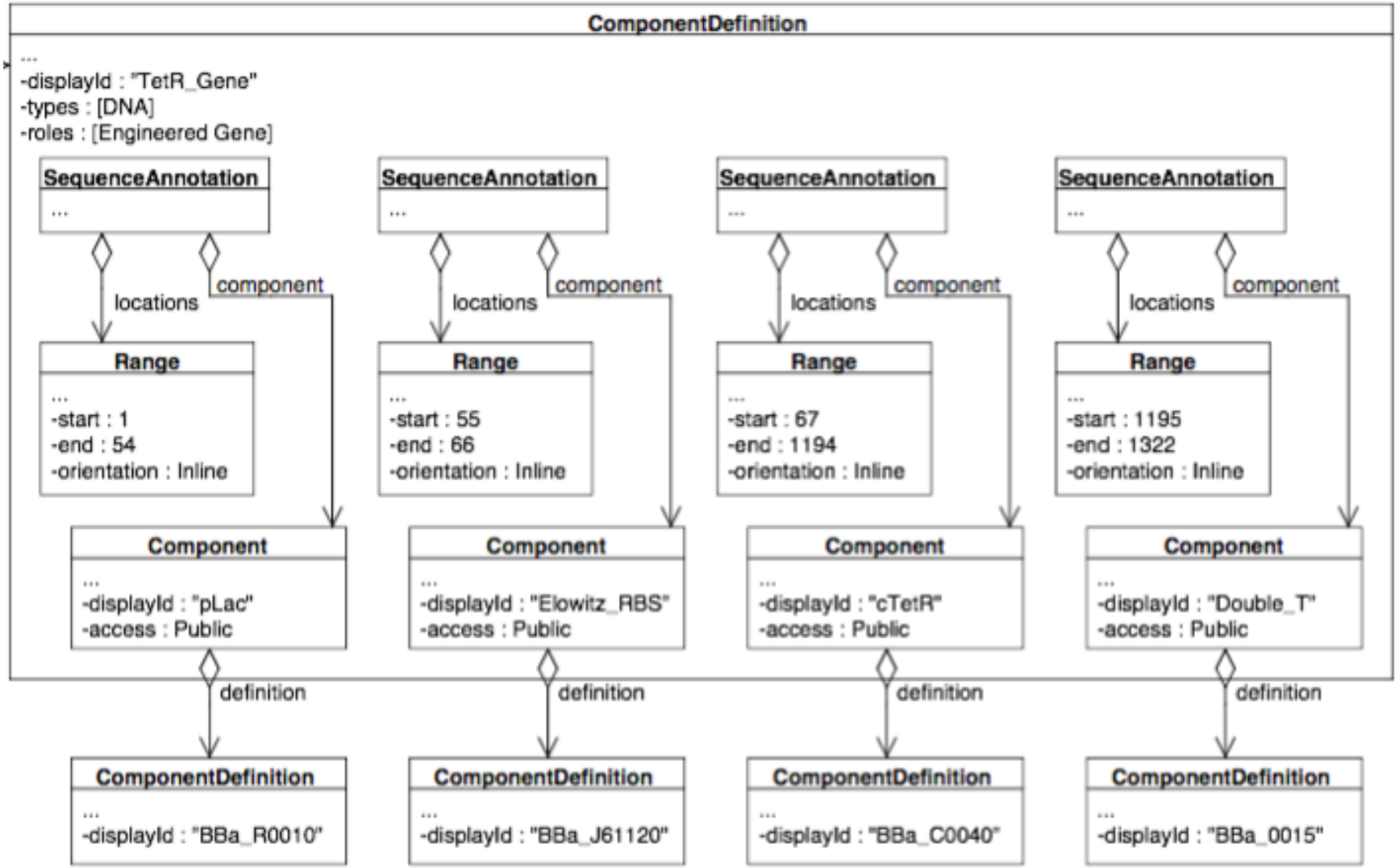
`http://sbols.org/v2#precedes`
`http://sbols.org/v2#sameOrientationAs`
`http://sbols.org/v2#oppositeOrientationAs`

Description

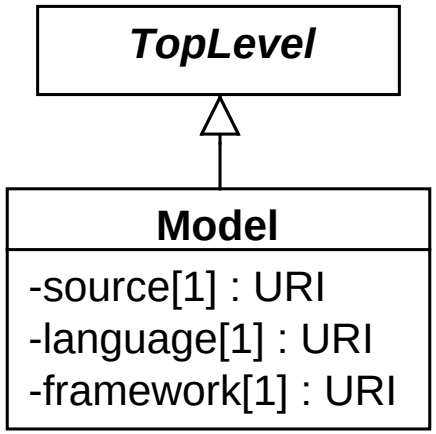
subject MUST precede *object* **Component**.
subject & *object* MUST have same orientation.
subject & *object* MUST have opposite orientations.

SequenceConstraint (Question)

QUESTION: Redrew the UML diagram below using SequenceConstraints rather than SequenceAnnotations.



Model



- *source* - URI reference to the source file for the model.
- *language* - URI that species language in which the model is implemented.

Model Language	URI for EDAM Term
SBML	`http://identifiers.org/edam/format_2585`
CellML	`http://identifiers.org/edam/format_3240`
BioPAX	`http://identifiers.org/edam/format_3156`

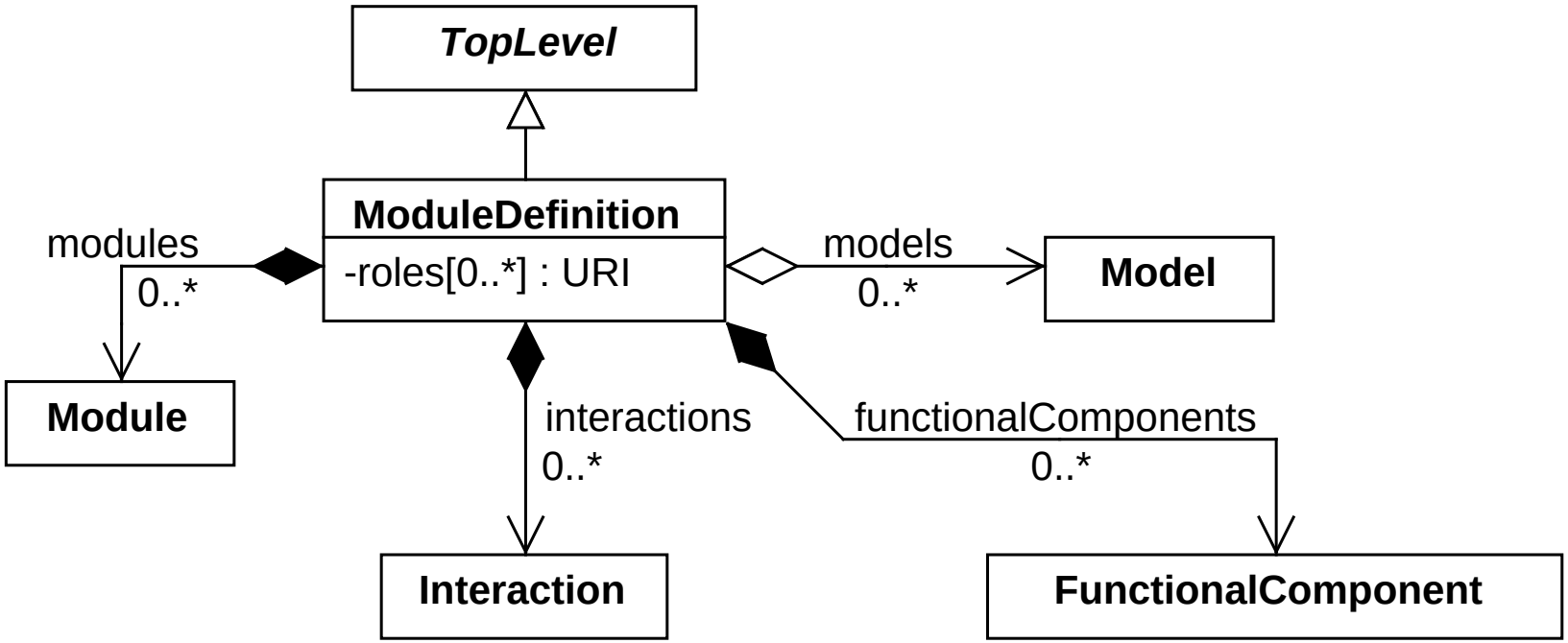
- *framework* - URI that species modeling framework used.

Framework	URI for SBO Term
Continuous	`http://identifiers.org/biomodels.sbo/SBO:0000062`
Discrete	`http://identifiers.org/biomodels.sbo/SBO:0000063`

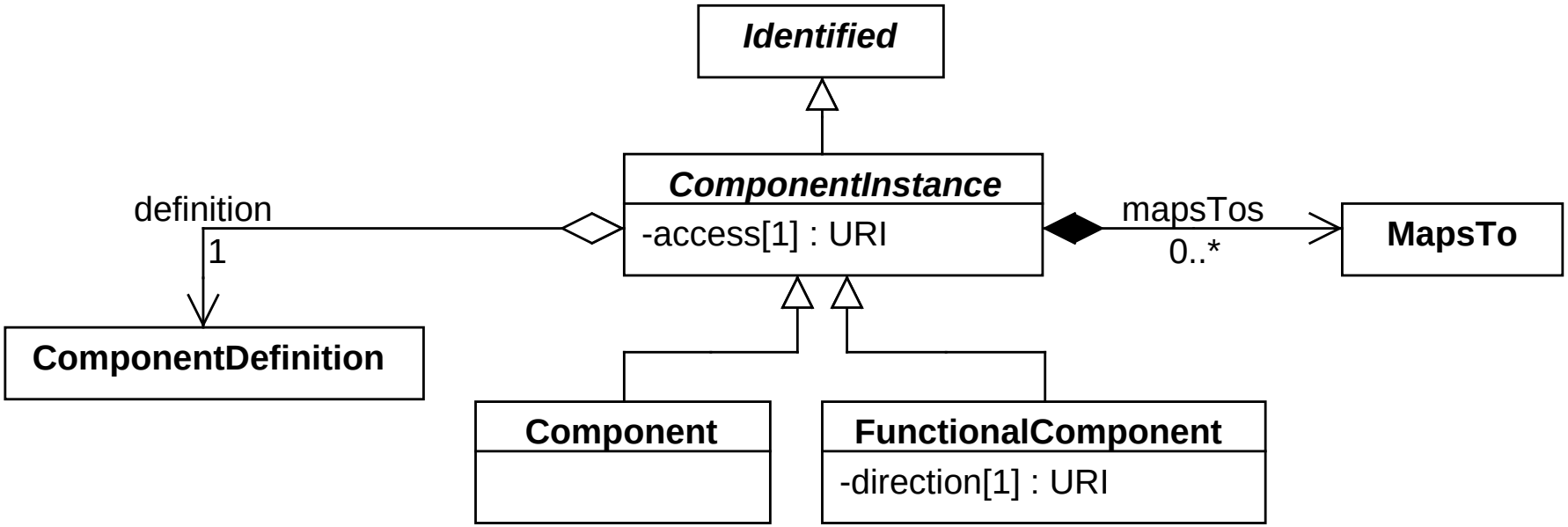
Model (Example)

```
<?xml version="1.0" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:dcterms="http://purl.org/dc/terms/"
 xmlns:prov="http://www.w3.org/ns/prov#"
 xmlns:sbol="http://sbols.org/v2#">
  <sbol:Model rdf:about="http://www.sbolstandard.org/examples/pIKE_Toggle_1">
    <sbol:persistentIdentity rdf:resource="http://www.sbolstandard.org/examples/pIKE_Toggle_1"/>
    <sbol:displayId> pIKE_Toggle_1 </sbol:displayId>
    <dcterms:title> pIKE_Toggle_1 toggle switch </dcterms:title>
    <sbol:source rdf:resource="http://virtualparts.org/part/pIKE_Toggle_1"/>
    <sbol:language rdf:resource="http://identifiers.org/edam/format_2585"/>
    <sbol:framework rdf:resource="http://identifiers.org/biomodels.sbo/SBO:0000062"/>
  </sbol:Model>
</rdf:RDF>
```

ModuleDefinition

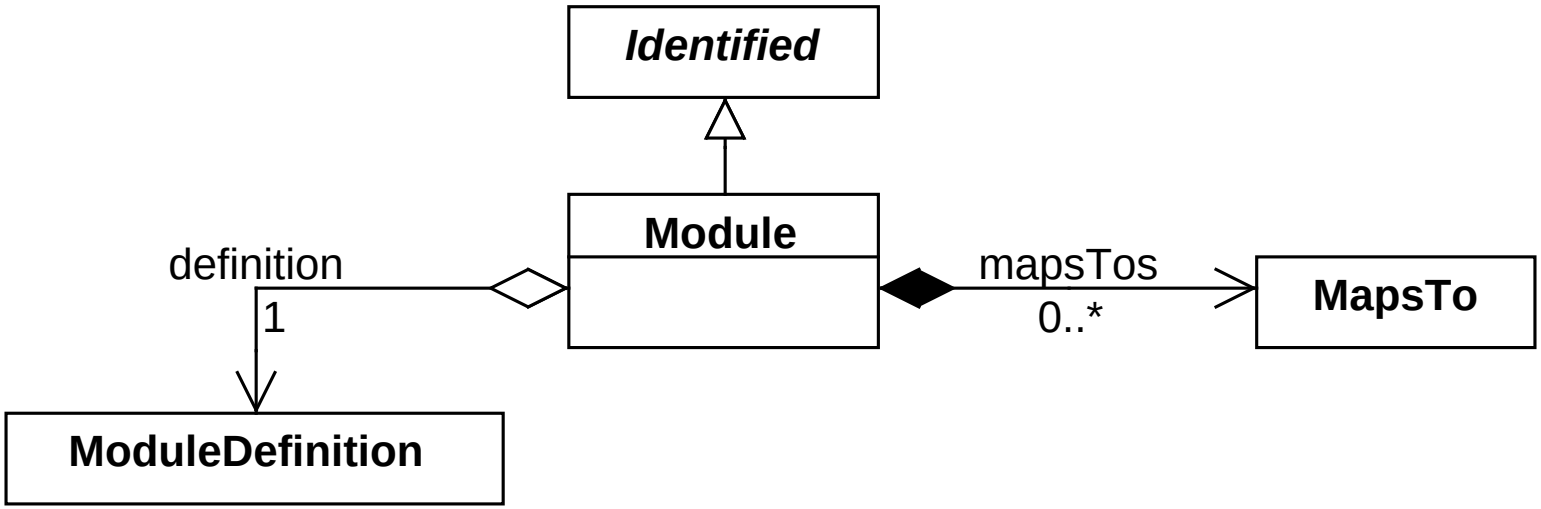


FunctionalComponent (Child of ModuleDefinition)

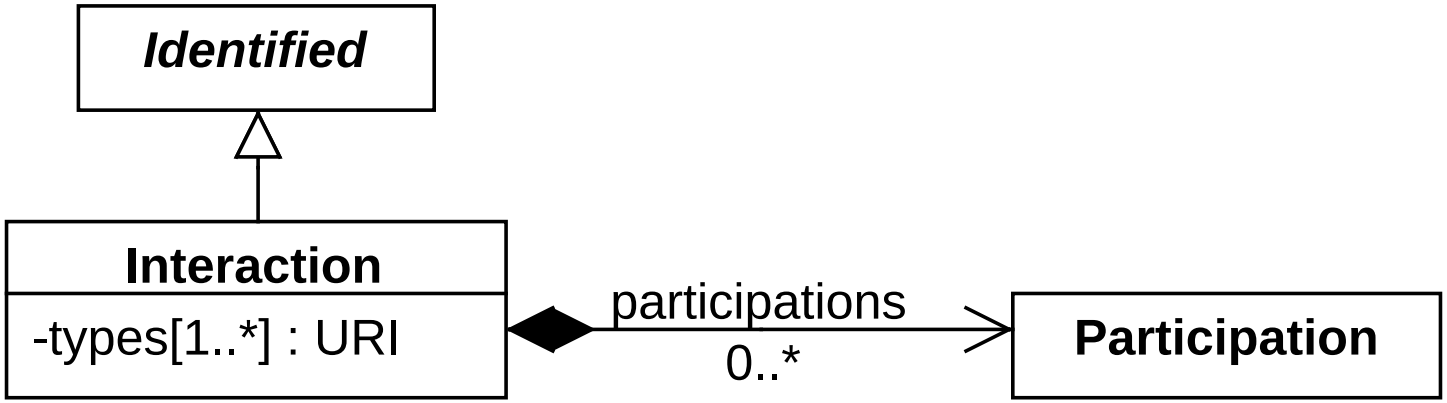


Direction URI	Description
`http://sbols.org/v2#in`	Indicates that it is an input.
`http://sbols.org/v2#out`	Indicates that it is an output.
`http://sbols.org/v2#inout`	Indicates that it is both an input and output
`http://sbols.org/v2#none`	Indicates that it is neither an input nor output.

Module (Child of ModuleDefinition)

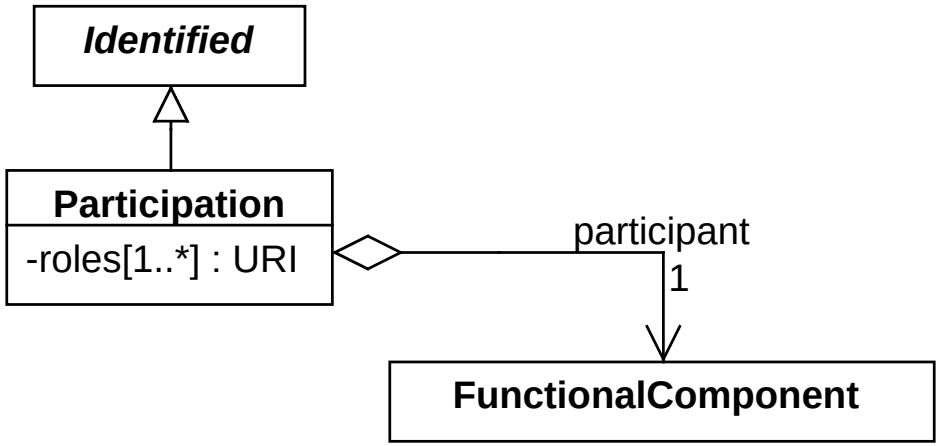


Interaction (Child of ModuleDefinition)



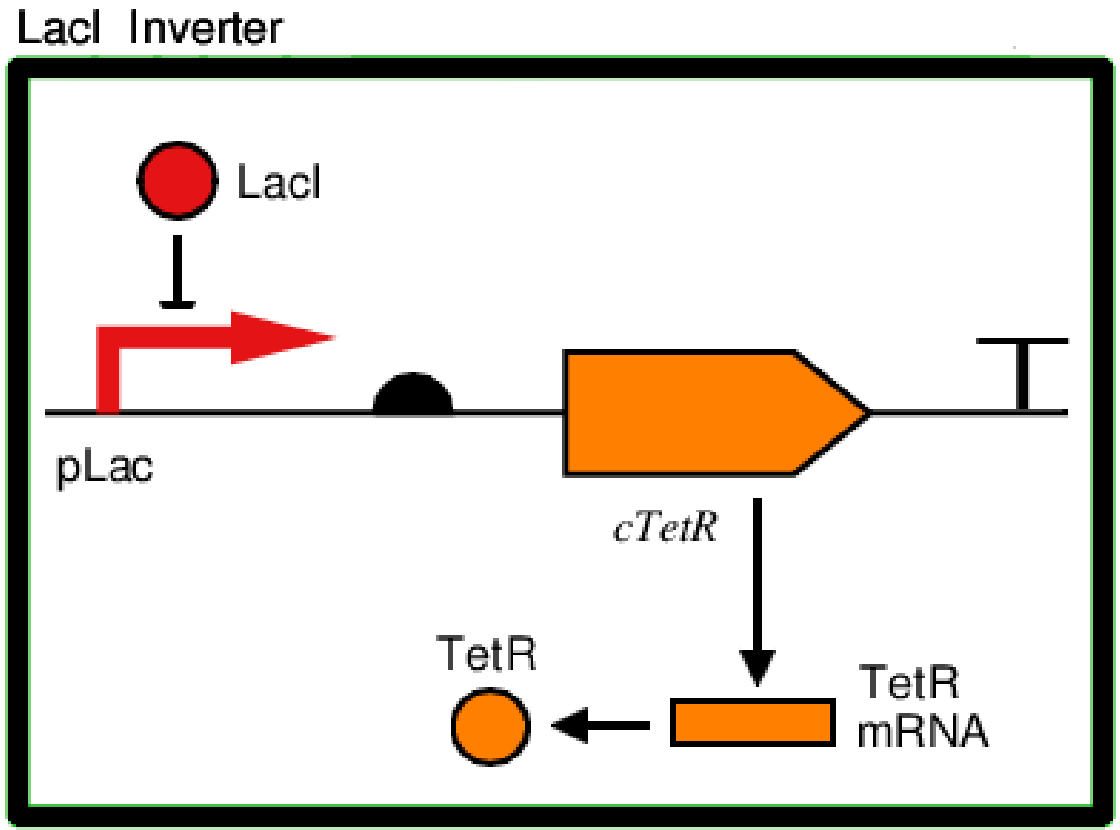
Interaction Type	**URI for SystemsBiologyOntology Term**
Inhibition	`http://identifiers.org/biomodels.sbo/SBO:0000169`
Stimulation	`http://identifiers.org/biomodels.sbo/SBO:0000170`
Biochemical Reaction	`http://identifiers.org/biomodels.sbo/SBO:0000176`
Non-Covalent Binding	`http://identifiers.org/biomodels.sbo/SBO:0000177`
Degradation	`http://identifiers.org/biomodels.sbo/SBO:0000179`
Genetic Production	`http://identifiers.org/biomodels.sbo/SBO:0000589`

Participation (Child of Interaction)

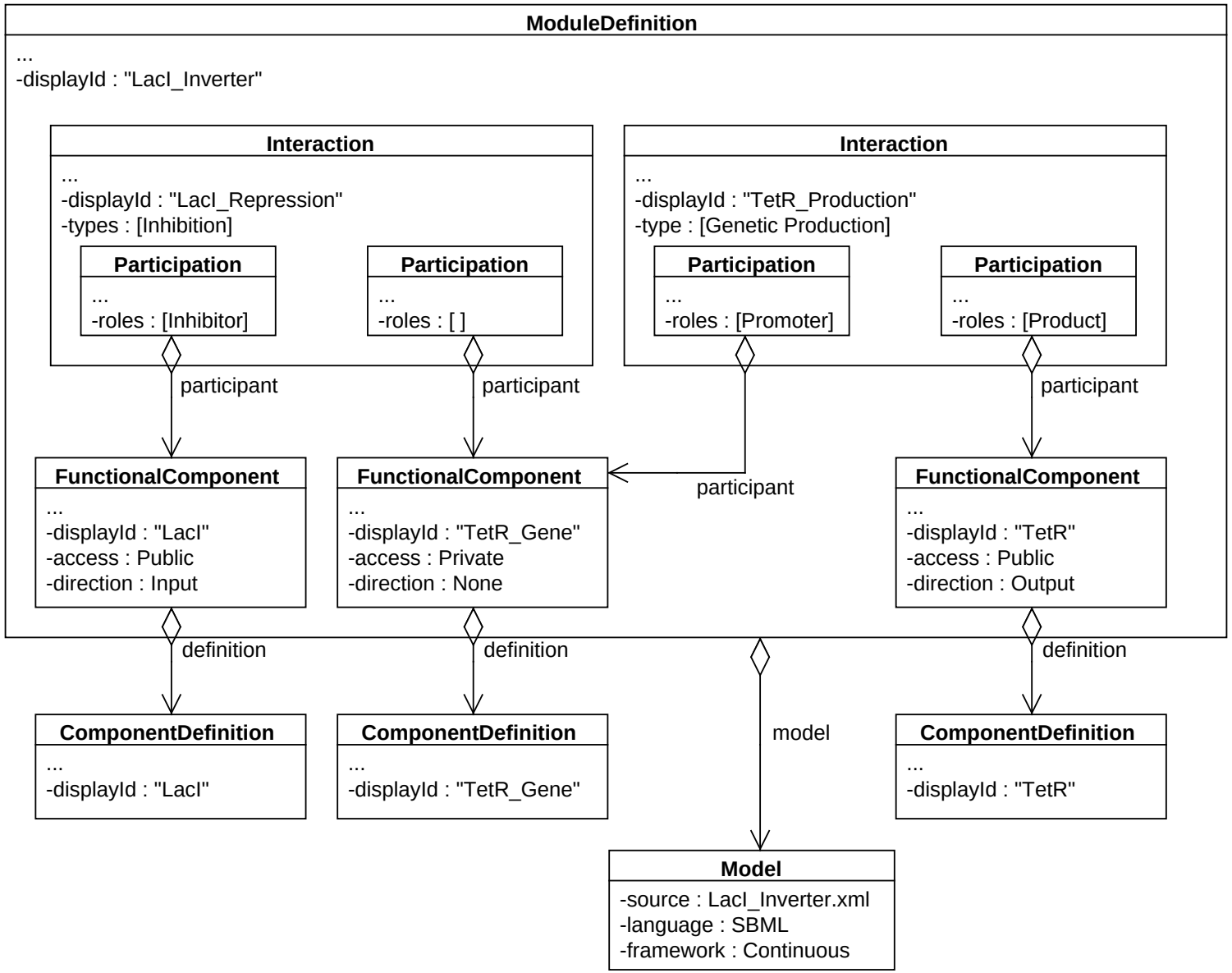


Part. Role	URI for SBO Term	Interaction Types
Inhibitor	http://identifiers.org/biomodels.sbo/SBO:0000020	Inhibition
Stimulator	http://identifiers.org/biomodels.sbo/SBO:0000459	Stimulation
Reactant	http://identifiers.org/biomodels.sbo/SBO:0000010	Non-Covalent Binding, Degradation Biochemical Reaction
Product	http://identifiers.org/biomodels.sbo/SBO:0000011	Non-Covalent Binding, Genetic Production, Biochemical Reaction
Promoter	http://identifiers.org/biomodels.sbo/SBO:0000598	Inhibition, Stimulation, Genetic Production
Modifier	http://identifiers.org/biomodels.sbo/SBO:0000019	Biochemical Reaction

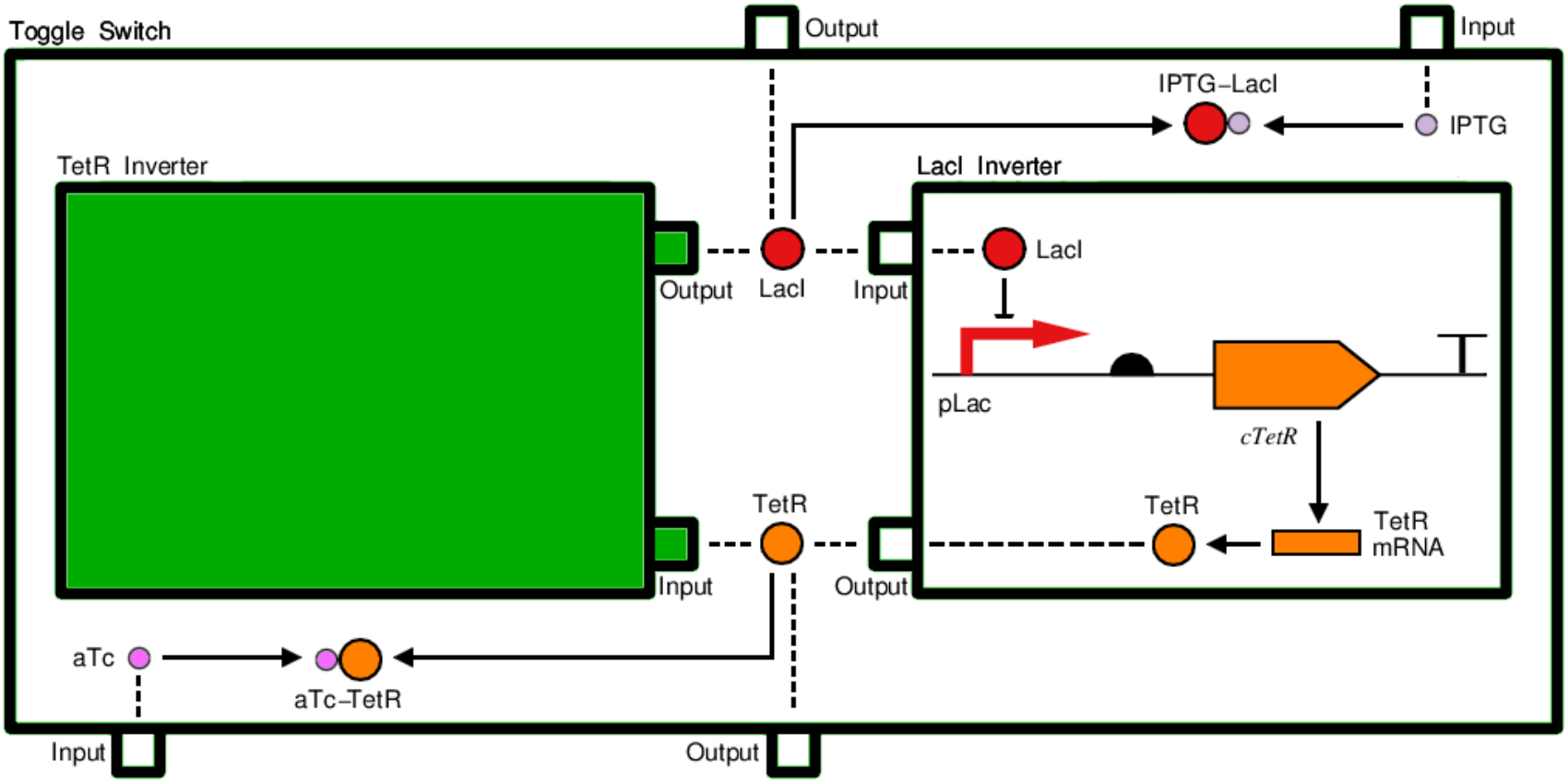
LacI Inverter



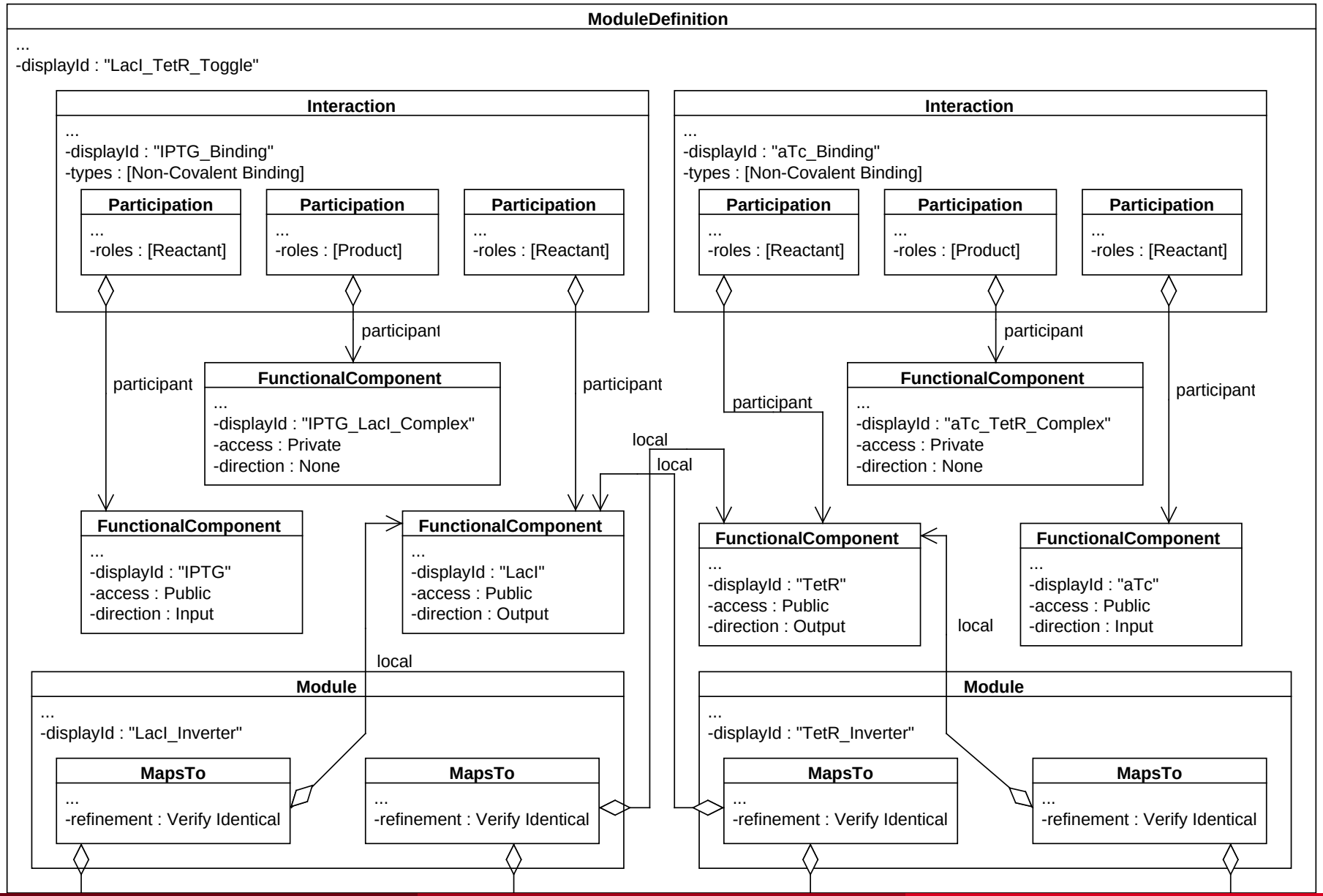
LacI Inverter (SBOL)



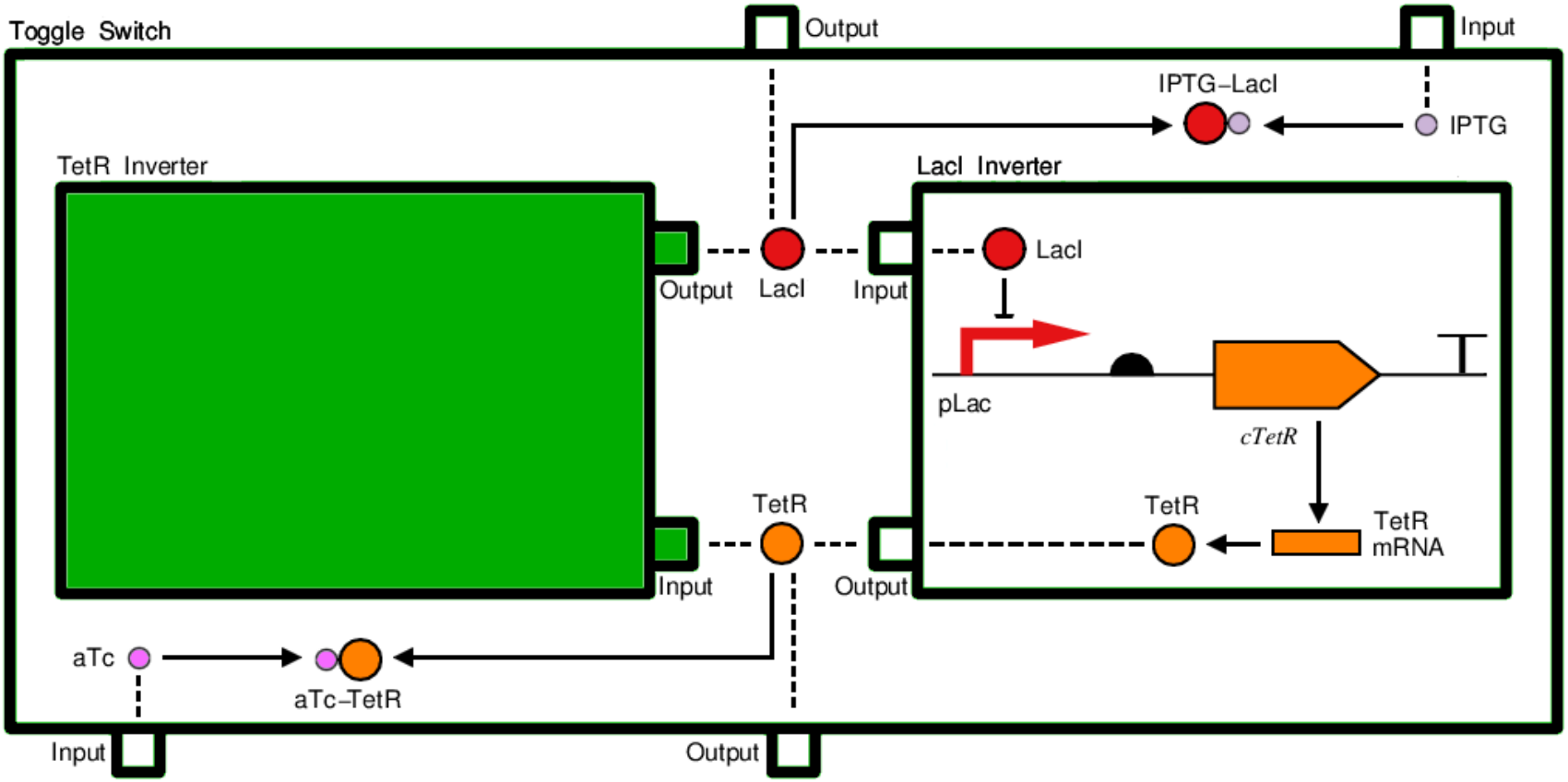
Genetic Toggle Switch



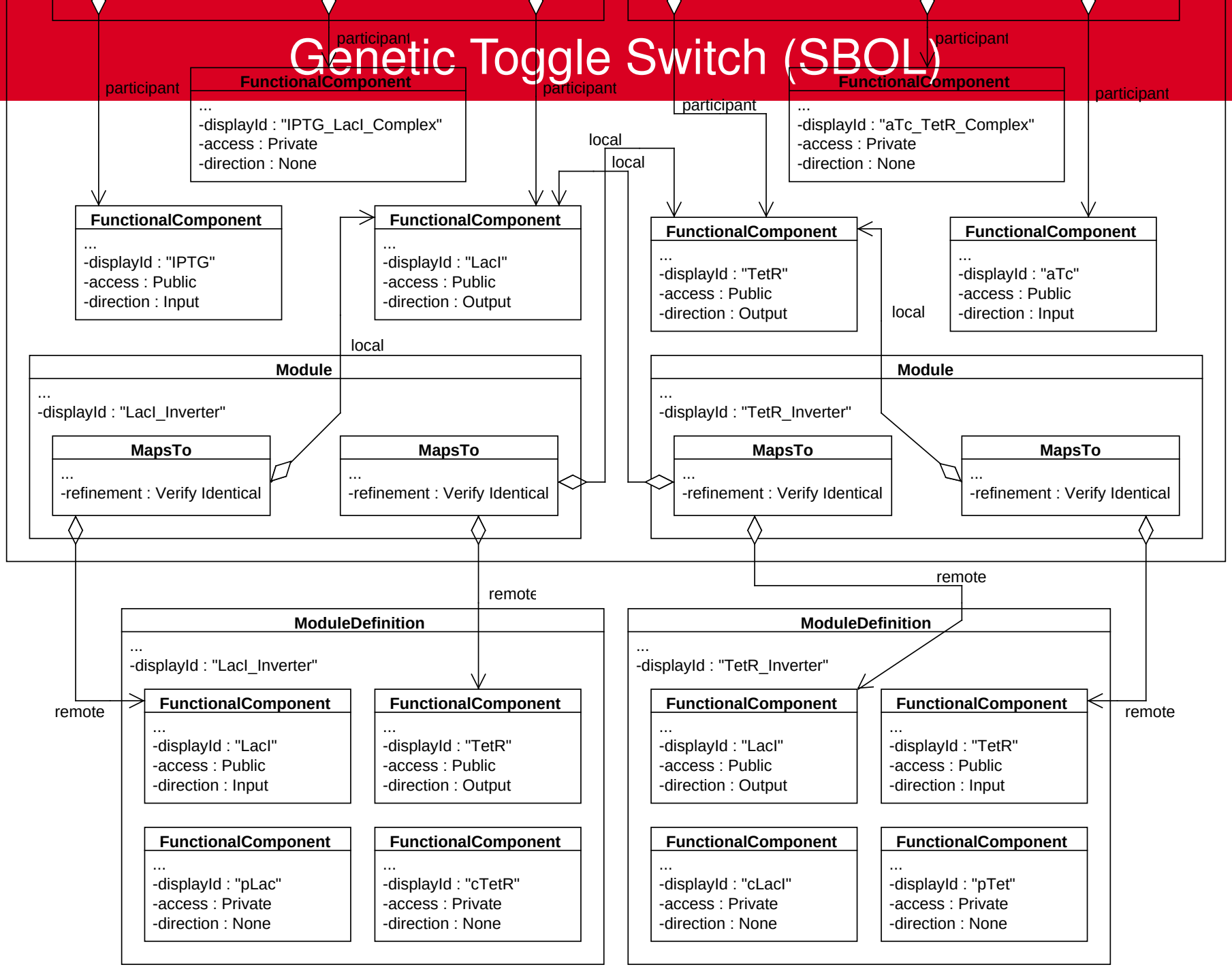
Genetic Toggle Switch (SBOL)



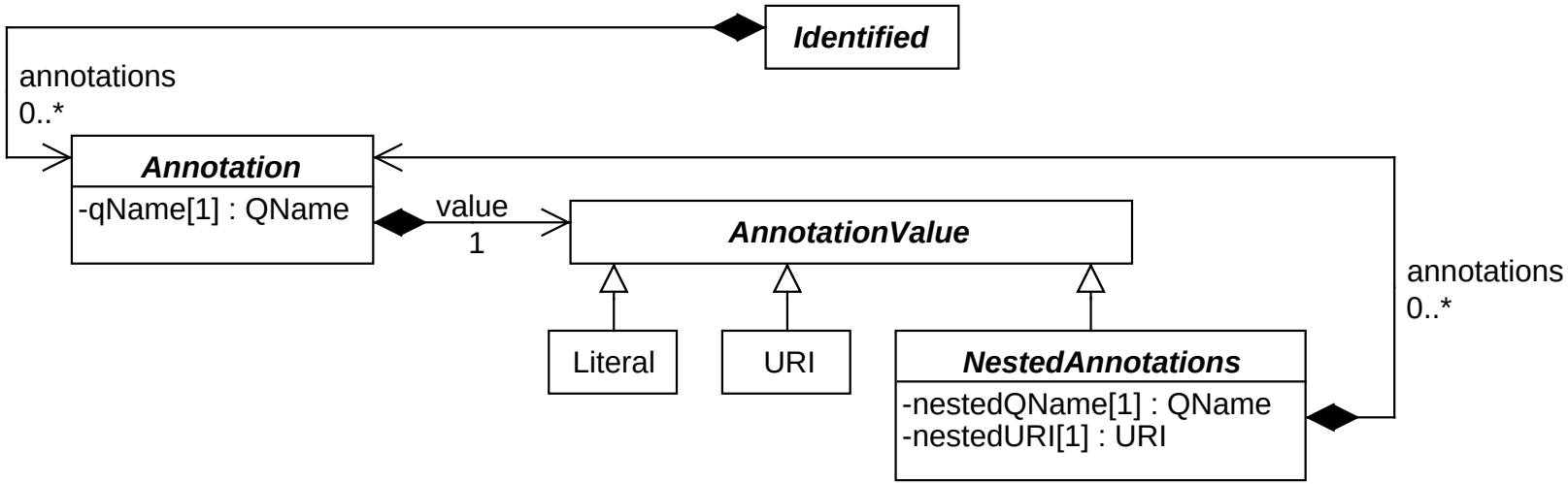
Genetic Toggle Switch



Genetic Toggle Switch (SBOL)



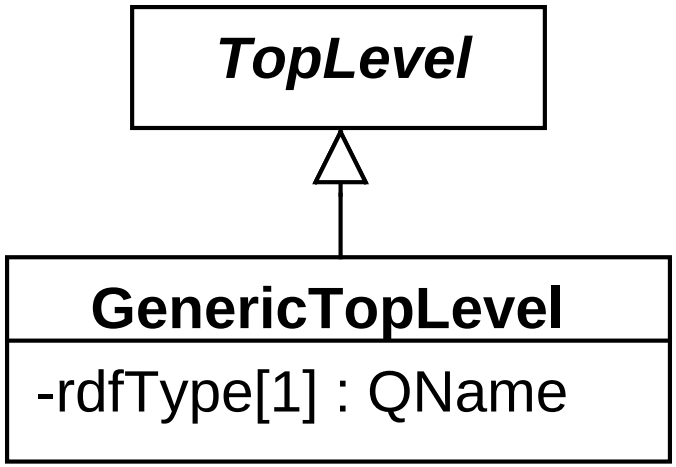
Annotations (Child of Identified)



Annotations (Example)

```
<?xml version="1.0" ?>
<rdf:RDF xmlns:pr="http://partsregistry.org"
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:dcterms="http://purl.org/dc/terms/"
 xmlns:prov="http://www.w3.org/ns/prov#"
 xmlns:sbol="http://sbols.org/v2#">
  <sbol:ComponentDefinition rdf:about="http://partsregistry.org/cd/BBa_J23119">
    <sbol:persistentIdentity rdf:resource="http://partsregistry.org/cd/BBa_J23119"/>
    <sbol:displayId> BBa_J23119 </sbol:displayId>
    <pr:group> iGEM2006_Berkeley </pr:group>
    <pr:experience rdf:resource="http://parts.igem.org/cgi/partsdb/part_info.cgi?part_name=BBa_J23
    <pr:information>
      <pr:Information rdf:about="http://parts.igem.org/cgi/partsdb/part_info.cgi?part_name=BBa_J23
        <pr:sigmafactor> //rnap/prokaryote/ecoli/sigma70 </pr:sigmafactor>
        <pr:regulation> //regulation/constitutive </pr:regulation>
      </pr:Information>
    </pr:information>
    <dcterms:title> J23119 </dcterms:title>
    <dcterms:description> Constitutive promoter </dcterms:description>
    <sbol:type rdf:resource="http://www.biopax.org/release/biopax-level3.owl#DnaRegion"/>
    <sbol:role rdf:resource="http://identifiers.org/so/SO:0000167"/>
  </sbol:ComponentDefinition>
</rdf:RDF>
```

GenericTopLevel



GenericTopLevel (Example)

```
<?xml version="1.0" ?>
<rdf:RDF xmlns:myapp="http://www.myapp.org/"
 xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
 xmlns:dcterms="http://purl.org/dc/terms/"
 xmlns:prov="http://www.w3.org/ns/prov#"
 xmlns:sbol="http://sbols.org/v2#">
  <sbol:ComponentDefinition rdf:about="http://www.partsregistry.org/cd/BBa_J23119">
    <sbol:persistentIdentity rdf:resource="http://www.partsregistry.org/cd/BBa_J23119"/>
    <sbol:displayId> BBa_J23119 </sbol:displayId>
    <prov:wasDerivedFrom rdf:resource="http://www.partsregistry.org/Part:BBa_J23119"/>
    <myapp:datasheet rdf:resource="http://www.partsregistry.org/gen/datasheet1"/>
    <dcterms:title> J23119 </dcterms:title>
    <dcterms:description> Constitutive promoter </dcterms:description>
    <sbol:type rdf:resource="http://www.biopax.org/release/biopax-level3.owl#DnaRegion"/>
    <sbol:role rdf:resource="http://identifiers.org/so/SO:0000167"/>
  </sbol:ComponentDefinition>
  <myapp:Datasheet rdf:about="http://www.partsregistry.org/gen/datasheet1">
    <sbol:persistentIdentity rdf:resource="http://www.partsregistry.org/gen/datasheet1"/>
    <sbol:displayId> datasheet1 </sbol:displayId>
    <myapp:characterizationData rdf:resource="http://www.myapp.org/measurement/1"/>
    <myapp:transcriptionRate> 1 </myapp:transcriptionRate>
    <dcterms:title> Datasheet 1 </dcterms:title>
  </myapp:Datasheet>
</rdf:RDF>
```

Conclusion (SBOL Compliant Software)

- Can either support all classes or only its structural subset.
- Can support import of SBOL, export of SBOL, or both (lossy/lossless).
- SBOL Test Suite is currently under development here:
  `https://github.com/SynBioDex/SBOLTestSuite`
- Validate SBOL files generated using the SBOL Validator found here:
  `http://www.async.ece.utah.edu/sbol-validator/`
- Report SBOL-compliant software to the SBOL editors (editors@sbolstandard.org), so it can be listed on the SBOL website.