University of
Southampton

COMP6256
Knowledge Graphs for AI Systems

Dr Nicholas Gibbins - nmg@ecs.soton.ac.uk

University of Southampton

Course Aims

- Understand the key ideas behind knowledge graphs and the Semantic Web
- Explain the state of the art in SW and knowledge graph technologies
- Gain practical experience of ontology design in OWL
- Understand the future directions of the Semantic Web, and its relationship with other Web developments

Lecturers

Dr Nicholas Gibbins
nmg@ecs.soton.ac.uk

Prof. Steffen Staab
srs2m14@ecs.soton.ac.uk

Dr Shoaib Jameel
M.S.Jameel@southampton.ac.uk

Course Structure

Three lectures per week:

- Monday 11.00 in 54/7035
- Wednesday 9.00 in 54/10031
- Thursday 11.00 in 13/3019

Teaching Schedule

Week 18:	Introduction, RDF and Linked Data (nmg)
Week 19:	Linked Data and SPARQL (nmg)
Week 20:	Ontologies, RDF Schema, Description Logics (nmg)
Week 21:	OWL, Protégé, Ontology Engineering (nmg)
Week 22:	Shacl, schema.org (srs)
Week 23:	Knowledge graphs and property graphs (nmg/sj)
Week 24:	RDF query processing, ontology alignment (srs)
Week 25:	Knowledge graph embedding (srs)

EASTER VACATION

Teaching Schedule

Week 30: *Rules, OWL2 Reasoning, OWL2 EL* (sj)

Week 31: *Open/closed world queries, OWL2 RL and Datalog* (sj)

Week 32: *OWL2 QL, Chase, query rewriting, ontology-based data integration* (sj)

Week 33: Review

Topics in *italics* are currently provisional

Assessment

Examination: 75% (120 minutes, 3 questions from 5)

Ontology design coursework: 25%

- Detailed specification published in week 21
- Submission due week 30
- Feedback due week 33

Industrial Action

There are 18 days of industrial action scheduled between now and Easter:

- Week 18: **1 Feb**
- Week 19: **9 Feb**, 10 Feb
- Week 20: 14 Feb, **15 Feb**, **16 Feb**
- Week 21: 21 Feb, **22 Feb**, **23 Feb**
- Week 22: **27 Feb**, 28 Feb, **1 Mar**, **2 Mar**
- Week 24: **16 Mar**, 17 Mar
- Week 25: **20 Mar**, 21 Mar, **22 Mar**

There are COMP6256 lectures scheduled on the **highlighted** dates

These lectures may not take place

Introduction to the Semantic Web

University of Southampton

The World Wide Web: Past, Present and Future

> a goal of the Web was that, if the interaction between person and hypertext could be so intuitive that the machine-readable information space gave an accurate representation of the state of people's thoughts, interactions, and work patterns, then machine analysis could become a very powerful management tool, seeing patterns in our work and facilitating our working together

Berners-Lee, T. (1996) The World Wide Web: Past, Present and Future.
https://www.w3.org/People/Berners-Lee/1996/ppf.html

University of Southampton

Weaving the Semantic Web

I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines.

T. Berners-Lee (1999) Weaving the Web. San Francisco, CA: Harper

What is the Semantic Web?

"The goal of the Semantic Web initiative is as broad as that of the Web: to create a universal medium for the exchange of data. It is envisaged to smoothly interconnect personal information management, enterprise application integration, and the global sharing of commercial, scientific and cultural data. Facilities to put machine-understandable data on the Web are quickly becoming a high priority for many organizations, individuals and communities.

The Web can reach its full potential only if it becomes a place where data can be shared and processed by automated tools as well as by people."

W3C (2013) Semantic Web Activity Statement. https://www.w3.org/2001/sw/Activity

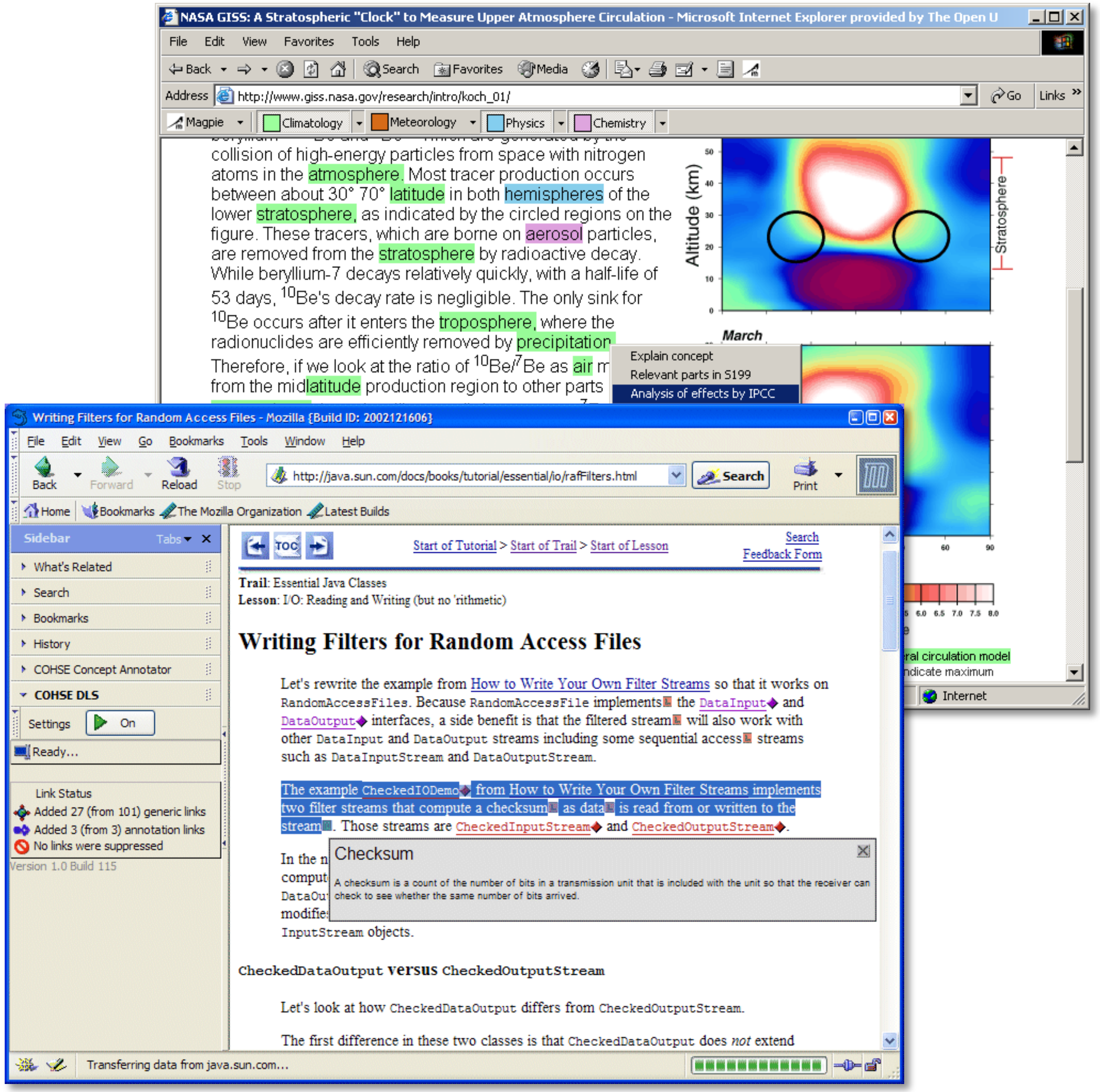
University of Southampton

The Annotated Web

Add structure to unstructured data

- Annotate existing web pages
- Classify web pages
- Use natural language techniques to extract information from web pages

Annotations enable enhanced browsing and searching



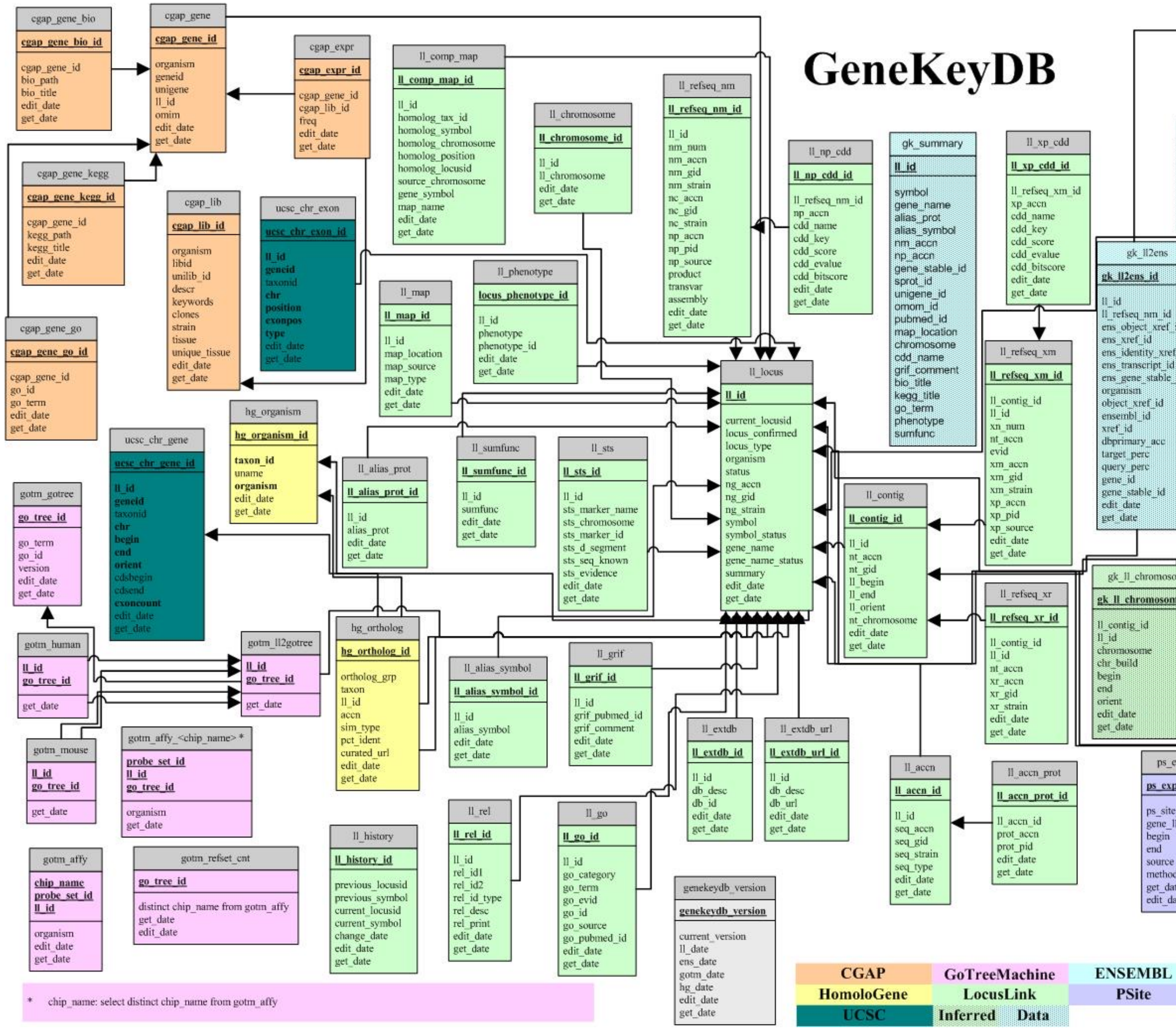
The Web of (Linked) Data

Make the most of the structure you already have

- Expose existing databases in a common format
- Express database schemas in a machine-understandable form

Common format allows the integration of data in unexpected ways

Machine-understandable schemas allow reasoning about data



University of Southampton

Rocket Science (not)

Is this rocket science? Well, not really. The Semantic Web, like the World Wide Web, is just taking well established ideas, and making them work interoperably over the Internet. This is done with standards, which is what the World Wide Web Consortium is all about. We are not inventing relational models for data, or query systems or rule-based systems. We are just webizing them.

Berners-Lee, T. (2001) Business Model for the Semantic Web, http://www.w3.org/DesignIssues/Business

Basic Concepts

The World Wide Web vs. the Semantic Web

The World Wide Web is the Web for people

- Information is predominantly textual
- Technologies include URI, HTTP, XML, HTML

The Semantic Web is the Web for machines

- Information needs to be structured
- Technologies include RDF, RDFS, OWL
  (in addition to those for the Web)

Machine readable vs. machine understandable

On the World Wide Web, information needs humans to give it interpretation

- Information is predominantly natural language
- Difficult to mediate by software agents

On the Semantic Web, information is structured so that it can be interpreted by machines

- Humans need not interact directly with Semantic Web information – mediation through agents

Formal meaning is critical to understanding

Machine readable vs. machine understandable

XML is a machine readable format:

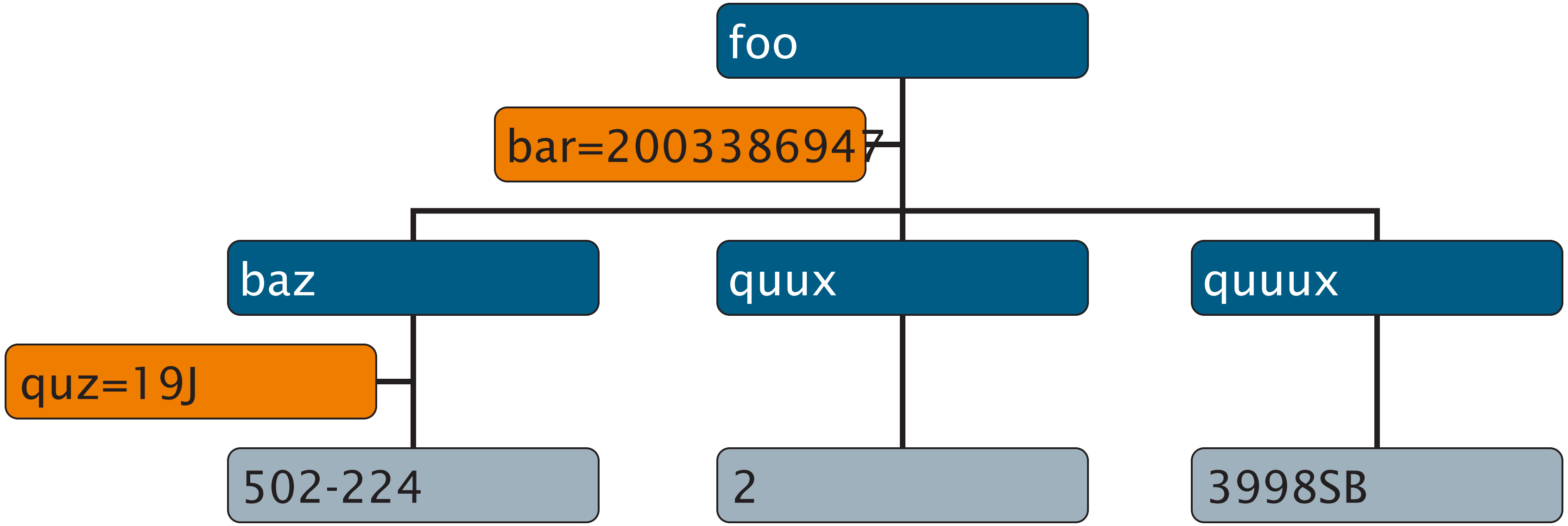
- It can be parsed to give an unambiguous document structure

but

- It has no formal meaning
- Meanings of XML interchange formats must be explicitly agreed

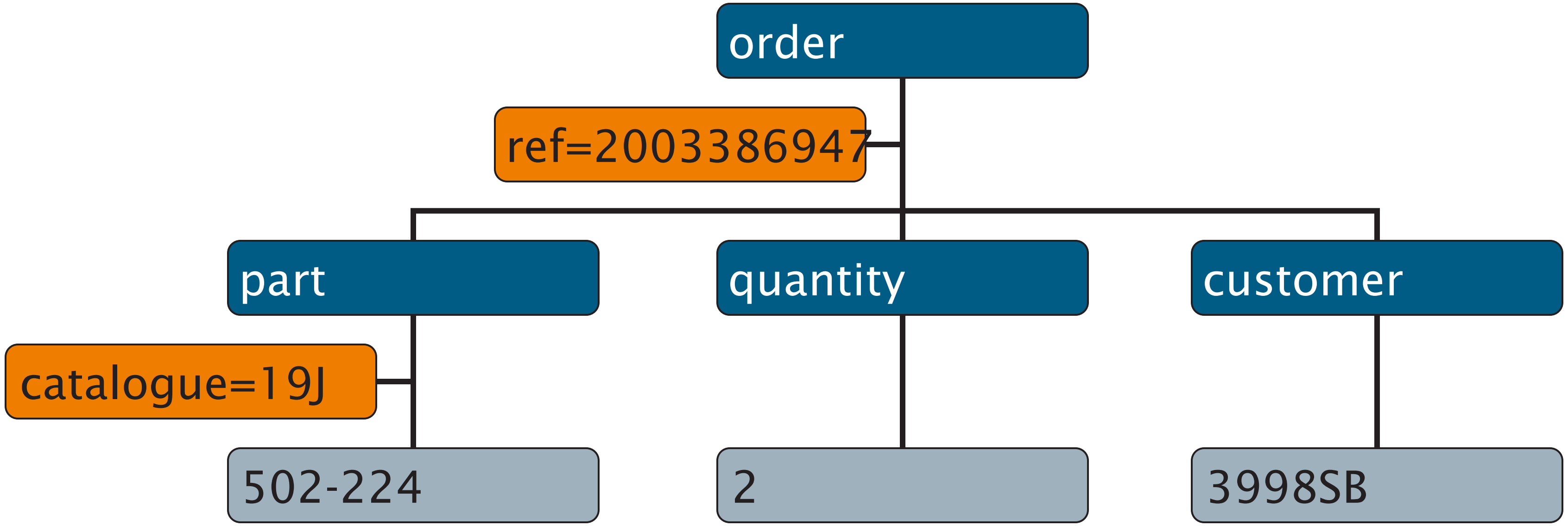
Machine readable: XML

```
<foo bar=“2003386947”>
  <baz qux=“19J”>502-224</baz>
  <quux>2</quux>
  <quuux>3998SB</quuux>
</foo>
```



Machine readable: XML

```
<order ref="2003386947">
  <part catalogue="19J">502-224</part>
  <quantity>2</quantity>
  <customer>3998SB</customer>
</order>
```



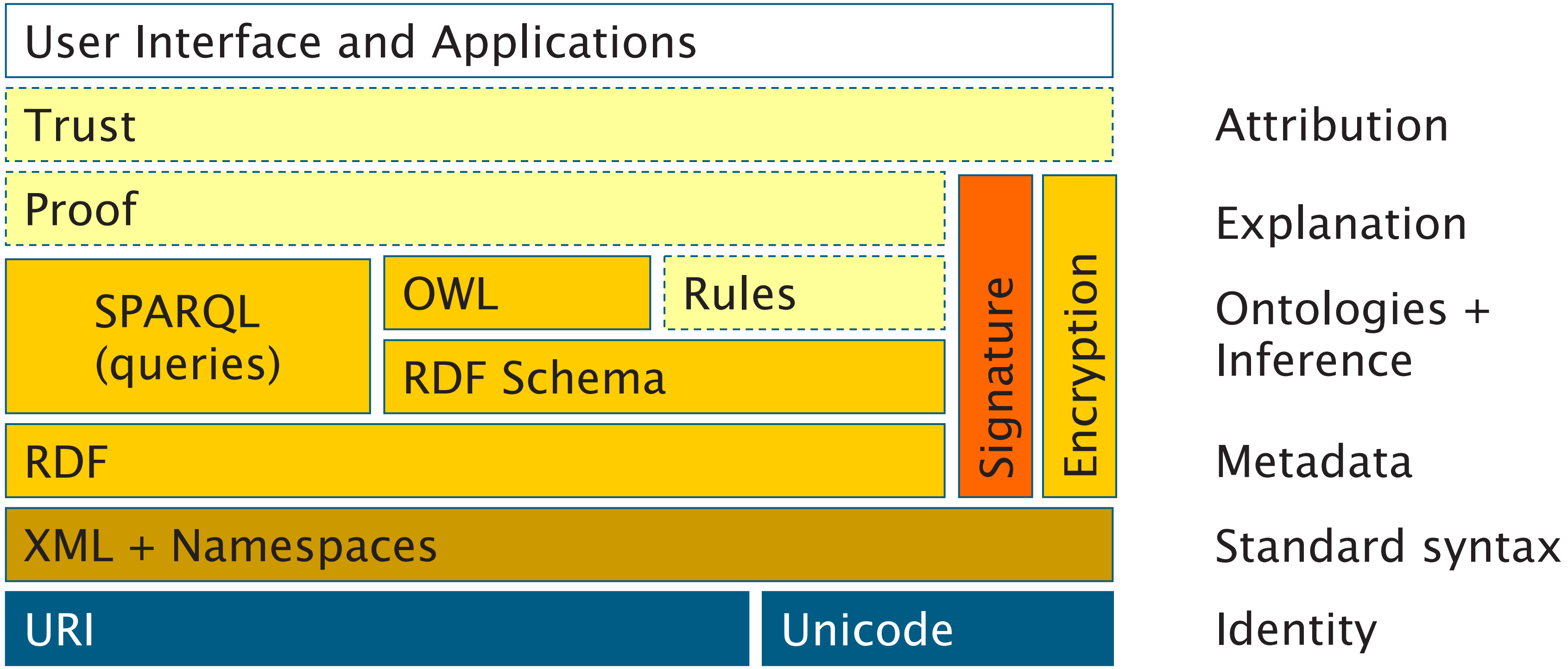
Machine readable vs. machine understandable

RDF is a machine understandable format

- The structures generated by an RDF parser have a formal meaning
- RDF is a framework for interchange formats that provides a base level of common understanding
- RDF provides basic notions of classes and properties
- RDF enables simple inference (certain types of deduction may be made from existing knowledge)

Semantic Web Technical Architecture

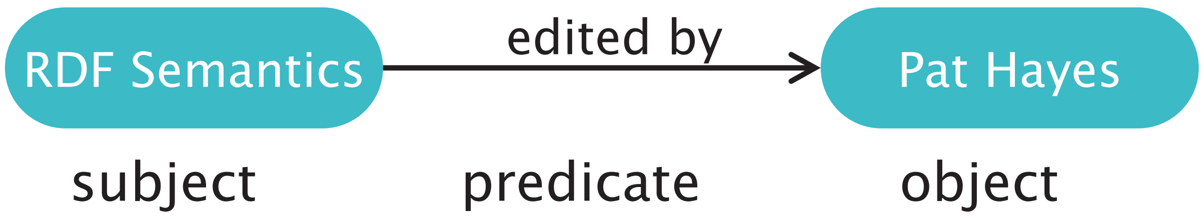
The Semantic Web layer cake



Resource Description Framework

Underlying model of triples used to describe the relations between entities

- Subject-Predicate-Object (compare Entity-Attribute-Value)
- Predicates are analogous to link types



Example

Take a citation:

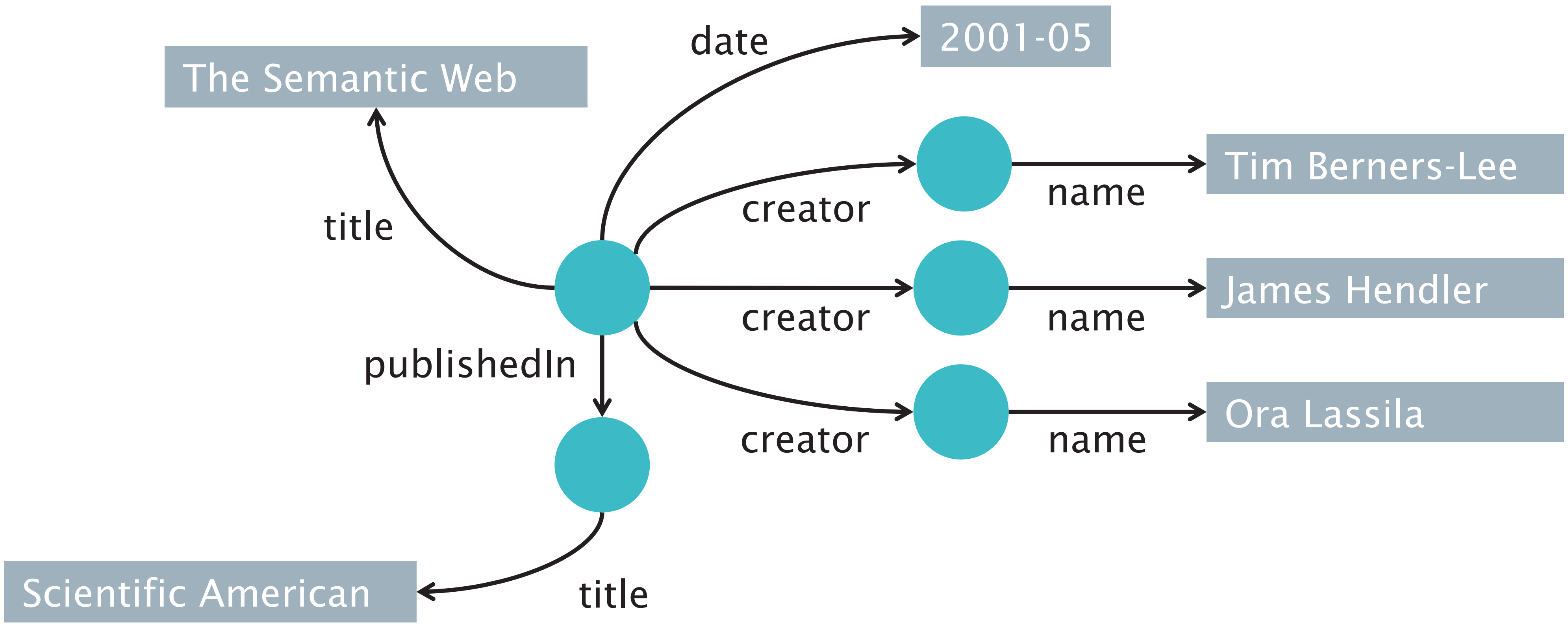
- Tim Berners-Lee, James Hendler and Ora Lassila. The Semantic Web. Scientific American, May 2001

We can identify a number of distinct statements in this citation:

- There is an article titled “The Semantic Web”
- One of its authors is a person named “Tim Berners-Lee” (etc)
- It appeared in a publication titled “Scientific American”
- It was published in May 2001

Example

We can represent these statements graphically:

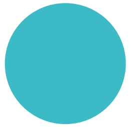


Example

There are two types of node in this graph:

- Literals, which have a value but no identity (a string, a number, a date)

- Resources, which represent objects with identity (a web page, a person, a journal)

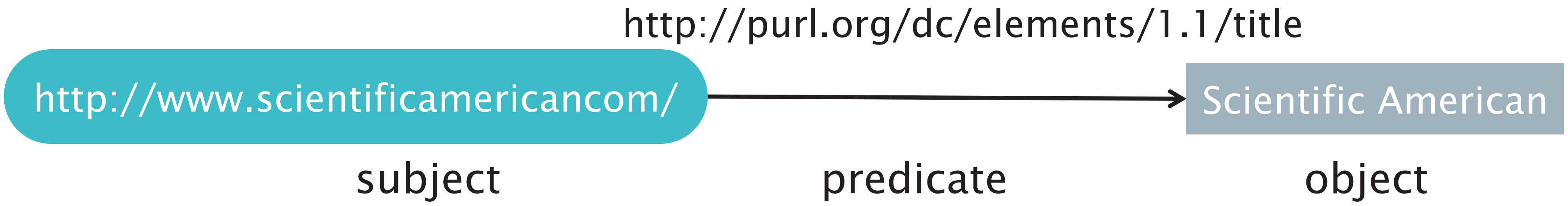


Example

Resources are identified by URIs

Properties are resources that are used as predicates

- Collection of properties constitutes a vocabulary (or ontology)



Resource Description Framework

RDF is a framework for representing information about resources

- Triple-based data model (abstract syntax)
- Uses URIs to identify resources and relations
- Model-theoretic semantics
- Various serialisation formats (RDF/XML, Turtle, JSON-LD, RDFa, etc)

RDF Vocabulary Description Language

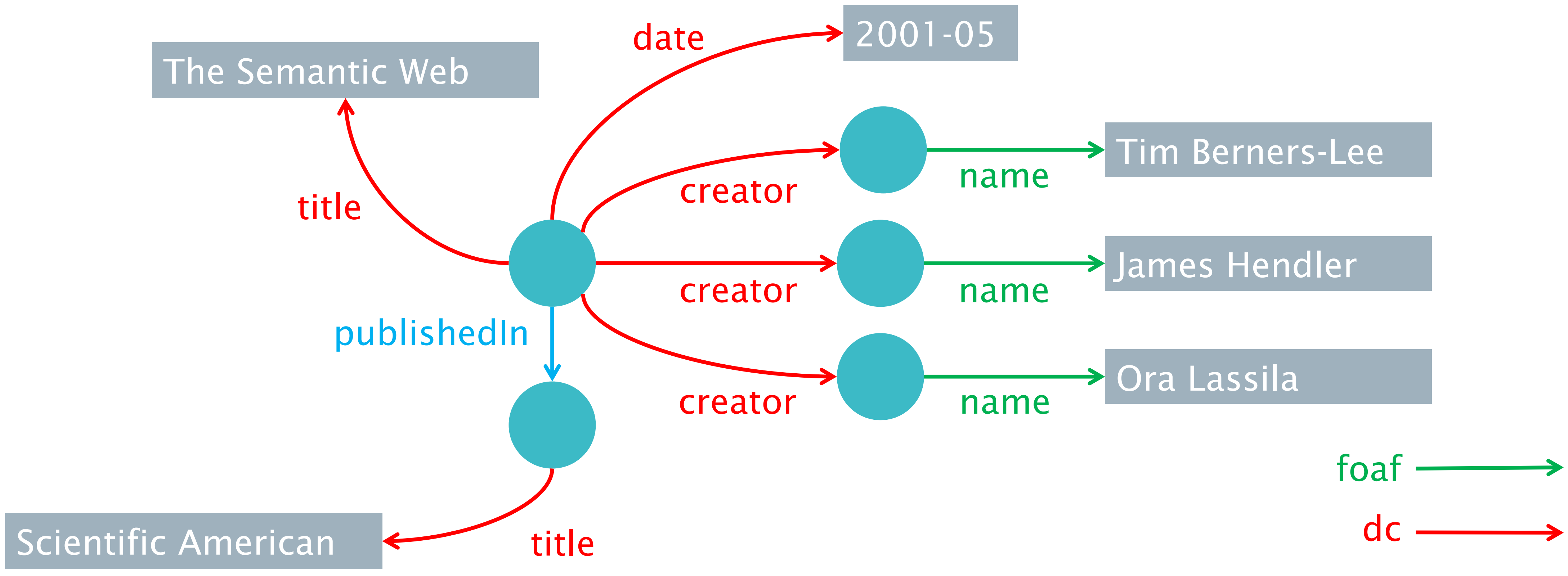
RDF lets us make assertions about resources using a given vocabulary

RDF does not let us define these domain vocabularies by itself

RDF Schema is an RDF vocabulary which we can use to define other vocabularies

- Define classes of objects and their relationship with other classes
- Define properties that relate objects together and their characteristics

Mixing Vocabularies



OWL Web Ontology Language

RDF Schema is not expressive enough for many applications

- Only supports explicit class/property hierarchies
- Only supports global range and domain constraints

OWL provides more expressive features:

- Property restrictions (local range/cardinality/value constraints)
- Equivalence and identity relations
- Property characteristics (transitive/symmetric/functional)
- Complex classes (set operators, enumerated classes, disjoint classes)

SPARQL

The SPARQL Protocol and RDF Query Language

- Expressive SQL-like language for querying RDF systems
- HTTP-based RESTful protocol

Next Lecture:
Vocabularies and Applications