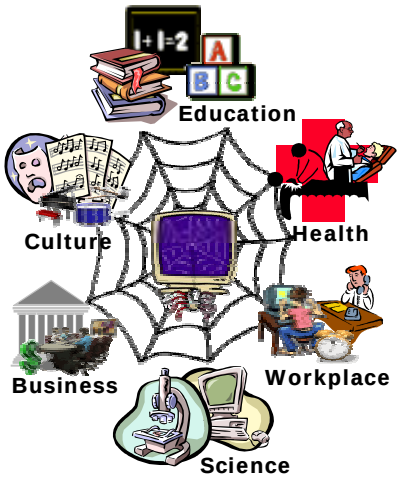
Benchmarking RDF Schemas for the Semantic Web

Aimilia Magkanaraki, Sofia Alexaki, Vassilis Christophides, Dimitris Plexousakis

Institute of Computer Science, FORTH
Heraklion, Crete, Greece

On Community Webs

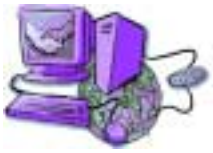
- **The Community Webs**
  - Groups of people sharing a **domain of discourse** and a set of **information resources** (e.g., documents, data, images etc)
  - Use of the **Web** as a medium to **exchange information** and **knowledge**
- **The need**
  - To **describe** information resources in a way that facilitates their **reuse** and **integration** by machines over the Web



ICS-FORTH June 2002

Semantic Web and RDF/S

- Communities can define their own **semantics** and share the same infrastructure for consistent **encoding**, **exchange** and **processing** of resource descriptions
- A standard **representation language** for resource description and domain modeling
  - **Humanly readable**/ **machine understandable** syntax
  - Interpreted within or across communities using **extensible descriptive schemas**
- RDF/S facilitates the development of metadata vocabularies **without central coordination**
  - **Reuse** and/or **refine** existing class and property definitions
  - **Instantiate** existing class and property definitions from different applications schemas

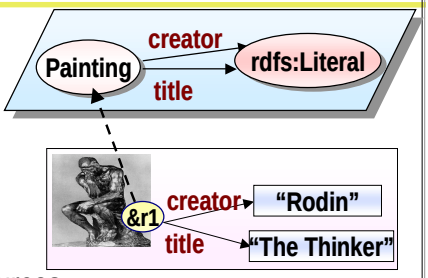


3

ICS-FORTH June 2002

Resource Description Framework/Schema

- **RDF**: Resource Descriptions
  - **Data Model**: Directed Label Graph
- **RDFS**: Schema Vocabularies
  - **Abstraction Mechanisms**
    - (multiple)class/property specialization
    - (multiple) classification of resources
  - **Restriction Mechanisms**
    - domain and range restriction
  - **Documentation Facilities**
    - label, comment, seeAlso, isDefinedBy



- **DAML+OIL** and the forthcoming **OWL** provide additional knowledge representation mechanisms

4

RDF/S & Knowledge Sharing/Reuse



Where can one find RDF Schemas?

xmlns.com



Dublin Core MEtadata Initiative
The Open Metadata Registry Prototypes

ULIS Open Metadata Registry

SWAG Dictionary

SCHEMAS

ZVON.org

DAML Ontology Library

- **Harvesting RDF/S Schemas** on the Web is a **time-consuming task**
- Need for **richer** RDF/S Schemas Registries
  - **Content**
  - **Services**
    - Querying and Browsing Interfaces

Benchmarking RDF Schemas

- Benchmarking RDF/S schemas is crucial
  - **Qualitative** and **quantitative** comparison of RDF/S Schemas

Benefits: **- Effective Registries**

- Efficient Implementations
Validation, Storage, Inference, Query Tools

Presentation Outline

- **Benchmark characteristics**
  - 28 RDF/S Schemas
  - Classification of RDF/S Schemas
    - Domain of discourse (9)
    - Semantic Depth (4)
- **Schema morphology observations**
- **Actual use of RDF/S modeling constructs**
- **Summary**



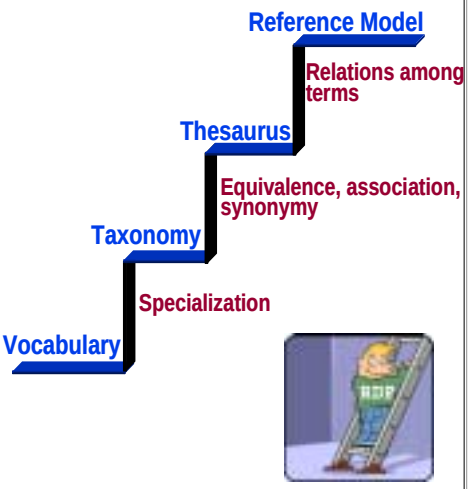
Application Domains

- **Cultural Heritage/ Archives/ Libraries** (3)
- **Educational/ Academic** (4)
- **Publishing/ News** (4)
- **Audio-Visual** (1)
- **Geospatial/ Environmental** (2)
- **Biology/ Medicine** (1)
- **E-Commerce** (3)
- **Ubiquitous/ Mobile/ Grid Computing** (4)
- **Cross-Domain** (6)

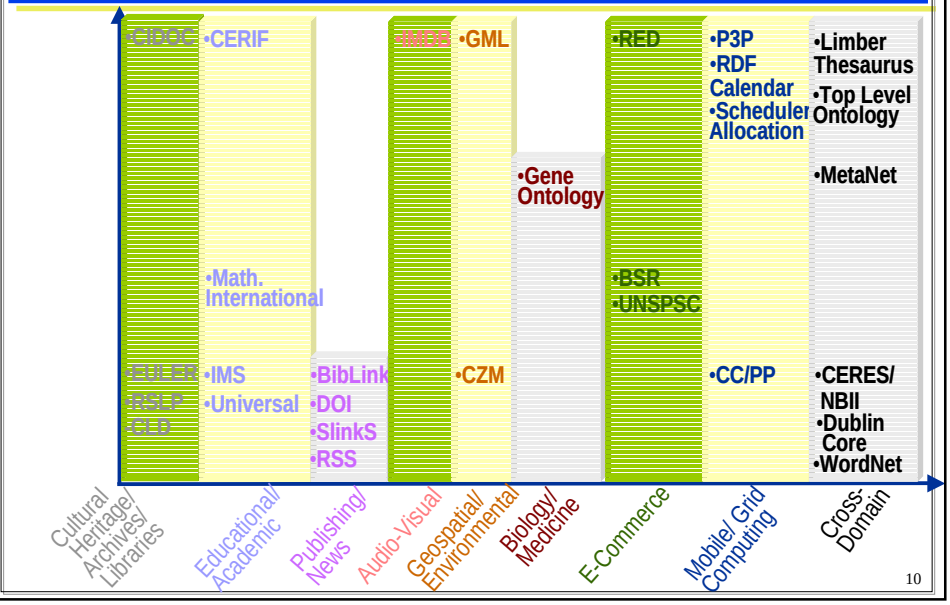


Semantic Depth of Resource Descriptions

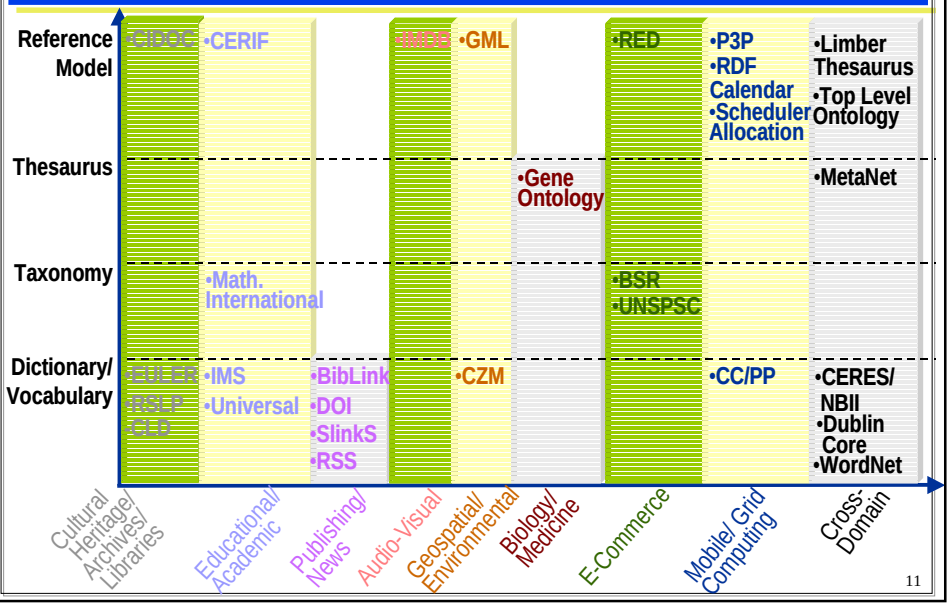
- **Dictionaries/ Vocabularies** (13)
  - simple lists of concepts and their definitions
- **Taxonomies** (3)
  - Specialization between concepts
- **Thesauri** (2)
  - Broader/narrower terms (definition of hierarchies), equivalence, association and synonymy relations
- **Reference Models** (10)
  - A representation vocabulary of the concepts in the subject area, the relations among the terms and the way the terms can or cannot be related to each other



Classification of RDF Schemas



Classification of RDF Schemas



Classes and Properties: General Data

Schema	Classes		Properties	
	Total	# H	Total	# H
CIDOC	77	3	205	20
Euler	20	2	22	4
RSLP-CLD	11	2	43	7
CERIF	42	2	142	3
IMS	17	1	8	1
Math. International	211	1	0	0
Universal	5	0	13	0
BibLink	14	2	20	2
DOI	13	1	13	0
SLinkS	20	1	56	4
RSS	6	0	9	3
IMDB	65	2	182	0
GML	20	3	33	1
CZM	77	1	66	0
Gene Ontol.	6993	175	9	0

Schema	Classes		Properties	
	Total	# H	Total	# H
BSR	2714	230	1754	0
RED	5073	5	285	1
UNSPSC	16506	57	2	0
CC/PP	18	3	3	1
P3P	414	14	365	4
RDF Calendar	57	17	92	2
Scheduler Allocation	16	1	23	2
Dublin Core	2	0	22	0
CERES/NBII	8	1	14	0
WordNet	9	1	5	0
Limber Thes	11	1	17	3
MetaNet	66	3	11	2
Top Level	189	1	141	1

Observations on General Data

- The **size** of schemas varies a lot
  - Most of the schemas define **few classes** and **properties**
  - The **richer schemas** in terms of number of classes and properties defined are mainly from the domain of **E-Commerce**
- **Class-centric vs. property-centric implementation**
  - When many classes are defined, the number of properties declared is relatively low
- **Small number** of class/property hierarchies
  - Formulation of few large hierarchies instead of many small hierarchies

13

Class Hierarchies (1)

Schema	Classes		subClassOf					
	Total	Hierarchies	Depth		SubNodes		SupNodes	
			Max	Avg	Max	Avg	Max	Avg
CIDOC	77	3	8	4.4	7	1.1	2	1.1
Euler	20	2	1	1	14	0.8	1	0.8
RSLP-CLD	11	2	1	1	3	0.4	1	0.4
CERIF	42	2	1	1	13	0.3	1	0.3
IMS	17	1	2	2	5	0.7	1	0.7
Math. International	211	1	11	7.9	43	1.6	**9**	1.6
Universal	5	0	-	-	-	-	-	-
BibLink	14	2	1	1	5	0.6	1	0.6
DOI	13	1	1	1	7	0.6	1	0.6
SLinkS	20	1	2	1.6	2	0.2	1	0.2
RSS	6	0	-	-	-	-	-	-
IMDB	65	2	2	1.8	37	0.9	1	0.9
GML	20	3	3	1.9	5	0.8	2	0.8
CZM	77	1	6	4.3	4	0.9	1	0.9

14

7

Class Hierarchies (2)

Schema	Classes		subClassOf					
	Total	Hierarchies	Depth		SubNodes		SupNodes	
			Max	Avg	Max	Avg	Max	Avg
Gene Ontology	6993	175	12	5	106	1.2	6	1.2
BSR	2714	230	4	1.7	62	0.6	1	1
RED	5073	5	5	2.4	763	1.9	5	1.9
UNSPSC	16506	57	3	3	63	1	1	1
CC/PP	18	3	2	1.2	4	0.7	1	0.8
P3P	414	14	3	2.1	245	1.6	4	1.7
RDF Calendar	57	17	3	1.3	4	0.6	3	0.6
Scheduler Allocation	16	1	1	1	3	0.2	1	0.3
Dublin Core	2	0	-	-	-	-	-	-
CERES/NBII	8	1	1	1	3	0.4	1	0.6
WordNet	9	1	2	1.3	4	0.6	1	0.6
Limber Thesaurus	11	1	2	1.3	3	0.4	1	0.5
MetaNet	66	3	2	1.6	17	0.9	2	1
Top Level Ontology	189	1	11	6.3	11	1	3	1.1

15

Class Hierarchies: Observations

- The **average number of subnodes** tends to be less than 1.0
  - Indicates the existence of nodes not attached to a hierarchy
- The schemas defining hierarchies are mostly **swallow** and they tend to be developed in **breadth** rather than **depth**
  - **Maximum depth**: 12 (Gene Ontology)
  - **Maximum breadth**: 763 (RED)

16

Property Hierarchies (1)

Schema	Properties		subPropertyOf					
	Total	Hierarchies	Depth		Sub		Super	
			Max	Avg	Max	Avg	Max	Avg
CIDOC	205	20	2	1.2	10	0.4	2	0.4
Euler	22	4	1	1	1	0.2	1	0.2
RSLP-CLD	43	7	1	1	7	0.5	1	0.5
CERIF	142	3	1	1	18	0.3	1	0.3
IMS	8	1	1	1	2	0.3	1	0.3
Math. International	0	0	-	-	-	-	-	-
Universal	13	0	-	-	-	-	-	-
BibLink	20	2	1	1	1	0.1	1	0.1
DOI	13	0	-	-	-	-	-	-
SLinkS	56	4	1	1	1	0.1	1	0.1
RSS	9	3	1	1	2	0.4	1	0.4
IMDB	182	0	-	-	-	-	-	-
GML	33	1	2	2	6	0.6	1	0.6
CZM	66	0	-	-	-	-	-	-

Property Hierarchies (2)

Schema	Properties		subPropertyOf					
	Total	Hierarchies	Depth		Sub		Super	
			Max	Avg	Max	Avg	Max	Avg
Gene Ontology	9	0	-	-	-	-	-	-
BSR	1754	0	-	-	-	-	-	-
RED	285	1	3	1.5	233	1.6	2	1.6
UNSPSC	2	0	-	-	-	-	-	-
CC/PP	3	1	1	1	1	0.3	1	0.3
P3P	365	4	1	1	312	0.9	1	0.9
RDF Calendar	92	2	1	1	3	0.1	1	0.1
Scheduler Allocation	23	2	1	1	2	0.2	1	0.2
Dublin Core	22	0	-	-	-	-	-	-
CERES/NBII	14	0	-	-	-	-	-	-
WordNet	5	0	-	-	-	-	-	-
Limber Thesaurus	17	3	1	1	4	0.6	1	0.6
MetaNet	11	2	1	1	2	0.3	1	0.3
Top Level Ontology	141	1	6	3	18	1	2	1

Property Hierarchies: Observations

- The number of schemas using the **subPropertyOf** construct is relatively small
  - The majority of schemas do not define property hierarchies or they do it to a limited extent
- The subPropertyOf construct is mainly used for **relationships** between classes rather than **attributes** of a class
- When domain restrictions are imposed, several properties were declared with **multiple domain definitions**

Multiple Inheritance

Schema	% Multiple Inheritance	
	Classes	Properties
RED	**0.840**	**0.757**
P3P	0.644	-
RDF Calendar	0.100	-
CIDOC	0.168	0.068
Mathematics International	**0.333**	-
GML	0.050	-
Gene Ontology	0.184	-
MetaNet	**0.030**	-
Top Level Ontology	0.068	0.035

- **Multiple inheritance** for classes is more frequent than for properties
  - **Repeated declaration** of *subClassOf/PropertyOf* of a class/property to all its ancestors in the hierarchy (no cycles were detected)

Additional Observations about RDF/S Modeling Constructs

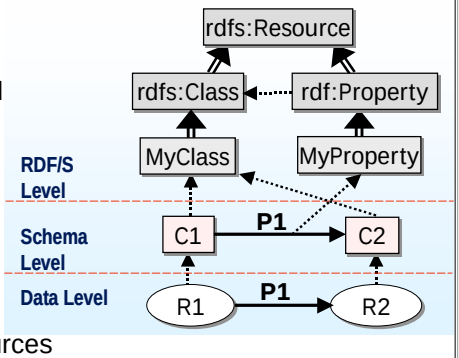
- The subPropertyOf construct is mainly used for **relationships** between classes rather than **attributes** of a class
- When the *subPropertyOf* is used, the **root property** is most of the times **unconstrained**
- When domain restrictions are imposed, several properties (mostly attributes) were declared with **multiple domain definitions**
- **Multiple classification of resources** was rarely used in the instance files examined
  - Inability to find a substantial number of instance files
- **Reification** not used
- Modeling techniques used by schema designers:
  - **Meta-schemas**
  - **Ternary or higher arity relations**
  - **Enumerated Types and Specialization of *rdfs:Literal***



21

Meta-Schemas

- **Extension** of the core RDF/S meta-model
  - Refine the classes *rdfs:Class* and *rdf:Property* with user-defined metaclasses.
- The **separation of abstraction levels** is not clear in RDF/S or the RDF M&T
  - A class can contain resources of different semantic levels, e.g., classes,properties and data resources



- Although **properties are first-class citizens**, they are not treated as equally as classes
  - A meta-property is a class and thus it cannot have a domain and range.
  - There is no instantiation mechanism for properties at data level.

22

Summary (1)

- RDF/S schemas were developed with **syntactic** and **semantic errors**
  - Need for generally accepted RDF/S authoring, parsing and validating tools
- **The size and the morphology of examined RDF/S schemas varies a lot**
  - The **characterization** of typical small, medium and large schemas depends on the semantic depth and the application domain
  - RDF/S schema **DAGs are not well-balanced** (significant difference between the max and the average depth and degrees)

Summary (2)

- **Need for detailed benchmarking in order to tune existing RDF/S implementations** (storage and querying)
  - Size and degree of schema DAGs is useful for choosing a good label encoding scheme to optimise schema querying.
  - Actual use of RDF/S constructs is useful for choosing better DB representation for RDF metadata.
- **Real scale applications require more expressive RDF/S modeling constructs** (meta-schemas, ternary relations)
  - Credible solutions should be incorporated into the core RDF/S standards
- **Qualitative vs. Quantitative schema evaluation**
  - **Taxonomic organization**≈ number of property and class hierarchies, in- and out- degrees of these hierarchies

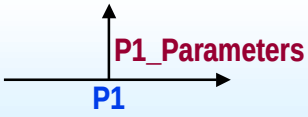
Thank you



Back Up Slides

Non-Binary Relations

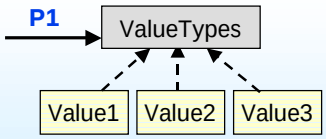
- The RDF/S data model is based on **binary relations** (relations between classes).



- There are modeling circumstances requiring the use of **ternary** or **higher arity relations**
  - Implicitly represented at data level by the use of *rdf:value* and an intermediate resource.
- The **inability** to model relations of higher arity **at schema level** stems from the fact that the domain of the property should always be a class.

Enumeration

- Need for **enumerated types** (definition of the possible values that a property can have)



  - **Implicitly** supported by RDF/S
  - **Not applicable** in the case of the ***rdfs:Literal***
  - **Specialization of rdfs:Literal** for richer data types.
  - ***XML data types*** can be used for **a richer set of data types**.