ÖAW AUSTRIAN ACADEMY OF SCIENCES

ACDH – AUSTRIAN CENTRE FOR DIGITAL HUMANITIES

Knowledge management in DH: from definition to fruition

Matteo Lorenzini 07.05.18

Outline

- Introduction
  - Knowledge definition
  - Data and metadata
- Semantic Web
- Web of Documents & Web of Data
  - Web 2.0
  - Web 3.0
- Formalism
  - RDF
  - OWL
- SPARQL
- Linked Data

Knowledge: DEFINITION

Knowledge is a familiarity, awareness, or understanding of someone or something, such as facts, information, descriptions, or skills, which is acquired through experience or education by perceiving, discovering, or learning.

from Wikipedia

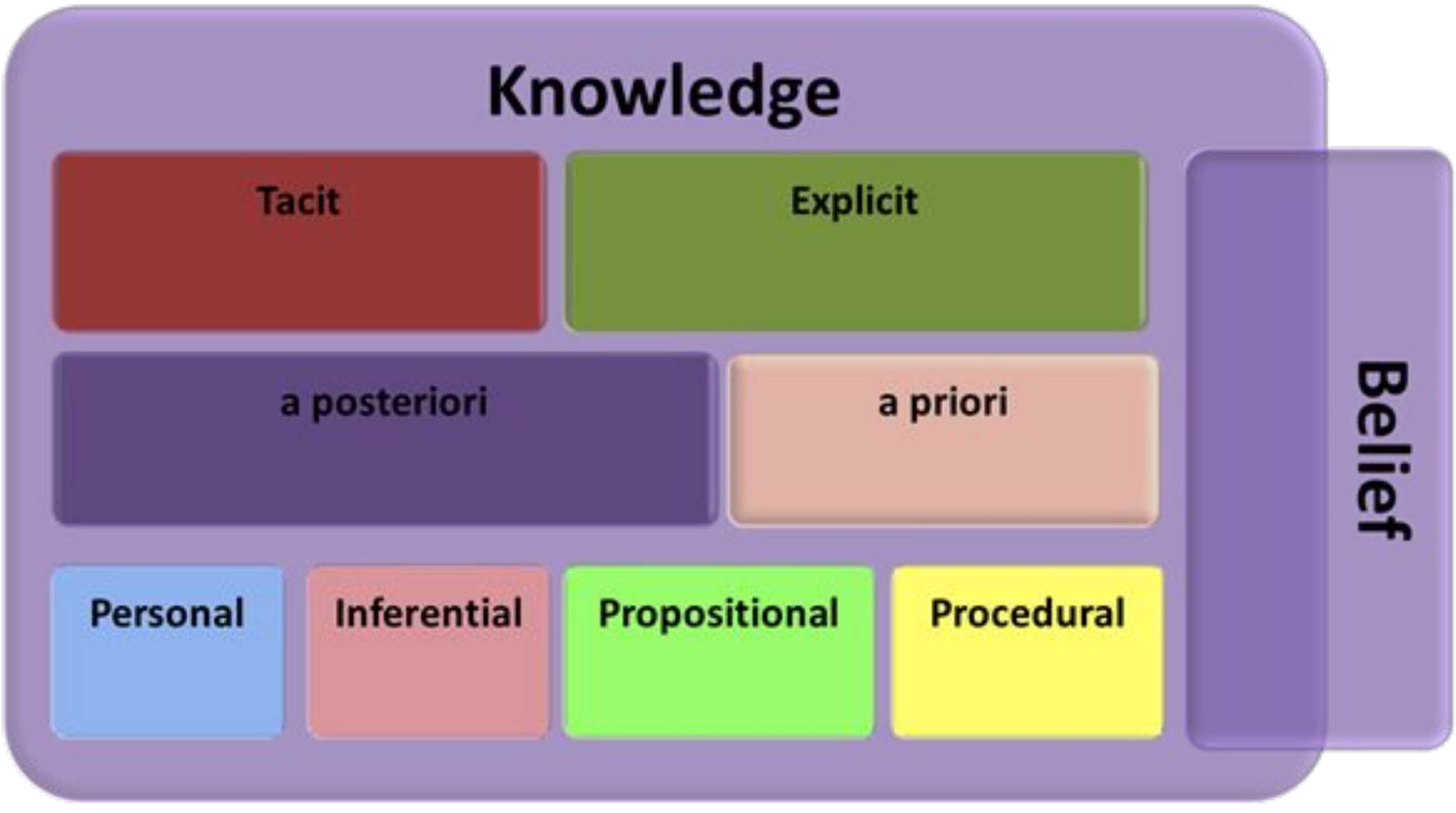
Types of knowledge acquisition

A priori	• is knowledge that is known independently of experience (that is, it is non-empirical, or arrived at beforehand).
A posteriori	• is knowledge that is known by experience (that is, it is empirical, or arrived at afterward).

Theories of knowledge acquisition

Empiricism	• A role of experience, especially experience based on perceptual observations by the five senses.
Rationalism	• A knowledge acquired by intuition or is innate
Constructivism	• all knowledge is "constructed" in as much as it is contingent on convention, human perception, and social experience

Types of knowledge

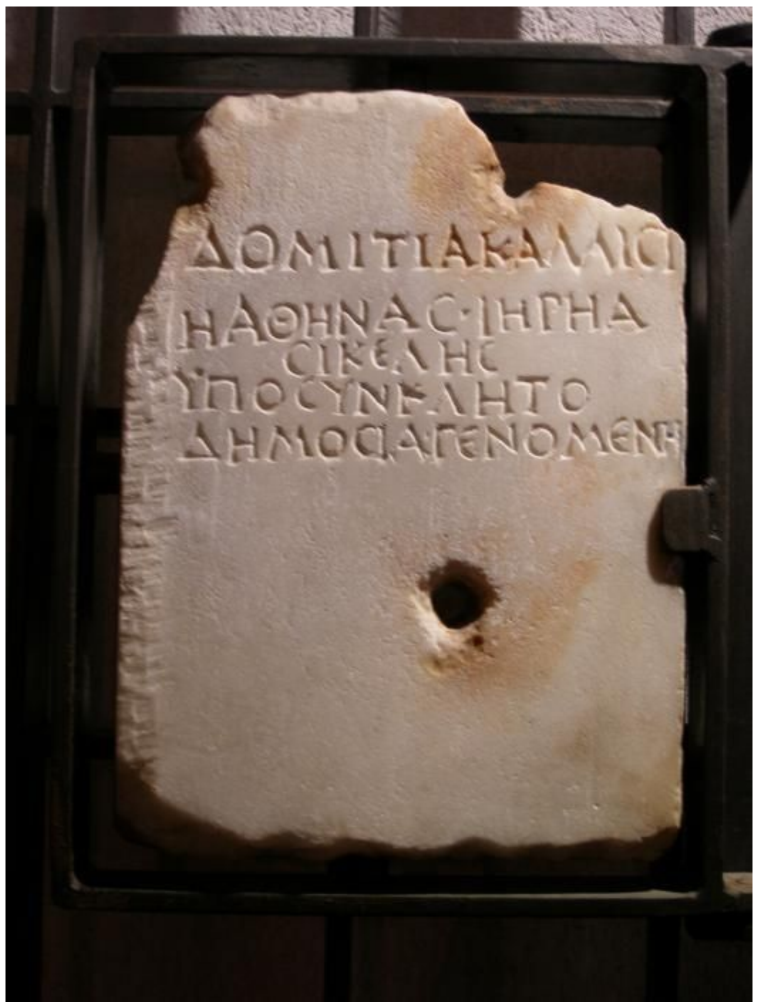


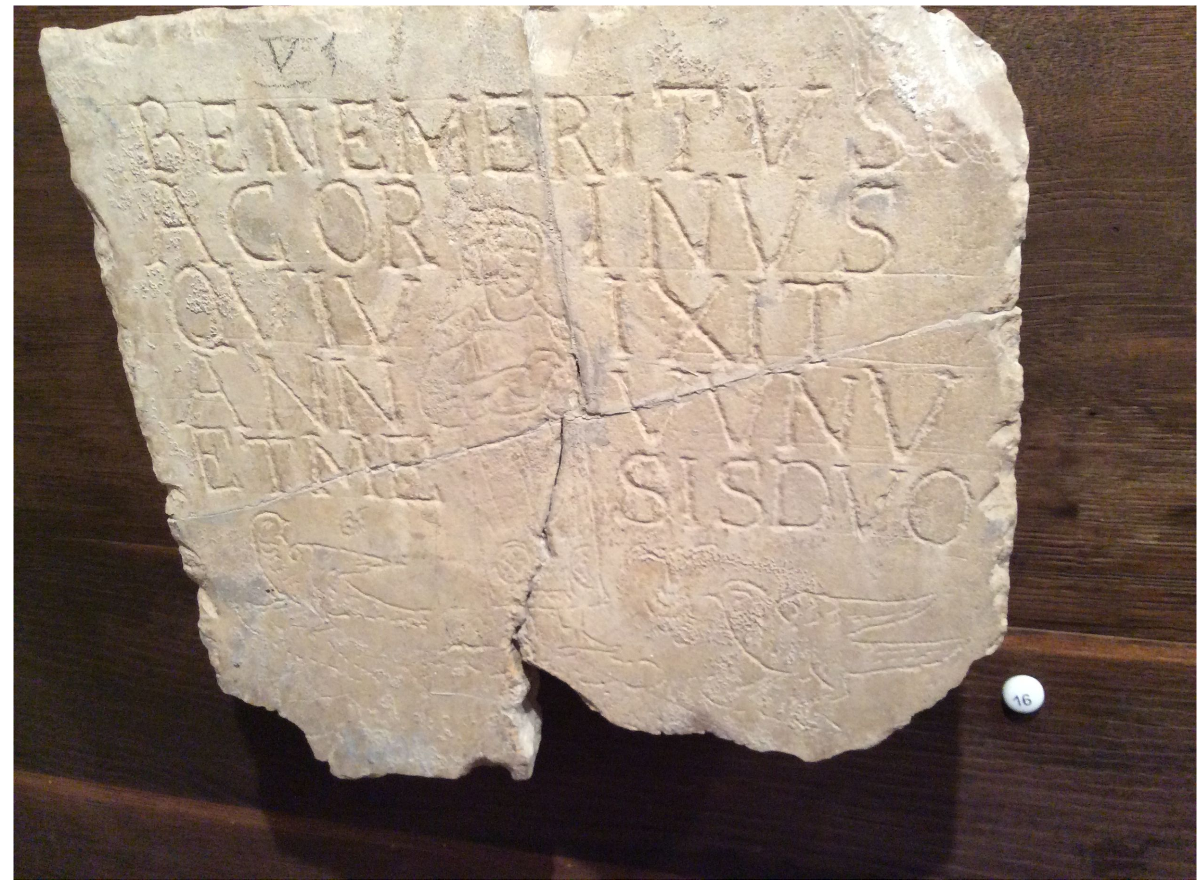
Types of knowledge

Propositional - What

A **proposition** is a sentence expressing something true or false

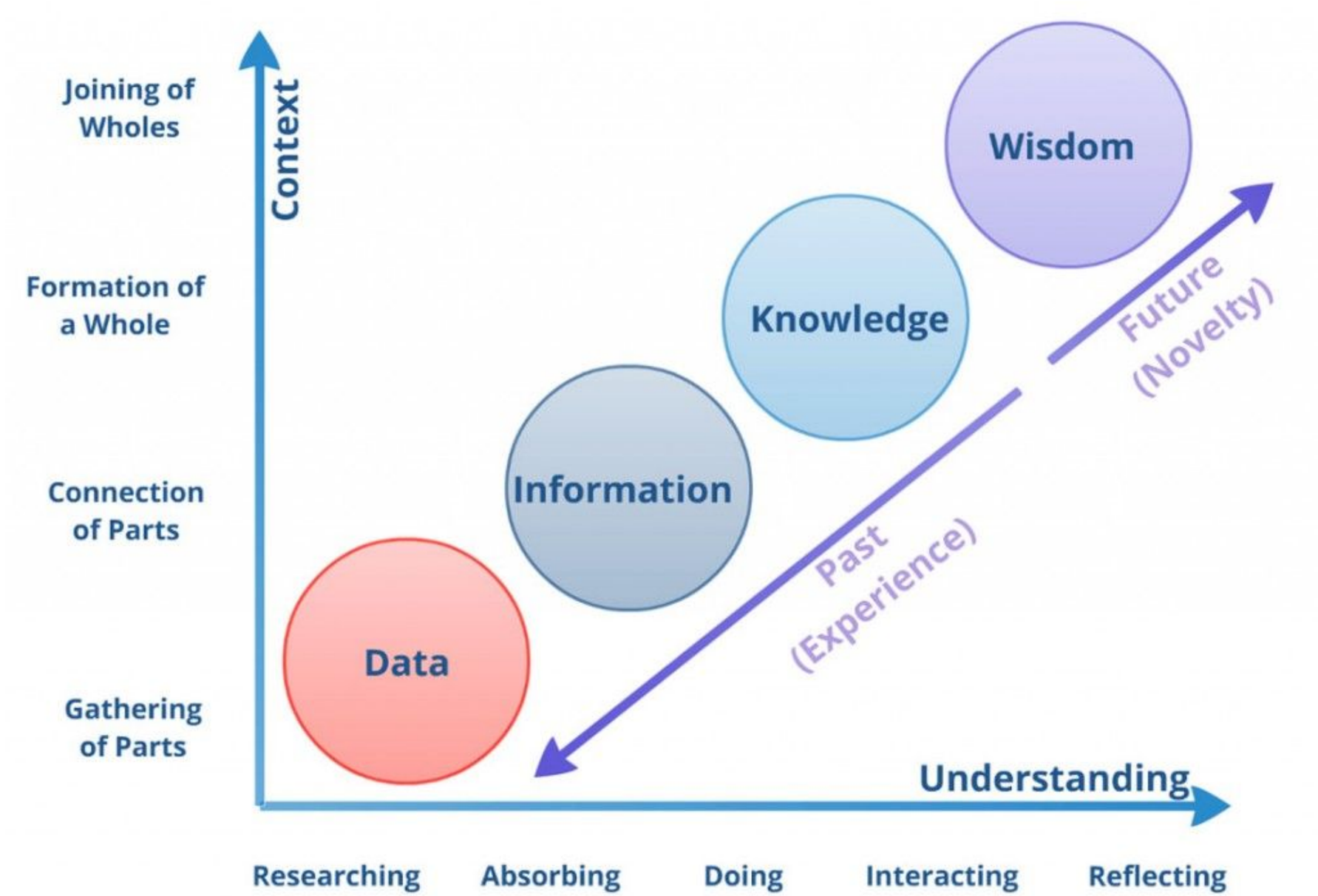
Procedural - How




BENEMERITVS
ACOR INVS
QVIV IXIT
ETME SISDVO

Interpretations and different points of view

- Linguistics ⟶ regards the language
- Philologist ⟶ regards its text
- Historians ⟶ as a source
- Archaeologist ⟶ as a material testimony of a event
- Conservationist ⟶ as a piece of matter to be preserved and restored


Joining of Wholes
Formation of a Whole
Connection of Parts
Gathering of Parts
Context
Wisdom
Knowledge
Information
Data
Future (Novelty)
Past (Experience)
Understanding
Researching
Absorbing
Doing
Interacting
Reflecting

SCENARIO

- Information is spread through different infrastructures archives.
- Digital archives are evenmore multi domain than ever and they manage different kinds of digital content and types of data.
- Resources are not always described using a common reference model.
- Resources are not created by archivist and librarians but by the experts or researchers.
- Often for the same object we have different conceptual definition.

Semantic Knowledge Management aims to

- Describe uniformly data present in the infrastructure contributing to the reuse of data and its interoperability
- Integrate research data also from different domains.
- Guarantee the semantic integration through mappings in the moment in which it‘s needed to use more than one achieve.

Metadata

...Data about data...Data over data...

Literal definition

<u>“Metadata is machine understandable information about web resources or other things”</u>

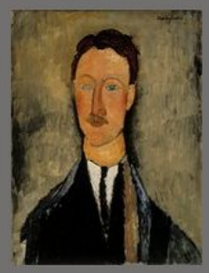
Tim Berners-Lee (1997-W3C)

Types of Metadata

- Structural
  - File extension
  - Type of object
  - ...
- Descriptive
  - Object
  - Subject
  - Title
  - ...
- Admninistrative
  - Right & Licence

Applications

- Archaeology
  - Archaeological items
  - Geographical entities
- 2D & 3D objects
  - …
- Digital Humanities
  - Digital Libraries
  - Showcase
  - …
- Digital Archieves
  - Repositories
  - …



Technical metadata

Title	**Taidemaalari Léopold Survagen muotokuva | Amedeo Modigliani** Portrait of the Artist Léopold Survage
Description	Inscription: oik.ylh.: Modigliani
People	**Creator:** Amedeo Modigliani (1884-07-12 — 1920-01-24)
Classifications) X 0 CancelCapture	**Type:** Physical Object , maalaus **Subject:** MUOTOKUVA , mies , taiteilija , Léopold Survage , rintakuva **Medium:** öljy kankaalle
Properties	**Size:** 61,5 x 46 cm
Time	**Creation Date:** 1918
Provenance	**Provenance:** Ateneumin taidemuseo **Identifier:** A IV 2971 **Institution:** Ateneumin taidemuseo **Provider:** National Library of Finland **Providing Country:** finland **First Published In Europeana:** 2014-05-02T08:21:09.172Z **Last Updated In Europeana:** 2014-05-02T08:21:09.172Z
Copyright	**Rights:** http://www.europeana.eu/rights/rr-f/
References And Relations	**Dataset:** 2021002_Ag_FI_NDL_fng

ÖAW AUSTRIAN ACADEMY OF SCIENCES

ACDH – AUSTRIAN CENTRE FOR DIGITAL HUMANITIES

Dublin Core Elements

Simple Dublin Core

identifier, creator, contributor, publisher, title, description, language, subject, coverage, format, type, date, relation, source, rights

- 15 elements
- Repeatable
- Domain Independent
- dc:namespace

ÖAW AUSTRIAN ACADEMY OF SCIENCES

ACDH – AUSTRIAN CENTRE FOR DIGITAL HUMANITIES

Dublin Core Elements

Qualified Dublin Core

abstract, accessRights, accrualMethod, accrualPeriodicity, accrualPolicy, alternative, audience, available, bibliographicCitation,conformsTo, **contributor**,**coverage**, created, **creator**, **date**, dateAccepted, dateCopyrighted, dateSubmitted, **description**, educationLevel, extent, **format**, hasFormat, hasPart, hasVersion, **identifier**, instructionalMethod, isFormatOf, isPartOf, isReferencedBy, isReplacedBy, isRequiredBy, issued, isVersionOf, **language**, license, mediator, medium, modified, provenance, **publisher**, references,**relation**, replaces, requires, **rights**, rightsHolder, source, spatial, **subject**, tableOfContents, temporal, **title**, **type**, valid

- 55 elements
- Repeatable
- Domain Independent
- dc:namespace
- Schema Validation

Formalisms

RDF -Resources Description Framework- http://www.w3.org/RDF/

DC –Dublin Core-

DC

```
<xml>
<rdf>
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc= "http://purl.org/dc/elements/1.1/">
<rdf:Description rdf:about="http://www.w3schools.com">
  <dc:description>W3Schools - Free tutorials>dc:description>
  <dc:publisher>Refsnes Data as</dc:publisher>
  <dc:date>2008-09-01</dc:date>
  <dc:type>Web Development>dc:type>
  <dc:format>text/html</dc:format>
  <dc:language>en</dc:language>
</rdf:Description>
</rdf>
```

Usefull links

The Dublin Core Metadata Initiative

- http://dublincore.org

DCMI Metadata Basics

- http://dublincore.org/metadatabasics

DCMI Metadata Terms

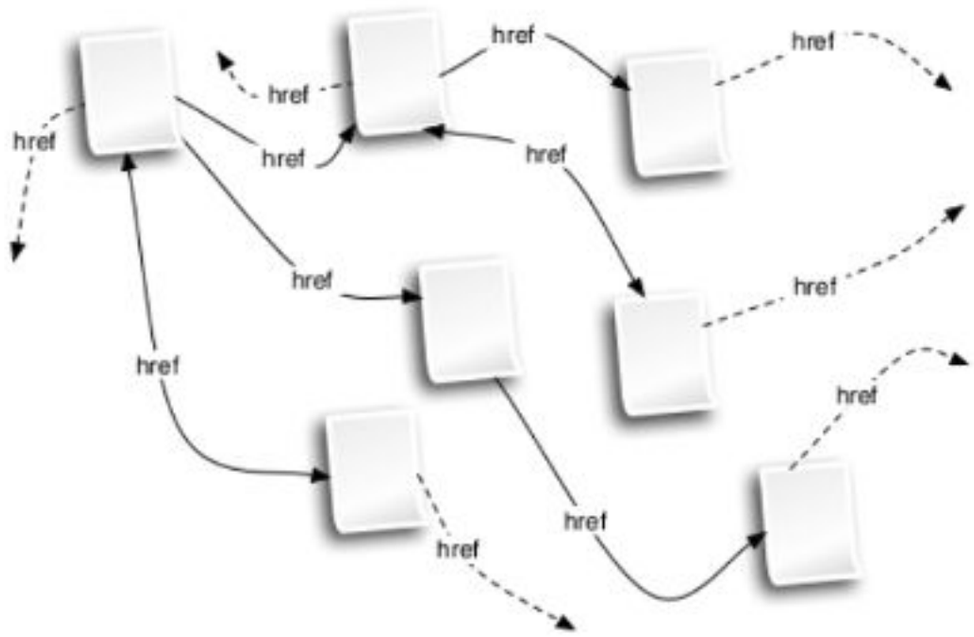
- http://dublincore.org/documents/dcmiterms

SEMANTIC WEB

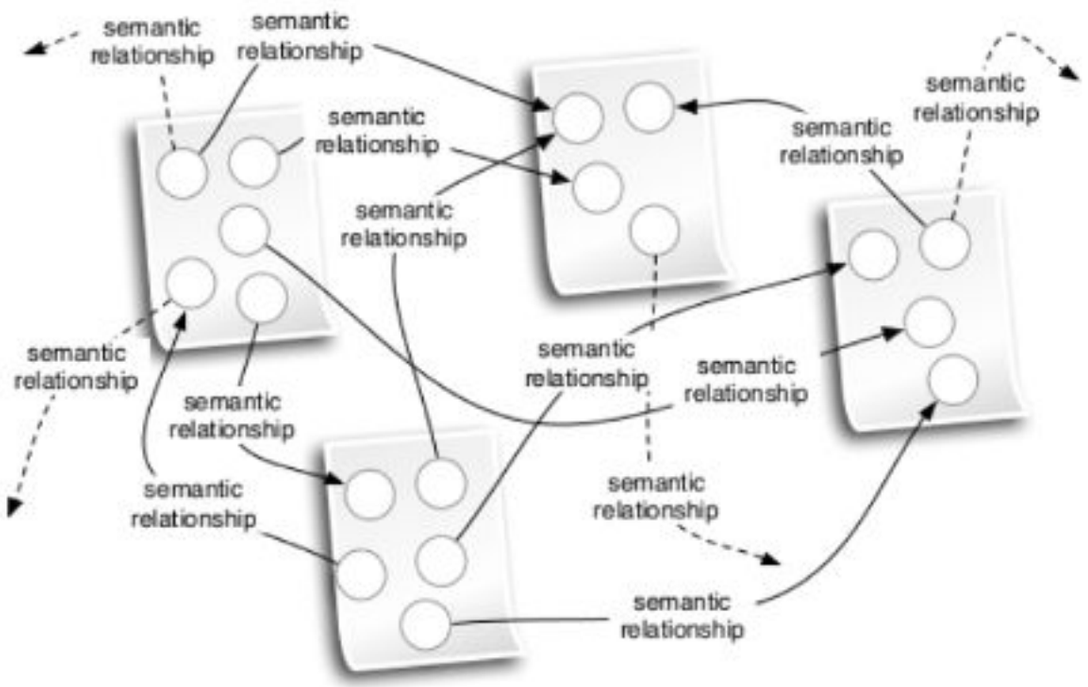
- Vision of Tim Berners-Lee **Web where the information assume a well defined significance**
- From Web of content to web of data...from web 2.0 to web 3.0 **common infrastructure able to allow the condivision and reuse of data**
- Set of rule and inference (common language)

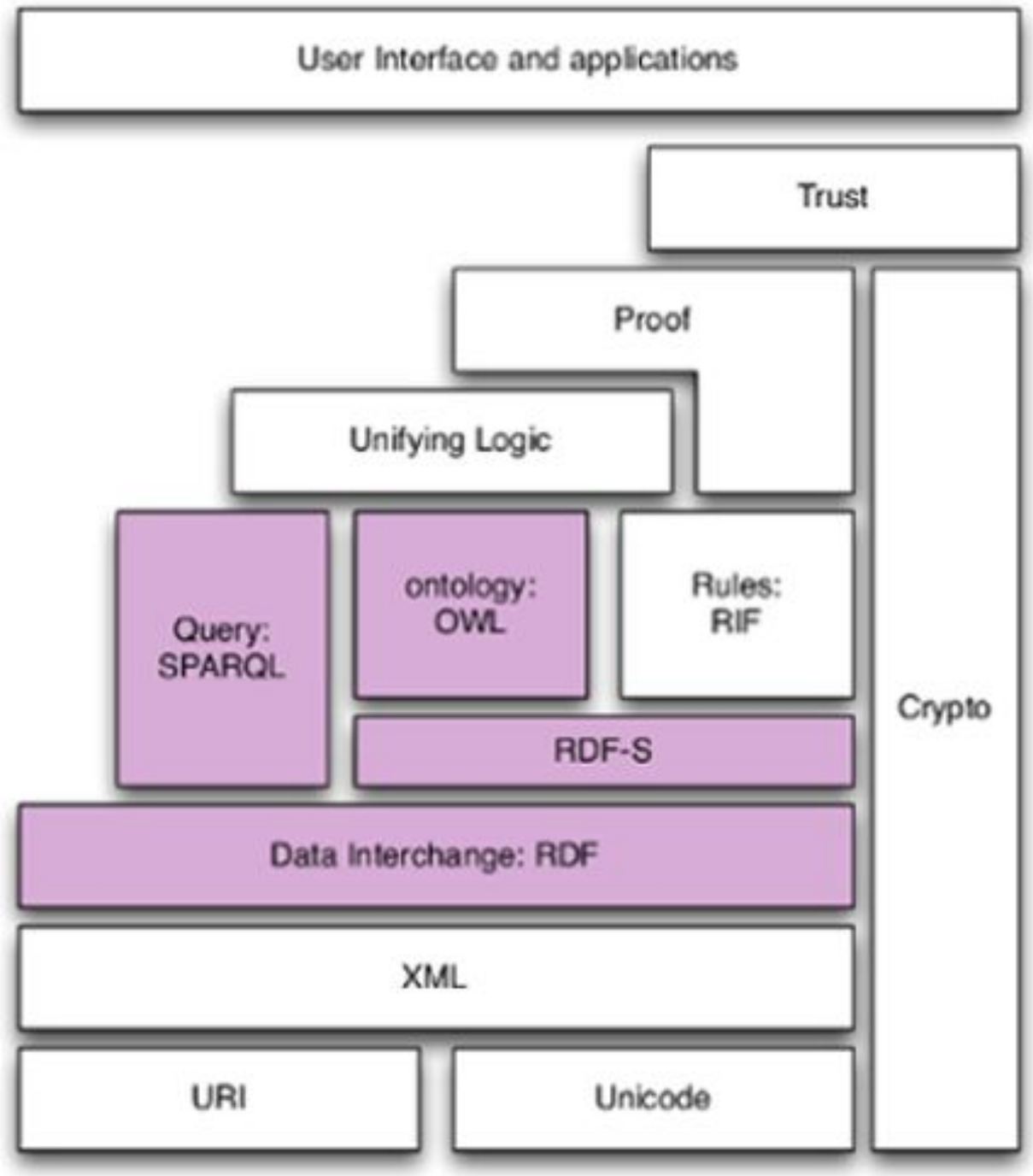
From web 2.0 to 3.0

The Web of Documents



The Web of Data




User Interface and applications
Trust
Proof
Unifying Logic
Query: SPARQL
ontology: OWL
Rules: RIF
RDF-S
Data Interchange: RDF
XML
URI
Unicode
Crypto

Resource Description Framework -RDF-

Self Describing!! - a framework for describing resources…

- Resources: Things that can be named with URIs (http, urn, doi…)
- Description: Statements about the *properties* of these resources
- Framework: A common model for diversity

- Builds on URI for identifiers, XML for syntax (file format)
  - URI: Uniform Resource Identifier
  - XML: Extensible Markup Language
- Standard serie ISO19100 developed by W3C consortium

https://www.w3.org/RDF/

Ontology Web Language -OWL-

- Extends existing Web standards
  - Such as XML, RDF, RDFS
- Easy to understand and use
  - Should be based on familiar KR idioms
- Formally specified - describes the meaning of knowledge precisely
- Support automated reasoning support

Ontology Web Language -OWL-

Three species of OWL

- OWL full is union of OWL syntax and RDF
- OWL DL restricted to FOL fragment
- OWL Lite is“easier to implement” subsetofOWLDL

OWL DL Benefits from many years of DL research

- Well defined semantics
- Formal properties well understood(complexity,decidability)
- Known reasoning algorithms
- Implemented systems (highly optimised)

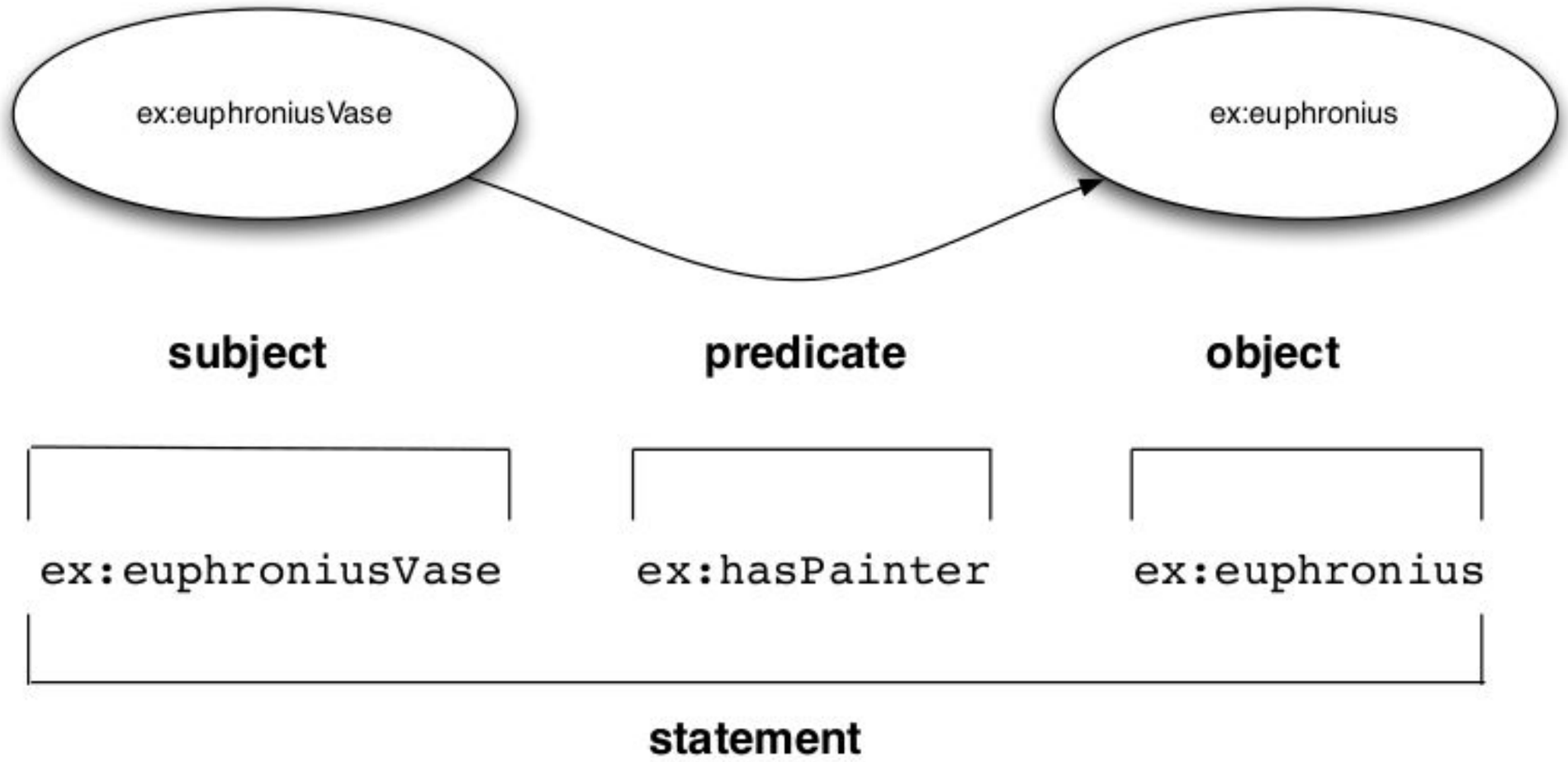
Description Logics

- A family of logic based Knowledge Representation formalisms
  - Descendants of **semantic networks** and **KL-ONE**
    - Describe domain in terms of **concepts** (classes), **roles** (properties, relationships) and **individuals**
- Distinguished by:
  - Formal Semantics (typically model theoretic)
    - Decidable fragments of FOL
  - Provision of inference services

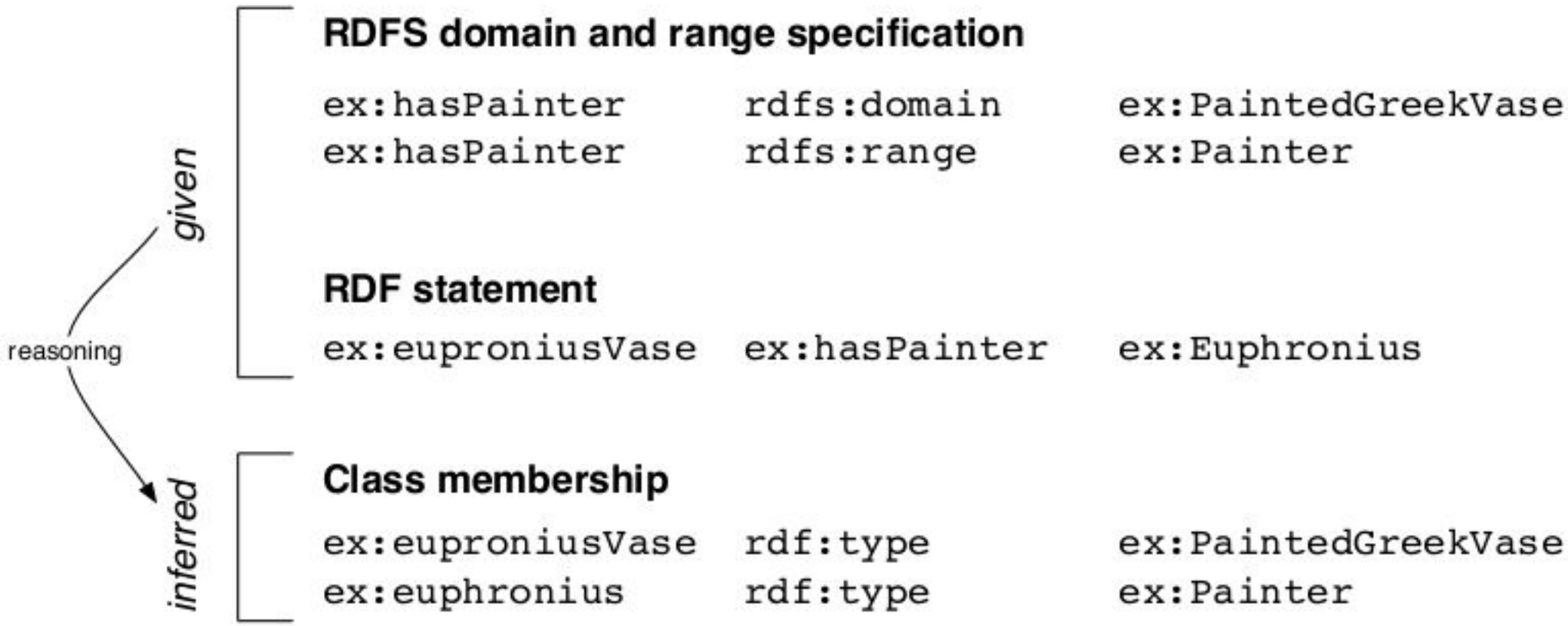
DL Basics

- Concept names are equivalent to unary predicates
  – In general, concepts equiv to formulae with one free variable
- Role names are equivalent to binary predicates
  – In general, roles equiv to formulae with two free variables
- Individual names are equivalent to constants
- Operators restricted so that:
  – Language is decidable and, if possible, of low complexity – No need for explicit use of variables
    - Restricted form of ∃ and ∀ (direct correspondence with ◊ and [])

Triple (S/O/P)


ex:euphroniusVase
ex:euphronius
subject
predicate
object
ex:euphroniusVase
ex:hasPainter
ex:euphronius
statement

Triple (inferred)

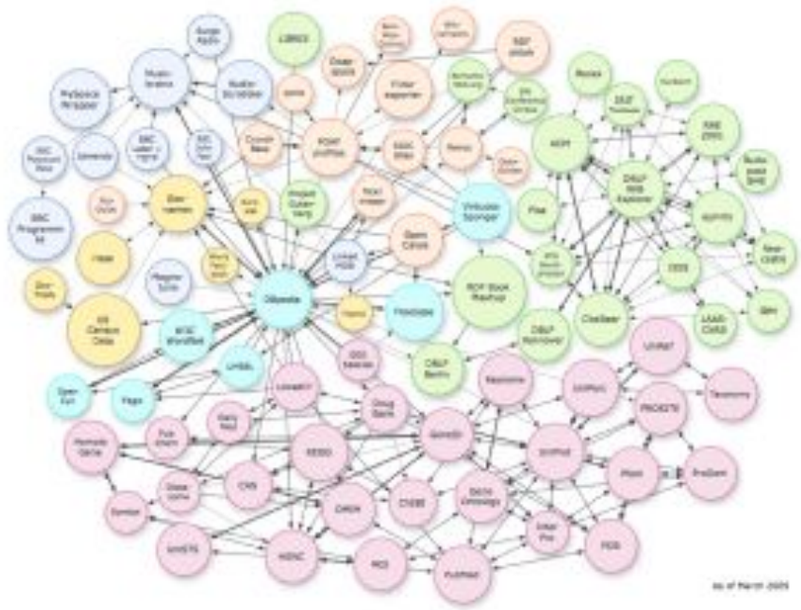


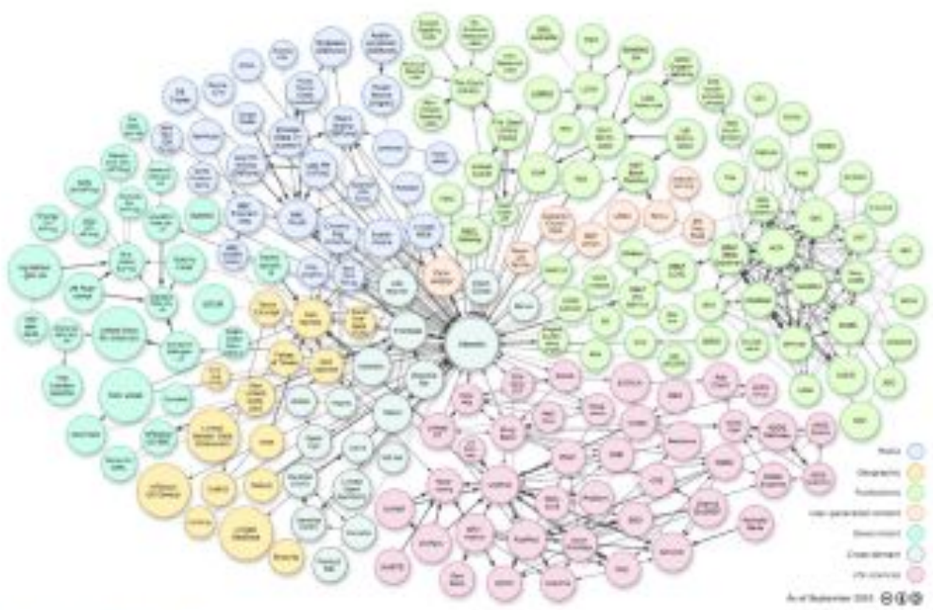
E.G “Give me all the authors from medieval Italy that have written about India” or “Give me all the context from north of Italy where I can find Terra Sigillata Italica”

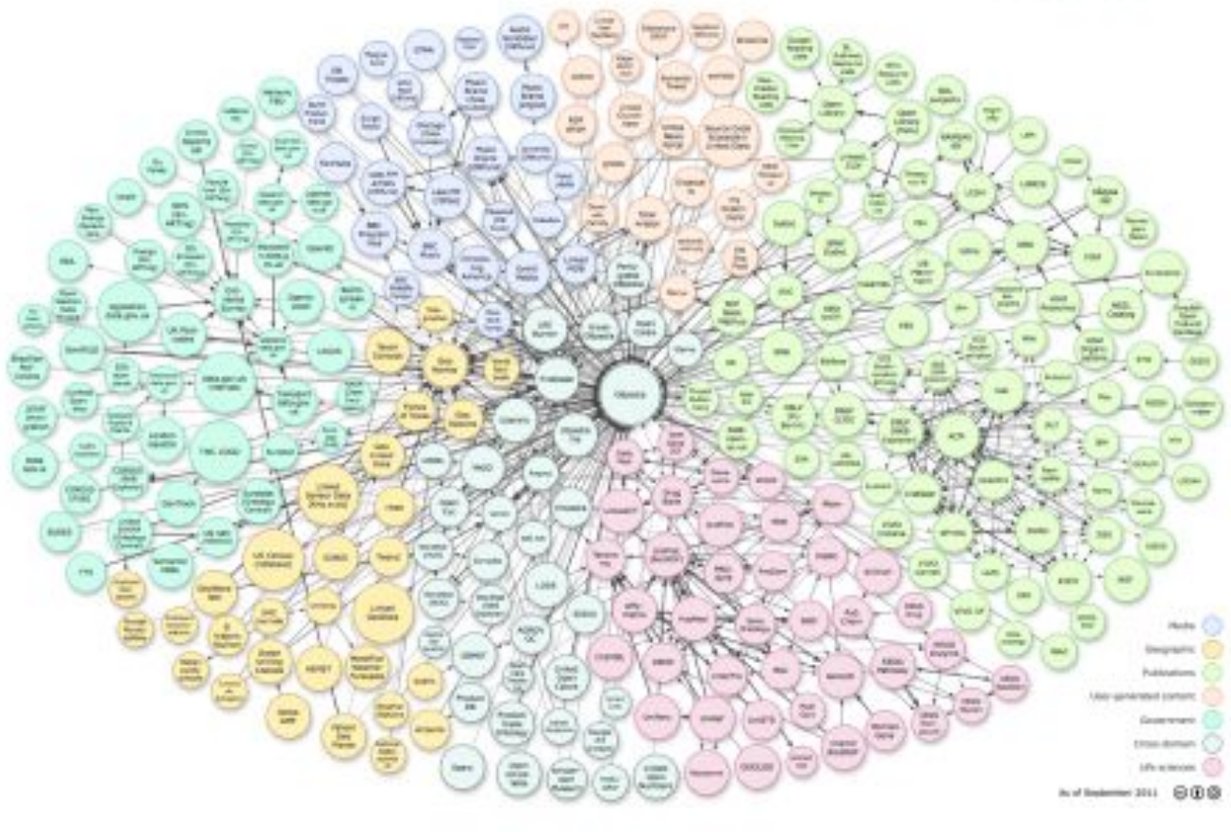
implies graphical, geographical and chronological resources that, if they are not defined to follow a homogeneous mapping, they will never be associated to the request made by the user.

Linked Data

- URI (Universal Resource Identifier) for the identification of each resource (concept/document)
- URI and HTTP prefix for finding the resource on the web
- RDF for providing usefull information about an object
- Including link between different resources






Media
Geographic
Publications
User-generated content
Government
Cross-domain
Life sciences

ÖAW AUSTRIAN ACADEMY OF SCIENCES

ACDH – AUSTRIAN CENTRE FOR DIGITAL HUMANITIES