

Data Quality in a semantic world

paul@proxml.be

Semantics ?

- EuropeanaTech tech conference
- Birth of Pablo Picasso

EVENT

This class comprises changes of states in cultural, social or physical systems, regardless of scale

Semantics ?

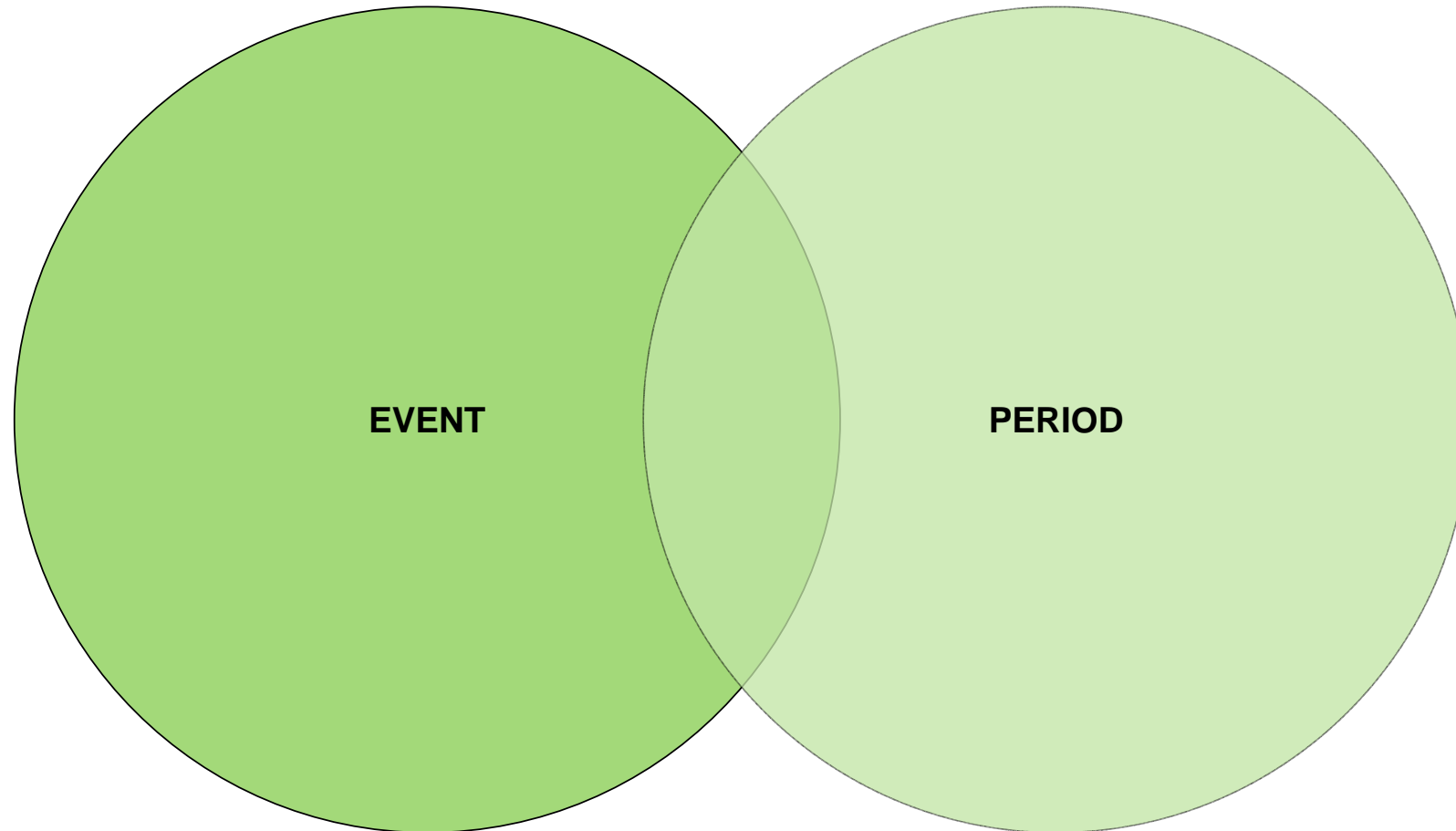
- Jugendstil

- Rococo Revival

PERIOD

This class comprises sets of coherent phenomena or cultural manifestations bounded in time and space.

Question



● My car accident of last week monday

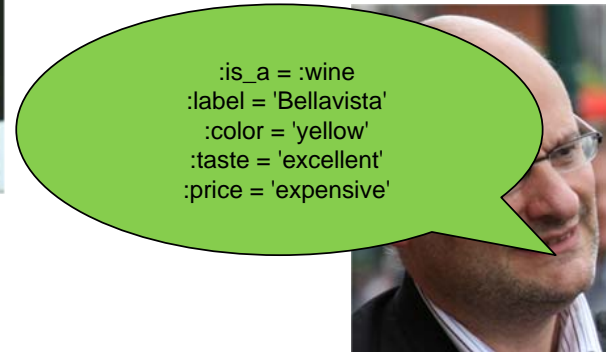
Characteristics of the semantic web

AAA

✿ Anyone can say Anything about Any Topic



```
:is_a = :sparklingWine  
:name = 'Bellavista'  
:colour = 'green'  
:body = 'full'
```



```
:is_a = :wine  
:label = 'Bellavista'  
:color = 'yellow'  
:taste = 'excellent'  
:price = 'expensive'
```

Open world

- You never know all information: new information can always pop up



[0x700x720x690x6A0x730x3A](#)
[0x4A0x610x630x710x750x650x73](#)
[0x620x650x740x610x610x6C0x74](#)

Non Unique Naming

- ✿ The same entity can be known by more than one name



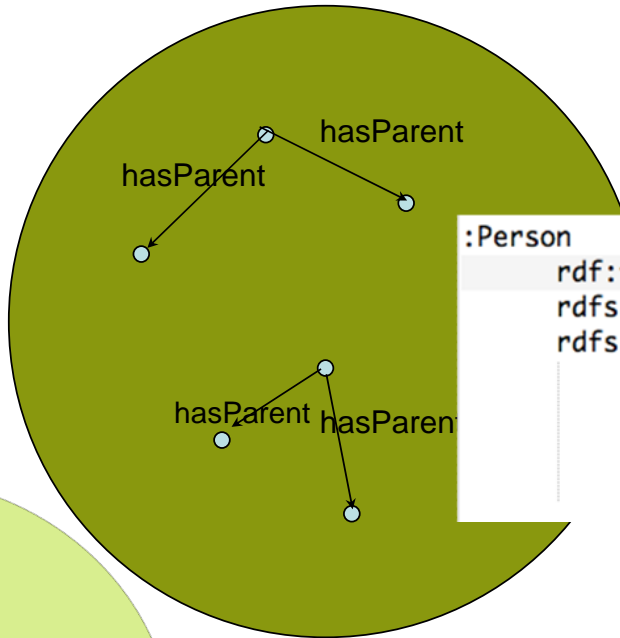
:id =
<http://www.gvh.be/wijkelder/fles213>



:id =
<http://www.bellavista.it/gc2005/B342632>

Checking Constraints

OWL classes



:Person

```
rdf:type owl:Class ;  
rdfs:label "Person"^^xsd:string ;  
rdfs:subClassOf  
  [ rdf:type owl:Restriction ;  
    owl:cardinality "2"^^xsd:nonNegativeInteger ;  
    owl:onProperty :hasParent  
  ] .
```

Person

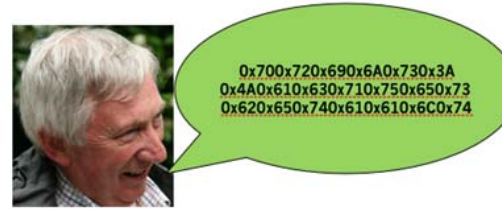
OWL classes

Open world

- * You never know all information: new information can always pop up

```
:Person_2
  rdf:type :Person ;
  rdfs:label "Person_2"^^xsd:string
  :hasParent :Person_3 .
```

SW model:
NOT about constraints
about inferences



Non Unique Naming

- * The same entity can be known by more than one name

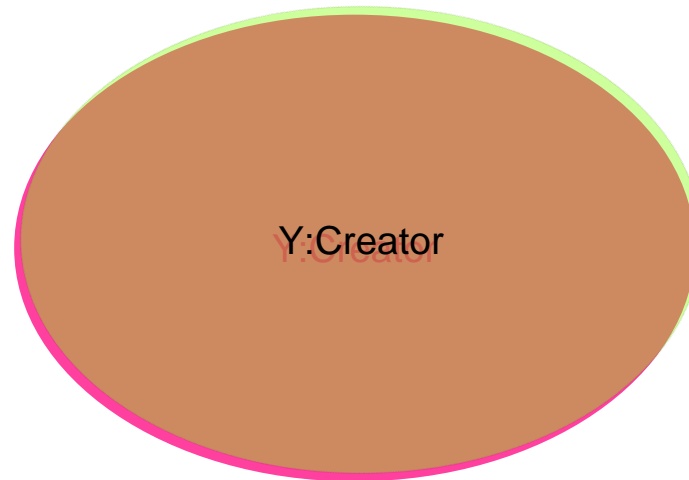
```
:Person_5
  rdf:type :Person ;
  rdfs:label "Person_5"^^xsd:string ;
  :hasParent :Person_8 , :Person_6 , :Pe
```



Metadata sets Reconciliation

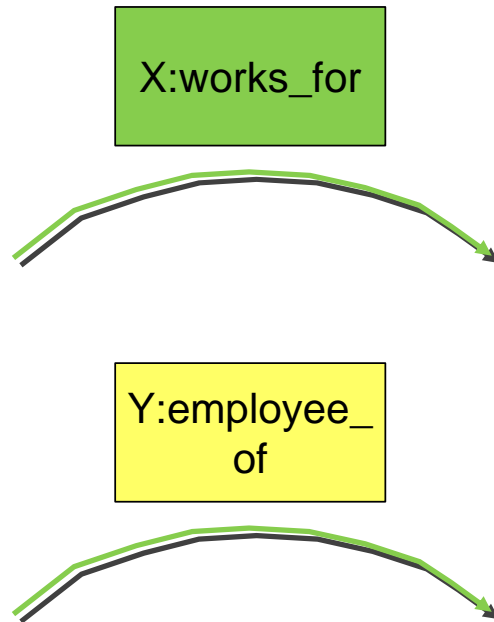
Equivalent classes

X:Author owl:equivalentClass Y:Creator



Equivalent properties

X:works_for owl:equivalentProperty Y:employee_of



Previously pre-reconciliation

```
<record>
  <preref>9644</preref>
  <edit.source>document&gt;book</edit.source>
  <edit.source>collect&gt;extern</edit.source>
  <input.source>boeken</input.source>
  <publisher>Editorial Prensa Ibérica</publisher>
  <edit.time>12:12:54</edit.time>
  <edit.time>17:51:32</edit.time>
  <title>siglo de Rubens en el Museo del Prado : catalogo razonado de pintura flamenca del siglo 17</title>
  <place_of_publication>Barcelona</place_of_publication>
  <pagination>3 dl. (1758 p.)</pagination>
  <edit.name>sa58678</edit.name>
  <edit.name>cs51037</edit.name>
  <edit.name>Nelly Verreydt</edit.name>
  <input.name>Nelly Verreydt</input.name>
  <material_type>book</material_type>
  <lead_word>El</lead_word>
  <year_of_publication>1996</year_of_publication>
  <isbn>84-87657-86-9 (Vol. I)</isbn>
  <isbn>84-87657-87-7 (Vol. II)</isbn>
  <isbn>84-87657-88-5 (Vol. III)</isbn>
  <isbn>84-87657-89-3 (Obra completa)</isbn>
  <copy.shelfmark>1 Q 2 / MAD 7.3 / 1996 / b / I-III / 1</copy.shelfmark>
  <copy.number>NC 97/13-15</copy.number>
  <site>centrale</site>
  <eg>centrale</eg>
  <loan_category>boek</loan_category>
  <edit.date>2010-03-15</edit.date>
  <edit.date>2007-11-15</edit.date>
  <edit.date>1997-01-28</edit.date>
  <input.date>1996-10-29</input.date>
  <author.name>Diaz Padrón, Matias</author.name>
  <author.name>Padrón Mérida, Aida</author.name>
  <object.creator>Rubens, Peter Paul</object.creator>
```



ESE



```
<record>
  <preref>17822</preref>
  <edit.source>document&gt;book</edit.source>
  <input.source>musea</input.source>
  <used_for>Galerij Willem V [s-Gravenhage]</used_for>
  <edit.time>16:02:30</edit.time>
  <edit.name>cs51037</edit.name>
  <input.name>Conversie AIS</input.name>
  <name.type
    option="INST"
    value="INST">
  <text language="0">institution</text>
  <text language="1">instelling</text>
  <text language="3">Institution</text>
  </name.type>
  <edit.date>2004-09-27</edit.date>
  <input.date>2003-12-22</input.date>
  <sigel>GRA 8.7</sigel>
  <address.place>s-Gravenhage</address.place>
  <name>s-Gravenhage, Galerij Willem V</name>
</record>
```



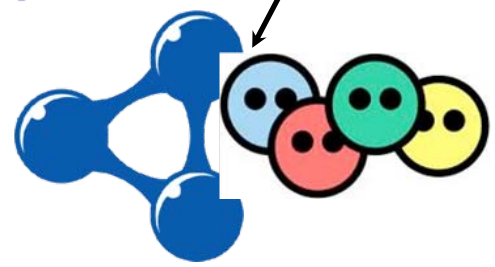
Post-reconciliation

```
<record>
<prifex>9644</prifex>
<edit.source>document&gt;book</edit.source>
<edit.source>collect&gt;extern</edit.source>
<input.source>boeken</input.source>
<publisher>Editorial Prensa Ibérica</publisher>
<edit.time>12:12:54</edit.time>
<edit.time>17:51:32</edit.time>
<title>siglo de Rubens en el Museo del Prado : catalogo razonado de pintura flamen
<place_of_publication>Barcelona</place_of_publication>
<pagination>3 dl. (1758 p.)</pagination>
<edit.name>sa58678</edit.name>
<edit.name>cs51037</edit.name>
<edit.name>Nelly Verreydt</edit.name>
<input.name>Nelly Verreydt</input.name>
<material_type>boek</material_type>
<lead_word>El</lead_word>
<year_of_publications>1996</year_of_publication>
<isbn>84-87657-86-9 (Vol. I)</isbn>
<isbn>84-87657-87-7 (Vol. II)</isbn>
<isbn>84-87657-88-5 (Vol. III)</isbn>
<isbn>84-87657-89-3 (Obra completa)</isbn>
<copy.shelfmark>1 Q 2 / MAD 7.3 / 1996 / b / I-III / 1</copy.shelfmark>
<copy.number>NC 97/13-15</copy.number>
<site>centrale</site>
<eg>centrale</eg>
<loan_category>boek</loan_category>
<edit.date>2010-03-15</edit.date>
<edit.date>2007-11-15</edit.date>
<edit.date>1997-01-28</edit.date>
<input.date>1996-10-29</input.date>
<author.name>Díaz Padrón, Matías</author.name>
<author.name>Padrón Mérida, Aida</author.name>
<object.creator>Rubens, Peter Paul</object.creator>
```



EDM

```
<record>
<prifex>17822</prifex>
<edit.source>document&gt;book</edit.source>
<input.source>musea</input.source>
<used_for>Galerij Willem V [s-Gravenhage]</used_for>
<edit.time>16:02:30</edit.time>
<edit.name>cs51037</edit.name>
<input.name>Conversie AIS</input.name>
<name.type
option="INST"
value="INST">
<text language="0">institution</text>
<text language="1">instelling</text>
<text language="3">Institution</text>
</name.type>
<edit.date>2004-09-27</edit.date>
<input.date>2003-12-22</input.date>
<sigel>GRA 8.7</sigel>
<address.place>s-Gravenhage</address.place>
<name>s-Gravenhage, Galerij Willem V</name>
</record>
```



Pre-reconciliation revisited

```
<record>
  <preref>9644</preref>
  <edit.source>document&gt;book</edit.source>
  <edit.source>collect&gt;extern</edit.source>
  <input.source>boeken</input.source>
  <publisher>Editorial Prensa Ibérica</publisher>
  <edit.time>12:12:54</edit.time>
  <edit.time>17:51:32</edit.time>
  <title>siglo de Rubens en el Museo del Prado : catalogo razonado de pintura flamenca del siglo 17</title>
  <place_of_publication>Barcelona</place_of_publication>
  <pagination>3 dl. (1758 p.)</pagination>
  <edit.name>sa58678</edit.name>
  <edit.name>cs51037</edit.name>
  <edit.name>Nelly Verreydt</edit.name>
  <input.name>Nelly Verreydt</input.name>
  <material_type>book</material_type>
  <lead_word>El</lead_word>
  <year_of_publication>1996</year_of_publication>
  <isbn>84-87657-86-9 (Vol. I)</isbn>
  <isbn>84-87657-87-7 (Vol. II)</isbn>
  <isbn>84-87657-88-5 (Vol. III)</isbn>
  <isbn>84-87657-89-3 (Obra completa)</isbn>
  <copy.shelfmark>1 Q 2 / MAD 7.3 / 1996 / b / I-III / 1</copy.shelfmark>
  <copy.number>NC 97/13-15</copy.number>
  <site>centrale</site>
  <eg>centrale</eg>
  <loan_category>boek</loan_category>
  <edit.date>2010-03-15</edit.date>
  <edit.date>2007-11-15</edit.date>
  <edit.date>1997-01-28</edit.date>
  <input.date>1996-10-29</input.date>
  <author.name>Díaz Padrón, Matías</author.name>
  <author.name>Padrón Mérida, Aida</author.name>
  <object.creator>Rubens, Peter Paul</object.creator>
```



EDM



```
<record>
  <preref>17822</preref>
  <edit.source>document&gt;book</edit.source>
  <input.source>musea</input.source>
  <used_for>Galerij Willem V [s-Gravenhage]</used_for>
  <edit.time>16:02:30</edit.time>
  <edit.name>cs51037</edit.name>
  <input.name>Conversie AIS</input.name>
  <name.type
    option="INST"
    value="INST">
    <text language="0">institution</text>
    <text language="1">instelling</text>
    <text language="3">Institution</text>
  </name.type>
  <edit.date>2004-09-27</edit.date>
  <input.date>2003-12-22</input.date>
  <sigel>GRA 8.7</sigel>
  <address.place>s-Gravenhage</address.place>
  <name>s-Gravenhage, Galerij Willem V</name>
</record>
```



Data Quality

Erfgoedplus

AAA

*Anyone can say Anything about Any Topic



is a: bottle
colour: green
thickness: 0.5



is a: wine
taste: excellent
price: expensive

Titel: Anoniem, fles met opschrift Gebroeders Blokken / Bierhandelaars / Hasselt, ijzer.

Object: flessen

Inventarisnummer: 2002.0036.00

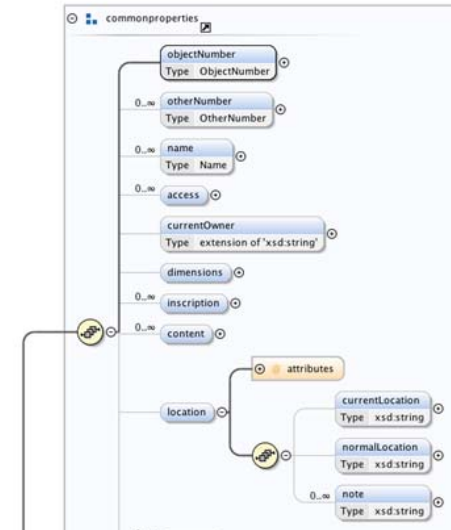
Collectie: Het Stadsmus, Hasselt

Beschrijving: Fles met opschrift: "Gebroeders Blokken / Bierhandelaars / Hasselt", met do

Materiaal: glas (groen)
ijzer

Afmetingen: geheel hoogte: 24,5 cm
geheel breedte: 6,2 cm

Artefact
Type: Artefact



Erfgoedplus

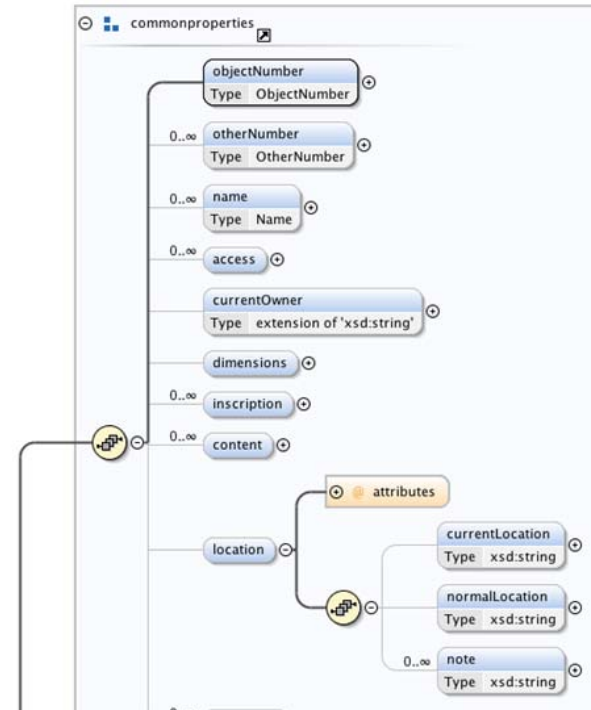
Open world

- * You never know all information: new information can always pop up

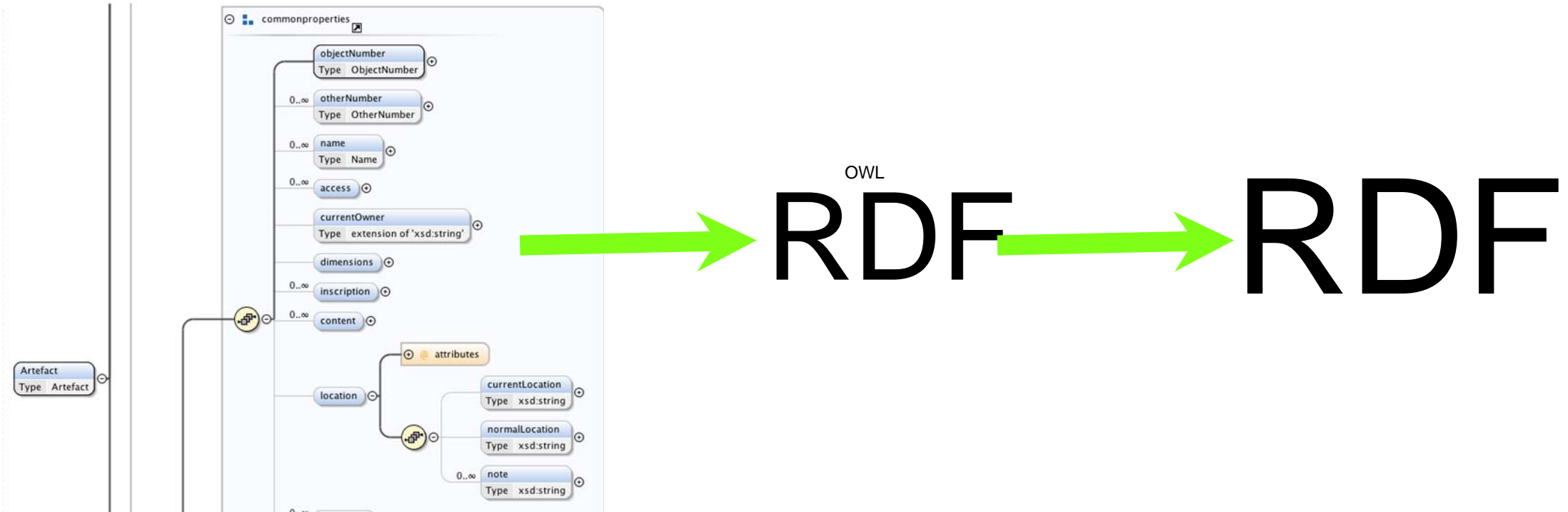


0x700x720x690x6A0x730x3A
0x4A0x610x630x710x750x650x73
0x620x650x740x610x610x6C0x74

Pre-reconciliation



Erfgoedplus semantic?



Actual possibilities



SPIN
SPARQL Inferencing Notation

Overview

SPIN is a [W3C Member Submission](#) that has become the de-facto industry standard to represent meta-modeling capabilities that allow users to define their own SPARQL functions and query

[SPIN in Five Slides](#)

SPIN in Five Slides

<http://spinrdf.org>

Holger Knublauch, TopQuadrant Inc.
holger@topquadrant.com



Example file: <http://topbraid.org/examples/spinsquare.ttl>
Open Source API: <http://topbraid.org/spin/api/>

Best thing since sliced bread



New: watch our introductory video

Clean up

Before linking, metadata always need to be cleaned up. Take out those hands out of your pockets and discover how to handle those embarrassing errors.

Reconcile

Match your metadata with controlled vocabularies connected to the Linked Data cloud and join the place which everyone knows but has never seen.

Event or Period revisited