Semantic Interoperability

Olle Olsson
Swedish W3C Office
Swedish Institute of Computer Science (SICS)

Stockholm Apr 27 2011

Background

Trends: from small/closed to large/open

History of computer use (IT)

- Canned applications
  - Software + data files
- Suites of applications (steps in a job)
  - Carefully crafted application programs sharing data
- Databases
  - Applications sharing local database
- Application environments
  - Harmonized APIs in applications and for application control
- Networks
  - Arms-length distance appl-to-appl and appl-to-data
- Internet
  - Loose coupling to (semi-)independent appls/data

Trends: from small/closed to large/open

History of computer use (IT)

- Canned applications
  - Software + data files
- Suites of applications (st
  - Carefully crafted applic
- Databases
  - Applications sharing loc
- Application environment
  - Harmonized APIs in applications and for application control
- Networks
  - Arms-length distance appl-to-appl and appl-to-data
- Internet
  - Loose coupling to (semi-)independent appls/data

· Increased dependence on "external" factors

· Increased dependence on shared resources

Interoperability spaces

Software interoperability

- Software-to-software interactions

Data/information interoperability

- Combination of data sources

Interoperability dimensions / 1

Interoperability – across "jurisdictions"

- Within an authority domain
- Between authority domains

Interoperation – across "space"

- Application to data: shared data storage
- Application to application: passive data exchanged
- Data to data: combining data repositories
- Application to some space of applications: (grid, cloud)

Interoperability dimensions / 2

Interoperability – across "time"

- Application to other version of application
- Application to old version of data

Drivers

Basic traditional drivers

- Cost
  - manage/maintain applications and data
- Time
  - Time to create / adapt to new requirements
- Quality
  - Provide needed functionality / data
- Quality assurance
  - “Proof” for a certain level of quality

Major trends

- Increased decentralisation of authority
- Increased interaction across administrative boundaries
- Increased speed of change
  - Technology
  - Business
- Less specific local development
  - Off-the-shelf solutions
  - Standards
  - Shared solutions

Political driver: PSI directive

- Provide public data for independent re-use
- Level of agreement about provisioning?
- Cost to live up to agreement
  - Provider ?
  - Re-user ?
- Data/information evolves
  - Work to adapt to changes ?

- “PSI solution” vs. “collaboration solutions”
  - Separate technology approaches?
  - Who is the consumer of tomorrow?

Semantic Web

Technology – Semantic Web

Web technology

- Formats (XML, WS-*, MathML, RDF, SVG, ...)
- Protocols (HTTP, SOAP, ...)
- Processing (XForms, DOM, Powder, Pipeline, ...)

- Semantic Web Technologies
  - Rich representation:
    - RDF, RDFS, OWL, SKOS, ...
  - Processing support:
    - OWL, RIF, SPARQL, ...

Technology – Semantic Web

Semantic web is about what?

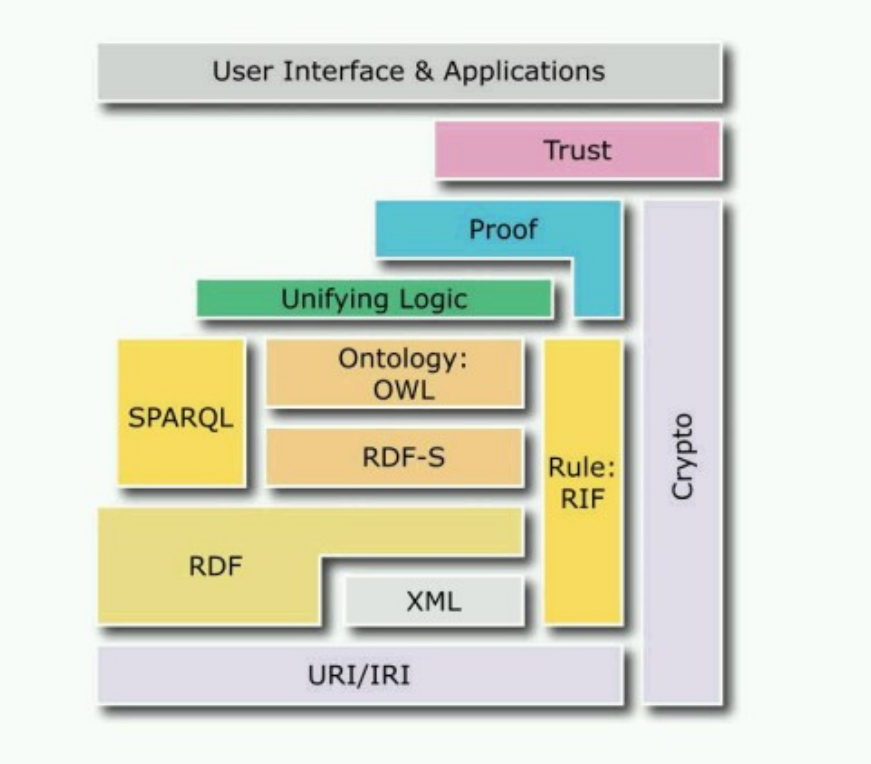
- about "meaning" and automation
  - "Meaning-based" automation

"Meaning" -- pragmatic approach in Semantic Web:

- a program "knows" what it *can* do with data
- self-describing data

No magic ... instead well-founded engineering

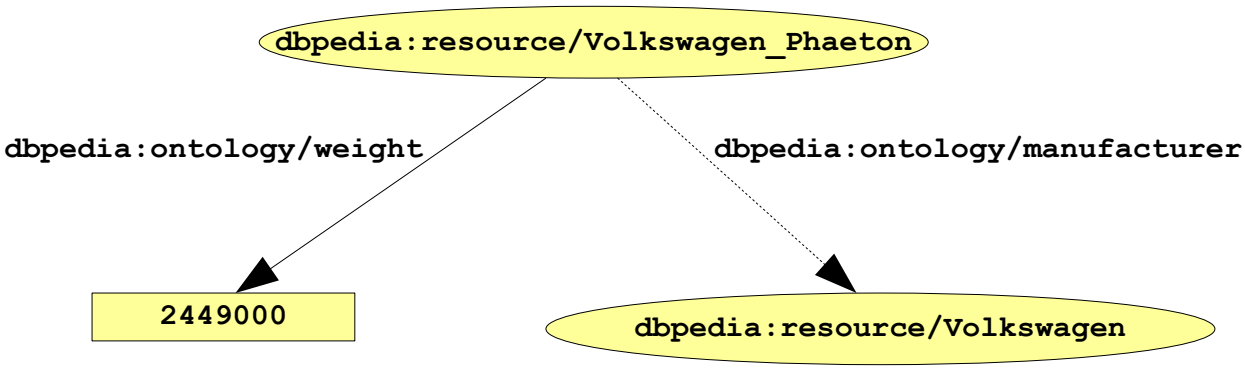
Semantic Web – building blocks



RDF – Resource Description Framework

- Basic data model – a "triple"
  - triple `(s,p,o)` :
    - "`s`" -- subject
    - "`p`" -- predicate
    - "`o`" -- object
  - Conceptually: "`s`" is related by "`p`" to "`o`"
- *RDF* is a general model for such triples
  - machine readable formats like RDF/XML, Turtle, n3, RXR

RDF – micro example




```
dbpedia:resource/Volkswagen_Phaeton
dbpedia:ontology/weight
2449000

dbpedia:resource/Volkswagen_Phaeton
dbpedia:ontology/manufacturer
dbpedia:resource/Volkswagen
```


OWL – Web Ontology Language

Define ontologies (conceptual model, ...) for data

Built on top of RDF

Basic components:

- *instance* – entity
- *class* – type
- *property* – relationship

- Ontology enables:
  - Checking consistency of instance graph
  - Inferring implicit statements about instance graph

SKOS – Simple Knowledge Organisation System

Define simple ontologies (conceptual model, ...)

Targeting traditional modelling approaches

- Taxonomies, classification schemes, thesauri, ...

Built on top of RDF

Compatible with modelling standards:

- NISO Z39.19 - 2005; ISO 5964:1985

SPARQL – RDF Query Language

Retrieve RDF data from RDF data graphs

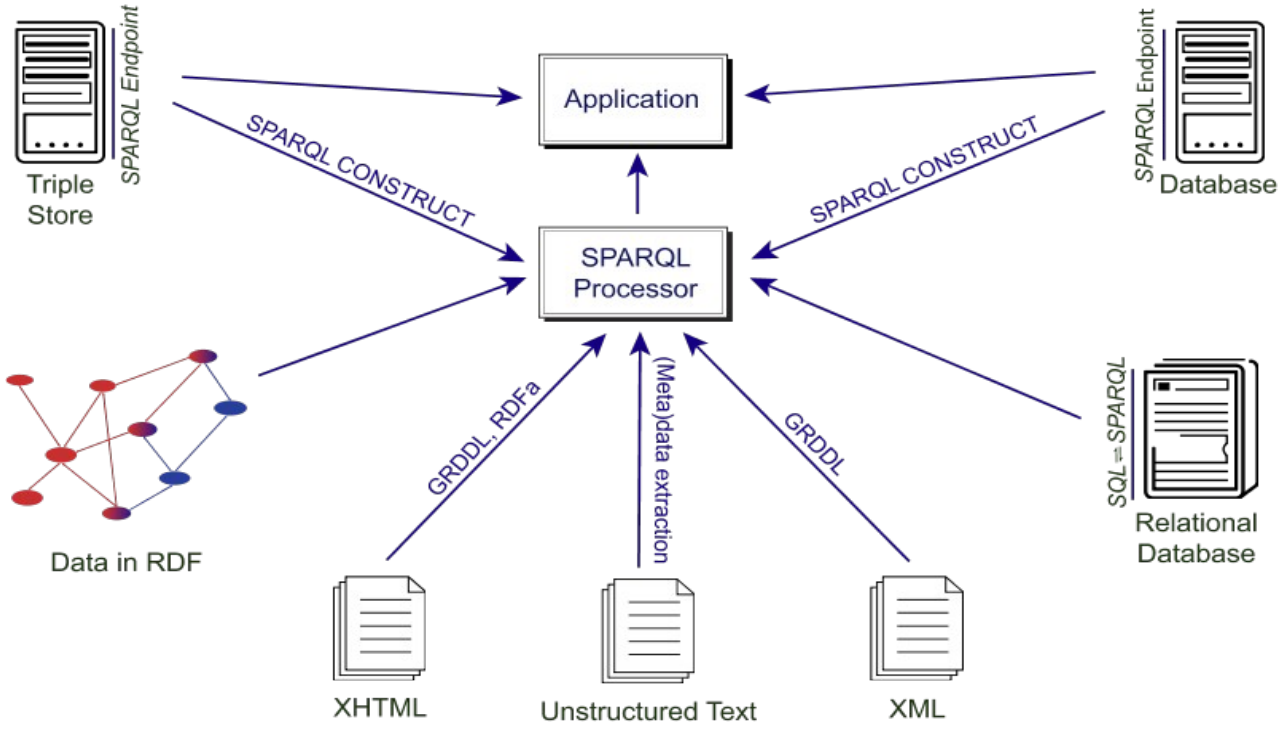
- RDF graph as answer to query

Syntax inspired by SQL

Example:

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX dbo: <http://dbpedia.org/ontology/>
SELECT ?manufacturer ?name ?car
WHERE {
    ?car skos:subject <http://dbpedia.org/resource/Category:Luxury_vehicles> .
    ?car foaf:name ?name .
    ?car dbo:manufacturer ?man .
    ?man foaf:name ?manufacturer
}
ORDER by ?manufacturer ?name
```

Integration via SPARQL



Approaches to interoperation

What can it mean?

Semantic Interoperability

- “increased independence among communicating parties”
- “integrating resources developed using different vocabularies and perspectives on data”
- “transmit data and exchange information while allowing each system to process information independently”
- “allow two independent programs to derive same conclusions from same data”

Some interoperability types

- Content-centric interoperability
  - Programs request data, and receive it as if from a data store
    - State-less
- Procedural interoperability
  - Programs interact with each other via structured dialogue
    - State-full
- Delegation-based interoperability
  - Programs request processing to be done by other programs
    - State-less
    - cf. DB-engines!

Models

Domain models

- Describes some domain, with respect to:
  - Terms
  - Entities (instances)
    - Properties, relations, ...
  - Types (classes)
    - Properties, relations, ...

Data models

- Describes stored/communicated data about a domain
  - Syntax
  - Structure
  - Values

Data standards

- Data standards
  - Documented data formats
  - Covering some slice of a domain
  - Committee managed
- Describing
  - Format of stored representation?
  - Format of exchanged data?
- Support for extensions/adaptations?

Terminological models

- Terminological models
  - Defining terms, “meanings”
  - Defining relations between terms
  - Restricted formalism – “cover common needs”
- Relationships to data?
- Use in software?

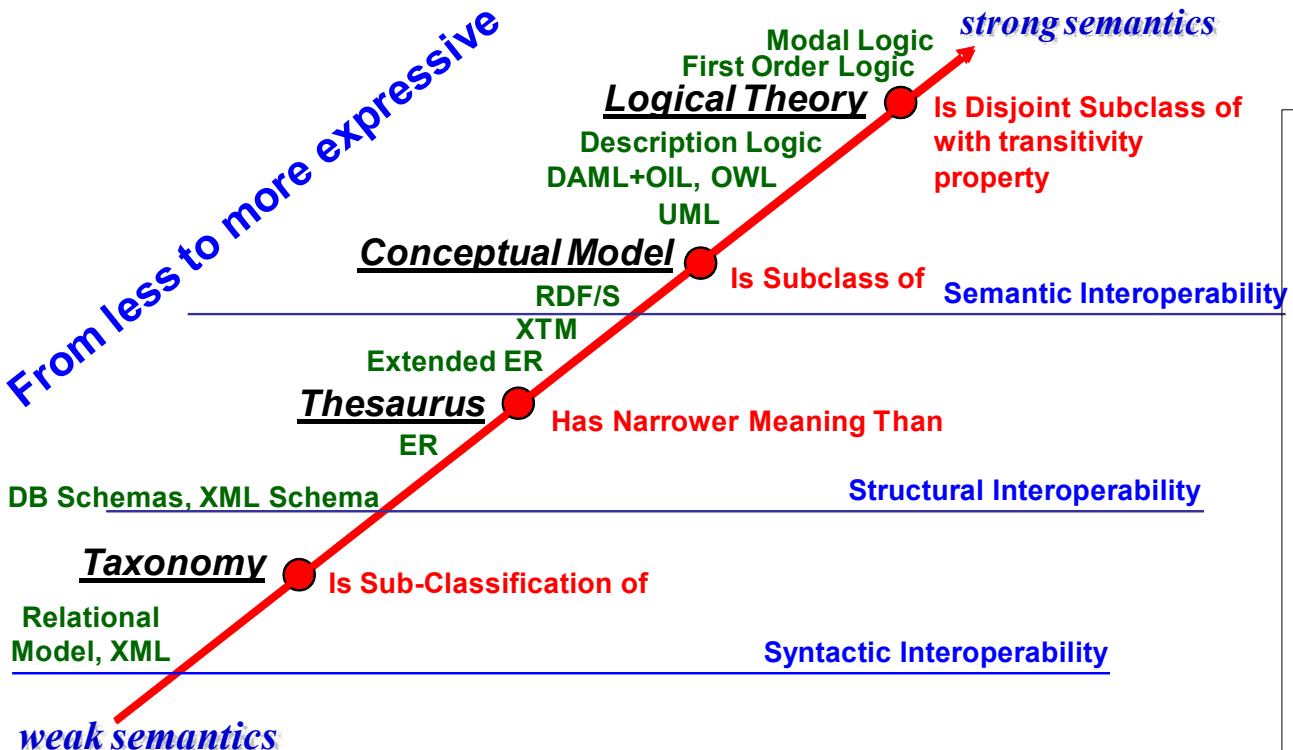
Ontology models

- A model of a domain
  - Cover aspects on instances/facts
  - Cover generic aspects of the domain
- Represented in standardised formats
  - RDF, OWL, SKOS
- Designed for use in software
  - In general domain-independent tools (cf. DB engines)
  - In specific application software

Models

- Terminological models
  - Focussed on terms (words, names)
  - Relationships between terms
- Conceptual models
  - Focussed on structure (entities)
  - Properties and relationships (on and between entities)
- Ontological models (semantic models)
  - Focussed on statements (about types)
  - Properties and relationships
  - Interdependencies, constraints, invariants, ...

Semantic Model Spectrum



Source: Obrst "The Ontology Spectrum and Semantics Models", 2006

Example

Linked Open Data Initiative

Web of data:

- Many open datasets on the web
- Interoperable when accessible as RDF

Examples:

- Wikipedia ("text") ==> dbpedia (RDF);
- Scientific data sets (experimental data)
- Public sector information (geodata, census data, statistics, ...)

- Different aims and coverage
  - But semantically interrelated
  - Increasingly so over time!

Linked Open Data

Linked Open Data (LOD) objective:

"expose" open datasets via RDF

- *set RDF links among the data items* from different datasets
  - a typical example is to set an owl:sameAs between two items in different datasets that refer to the same "thing"
- set up query endpoints (usually SPARQL)

Altogether billions of triples, millions of links…

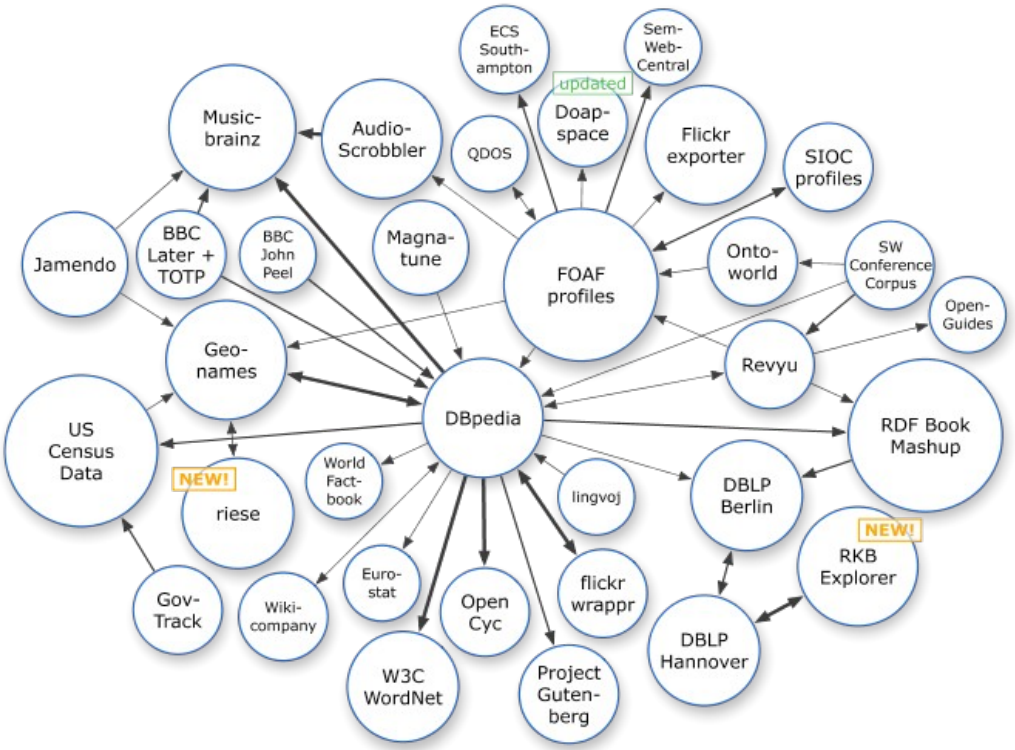
The "seed" for a general Web of Data

Linked Open Data

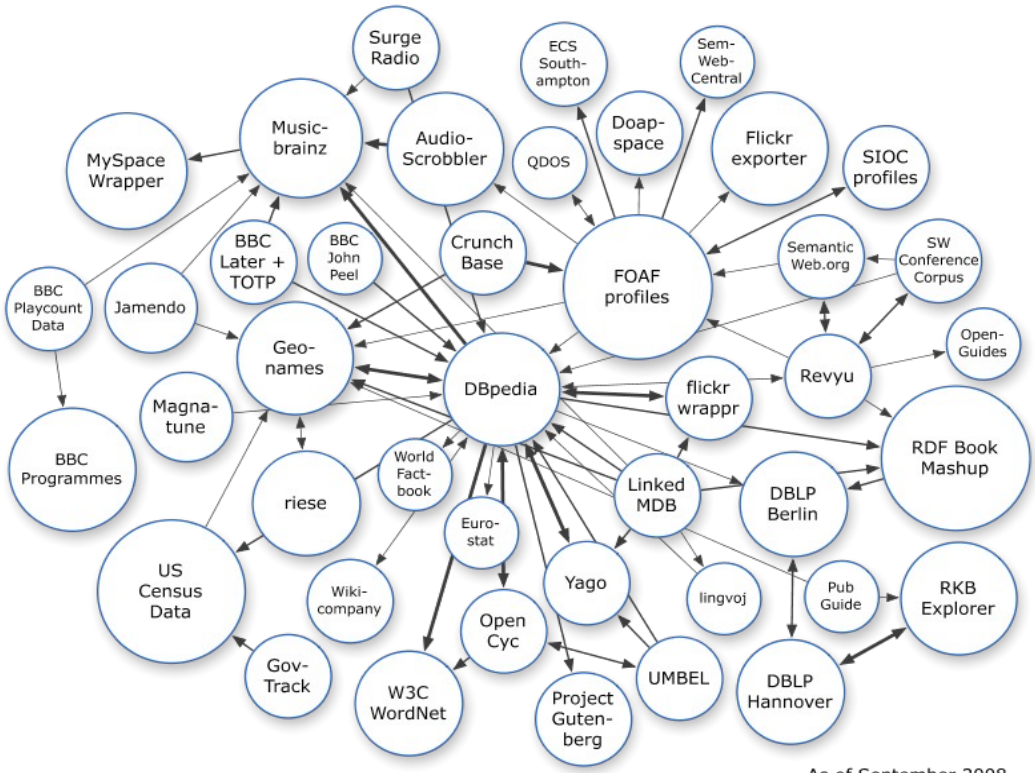
Technologies:

- URIs (specifically, of the dereferenceable variety)
- HTTP
- Resource Description Framework (RDF)
- Serialization formats (RDFa, RDF/XML, N3, Turtle, and others)

The LOD “cloud”, March 2008

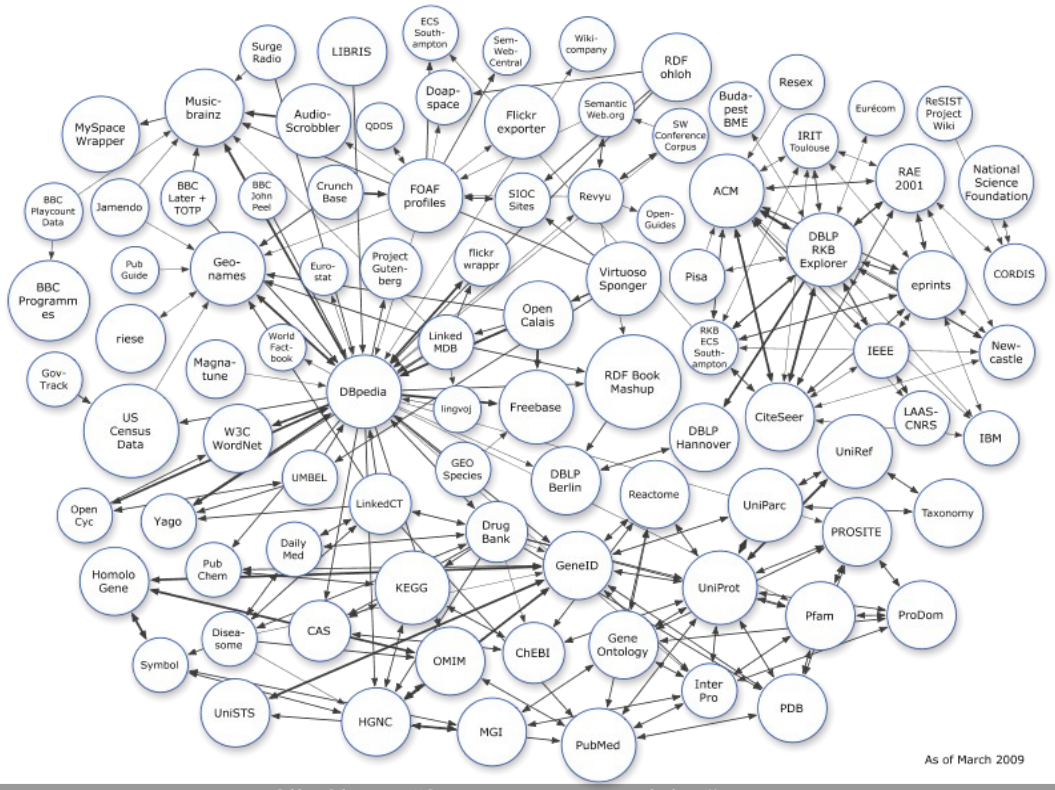


The LOD “cloud”, September 2008

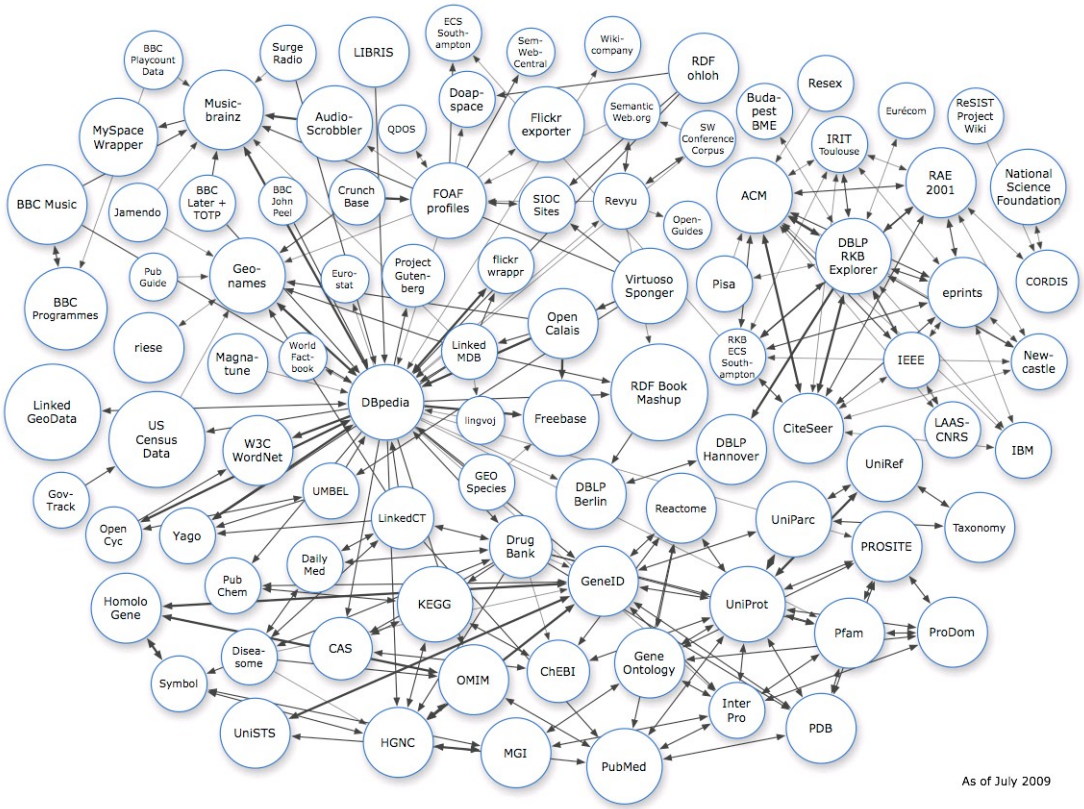


As of September 2008

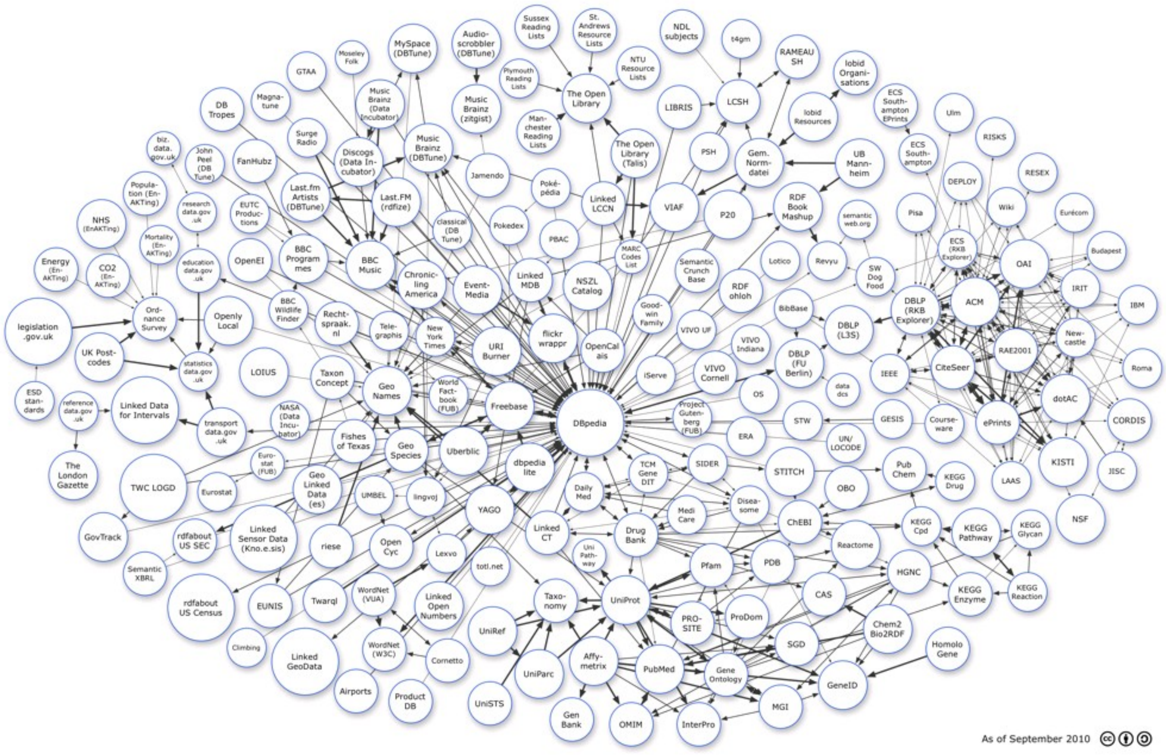
The LOD “cloud”, March 2009



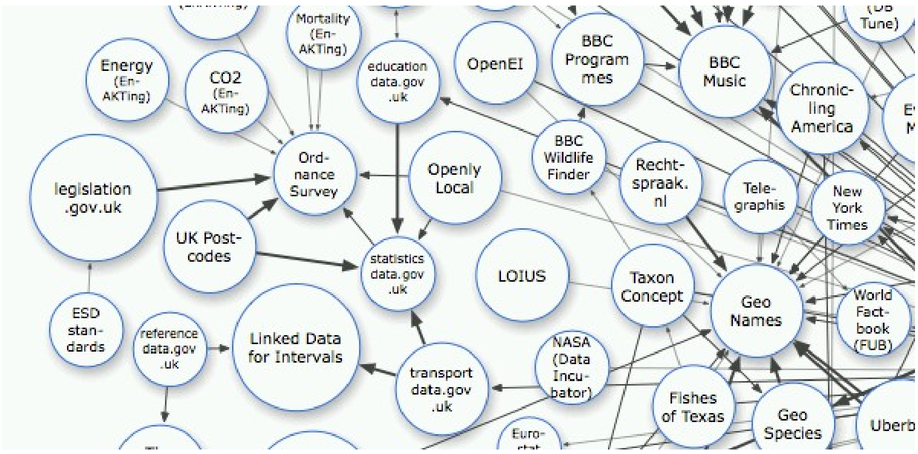
The LOD “cloud”, July 2009



The LOD “cloud”, March 2011

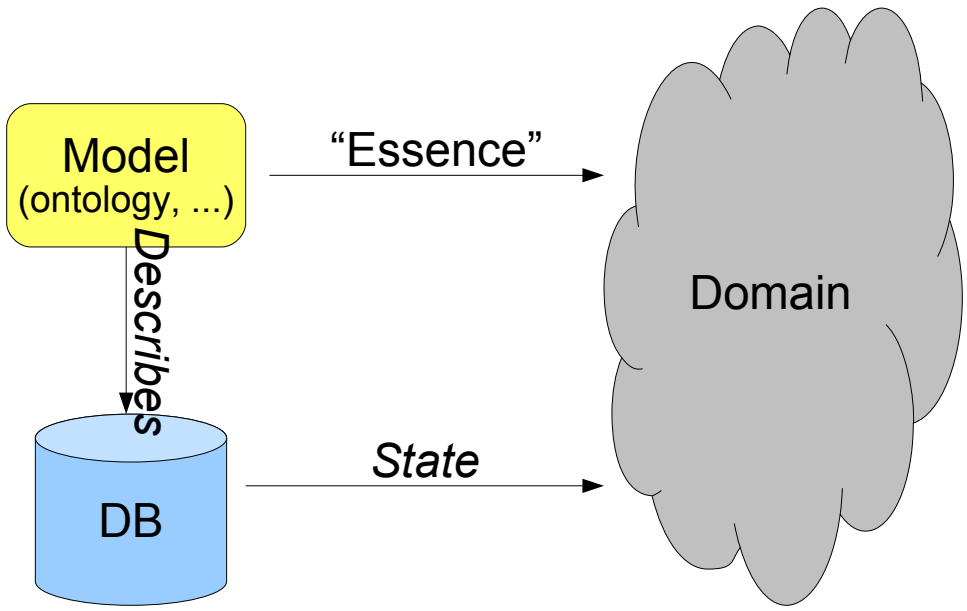


Linked Open Data

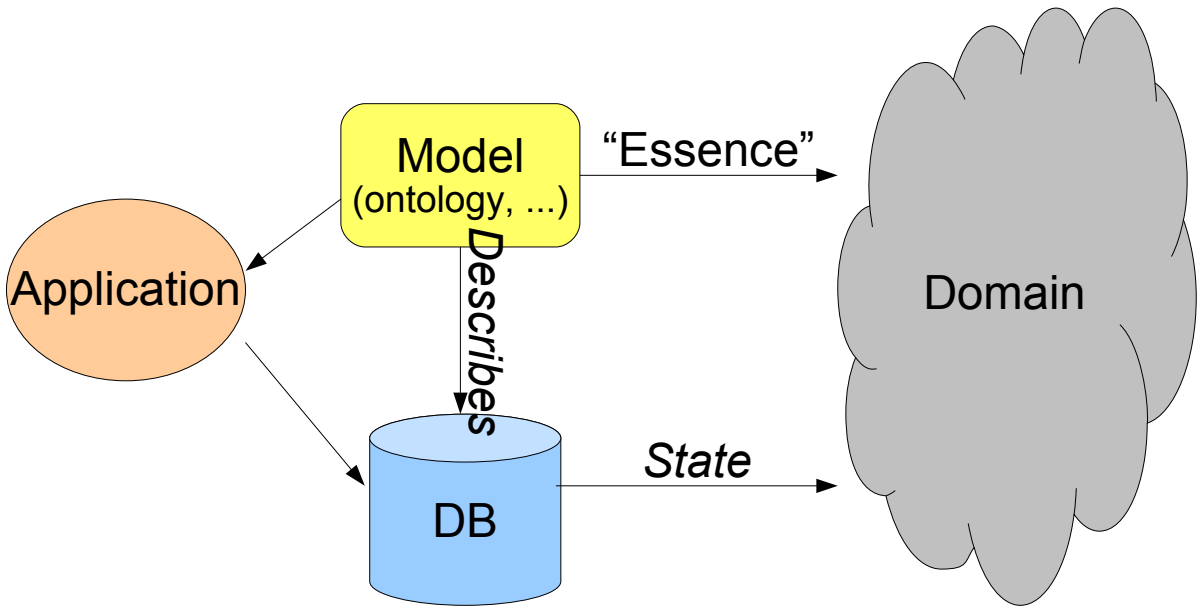


Models and data and programs

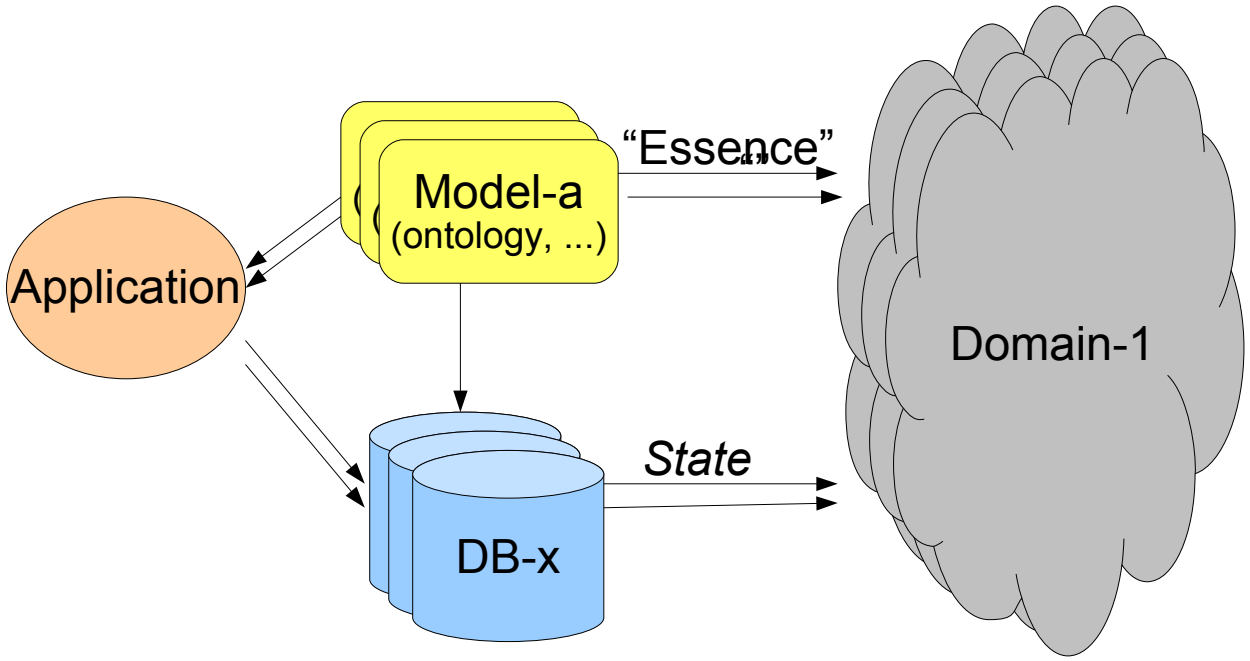
Models and data / 1



Models and data / 2



Models and data / 3



Conclusion

Semantic interoperability

- Rich semantics …
- Basic technologies in place
  - RDF, OWL, SPARQL, ...
- Support tools emerging
- Methodology emerging
- Challenges:
  - Ontology construction
    - Manual work … specific competence
    - But: re-use conceptual models, taxonomies, etc
  - Ontology alignment
    - Tool-supported approaches are explored
  - All modelling is difficult

---end---