

## **W3C°** Towards the Semantic Web



|                         | 26. RDF in Programm    | 51. Classes, Resource  | 76. Ontologies (cont   | 101. OWL Lite            | 126. SW Application   |
|-------------------------|------------------------|------------------------|------------------------|--------------------------|-----------------------|
|                         | 27. Python Example     | 52. Classes, Resource  | 77. W3C's Ontology     | 102. "Description Log    | 127. SW Application   |
| 3. Towards a Seman      | 28. Use of RDF in Ot   | 53. Schema Example     | 78. Classes in OWL     | 103. Ontology Exam       | 128. SW Application   |
| 4. Towards a Seman      | 29. Merging            | 54. Schema Example     | 79. Need for Enumer    |                          | 129. SW Application   |
| 5. However              | 30. Merge Shown as     | 55. Inferred Propertie | 80. (OWL) Classes c    | 105. Semantic Web 1      | 130, SW Application   |
| 6. The Semantic Wel     | 31. Merge in Practice  | 56. Properties (Predi  | 81. Same in RDF/XM     | 106. Semantic Web /      | 131. SW Application   |
| 7. What Is Needed (1    | 32. Adding New State   | 57. Properties (cont.) | 82. Union of Classes   | 107. "Best Practices"    | 132. SW Application   |
| 8. The Semantic Wel     | 33. Blank Nodes        | 58. Property Specific  | 83. Same in RDF/XM     | 108. RDF Data Acce       | 133. Further Informat |
| 9. This Course Will     | 34. Blank Nodes: Tui   | 59. Property Specific  | 84. Property Restricti | 109. Data Access Ex      |                       |
|                         | 35. Blank Nodes: Let   | 60. Literals           | 85. Property Restricti | 110. Rules               |                       |
| 11. Problem Example     | 36. Blank Nodes: Sor   | 61. Literals in RDF/X  | 86. Cardinality Restri | 111. RDF API-s           |                       |
| 12. Statements          | 37. Typed Nodes        | 62. Literals in RDF/X  | 87. Same in RDF/XM     | 112. Trust               |                       |
| 13. Resource Descrip    | 38. Typed Nodes (co    |                        | 88. Property Charact   | 113. A Number of Re      |                       |
| 14. RDF is a Graph      | 39. Sequences          | 64. Small Practical Is | 89. Characterization   |                          |                       |
| 15. A Simple RDF Ex     | 40. Sequences (cont    | 65. Binding RDF to a   | 90. Same in RDF/XM     | 115. Available Specif    |                       |
| 16. URI-s Play a Fun    | 41. Sequences (cont    | 66. RDF/XML with XI    | 91. OWL: Additional    | 116. Available Specif    |                       |
| 17. RDF/XML Princip     | 42. Sequences (cont    | 67. RDF Can Also Be    | 92. Term Equivalenc    | 117. Available Specif    |                       |
| 18. RDF/XML Princip     | 43. An Aside: Typed    | 68. RDF/XML has its    | 93. Example: Connec    | 118. Some Books          |                       |
| 19. RDF/XML Princip     | 44. Other Containers   | 69. Programming Pra    | 94. Another Use of E   | 119. Further Information |                       |
| 20. Several Propertie   | 45. Collections (Lists | 70. Programming Pra    | 95. Versioning, Anno   | 120. Further Informat    |                       |
| 21. Several property    | 46. The Same in RDI    | 71. Jena               | 96. OWL and Logic      | 121. Public Fora at V    |                       |
| 22. Adding a New pro    | 47. Our Graphical Sh   | 72. Jena (cont)        | 97. Examples for Log   | 122. Some Tools          |                       |
| 23. Adding a New pro    | 48. Some Words of V    | 73. Lots of Other tool | 98. However: Ontolog   | 123. Some Tools (Co      |                       |
| 24. A Very Useful Sir   |                        |                        | 99. OWL Full           |                          |                       |
| 25. Simplification in C | 50. Back to Typing: F  | 75. Ontologies         | 100. OWL Descriptio    | 125. SW Applications     |                       |





**PART I: Introduction** 

**PART II: Basic RDF** 

PART III: RDF Vocabulary Description Language (RDFS)

PART IV: RDF(S) in Practice

PART V: Ontologies (OWL)

**PART VI: Future Developments** 

**PART VII: Available Documents, Tools** 

**PART VIII: Some Application Examples** 





### **PART I: Introduction**

#### Introduction



### **Towards a Semantic Web**



- The current Web represents information using
  - natural language (English, Hungarian, Finnish,...)
  - o graphics, multimedia, page layout
- Humans can process this easily
  - can deduce facts from partial information
  - can create mental associations
  - are used to various sensory information
    - (well, sort of... people with disabilities may have serious problems on the Web with rich media!)

#### Introduction



### **Towards a Semantic Web**



- Tasks often require to combine data on the Web:
  - hotel and travel infos may come from different sites
  - searches in different digital libraries
  - o etc.
- Again, humans combine these information easily
  - even if different terminologies are used!

## Introduction However...





- However: machines are ignorant!
  - partial information is unusable
  - difficult to make sense from, e.g., an image
  - drawing analogies automatically is difficult
  - difficult to combine information
    - o is <foo:creator> same as <bar:author>?
    - o how to combine different XML hierarchies?
  - 0 ...
- But you know that better than I do...

## W3C®

#### Introduction

### The Semantic Web Approach



- A resource should provide information about itself
  - also called "metadata"
  - metadata should be in a machine processable format
  - agents should be able to "reason" about (meta)data
  - metadata vocabularies should be defined

#### Introduction



## What Is Needed (Technically)?



- To make metadata machine processable, we need:
  - unambiguous names for resources (URIs)
  - a common data model for expressing metadata (RDF)
    - o and ways to access the metadata on the Web
  - common vocabularies (Ontologies)
     The "Semantic Web" is a metadata based infrastructure for reasoning on the Web
- It extends the current Web (and does not replace it)

#### Introduction



### The Semantic Web is Not



### "Artificial Intelligence on the Web"

- although it uses elements of logic...
- ... it is much more down-to-Earth (we will see later)
- o it is all about properly representing and characterizing metadata
- of course: AI systems may use the metadata of the SW
  - o but it is a layer way above it

### "A purely academic research topic"

- SW is out of the university labs now
- lots of applications exist already (see examples later)
- big players of the industry use it (Sun, Adobe, HP, IBM,...)
- of course, much is still be done!

## W3C®

## Introduction This Course Will



- Present the basic model used in the Semantic Web (RDF)
- Show how to represent RDF in XML for the Web
- Introduce the usage of Ontologies on the top of RDF
- Give an idea on how SW applications can be programmed
- Give some examples of SW applications
- Hints for further study





### **PART II: Basic RDF**

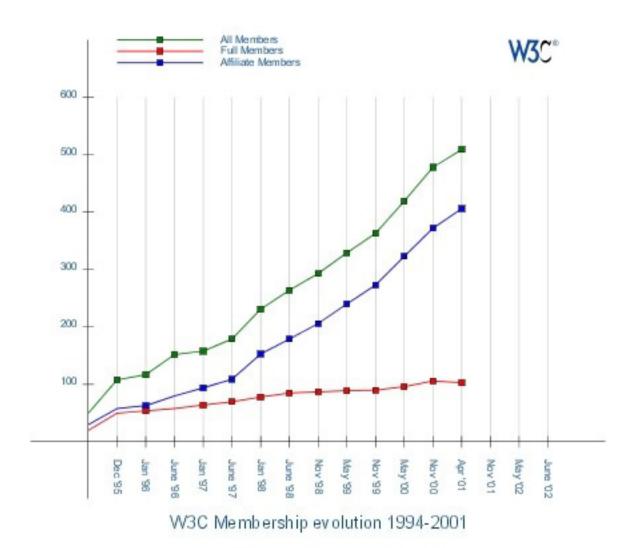


#### **Basic RDF**

## **Problem Example for the Course**



- Convey the meaning of a figure through text (important for accessibility)
  - o add *metadata* to the image describing the content
  - let a tool produce some simple output using the metadata
  - use a standard metadata formalism





## Statements



- The metadata is a set of statements
- In our example:
  - "the type of the full slide is a chart, and the chart type is «line»"
  - "the chart is labeled with an (SVG) text element"
  - "the legend is also a hyperlink"
  - "the target of the hyperlink is «URI»"
  - "the full slide consists of the legend, axes, and data lines"
  - "the data lines describe full and affiliate members, all members"
- The statements are about resources:
  - SVG elements, general URI-s, ...



#### **Basic RDF**

### **Resource Description Framework**



- Statements can be modeled (mathematically) with:
  - Resources: an element, a URI, a literal, ...
  - Properties: directed relations between two resources
  - Statements: "triples" of two resources bound by a property
    - o usual terminology: (s,p,o) for subject, properties, object
- RDF is a general model for such statements
  - with machine readable formats (e.g., RDF/XML, n3, Turtle, RXR)
  - RDF/XML is the "official" W3C format



# RDF is a Graph

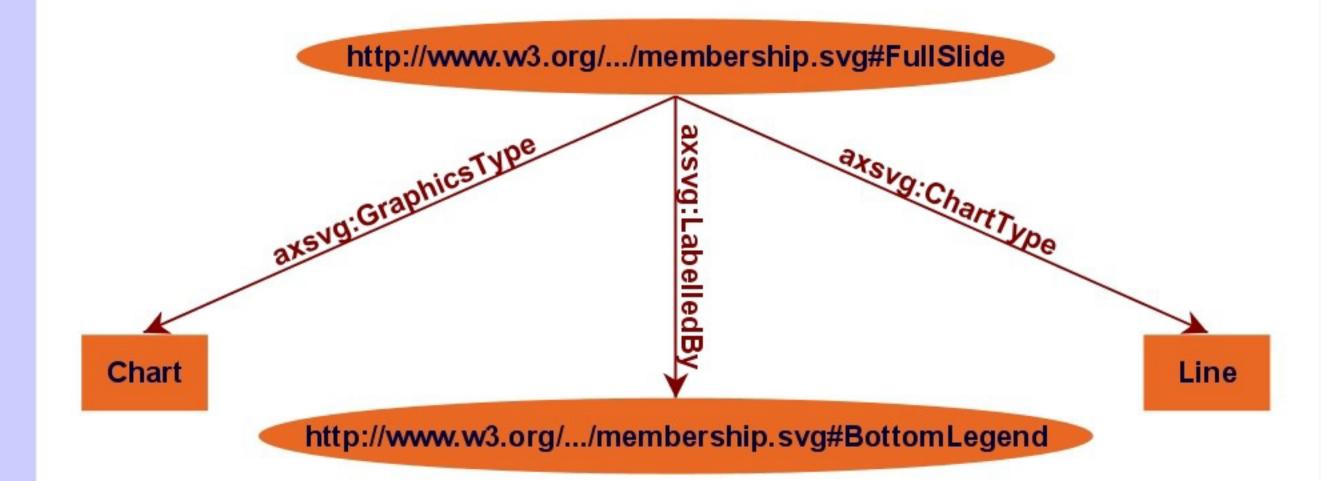


- An (s,p,o) triple can be viewed as a labelled edge in a graph
  - i.e., a set of RDF statements is a directed, labelled graph
    - both "objects" and "subjects" are the graph nodes
    - o "properties" are the edges
  - the formal semantics of RDF is also described using graphs (see the RDF Semantics document)
- One should "think" in terms of graphs, and...
   ...XML or n3 syntax are only the tools for practical usage!
  - the term "serialization" is often used for encoding
- RDF authoring tools usually work with graphs, too (XML or n3 is done "behind the scenes")



# A Simple RDF Example







#### Basic RDF

### **URI-s Play a Fundamental Role**

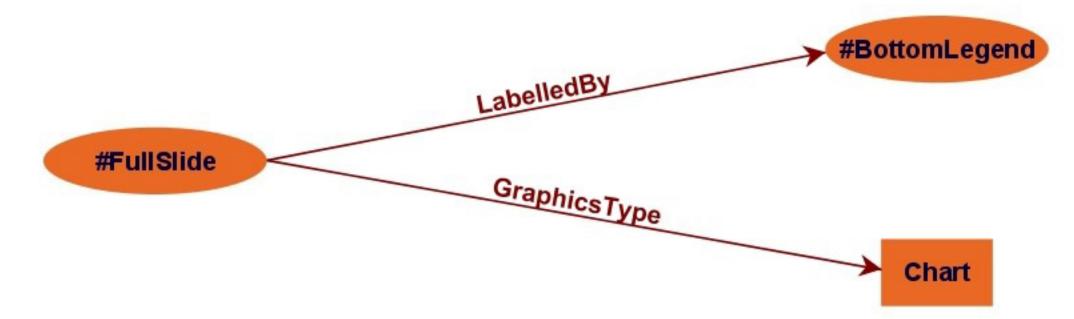


- One can uniquely identify all resources on the web
- Uniqueness is vital to make consistent statements
- Anybody can create metadata on any resource on the Web
  - e.g., the same SVG file could be annotated through other terms
- It becomes easy to merge metadata
  - e.g., applications may merge the SVG annotations
  - this can be done because they refer to the same URI-s!
- URI-s ground RDF into the Web
  - e.g., information can be retrieved using existing tools



# RDF/XML Principles





Encode nodes and edges as XML elements or with literals:

```
«Element for #FullSlide»
    «Element for LabelledBy»
    «Element for #BottomLegend»
    «/Element for LabelledBy»

«/Element for #FullSlide»

«Element for #FullSlide»

«Element for GraphicsType»

Chart
    «/Element for GraphicsType»

«/Element for #FullSlide»
```

# RDF/XML Principles (cont)





Encode the resources (i.e., the nodes):

Note the usage of namespaces!

# RDF/XML Principles (cont)



#FullSlide axsvg:LabelledBy #BottomLegend

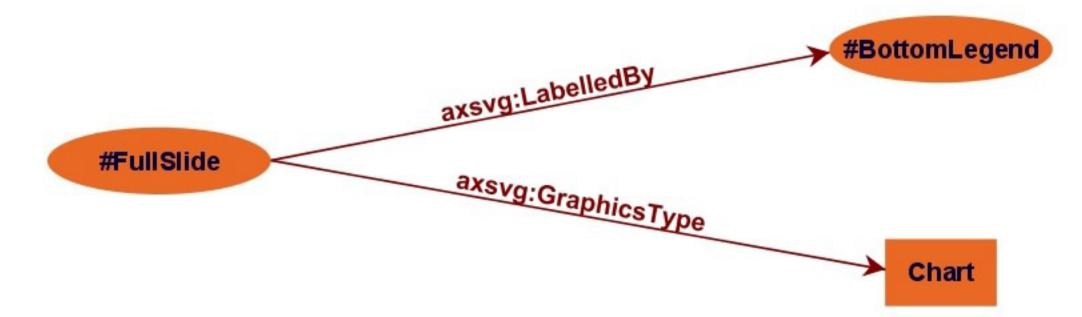
Encode the property (i.e., edge) in its own namespace:



#### Basic RDF

### Several Properties on the Same Node





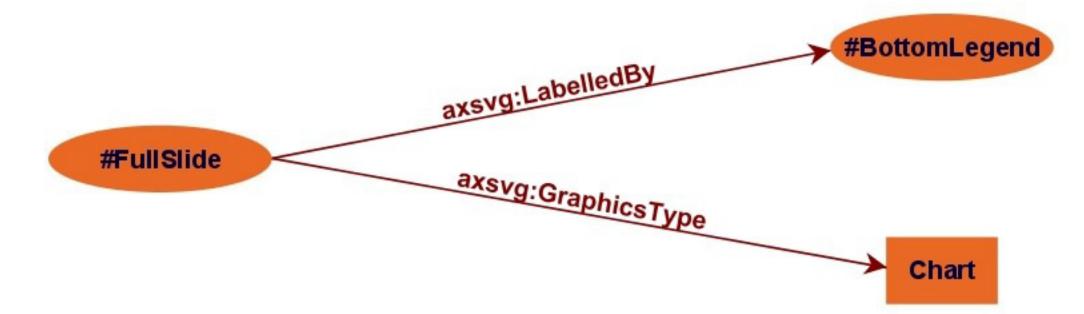
• The "canonical" solution:



#### Basic RDF

### Several property on the same node





The "simplified" version:

There are lots of other simplification rules, see later

# Adding a New property





- (Note: the subject became also an object!)
- The "canonical" solution:





• The "alternative" solution:

Which version is used is a question of taste

### **(**-(i) (→)

## **A Very Useful Simplification**

• The following structure:



# Simplification in Our Example



#FullSlide axsvg:LabelledBy #BottomLegend

Can be expressed by:



#### Basic RDF

## **RDF** in Programming Practice



- For example, using Python+RDFLib:
  - a "Triple Store" is created
  - the RDF file is parsed and results stored in the Triple Store
  - the Triple Store offers methods to retrieve:
    - triples
    - (property,object) pairs for a specific subject
    - (subject,property) pairs for specific object
    - o etc.
  - the rest is conventional programming...
- Similar tools exist in PHP, Java, etc. (see later)



## **Python Example**



#### In Python syntax:

```
# import the libraries
from rdflib.TripleStore import TripleStore
from rdflib.URIRef import URIRef
# resource for a specific URI:
subject = URIRef("URI_of_Subject")
# create the triple store
triples = TripleStore()
# parse an RDF file and store it in the triple store
triples.load("membership.rdf")
# do something with (p,o) pairs
for (p,o) in triples.predicate_objects(subject) :
    do something(p,o)
```



#### **Basic RDF**

### **Use of RDF in Our Example**



#### The tool:

- 1. Uses an RDF parser to extract metadata
- 2. Resolves the URI-s in RDF to access the SVG elements
- 3. Extracts information for the output
  - e.g., text element content, hyperlink data, descriptions
- 4. Combines this with a general text
- 5. Produces a (formatted) text for each RDF statement



# Basic RDF Merging

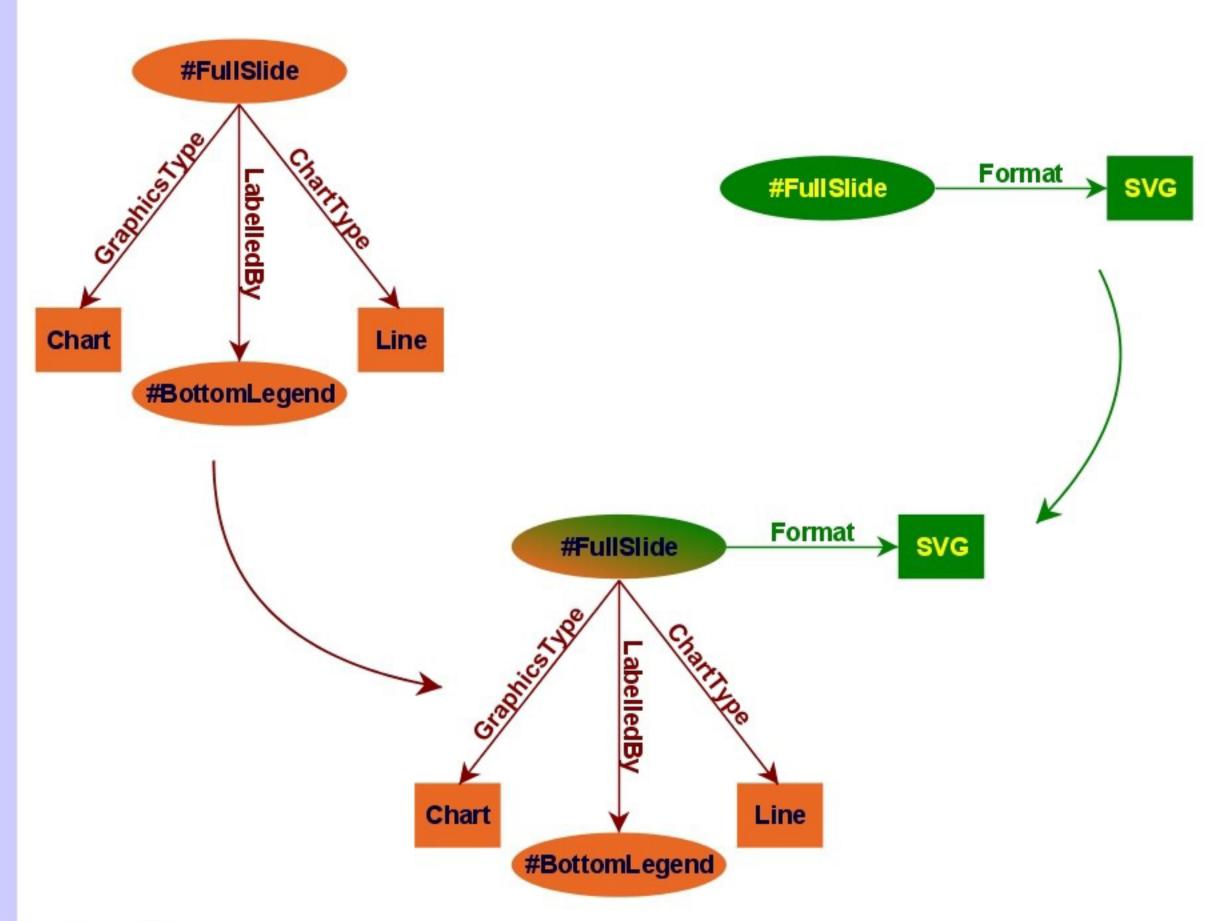


- RDF statements are made on any URI-s
- There may be several graphs using identical URI-s
- An application merges these graphs (conceptually)
  - nodes with identical URI-s are considered identical
  - the rest is quite obvious
- Merging is a very powerful feature of RDF
  - metadata may be defined by several (independent) parties...
  - ...and combined by an application



# Merge Shown as Graphs







## Merge in Practice



- Development environments merge graphs automatically
  - e.g., in Python, the Triple Store can "load" several files
  - the load merges the new statements automatically
- Merging the RDF/XML files into one is also possible
  - but not really necessary, the tools will merge them for you
  - keeping them separated may make maintenance easier
  - some of the files may be on a remote site anyway!



# Adding New Statements



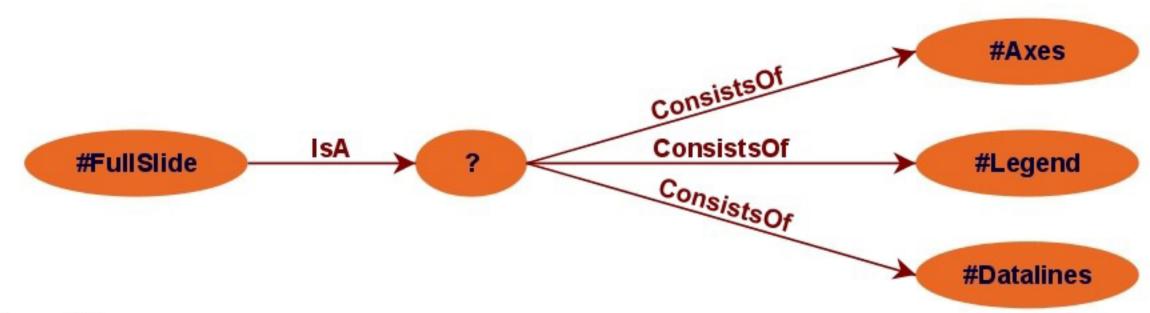
- Adding a new statement is also very simple
  - e.g., in Python+RDFLib: store.add((s,p,o))
- In fact, it can be seen as a special case of merging
- This is a very powerful feature, too
  - managing data in RDF makes it very flexible indeed...



## Blank Nodes



- Consider the following statement:
  - "the full slide is a «thing» that consists of axes, the legend and the datalines"
- Until now, nodes were identified with a URI. But...
- ...what is the URI of «thing»?





# Blank Nodes: Turn Them Into Regulars



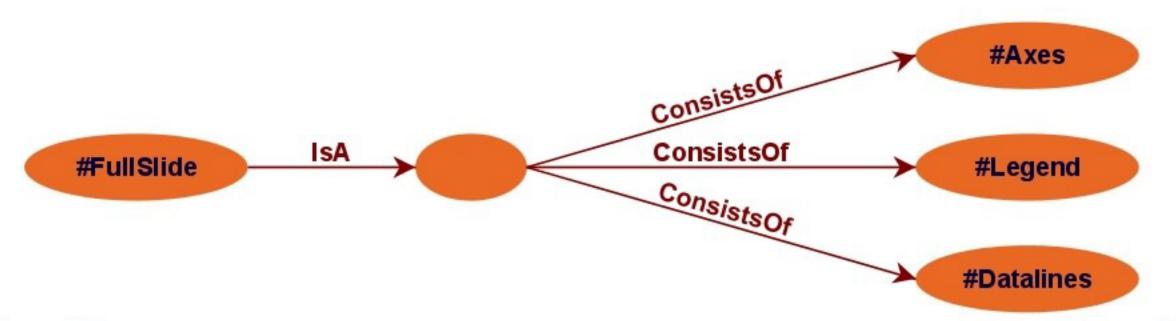
• In the XML serialization: give an id with rdf: ID

- Defines a fragment identifier within the RDF portion
- Identical to the id in HTML, SVG, ...
- Can be referred to with regular URI-s from the outside

## Blank Nodes: Let the System Do It



Let the system create a nodeID internally



#### Basic RDF

#### **Blank Nodes: Some More Remarks**



- Blank nodes require attention when merging
  - blanks nodes in different graphs are different
  - the implementation must be be careful with its naming schemes
- The XML Serialization introduces a simplification



# Typed Nodes



- To emphasize that a node is of a specific class
  - i.e., it is part of a possible set of individuals
  - e.g., #Datalines node is an "SVG entity"
- There is a separate document on how to define classes
  - "RDF Vocabulary Description Language", a.k.a. "RDF Schemas"
  - see later in this tutorial
- We can use the special RDF property rdf:type:

```
<rdf:Description rdf:about="#Datalines">
  <rdf:type
   rdf:resource="http://.../axsvg-schema.rdf#SVGEnti
   ...
</rdf:Description/>
```



# Typed Nodes (cont)



A resource may belong to several classes

```
(rdf:type is just a property...)
```

- The type information may be very important for applications
  - e.g., it may be used for a categorization of possible nodes
- The rdf namespace contains predefined classes
  - see later...



## Sequences



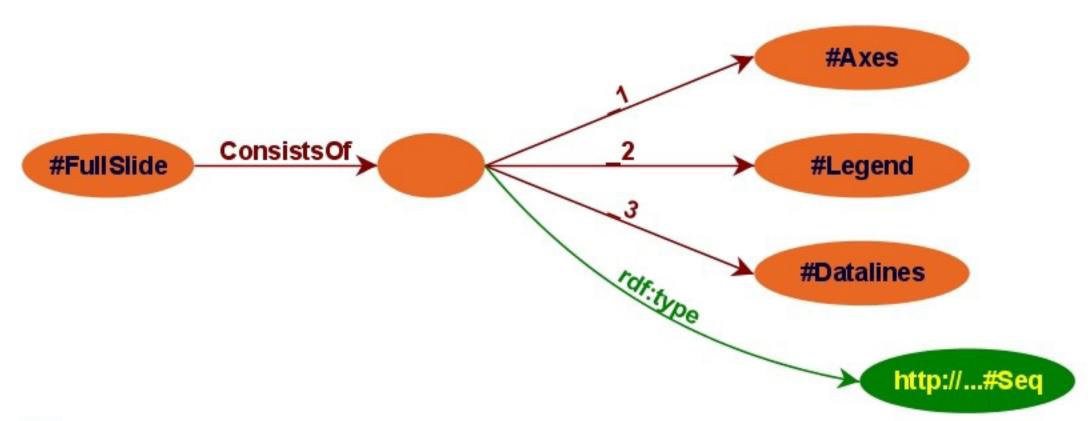
- We used the following statement:
  - "the full slide is a «thing» that consists of axes, the legend and the datalines"
- But we also want to express the constituents in this order
- Using blank nodes is not enough



# Sequences (cont)



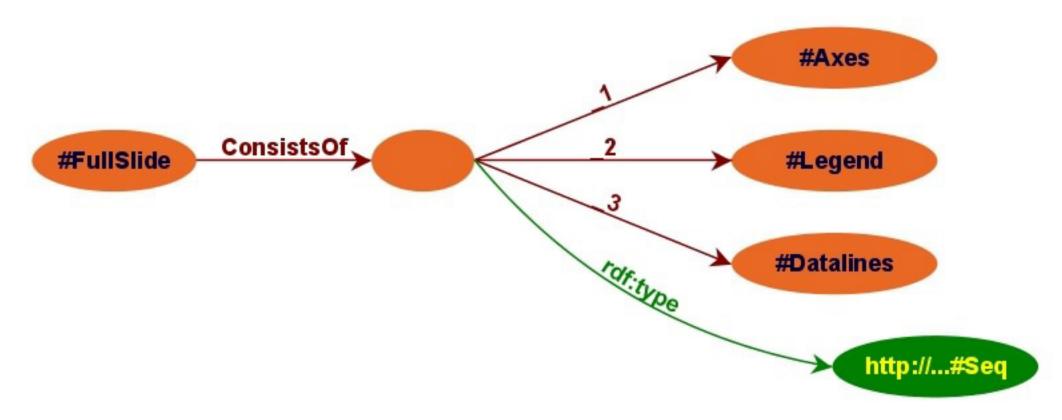
- One can use the predefined:
  - RDF class Seq
  - RDF properties rdf:\_1, rdf:\_2, ...
- The agreed semantics is of a sequential containment





# Sequences (cont)



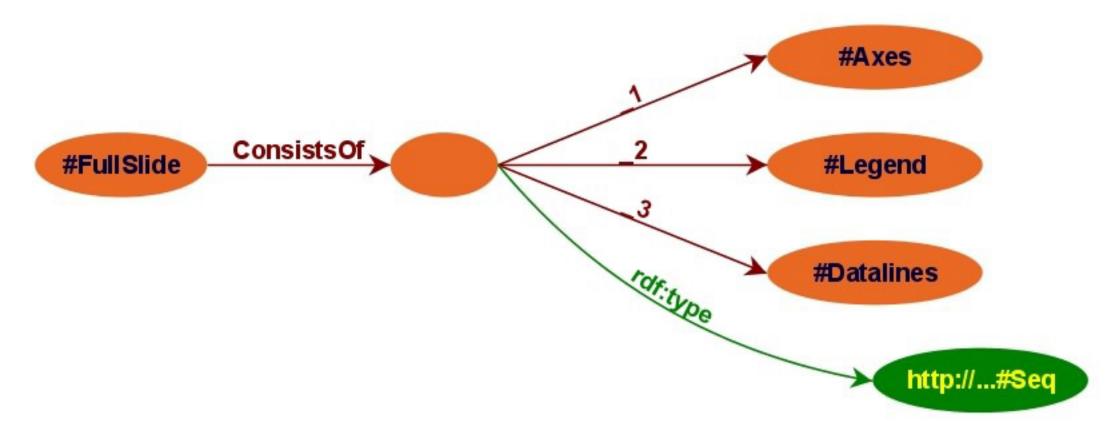


#### In RDF/XML:



# Sequences (cont)





A simplified alternative (this is only syntax...):



#### Basic RDF



### An Aside: Typed Nodes in RDF/XML

A frequent simplification rule: instead of:

Usage of rdf:Seq is based on this simplification rule



## Other Containers



rdf:Bag

a general bag, no particular semantics attached

• rdf:Alt

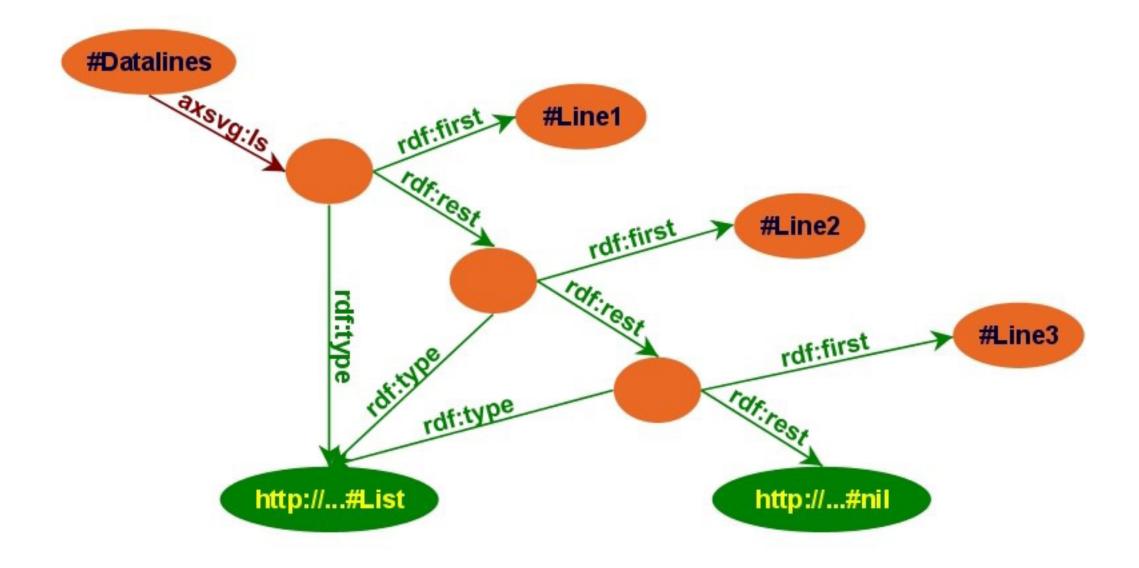
attached semantics: only one of the constituents is "valid"



# Basic RDF Collections (Lists)



- RDF also includes lists
  - familiar structure for Lisp programmers...



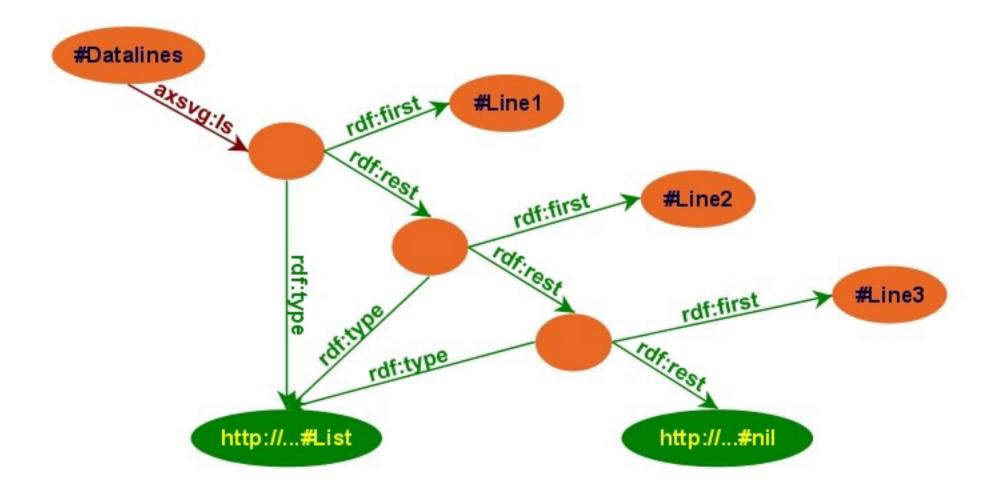
#### **Basic RDF**

## The Same in RDF/XML



#### List in terms of XML:

```
<rdf:Description rdf:about="#Datalines">
    <axsvg:Is rdf:parseType="Collection">
        <rdf:Description rdf:about="#Line1"/>
        <rdf:Description rdf:about="#Line2"/>
        <rdf:Description rdf:about="#Line3"/>
    </axsvg:Is >
</rdf:Description/>
```





### **Our Graphical Shorthand**



(To simplify the images...)

```
<rdf:Description rdf:about="#Datalines">
    <axsvg:Is rdf:parseType="Collection">
        <rdf:Description rdf:about="#Line1"/>
        <rdf:Description rdf:about="#Line2"/>
        <rdf:Description rdf:about="#Line3"/>
    </axsvg:Is >
</rdf:Description/>
                                           #Line1
              axsvg:ls
   #Datalines
                                           #Line2
                         List
                                           #Line3
```



## Some Words of Warning



- RDF/XML introduces a number of simplifications
  - usage of rdf:li instead of rdf:\_1, rdf:\_2, ...
  - usage of rdf:parseType instead of rdf:first, rdf:rest, ...
  - o etc.
- This can be deceptive when using, e.g., RDFLib:
  - the triples in the Triple Store are the "real" ones!
    - o i.e., rdf:\_1, rdf:\_2 and not rdf:li
  - rdf:Seq does not appear directly
    - instead, a (possibly blank) node with a rdf:type property
  - o etc.
- Never forget: only the graph is "real", the rest is convenience!





#### PART III: RDF Vocabulary Description Language

(a.k.a. RDFS)





### **Back to Typing: RDF Schemas**



- Adding metadata and using it from a program works...
- ... provided the program knows what terms to use!
- We used terms like:
  - Chart, LabelledBy, IsAnchor, ...
  - ChartType, GraphicsType, ...
  - etc
- Are they all known? Are they all correct?
- It is a bit like defining record types for a database
- This is where RDF Schemas come in
  - officially: "RDF Vocabulary Description Language"



### Classes, Resources, ...



#### Think of well known in traditional ontologies:

- use the term "mammal"
- "every dolphin is a mammal"
- "Flipper is a dolphin"
- o etc.

#### RDFS defines the terms of resources and classes:

- everything in RDF is a "resource"
- "classes" are also resources, but...
- they are also a collection of possible resources (i.e., individuals)
   (e.g., "mammal", "dolphin")

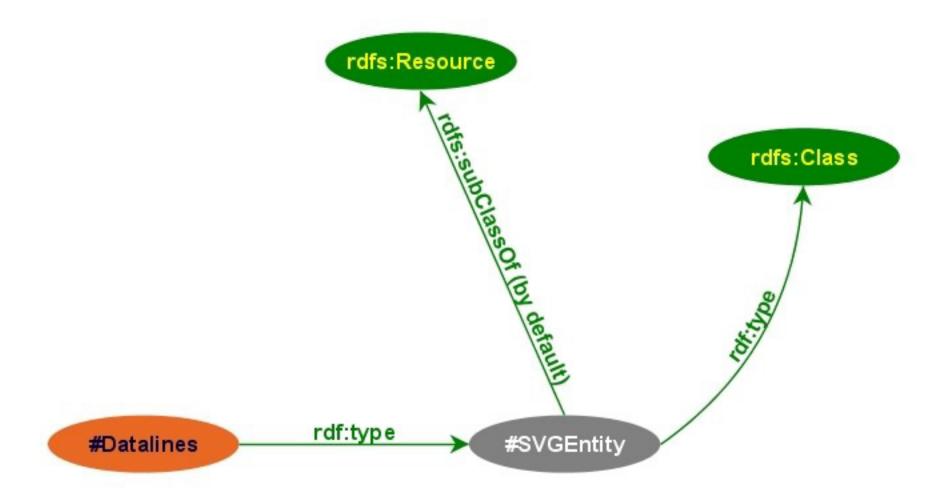
#### Relationships are defined among classes/resources:

- "typing": an individual belongs to a specific class (e.g., "Flipper is a dolphin")
- "subclassing": instance of one is also the instance of the other (e.g., "every dolphin is a mammal")

#### RDF Schemas

### Classes, Resources in RDF





- RDFS defines rdfs:Resource, rdfs:Class as nodes, ...
   ... rdf:type, rdfs:subClassOf as properties
- User should create RDF Schema file for the user types (Note: RDFS is also RDF!)



### Schema Example in RDF/XML



In axsvg-schema.rdf ("application's data types"):

```
<rdf:Description rdf:ID="SVGEntity">
    <rdf:type
    rdf:resource="http://www.w3.org/2000/01/rdf-schema#(
    />
    </rdf:Description>
```

In the rdf data on a specific graphics ("using the type"):

```
<rdf:Description rdf:about="#Datalines">
     <rdf:type rdf:resource="axsvg-schema.rdf#SVGEntity'
</rdf:Description>
```



### Schema Example in RDF/XML (alt.)



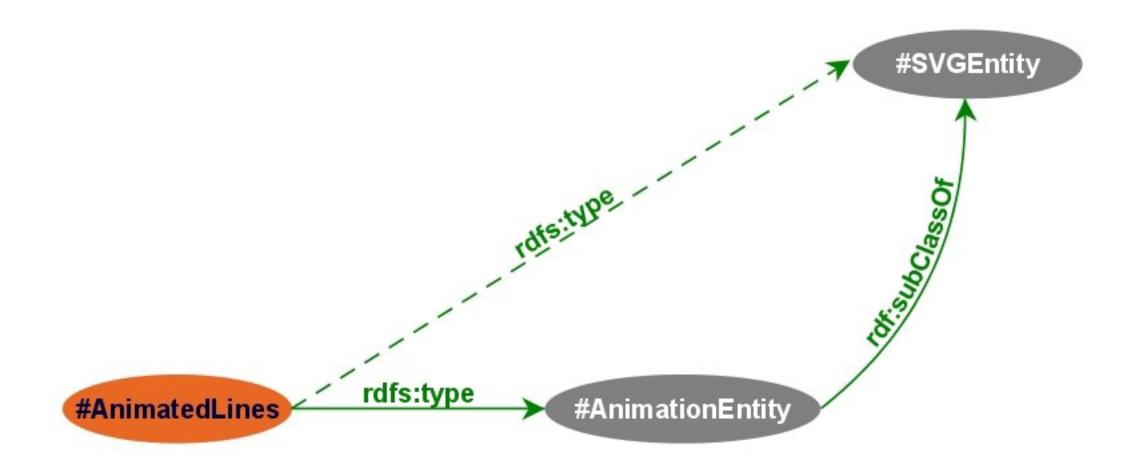
• In axsvg-schema.rdf (remember the simplification rule):

```
<rdfs:Class rdf:ID="SVGEntity">
...
</rdfs:Class>
```

• In the rdf data on a specific graphics:

# Inferred Properties





(#AnimatedLines rdf:type #SVGEntity)

- is *not* in the original RDF data...
- ...but can be inferred from the RDFS rules
- Better RDF environments will return that triplet, too

#### **RDF Schemas**



### **Properties (Predicates)**



- Property is a special class (rdf:Property)
  - i.e., properties are also resources
- Properties are constrained by their range and domain
  - i.e., what individuals can be on the "left" or on the "right"
- There is also a possibility for a "sub-property"
  - all resources bound by the "sub" are also bound by the other



### Properties (cont.)



- Properties are also resources...
- So properties of properties can be expressed as...
  ...RDF properties ©
  - this twists your mind a bit, but you will get used to it
- For example:

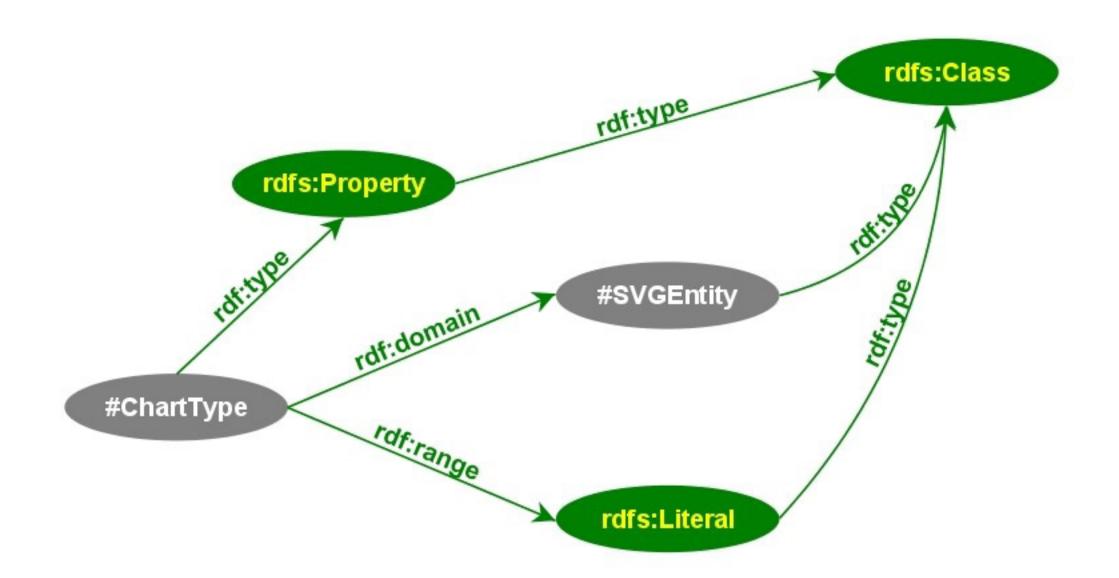
```
(P rdfs:range C) means:
```

- 1. P is a property
- 2. C is a class instance
- 3. when using P, the "object" must be an individual in C
- this is an RDF statement with subject P, object C
   and property rdfs:range



### **Property Specification Example**





- Note that one cannot define what literals can be used
- This requires ontologies (see later)



#### **Property Specification in XML**



#### Same example in XML/RDF:

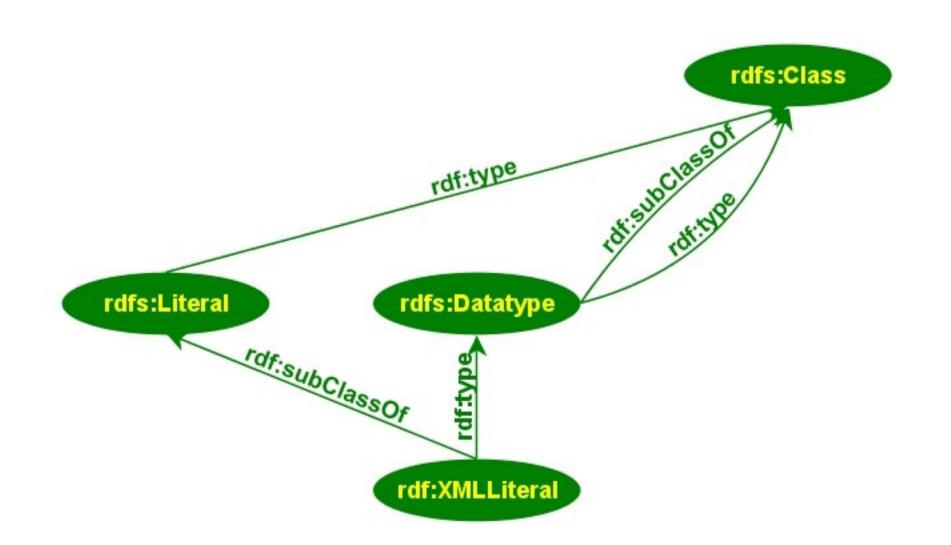
```
<rdfs:Property rdf:ID="ChartType">
     <rdf:domain rdf:resource="#SVGEntity"/>
     <rdf:range rdf:resource="http://...#Literal"/>
</rdfs:Property>
```



#### RDF Schemas Literals



- Literals may have a data type
  - floats, int, etc.
  - all types defined in XML Schemas
- (Natural) language can be specified
- Formally, data types are separate RDFS classes
- Full XML fragments may also be literals



## Literals in RDF/XML



Typed literals:



### Literals in RDF/XML (cont.)



- XML Literals:
  - makes it possible to "bind" RDF resources with XML vocabularies:





### PART IV: RDF(S) in Practice



## Small Practical Issues



• RDF/XML files have a registered Mime type:

application/rdf+xml

Recommended extension: .rdf



#### RDF(S) in Practice

#### Binding RDF to an XML Resource



- You can use the rdf:about as a URI for external resources
  - i.e., store the RDF as a separate file
- You may add RDF to XML directly (in its own namespace)
  - ∘ e.g., in SVG:



## RDF(S) in Practice RDF/XML with XHTML



- XHTML is still based on DTD-s (lack of entities in Schemas)
- RDF within XHTML's header does not validate...
- Currently, people use
  - link/meta in the header (perfectly o.k.!)
    - using conventions instead of namespaces in metas
  - put RDF in a comment (e.g., Creative Commons)
- XHTML 2.0 will have a separate 'metadata' module
  - essentially, the current meta/link elements are extended
  - one can define "triplets" using this formalism
  - in fact, a new RDF serialization... (like RDF/XML and n3)



#### RDF(S) in Practice

#### **RDF Can Also Be Generated**



- There might be conventions to use in XHTML...
  - e.g., by using class names
- ... and then generate RDF automatically
- There are tools and developments in this direction



## RDF(S) in Practice RDF/XML has its Problems



- RDF/XML was developed in the "prehistory" of XML
  - e.g., even namespaces did not exist!
- Coordination was not perfect, leading to problems
  - the syntax cannot be checked with XML DTD-s
  - XML schemas are also a problem
  - encoding is verbose and complex
    - (e.g., simplifications lead to confusions)

but there is too much legacy code 😌

- Don't be influenced (and set back...) by the XML format
  - the important point is the model, XML is just syntax
  - other "serialization" methods may come to the fore



#### **Programming Practice**



We have already seen how to retrieve triples in RDFLib:

```
# import the libraries
from rdflib.TripleStore import TripleStore
from rdflib.URIRef import URIRef
# resource for a specific URI:
subject = URIRef("URI_of_Subject")
# create the triple store
triples = TripleStore()
# parse an RDF file and store it in the triple store
triples.load("membership.rdf")
# do something with (p,o) pairs
for (p,o) in triples.predicate_objects(subject) :
    do something(p,o)
```



### **Programming Practice (cont)**



One can also edit triples, save it to an XML file, etc:

```
# add a triple to the triple store
triples.add((subject,pred,object))
# remove it
triples.remove_triples((subject,pred,object))
# save it in a file in RDF/XML
triples.save("filename.rdf")
```

- It is very easy to start with this
- Does not have (yet) powerful schema processing
  - no "inferred" properties, for example
- You can get RDFLib at: http://rdflib.net





- RDF toolkit in Java from HP's Bristol lab
- The RDFLib features are all available:

```
// create a model (a.k.a. Triple Store in python)
Model model=new ModelMem();
Resource subject=model.createResource("URI_of_Subject")
// 'in' refers to the input file
model.read(new InputStreamReader(in));
StmtIterator iter=model.listStatements(subject,null,nul)
while(iter.hasNext()) {
    st = iter.next();
    p = st.getProperty();
    o = st.getObject();
    do_something(p,o);
}
```



# PRDF(S) in Practice Jena (cont)



- But Jena is much more than RDFLib
  - it has a large number of classes/methods
    - listing, removing associated properties, objects
    - comparing full RDF graphs
    - manage typed literals
    - o mapping **Seq**, **Alt**, etc. to Java constructs
    - o etc.
  - it has an "RDFS Reasoner"
    - a new model is created with an associated RDFS file
    - o all the "inferred" properties, types are accessible
    - errors are checked
  - and more...
- Of course, it is much bigger and more complicated...
- Is available at: http://jena.sourceforge.net/



#### RDF(S) in Practice

#### **Lots of Other tools**



- There are other tools:
  - RDFSuite: another Java environment (from ICS-FORTH)
  - RDFStore: RDF Framework for Perl
  - Redland: RDF Framework for C
  - RAP: RDF Framework for PHP
  - SWI-Prolog: RDF Framework for Prolog
  - 0 ...
  - Sesame: Java based storage and query for RDF and RDFS
  - Kowari and Tucana: triple based database systems
    - they have Jena interfaces, too
  - o etc.
- You can always start by:

http://www.w3.org/RDF/#developers





#### **PART V: Ontologies (OWL)**



### **Ontologies**



- RDFS is useful, but does not solve all the issues
- Complex applications may want more possibilities:
  - can a program reason about some terms? E.g.:
    - o "if «A» is left of «B» and «B» is left of «C», is «A» left of «C»?"
    - obviously true for humans, not obvious for a program ...
    - ... programs should be able to deduce such statements
  - if somebody else defines a set of terms: are they the same?
    - obvious issue in an international context
  - construct classes, not just name them
  - restrict a property range when used for a specific class
  - o etc.



### Ontologies (cont.)



- The Semantic Web needs a support of *ontologies*: "defines the concepts and relationships used to describe and represent an area of knowledge"
- We need a Web Ontologies Language to define:
  - the terminology used in a specific context
  - more constraints on properties
  - the logical characteristics of properties
  - the equivalence of terms across ontologies
  - o etc.
- Language should be a compromise between
  - rich semantics for meaningful applications
  - feasibility, implementability



# W3C's Ontology Language (OWL)



- A layer on top of RDFS with additional possibilities
- Outcome of various projects:
  - 1. a DARPA project: DAML
  - 2. a EU project: OIL
  - 3. an attempt to merge the two: DAML+OIL
  - 4. the latter was submitted to W3C
  - 5. lots of coordination with the core RDF work
  - 6. recommendation since early 2004

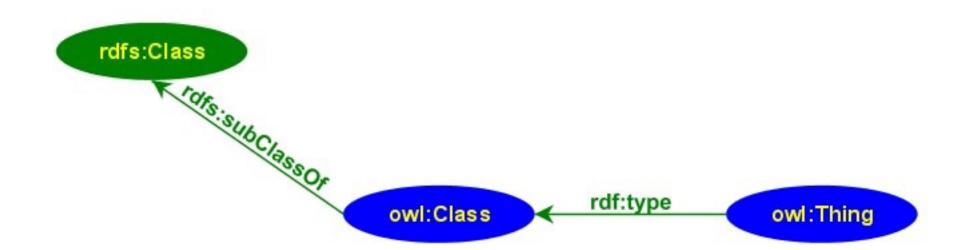




#### Classes in OWL



- In RDFS, you can subclass existing classes...
  - ... but, otherwise, that is all you can do
- In OWL, you can construct classes from existing ones:
  - enumerate its content
  - through intersection, union, complement
  - through property restrictions
- To do so, OWL introduces its own Class...
  - ... and Thing to differentiate the individuals from the classes



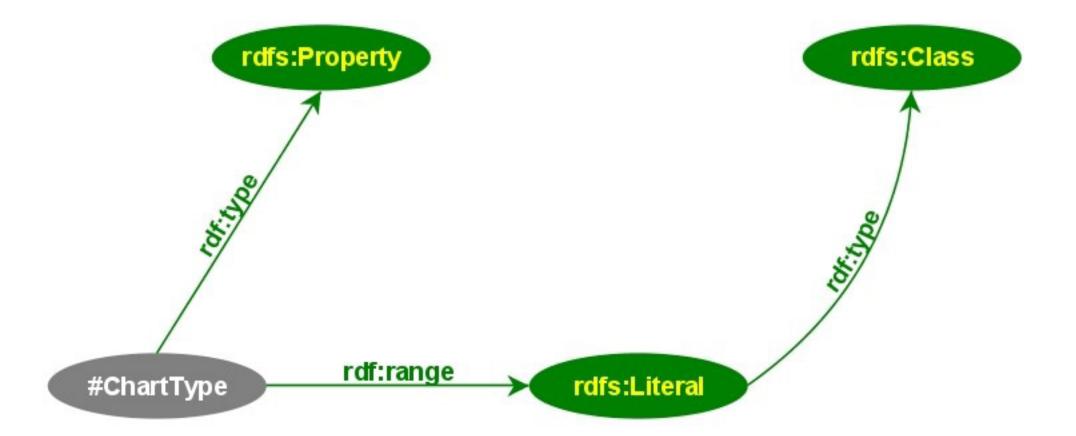


#### Need for Enumeration



#### • Remember this issue?

- one can use XML Schema types to define an enumeration for ChartType, but...
- ...wouldn't it be better to do it within RDF?

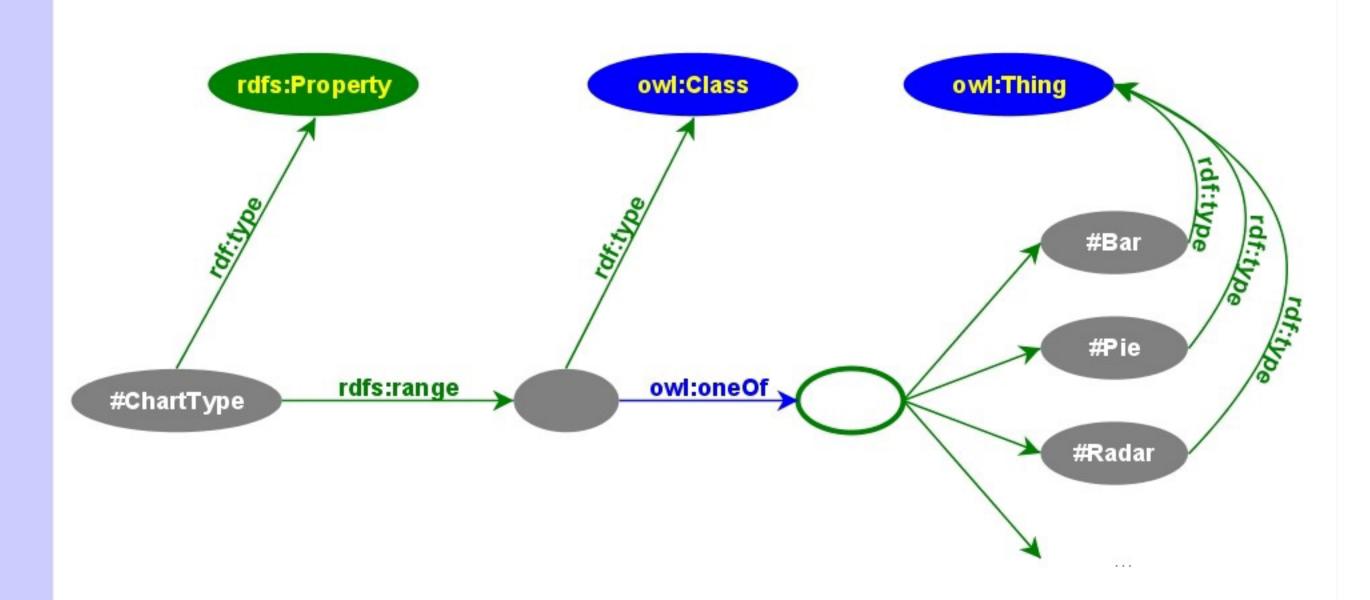




# (OWL) Classes can be Enumerated



• The OWL solution, where possible content is explicitly listed:





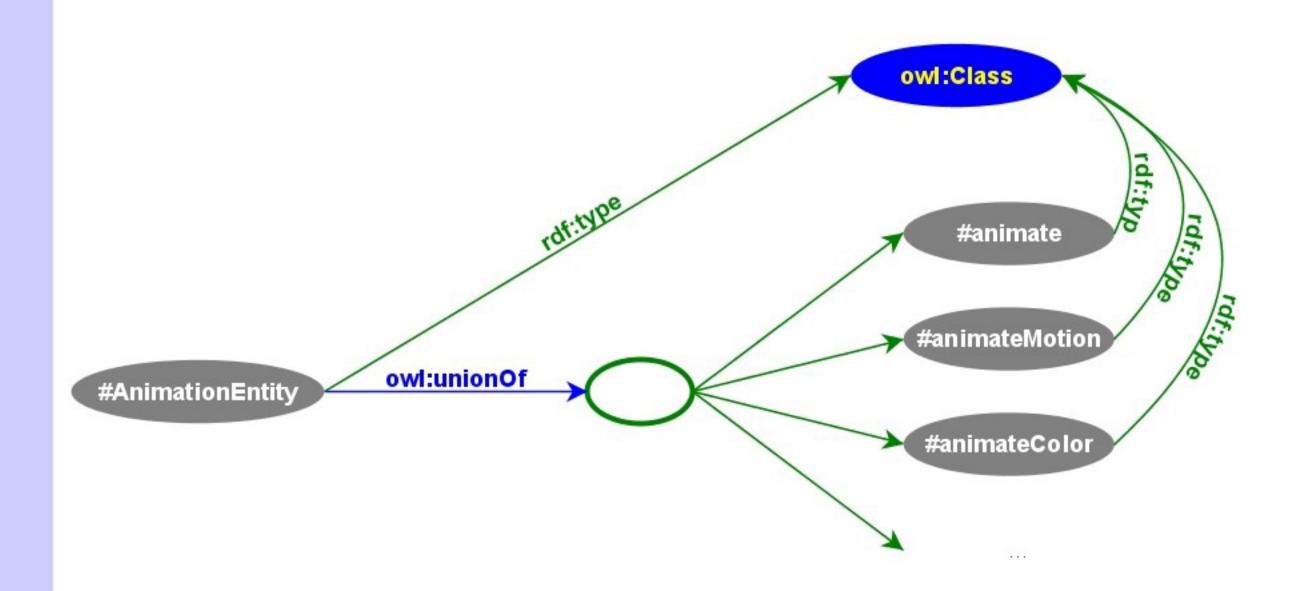
#### **Enumeration in XML:**



### Union of Classes



• Essentially, set-theoretical union:





# Same in RDF/XML



#### Union in XML:

Other possibilities: complementOf, intersectionOf



## Property Restrictions



- (Sub)classes can be created by restricting the behavior of a property on that class
  - "a dolphin is a mammal living in water"
  - we restrict the value of "living in"
- Restriction may be by:
  - value constraints (i.e., further restrictions on the range)
    - o all values must be from a class
    - o at least one value must be from a class
  - cardinality constraints
    - (i.e., how many times the property can be used on an instance?)
    - minimum cardinality
    - maximum cardinality
    - exact cardinality



# Property Restrictions (cont.)



#### Formally:

- owl: Restriction defines a blank node with restrictions
  - refer to the property that is constrained
  - o define the restriction itself
- one can, e.g., subclass from this node

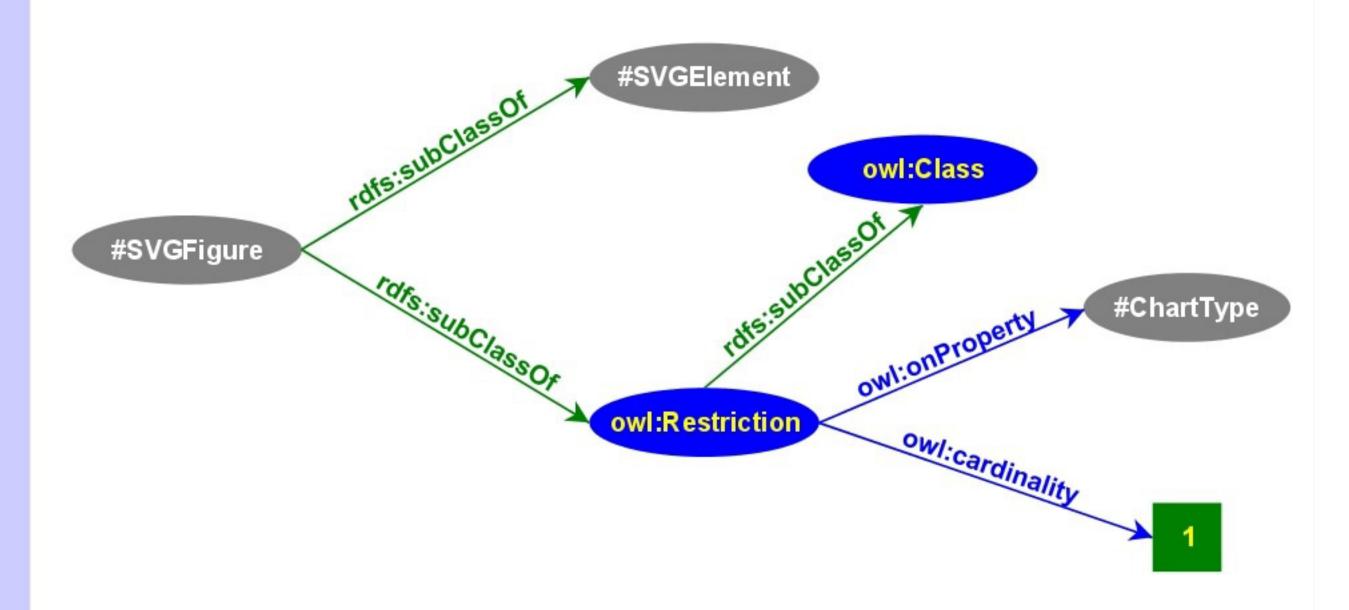


#### OWL

#### **Cardinality Restriction Example**



 "An SVG figure is an SVG element that have a single chart type":





# Same in RDF/XML



#### Cardinality constraint in XML:

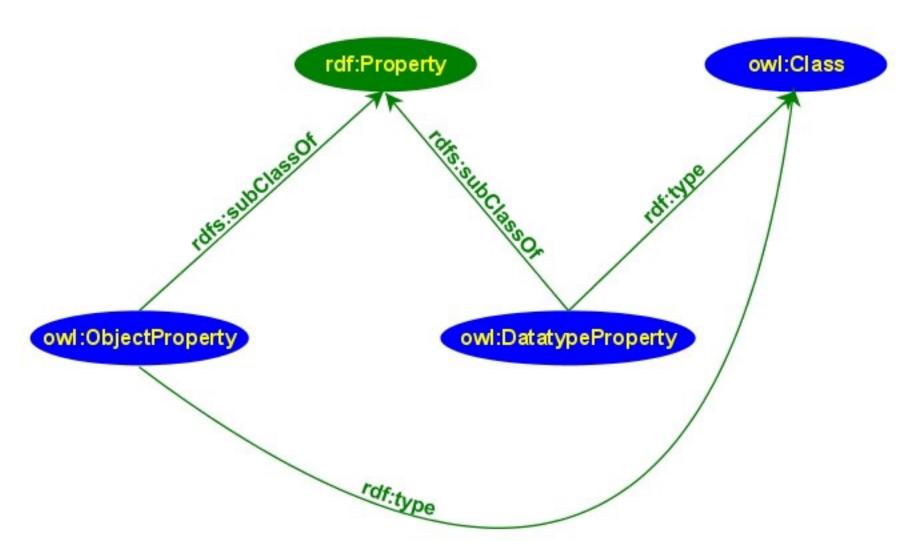
- Note the usage of a typed literal
- cardinality could be replaced by:
  - minCardinality, maxCardinality
  - someValuesFrom, allValuesFrom



# Property Characterization



- In RDFS, properties are constrained by domain and range
- In OWL, one can also characterize their behavior
  - symmetric, transitive, functional, etc
- OWL separates data properties
  - "datatype property" means that its range are typed literals

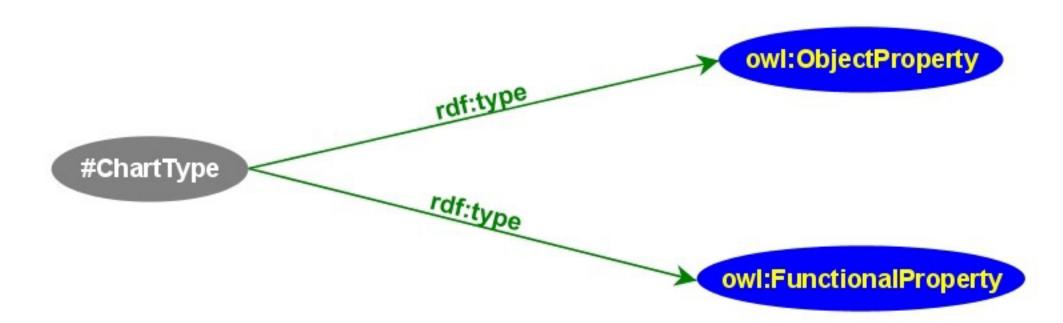




# **Characterization Example**



An alternative for the cardinality=1 setting:





# Same in RDF/XML



#### Characterization in XML:

```
<owl:ObjectProperty rdf:ID="ChartType">
     <rdf:type rdf:resource="..../#FunctionalProperty/>
</owl:ObjectProperty>
```

- Similar characterization possibilities:
  - InverseFunctionalProperty
  - TransitiveProperty, SymmetricProperty
- Range of DatatypeProperty can be restricted (using XML Schema)
- These features can be extremely useful for ontology based applications!



#### **OWL: Additional Requirements**



#### Ontologies may be extremely a large:

- their management requires special care
- they may consist of several modules
- come from different places and must be integrated
- Ontologies are on the Web. That means
  - applications may use several, different ontologies, or...
  - ... same ontologies but in different languages
  - equivalence of, and relations among terms become an issue



## Term Equivalence/Relations



- For classes:
  - owl:equivalentClass: two classes have the same individuals
  - owl:disjointWith: no individuals in common
- For properties:
  - owl:equivalentProperty: equivalent in terms of classes
  - owl:inverseOf: inverse relationship
- For individuals:
  - owl:sameAs: two URI refer to the same individual (e.g., concept)
  - owl:differentFrom: negation of owl:sameAs



# **Example: Connecting to Finnish**





#SVGEntity

owl:equivalentClass

→ http://..../SVGEntiteetti





#### **Another Use of Equivalence**



 Equivalence can also be used for a complete specification of a class:



#### Versioning, Annotation



- Special class owl:Ontology with special properties:
  - owl:imports, owl:versionInfo, owl:priorVersion
  - owl:backwardCompatibleWith,owl:incompatibleWith
  - rdfs:label, rdfs:comment can also be used
- One instance of such class is expected in an ontology file
- Deprecation control:
  - owl:DeprecatedClass, owl:DeprecatedProperty types



# OWL and Logic



- OWL expresses a small subset of First Order Logic
  - it has a "structure" (class hierarchies, properties, datatypes...),
     and "axioms" can be stated within that structure only
  - i.e., OWL uses FOL to describe "traditional" ontology concepts...
     ...but it is not a general logic system per se!
- Inference based on OWL is within this framework only
  - it seems modest, but has proven to be remarkably useful...
  - people in knowledge representation know that!



#### Examples for Logic Formalism



• The transitivity of leftOf is:

```
\forall x,y,z: (x \text{ leftOf } y \land (y \text{ leftOf } z)) \Rightarrow (x \text{ leftOf } z))
```

Cardinality restriction:

```
\forall x: ((x \in X) \land (X \subseteq dom(prop))) \Rightarrow (\exists !y: x prop y)
```

- Union, intersection, etc., can be trivially formalized, too
- etc.
- But, again: this is a restricted form of FOL only!



#### OWL

#### **However: Ontologies are Hard!**



- A full ontology-based application is a very complex system
- Hard to implement, may be heavy to run...
- ... and not all applications may need it!
- Three layers of OWL are defined: Lite, DL, and Full
  - increasing level of complexity and expressiveness
    - "Full" is the whole thing
    - "DL (Description Logic)" restricts Full in some respects
    - "Lite" restricts DL even more



#### OWL Full



- No constraints on the various constructs
  - owl:Class is equivalent to rdfs:Class
  - owl:Thing is equivalent to rdfs:Resource
- This means that:
  - Class can also be an individual
    - it is possible to talk about class of classes, etc.
  - one can make statements on RDFS constructs
    - declare rdf:type to be functional...
  - o etc.
- A real superset of RDFS



# OWL Description Logic (DL)



- Goal: maximal subset of OWL Full against which current research can assure that a decidable reasoning procedure is realizable
- owl:Class, owl:Thing, owl:ObjectProperty, and owl:DatatypePropery are strictly separated
  - i.e., a class cannot be an individual of another class
  - object properties' values must be an owl: Thing
    - except for rdf:type, rdfs:subClassOf, ...
- No mixture of owl:Class and rdfs:Class in definitions
  - essentially: use OWL concepts only!
- No statements on RDFS resources
- No characterization of datatype properties possible
- No cardinality constraint on transitive properties
- Some restrictions on annotations



#### OWL Lite



- Goal: provide a minimal useful subset, easily implemented
  - simple class hierarchies can be built
  - property constraints and characterizations can be used
- All of DL's restrictions, plus some more:
  - class construction can be done only through:
    - o intersection
    - property constraints



### "Description Logic"



- The term refers to an area in knowledge representation
  - a special type of "structured" First Order Logic
  - there are several variants of Description Logic
  - i.e., OWL DL is an embodiment of a Description Logic
- Traditional DL terms sometimes used (by experts...):
  - o "named objects, concepts": definition of classes, individuals, ...
  - "axioms": e.g., subclass or subproperty relationships, ...
  - "facts": statements about individuals (owl: Thing-s)

none of these are "standardized" in W3C...
but you may see them in papers, references



## Ontology Examples



- A possible ontology for our graphics example
  - on the borderline of DL and Full
- International country list
  - example for an OWL Lite ontology
- The hard work is to create the ontologies
  - requires a good knowledge of the area to be described
  - some communities have good expertise already (e.g., librarians)

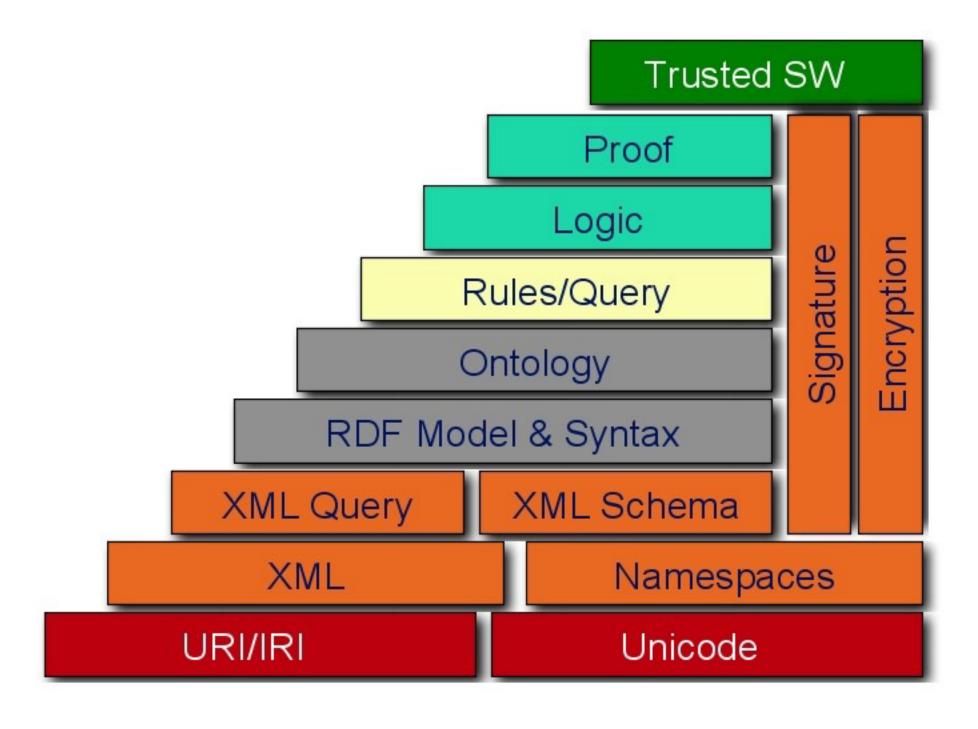




#### **PART VI: Future Developments**









#### **Semantic Web Activity Phase 2**



- First phase (completed): core infrastructure
- Second phase: promotion and implementation needs
  - relevant working groups
  - outreach to user communities
    - life sciences
    - geospatial information systems
    - libraries and digital repositories
    - 0 ...
  - intersection of SW with other technologies
    - Semantic Web Services
    - privacy policies
    - 0 ...





#### "Best Practices" Work



- "Semantic Web Best Practices and Deployment"
  - recommendations for practical deployment
  - engineering guidelines
  - ontology/vocabulary development practices
  - educational material
  - effective demonstrations
  - information on applications
  - o etc.
- Goal is to increase awareness on SW
- W3C started work in this area recently
  - some initial drafts are already available



# RDF Data Access (a.k.a. Queries)



In Python, for example, one uses:

```
# do something with (p,o) pairs
for (p,o) in triples.predicate_objects(subject) :
    do_something(p,o)
"predicate_objects" returns a subgraph
```

- Applications may want more
  - i.e., return complex subgraph with parts missing
- Very important for large and distributed RDF depositories
- There are more than 20 RDF Query languages



# **Data Access Example**



One may want something like:

```
SELECT (a,b)
WHERE [?x 'parent' a] and [b 'brother' ?x]
(i.e., 'b is the uncle of a')
```

- W3C started a standardization work in this area recently
  - precise relationships to XML Query has to be defined
  - concentrates also on *protocols* to extract subgraphs
     e.g., using SOAP
- Such facilities already implemented in Jena, RAP,...





- OWL can be used for simple inferences
- Applications may require more, e.g., Horn clauses:
  - ∘ (ant-1  $\wedge$  ant-2  $\wedge$  ...) ⇒ (cons-1  $\wedge$  cons-2  $\wedge$  ...)
  - ∘ e.g.:
    - for any «X», «Y» and «Z»:
       "if «Y» is a parent of «X», and «Z» is a brother of «Y» then «Z» is the uncle of «X»"
    - using a logic formalism:
       ∀x,z: ((∃y: (y parent x) ∧ (y brother z)) ⇒ (z uncle x))
- Lots of research is happening to extend RDF/OWL (RuleML, SWRL, cwm, ...)
- W3C may initiate a standardization work in this area, too
  - question is whether results are "ripe" for standardization
  - o and whether the necessary manpower is available



# RDF API-S



- We have seen Jena and RDFLib
- There are lots of other programming environments
  - Redland, RDFStore, RAP, etc.
- Each use their own "view" on binding RDF to programming concepts
- A standardization would enhance interoperability
  - similar to the DOM Specification for XML:
    - common vocabulary is developed in terms of OMG's IDL
    - there are IDL "bindings" to C, C++, Python, etc.
- W3C may initiate a standardization work in this area, or ...
- ... leave it to others to standardize in practice
  - (it is not clear whether this is the task of W3C)



# Future Developments Trust



#### Can I trust a metadata on the Web?

- is the author the one who claims he/she is?
- can I check the credentials?
- can I trust the inference engine?
- what about IPR of the metadata?
- o etc.

## Some of the basic building blocks are available:

- XML Signature/Encryption
- XML based Key Management is in preparation

## Much is missing, e.g.:

- a "canonical" form of RDF/XML
  - necessary for unambiguous signatures
- exhaustive tests for inference engines
- o protocols to check, for example, a signature
- It is on the "future" stack of W3C...



## A Number of Research Issues Still...



- Knowledge representation is an active R&D area:
  - temporal & spatial reasoning
  - fuzzy logic
  - improve the inference algorithms and implementations
  - improve scalability
  - reasoning with OWL Full
  - 0 ...
- They usually happen outside of W3C, though
  - W3C is not a research entity...





# **PART VII: Available Documents, Tools**



# **Available Specifications: Primers**



#### **RDF Primer**

URI: http://www.w3.org/TR/rdf-primer

#### **OWL** Guide

URI: http://www.w3.org/TR/owl-guide/

#### **RDF Test Cases**

URI: http://www.w3.org/TR/rdf-testcases/

#### **OWL Test Cases**

URI: http://www.w3.org/TR/owl-test/



## **Available Specifications: RDF**



## RDF: Concepts and Abstract Syntax

URI: http://www.w3.org/TR/rdf-concepts/

Note: there is a previous Recommendation of 1999 that is

superseded by these

#### **RDF Semantics**

URI: http://www.w3.org/TR/rdf-mt/

Precise, graph based definition of the semantics

This is primarily for implementers

## **RDF/XML Serialization**

URI: http://www.w3.org/TR/rdf-syntax-grammar/

#### **N3 Serialization Primer**

URI: http://www.w3.org/2000/10/swap/Primer

Note: this is not part of the W3C Recommendation track!



## **Available Specifications: Ontology**



## RDF Vocabulary Description Language (RDF Schema)

URI: http://www.w3.org/TR/rdf-schema/

#### **OWL Overview**

URI: http://www.w3c.org/TR/owl-features/

#### **OWL Reference**

URI: http://www.w3c.org/TR/owl-ref/

## **OWL Semantics and Abstract Syntax**

URI: http://www.w3c.org/TR/owl-semantics/

## **OWL Use Cases and Requirements**

URI: http://www.w3.org/TR/webont-req/



# Some Books



- M. Dertouzos: The Unfinished Revolution (1995)
  - an early "vision" book (not only on the Semantic Web)
- T. Berners-Lee: Weaving the Web (1999)
  - another "vision" book
- J. Davies, D. Fensel, F. van Harmelen: Towards the Semantic Web (2002)
- S. Powers: Practical RDF (2003)
- D. Fensel, J. Hendler: Spinning the Semantic Web (2003)
- G. Antoniu, F. van Harmelen: Semantic Web Primer (2004)



## **Further Information**



## Bristol University

- http://www.ilrt.bristol.ac.uk/discovery/rdf/resources/
- huge list of documents, publications

## Semantic Web Community Portal

- http://www.semanticweb.org/
- "Business model IG" (part of the portal)
- huge set of links to documents, software, ...

#### SemWeb Central

- http://semwebcentral.org
- Open Source development archive

## W3C team public presentations:

http://www.w3.org/2001/sw/EO/talks

## W3C's Semantic Web home page:

http://www.w3.org/2001/sw/



# Further Information (cont)



- Full, interactive view of the RDFS and OWL definitions
  - requires an SVG client
- References on Description logic:
  - Online courses: http://dl.kr.org/courses.html
  - A general introduction: http://www.inf.unibz.it/~franconi/dl/ course/dlhb/dlhb-01.pdf
- Ontology Development 101
  - URI: http://protege.stanford.edu/publications/
     ontology\_development/ontology101-noy-mcguinness.html
- OWL Reasoning Examples:
  - URI: http://owl.man.ac.uk/2003/why/latest/
- Lots of papers at WWW2003 and WWW2004



## **Public Fora at W3C**



## **Semantic Web Interest Group**

a forum for discussions on applications URI: http://www.w3.org/RDF/Interest

## RDF Logic

public (archived) mailing list for technical discussions URI: http://lists.w3.org/Archives/Public/www-rdf-logic/



## **Some Tools**



## (Graphical) Editors

- IsaViz (Xerox Research/W3C)
- RDFAuthor (Univ. of Bristol)
- Longwell (MIT)
- Protege 2000 (Stanford Univ.)
- SWOOP (Univ. of Maryland)
- Orient (IBM Alphawork)

0 ...

Further info on RDF/OWL tools at:

http://www.w3.org/2001/sw/WebOnt/impls, or

http://semwebcentral.org

## **Programming environments**

We have already seen some but Jena 2 and SWI-Prolog do OWL reasoning, too!



# Some Tools (Cont.)



#### **Validators**

- For RDF:
  - http://www.w3.org/RDF/Validator/
- For OWL:
  - http://owl.bbn.com/validator/
  - http://phoebus.cs.man.ac.uk:9999/OWL/Validator
  - http://www.mindswap.org/2003/pellet/demo.shtml

## Ontology converter (to OWL)

at http://www.mindswap.org/2002/owl.html

## Schema/Ontology registries

e.g., SchemaWeb, SemWeb Central, ...





## **PART VII: Some Application Examples**



# Some Application Examples SW Applications



## Large number of applications emerge

- some applications use RDF only
- others use ontologies, too
  - huge number of ontologies exist, using proprietary formats
  - converting them to RDF/OWL will be a major task (but there are converters)
  - but it will be worth it!

## SWAD-Europe survey:

- URI: http://www.w3.org/2003/11/SWApplSurvey
- lists more than 50 applications in 12 categories...
- and is already more than a years old!

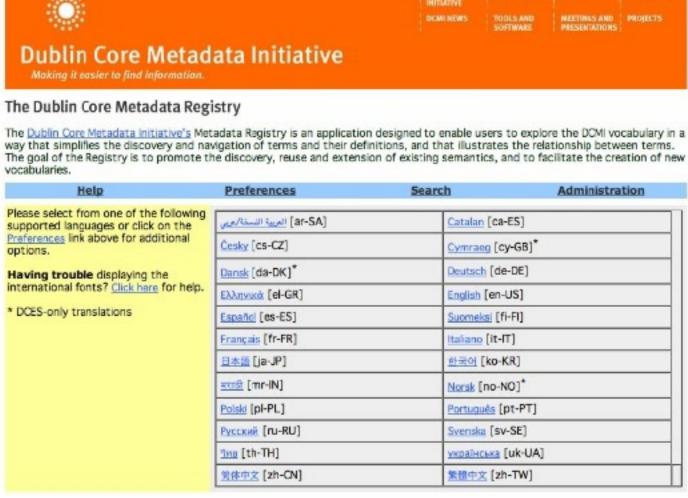
#### Some Application Examples

# **SW** Application Examples



#### **Dublin Core**

- vocabularies for distributed Digital Libraries
- one of the first metadata vocabularies in RDF
- URI: http://www.dublincore.org
- extensions exist, eg, PRISM that includes digital right tracking



©1994-2004, W3C (MIT, ERCIM, Keio)

Ivan Herman, W3C 126 (134)

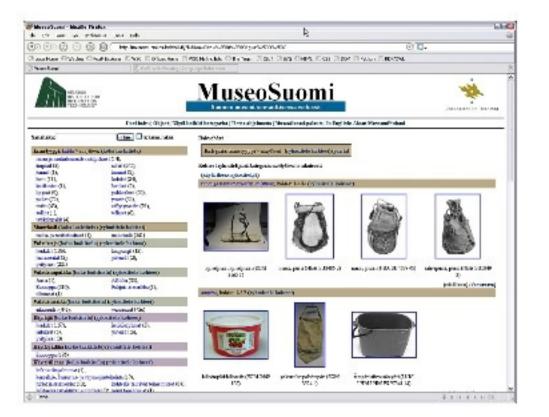


# SW Application Examples (cont)



## **Data integration**

- achieve semantic integration of corporate resources or different databases
- RDF/RDFS/OWL based vocabularies as an "interlingua" among system components
- Boeing example: http://www.cs.rutgers.edu/~shklar/www11/ final\_submissions/paper3.pdf
- similar approaches: Artiste project, MITRE Corp., MuseoSuomi, ...
- there are companies specializing in the area





# SW Application Examples (cont)



#### Sun's SwordFish

- Sun provides assisted support for its products, handbooks, etc
- Public queries go through an internal RDF engine for, eg:
  - Sun's White Papers collection
     (http://www.sun.com/servers/wp.html/)
  - Sun's System Handbooks collection
     (http://sunsolve.sun.com/handbook\_pub/)

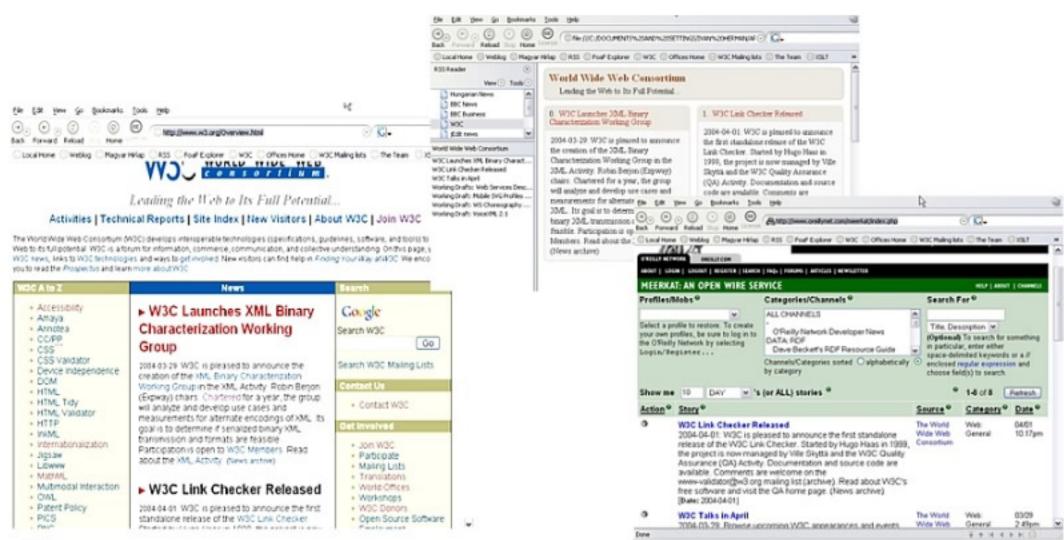
#### Some Application Examples

# SW Application Examples (cont)



## Web Content Syndication (RSS)

- can be used to specify the important content of a page
- there is a Yahoo discussion group and (non-W3C) working group
- URI: http://purl.org/rss/
- widely used in the weblog world!
- example: W3C home page syndicated



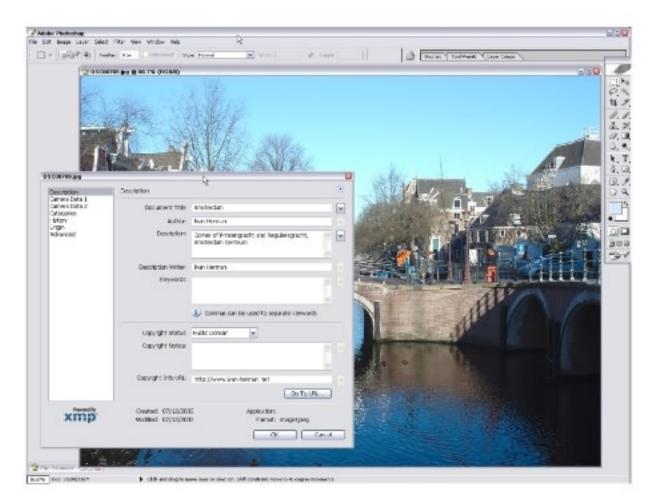
#### **Some Application Examples**

# SW Application Examples (cont)



#### **XMP**

- Adobe's tool to add RDF-based metadata to all their file formats
  - o eg, Photoshop in Creative Suite
  - millions of people use RDF without knowing it...
- the tool is available for all!
- URI: http://www.adobe.com/products/xmp/main.html





# SW Application Examples (cont)



#### Mozilla

internal data are stored in RDF (eg, bookmarks, conf. files)

#### **Brandsoft**

- entreprise Web Management
- all business models are stored in RDF
- easy to set up internal rules

#### **Creative Commons**

- an environment to express rights of digital content on the Web
  - legal constraints referred to in RDF, added to pages
- there are specialized browsers, browser plugins
- more than 1,000,000 users worldwide(!)
  - without knowing that they use RDF...



# SW Application Examples (cont)



## **Baby CareLink**

- centre of information for the treatment of premature babies
- provides an OWL service as a Web Service
  - combines disparate vocabularies like medical, insurance, etc
  - remember: ontology is hard!
  - users can add new entries to ontologies
  - complex questions can be asked through the service
- perfect example for the synergy of Web Services and the Semantic Web!





## **Further Information**



#### These slides are at:

http://www.w3.org/2004/Talks/0209-Helsinki-IH/

## Semantic Web homepage

http://www.w3.org/2001/sw/

#### More information about W3C:

http://www.w3.org//Consortium/

#### Finnish Office of W3C

http://www.w3c.tut.fi/

#### Mail me:

ivan@w3.org





# **Questions?**