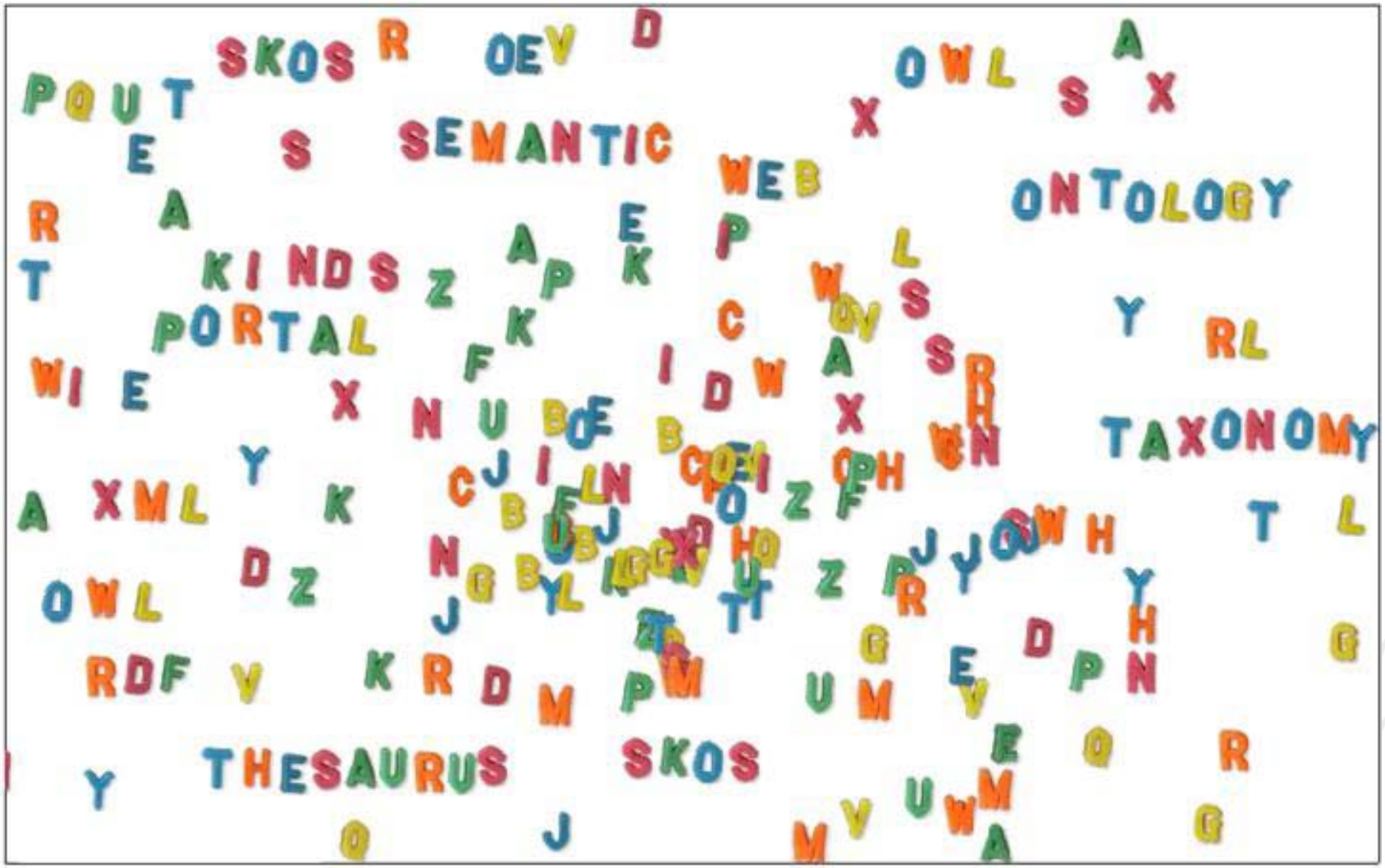
Taxonomies and the Semantic Web

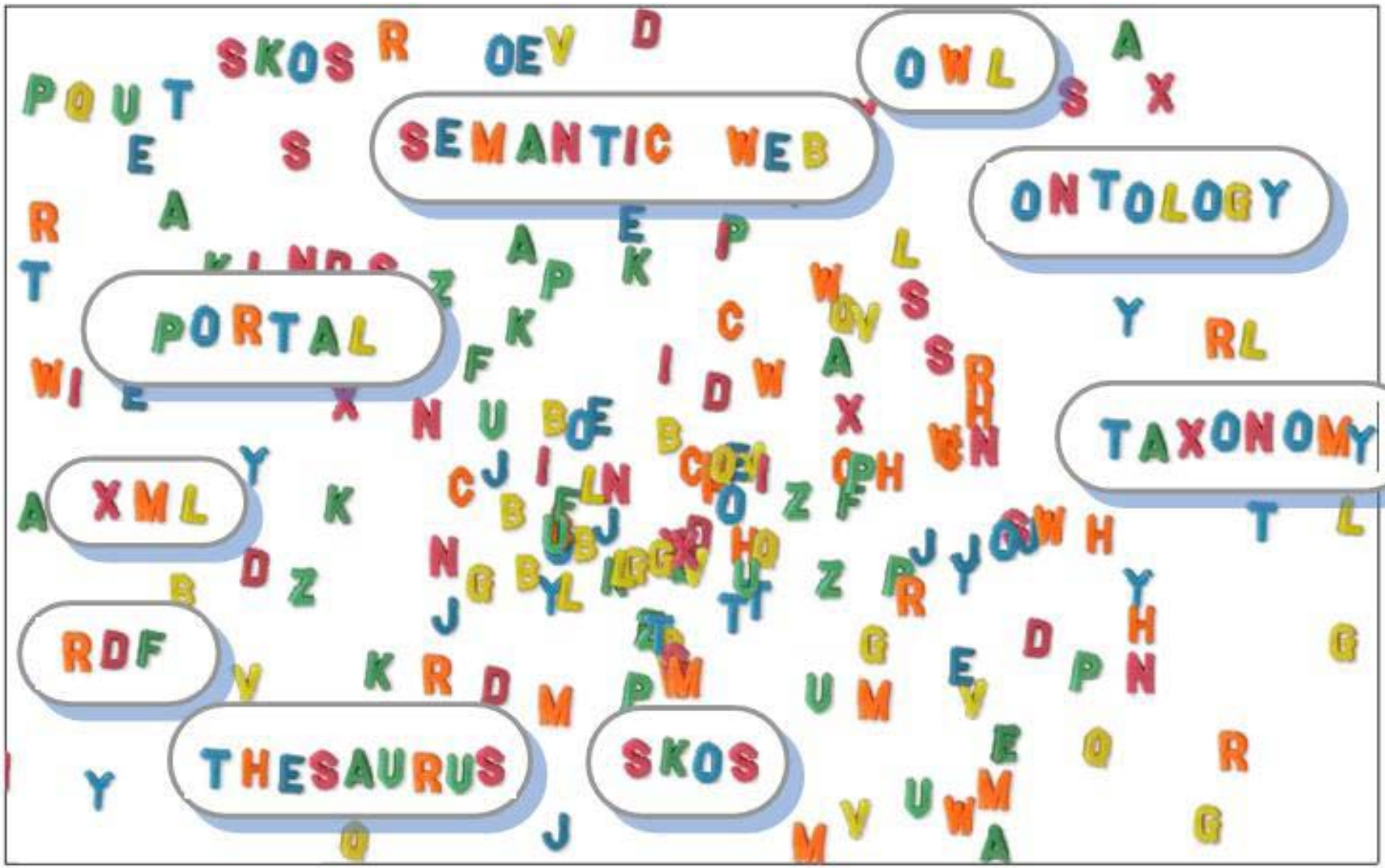
CISTRANA Workshop, Feb 2006

Alistair Miles
Business and Information Technology
CCLRC Rutherford Appleton Laboratory

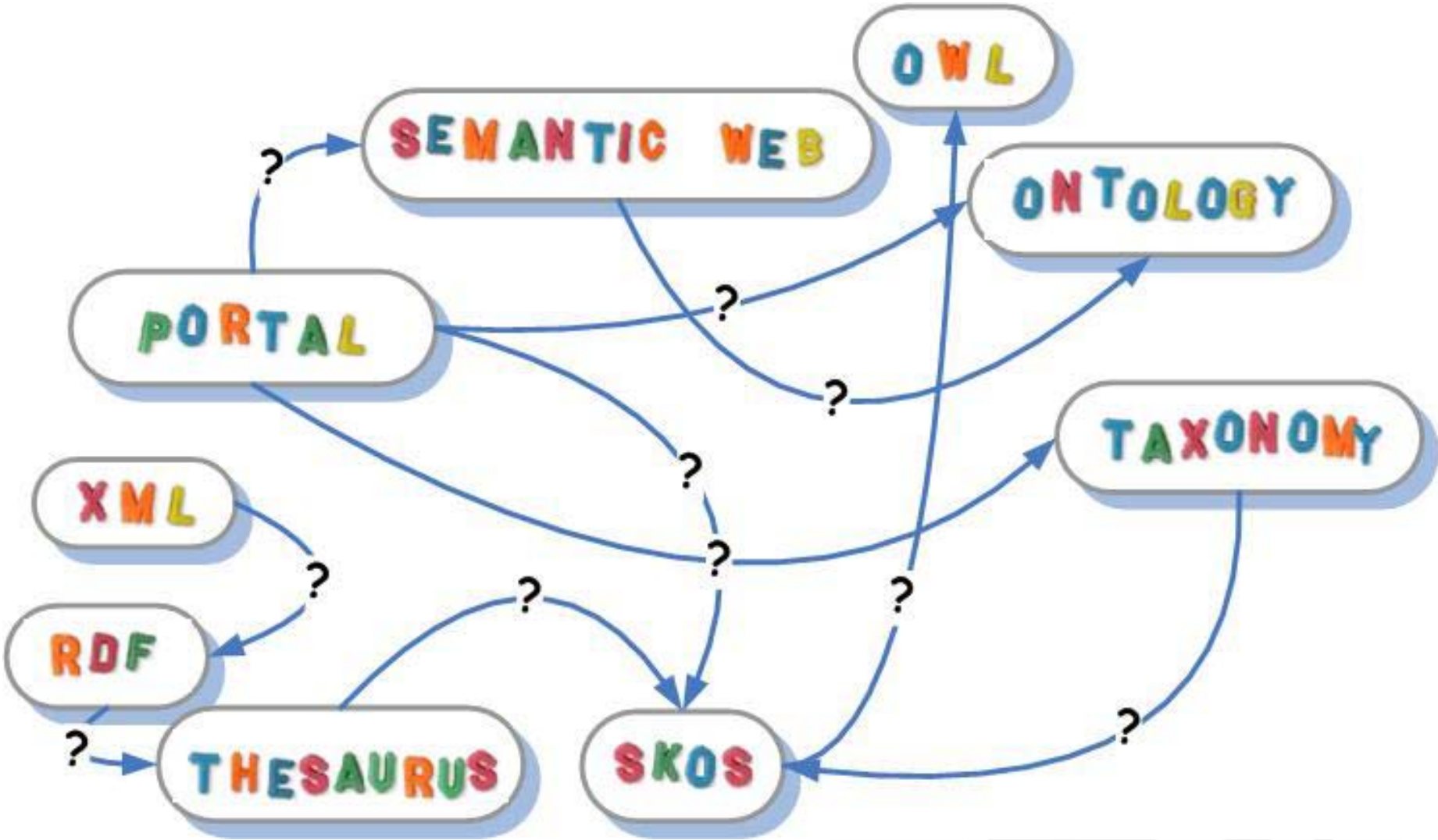
About this talk



About this talk



About this talk



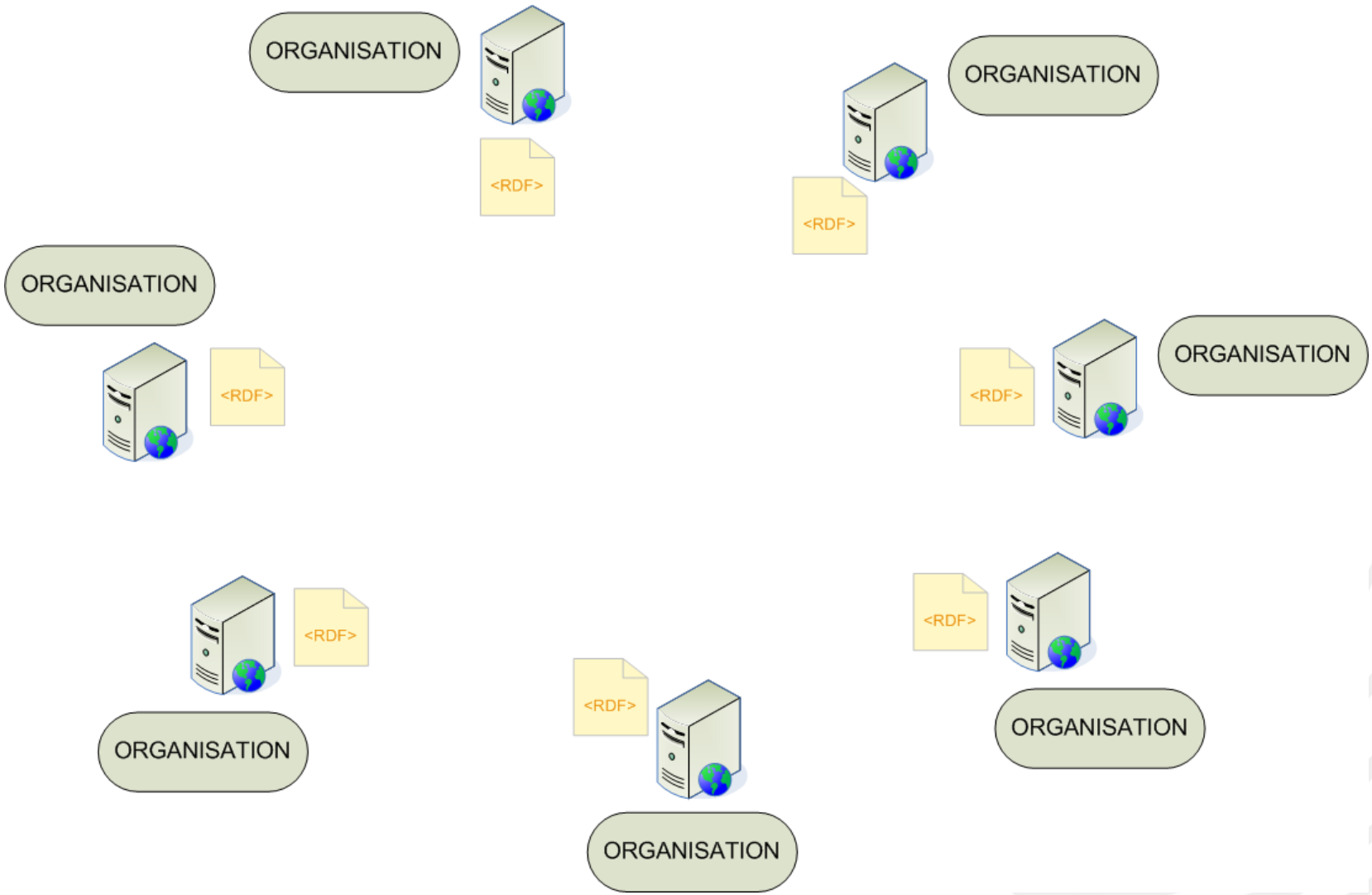
About this talk

- Technologies
  - XML, RDF, OWL, SKOS
- Modelling frameworks
  - Ontology, taxonomy, thesaurus, classification scheme
- Applications
  - Portal
- Big ideas
  - Semantic Web

Ideas

- Semantic Web
  - … is an elephant …

SWED


ORGANISATION
<RDF>
ORGANISATION
<RDF>
ORGANISATION
<RDF>
ORGANISATION
<RDF>
ORGANISATION
<RDF>
<RDF>
ORGANISATION
<RDF>
ORGANISATION

‘A Semantic Web’

- SWED is a portal that harvests it’s content from ***a semantic web***.
- A semantic web is a ***web of machine-understandable content***.
- Computers can do a lot more with machine-understandable content…
  - Merge
  - Analyse
  - Repurpose
  - E.g. SWED functionality.

‘The Semantic Web’

- ‘The [World-Wide] Semantic Web’
  - cf. the World-Wide Web

RDF

- RDF
  - Resource Description Framework
  - W3C Recommendation Feb 2004

- RDF is about ***distributed information*** … merging information from multiple sources.

RDF semantics

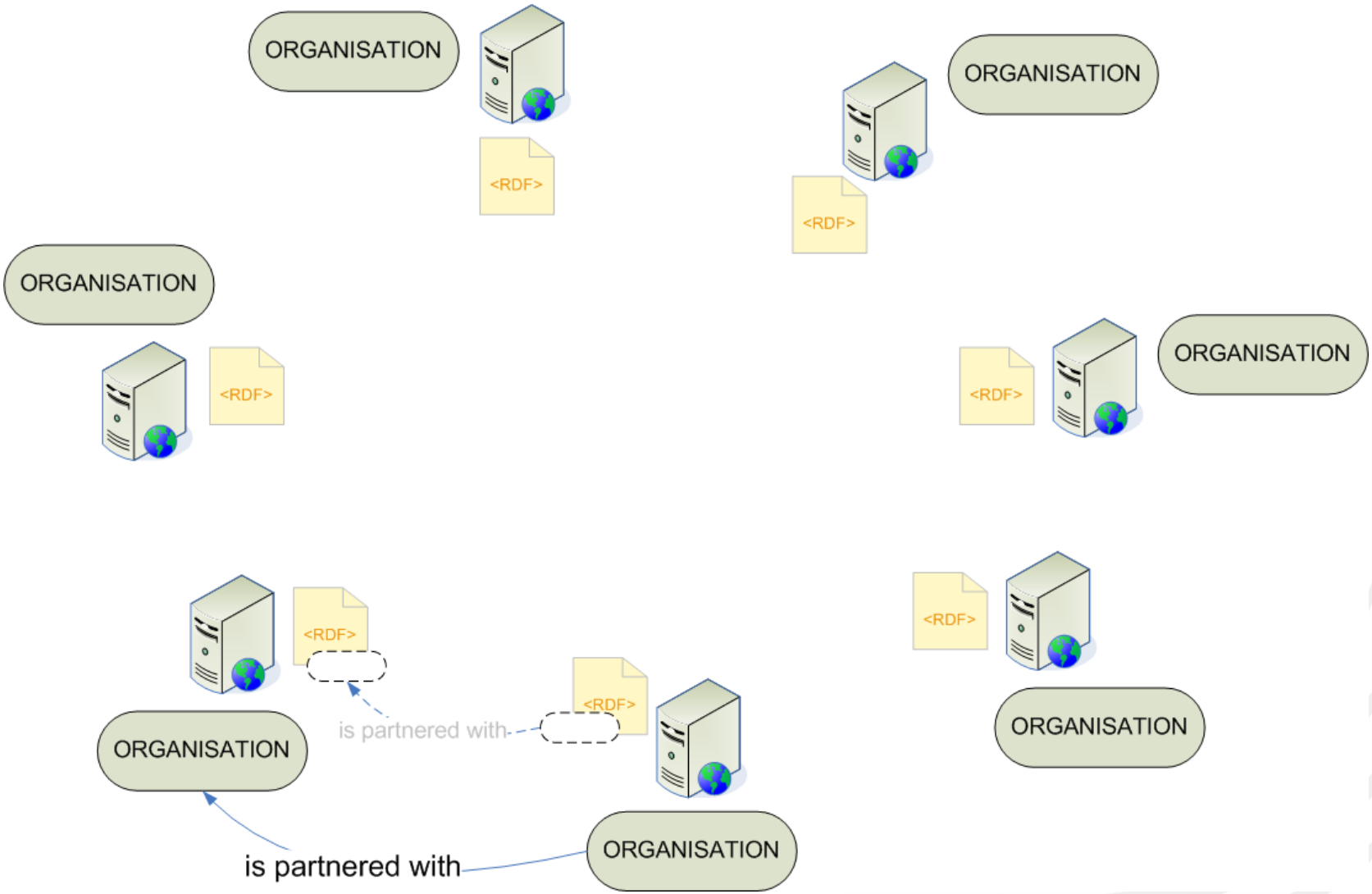
- Because RDF has a ***formal semantics***, it is possible to ***merge information*** obtained from ***multiple sources*** in a ***consistent*** and ***meaningful*** way.

- Formal semantics … machine-understandable.

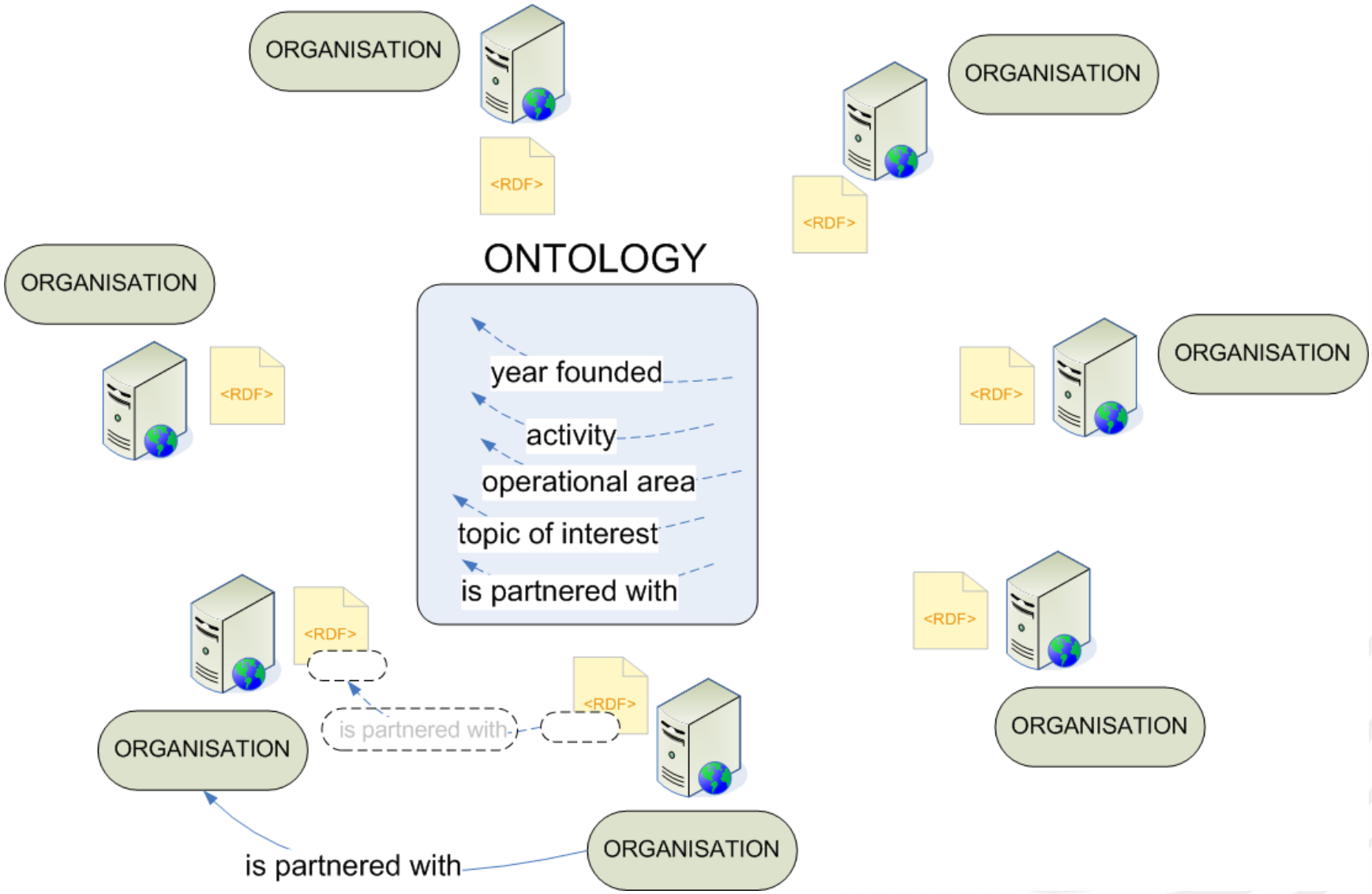
RDF analogy

- RDF is to machines what HTML is to humans.
  - HTML allows human-understandable content to be distributed across the web and 'hyperlinked'.
  - RDF allows machine-understandable content to be distributed across the web and 'hyperlinked'.

E.g.


ORGANISATION
<RDF>
ORGANISATION
<RDF>
ORGANISATION
<RDF>
ORGANISATION
<RDF>
<RDF>
is partnered with
ORGANISATION
<RDF>
ORGANISATION
is partnered with
ORGANISATION

E.g. Ontology


ORGANISATION
<RDF>
ORGANISATION
<RDF>
ONTOLOGY
ORGANISATION
<RDF>
year founded
activity
operational area
topic of interest
is partnered with
<RDF>
ORGANISATION
<RDF>
<RDF>
<RDF>
ORGANISATION
is partnered with
ORGANISATION
is partnered with
ORGANISATION

Ontology

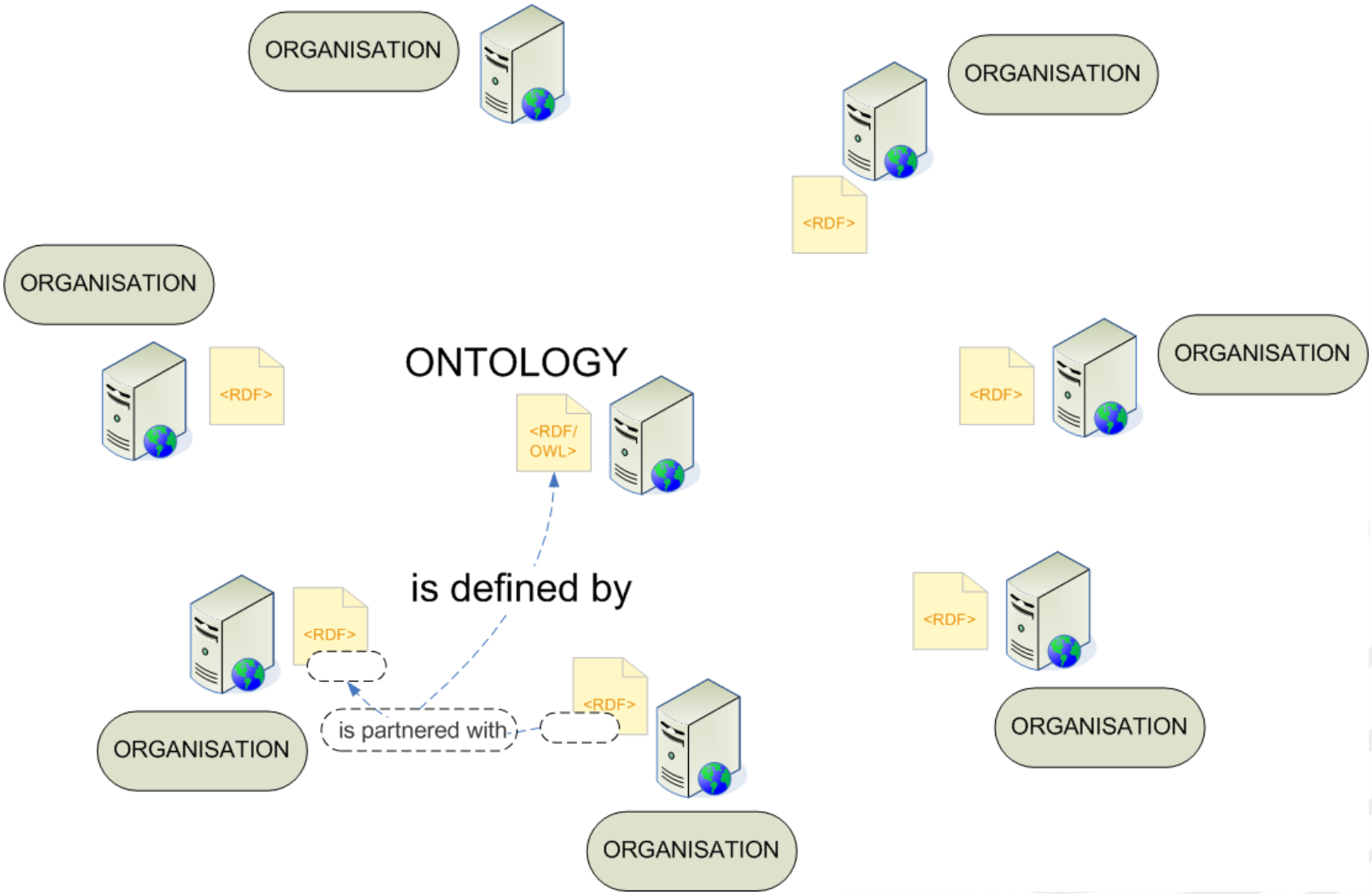
- Ontology
  - An ontology is a way of declaring what ***types of things*** exist, and what ***types of relationships*** they have with each other.
  - Provides ***common meaning*** to use in machine-understandable content.
- E.g. SWED Ontology
  - Types of things (classes): **Organisation**, **Project**
  - Relationships (properties): **topic of interest**, **operational area**, **activity**, **year founded** …

OWL

- OWL
  - Web Ontology Language.
  - W3C Recommendation Feb 2004.

- OWL is about declaring and publishing ***web ontologies***.
  - A ***web ontology*** is an ontology that can be used in a distributed, decentralised, information system (I.e. a semantic web).

Web Ontology



OWL and RDF

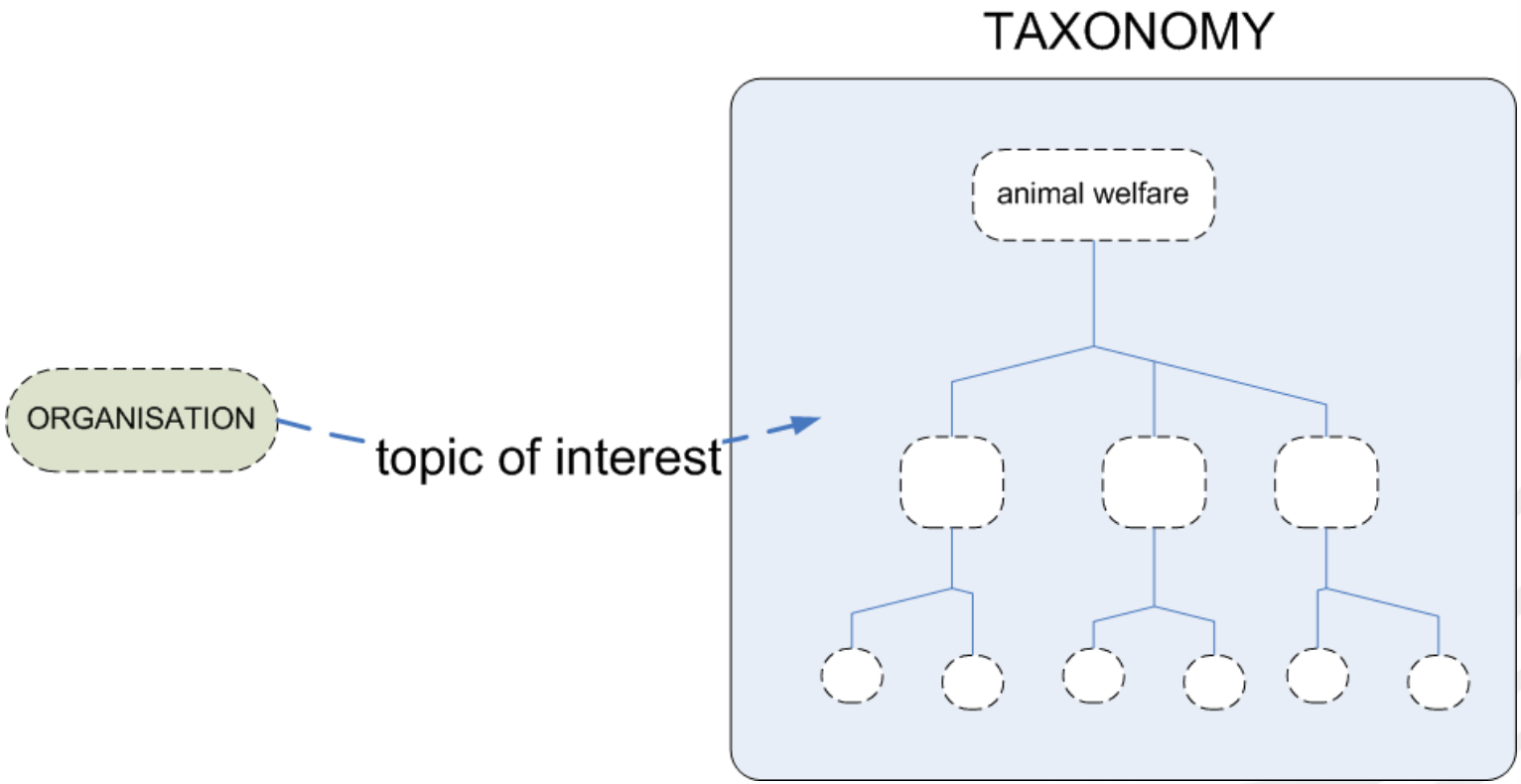
- OWL is a semantic ***extension of RDF***, allowing you to specify additional ***logical dependencies*** between information structures.

So far …

- Semantic Web
- RDF
- Web Ontology
- OWL
- …

Taxonomy

- SWED E.g.
  - **topic of interest**: animal welfare > captive animals (welfare of)



SKOS

- SKOS
  - Simple Knowledge Organisation System(s)
  - W3C 2nd Public Working Draft – work in progress!
  - (probably Recommendation within 1-2 years).

- SKOS is about ***declaring*** and ***publishing taxonomies***, ***thesauri*** or ***classification schemes***, for use in a distributed, decentralised information system (I.e. a semantic web).
  - The application of library science to the semantic web.

SKOS and RDF

- SKOS is an ***application of RDF*** that allows you to construct a simple hierarchy and/or network of concepts, provide multilingual labels and documentation for those concepts, and publish this information in a machine-understandable form.

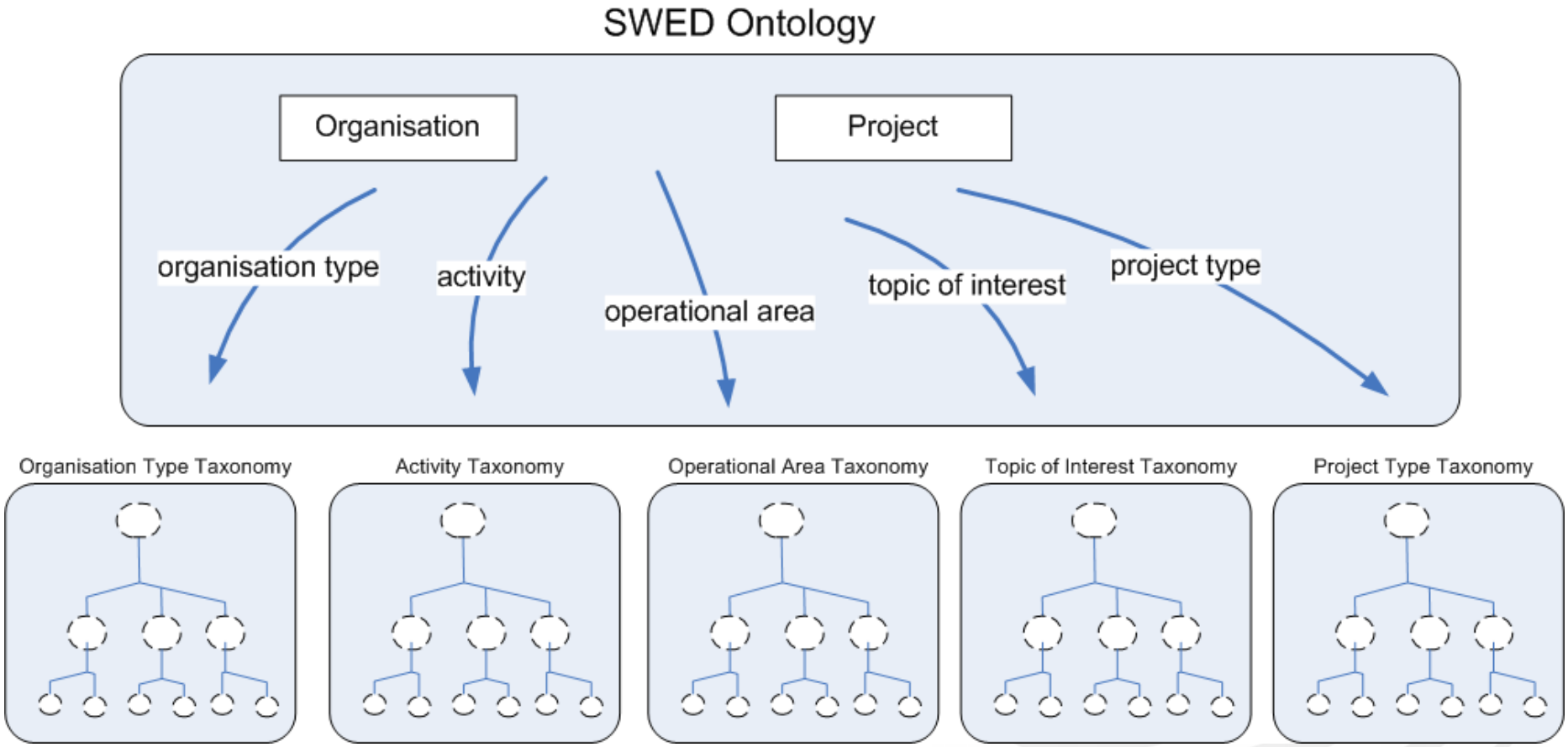
SKOS and OWL

- SKOS
  - Ideally suited for simple generalisation hierarchies and networks of 'topics' or 'subjects of interest', such as found in a taxonomy, thesaurus or classification scheme.
    - E.g. 'animal welfare', 'human health', 'disasters', 'resource management' ….
- OWL
  - Ideally suited for describing highly structured information in a logical way.
    - E.g. The class 'Organisation' has the properties 'topic of interest' and 'operational area' and may be related to other organisations by the properties 'partnered with' …

SKOS and OWL

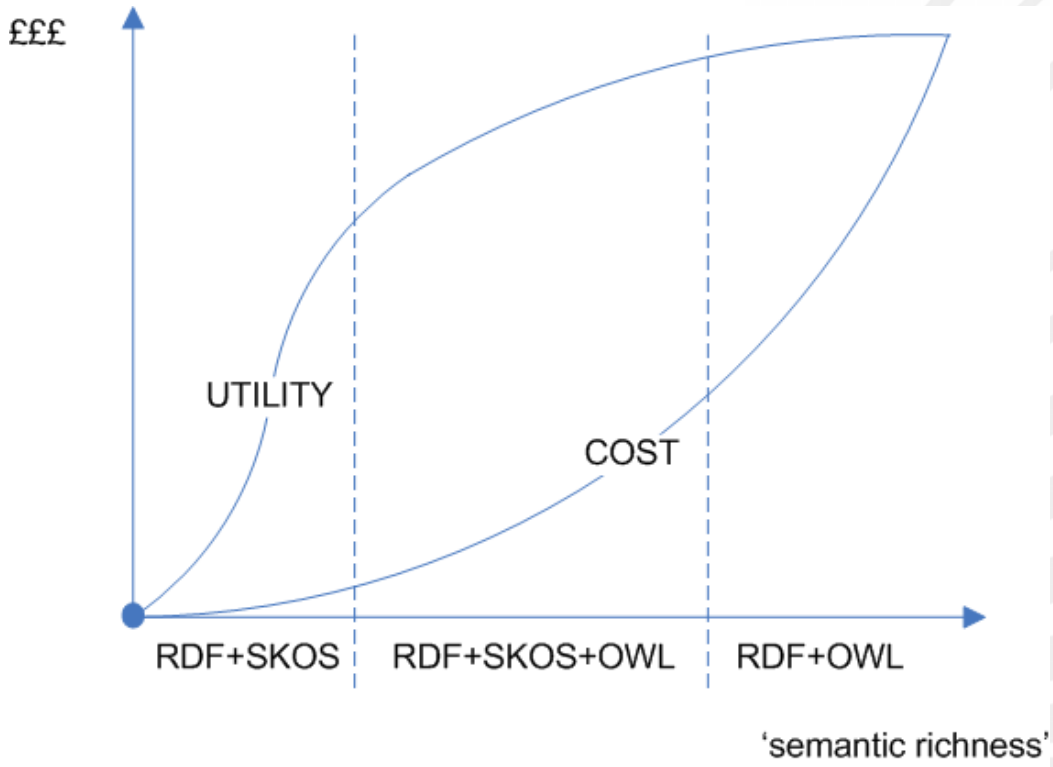
- SWED uses a combination of SKOS and OWL to define one ontology and 5 taxonomies …
  - (which are all published on the SWED site btw.)

SKOS and OWL


SWED Ontology
Organisation
Project
organisation type
activity
operational area
topic of interest
project type
Organisation Type Taxonomy
Activity Taxonomy
Operational Area Taxonomy
Topic of Interest Taxonomy
Project Type Taxonomy

SKOS and OWL

- SKOS and OWL allow you to explore the cost/benefit tradeoffs involved in investing in semantics …



XML

- XML
  - eXtensible Markup Language
  - W3C Recommendation when?

- XML is about ***transport*** … passing documents between a sender and a receiver, where the receiver has an implicit understanding of the XML language(s) used.
  - E.g. (X)HTML, SVG.
  - RDF/XML is the serialisation format for RDF.
  - RDF/XML is also the serialisation format for OWL and SKOS.

XML, RDF and OWL

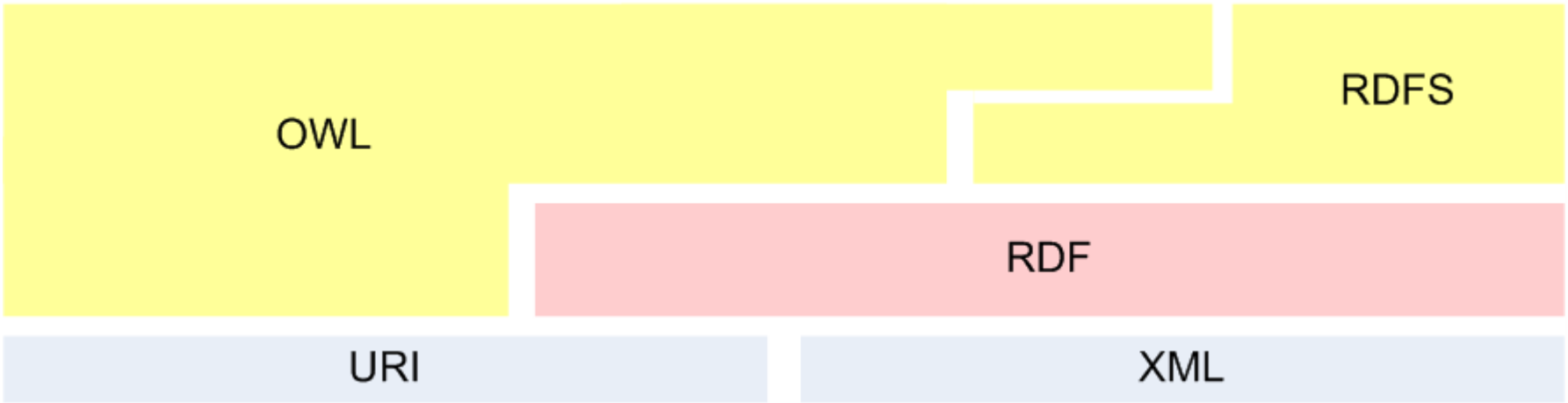
- XML and RDF do different jobs.
- XML
  - structuring text
- RDF
  - structuring information
- XML Schema
  - document model (a.k.a. content model)
- RDF Schema/OWL
  - information model (a.k.a. logical model)

XML semantics

- Because XML has ***no formal semantics***, it is impossible to know how information represented in one XML document relates to information represented in another … which means you ***cannot meaningfully merge*** the ***information content*** of two XML documents.

- N.B. This is ***not a shortcoming*** of XML, because it allows each XML language to define its own semantics (if any).

Technology stack



Summary

- Semantic Web
- RDF
- Web Ontology
- OWL
- Taxonomy/thesaurus/classification scheme
- SKOS
- Technology stack

Tools

- Basic application toolkits for storing, querying and manipulating RDF are good and stable, but …
- Collaborative development environments for ontologies, taxonomies, thesauri and classification schemes are missing.

Open Areas

- Semantic mapping
  - Semantics
  - Representation
  - Tools
- Change management
  - Semantics
  - Representation
  - Processes
  - Tools

Discussion

- ‘For example, a portal may offer the ability to search for countries with national research programmes in nanoelectronics, or nationally funded projects in the same area, or to present a display of institutes in different countries that are working together.’

- How many ontologies?
- How many taxonomies?