KNOWLEDGE GRAPH COMPLETION
PART 1: INTRODUCTION

FATIHA SAÏS [1]
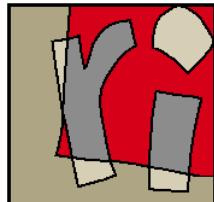
NATHALIE PERNELLE[1]

DANAI SYMEONIDOU[2]
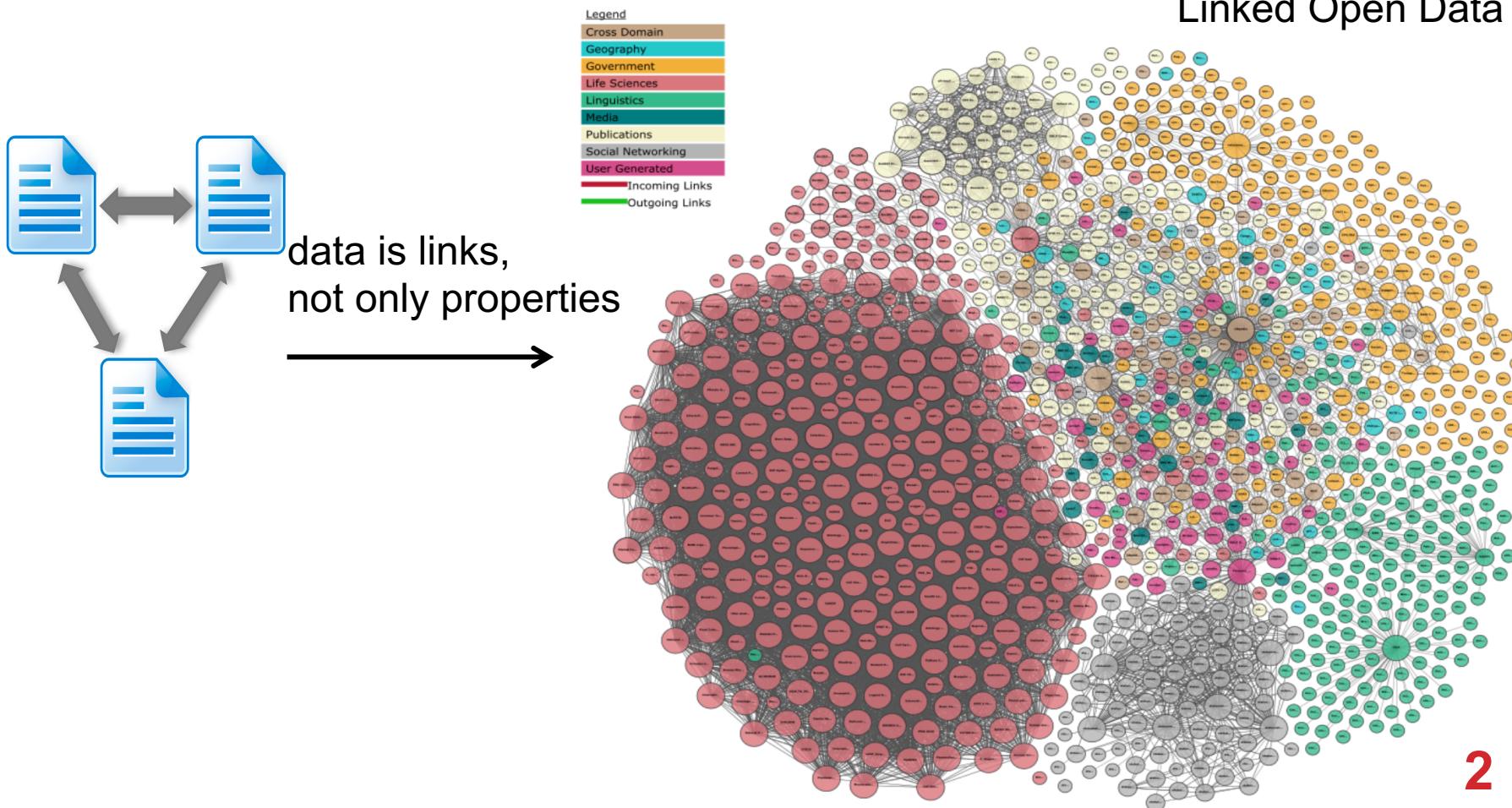
[1] LRI, PARIS SUD UNIVERSITY, CNRS, PARIS SACLAY UNIVERSITY

[2] INRA, GAMMA TEAM

1

FROM THE WWW TO THE WEB OF DATA

- applying the principles of the WWW to data

data is links,
not only properties

Linked Open Data

Legend
- Cross Domain
- Geography
- Government
- Life Sciences
- Linguistics
- Media
- Publications
- Social Networking
- User Generated
- Incoming Links
- Outgoing Links

LINKED DATA PRINCIPLES

Tim Berners Lee, 2006

① **Use HTTP URIs as identifiers for resources**

→ so people can look up the data

② **Provide data at the location of URIs**

→ to provide data for interested parties

③ **Include links to other resources**

→so people can discover more information

→bridging disciplines and domains
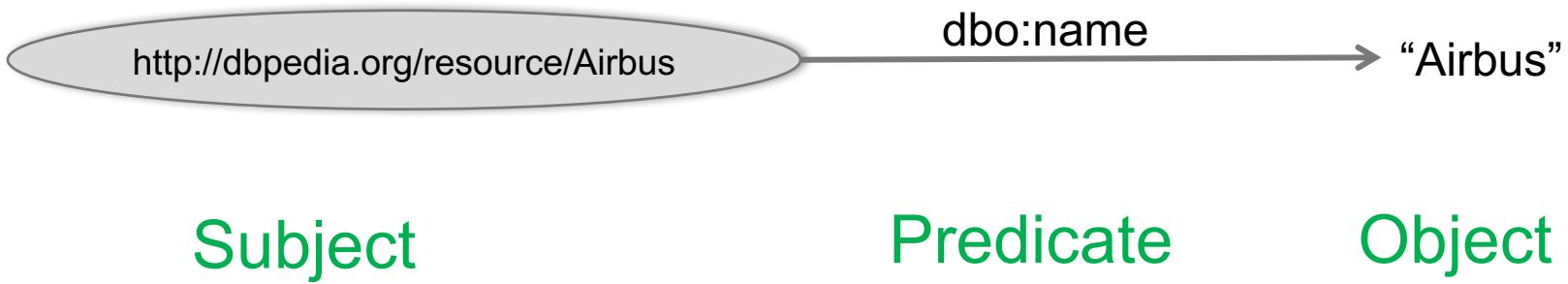
→ Unlock the potential of isolated repositories (islands)

3

RDF – RESOURCE DESCRIPTION FRAMEWORK

- **Statements of < subject predicate object >**

Subject   Predicate   Object
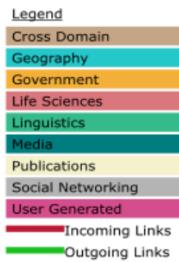
… is called a triple

LINKED OPEN DATA

Linked Open Data (LOD)
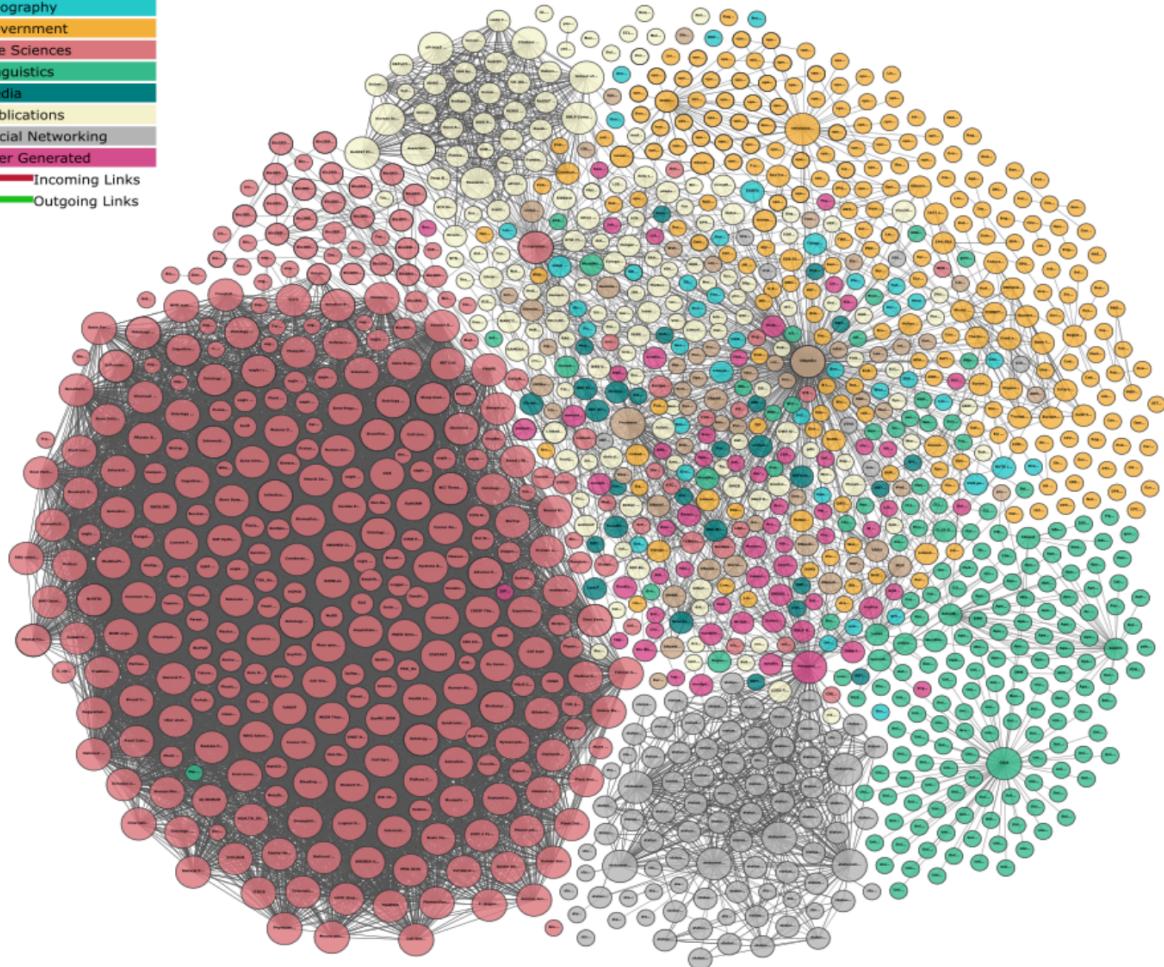
Linked Data - Datasets under an open access

- **1,139 datasets**

- **over 100B triples**

- **about 500M links**

- **several domains**

Ex. DBPedia : 1.5 B triples

Legend
Cross Domain
Geography
Government
Life Sciences
Linguistics
Media
Publications
Social Networking
User Generated
Incoming Links
Outgoing Links

"Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/"

NEED OF KNOWLEDGE

THE ROLE OF KNOWLEDGE IN AI

[Artificial Intelligence 47 (1991)]

ON THE THRESHOLDS OF KNOWLEDGE

Douglas B. Lenat

MCC
3500 W. Balcones Center
Austin, TX 78759

Edward A. Feigenbaum

Computer Science Department
Stanford University
Stanford, CA 94305

Abstract

We articulate the three major fmdings of AI to date: (1) The Knowledge Principle: if a program is to perform a complex task well, it must know a great deal about the world in which it operates. (2) A plausible extension of that principle, called the Breadth Hypothesis: there are two additional abilities necessary for intelligent behavior in unexpected situations: falling back on increasingly general knowledge, and analogizing to specific but far-flung knowledge. (3) AI as Empirical Inquiry: we must test our ideas experimentally, on large problems. Each of these three hypotheses proposes a particular threshold to cross, which leads to a qualitative change in emergent intelligence. Together, they determine a direction for future AI research.

opponent is Castling.) Even in the case of having to search for a solution, the *method* to carry out the search may be

*The knowledge principle: "if a program is to perform a complex task well, **it must know a great deal about the world** in which it operates."*

there is some minimum knowledge needed for one to even formulate it.

7

ONTOLOGY, A DEFINITION

"An ontology is an **explicit**, **formal specification** of a **shared conceptualization."**

[Thomas R. Gruber, 1993]

Conceptualization: abstract model of domain related expressions

Specification: domain related

Explicit: semantics of all expressions is clear

Formal: machine-readable

**Shared:** consensus (different people have different perceptions)
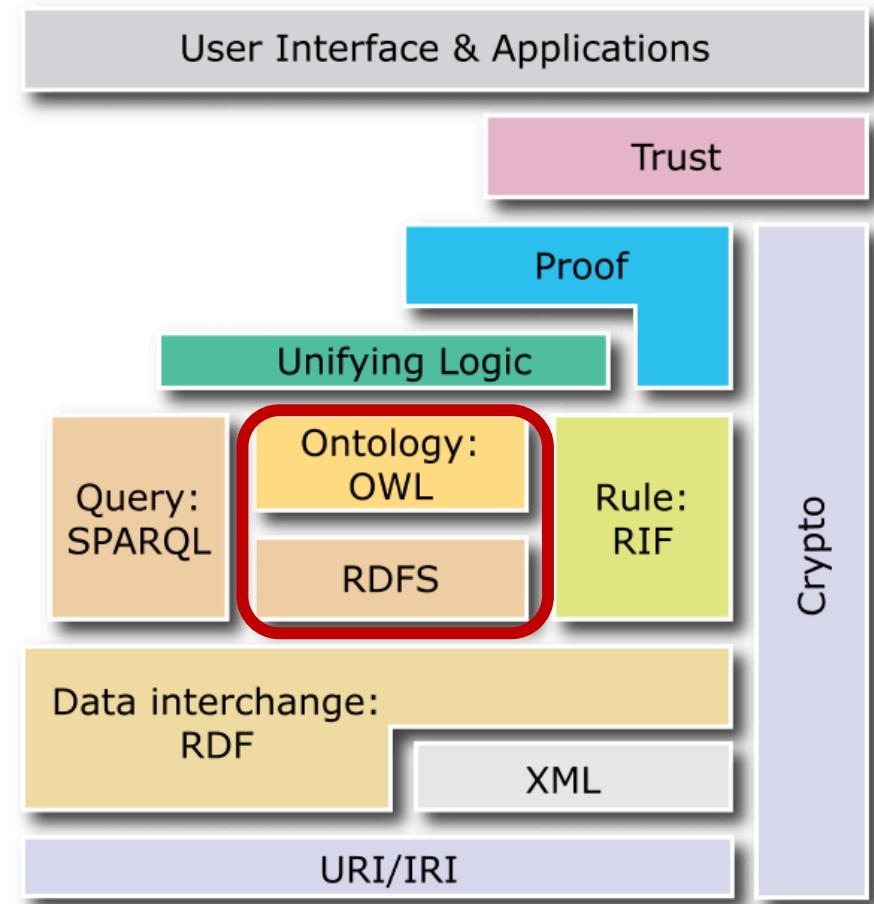
SEMANTIC WEB: ONTOLOGIES

RDFS – Resource Description Framework Schema

- Lightweight ontologies

OWL – Web Ontology Language

- Expressive ontologies
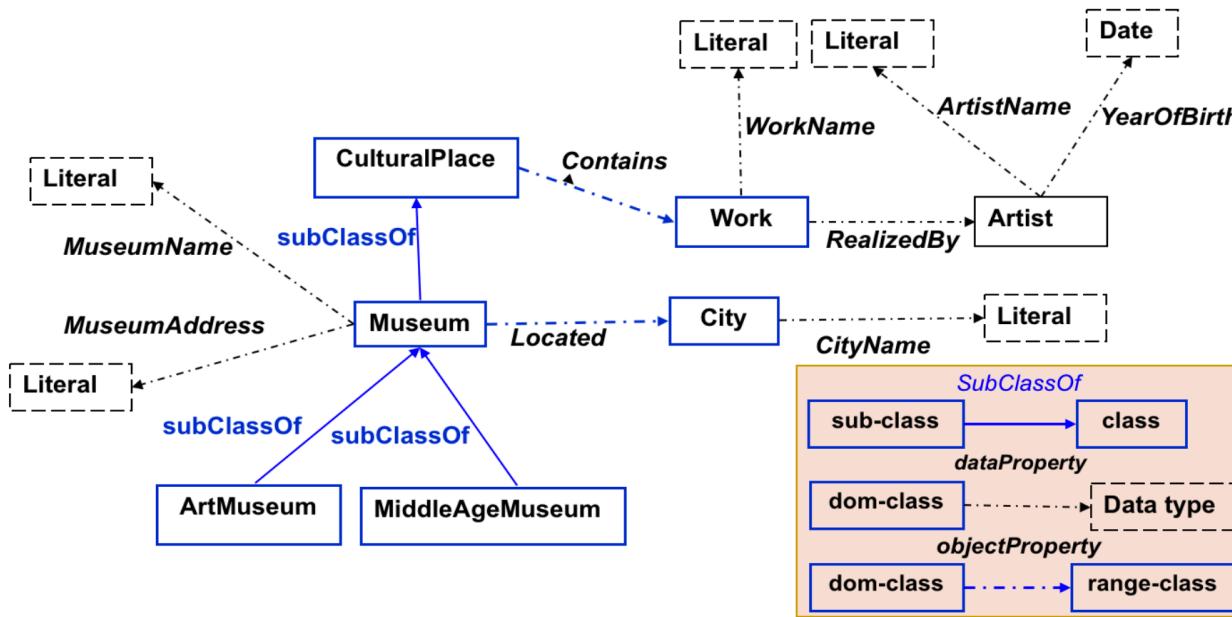
OWL ONTOLOGY

OWL – Web Ontology Language

- Represents rich and complex knowledge about things

- Based on Description Logic

- Can be used to verify the consistency of knowledge

- Can make implicit knowledge explicit
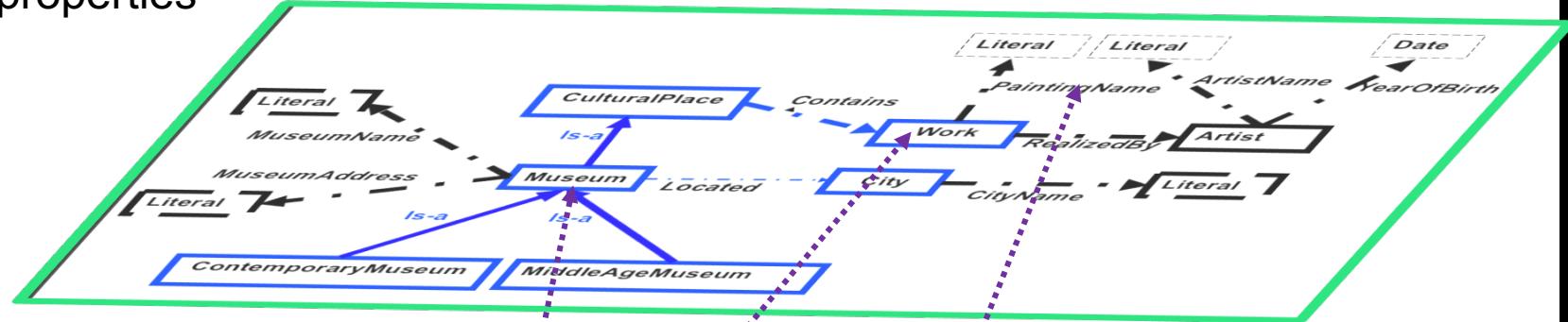
- **Classes:** concepts or collections of objects (individuals)

- **Properties:**
  - owl:DataTypeProperty (attribute)
  - owl:ObjectProperty (relation)

- **Hierarchy:**
  - owl:subClassOf
  - owl:subPropertyOf

- **Individuals:** ground-level of the ontology (instances)
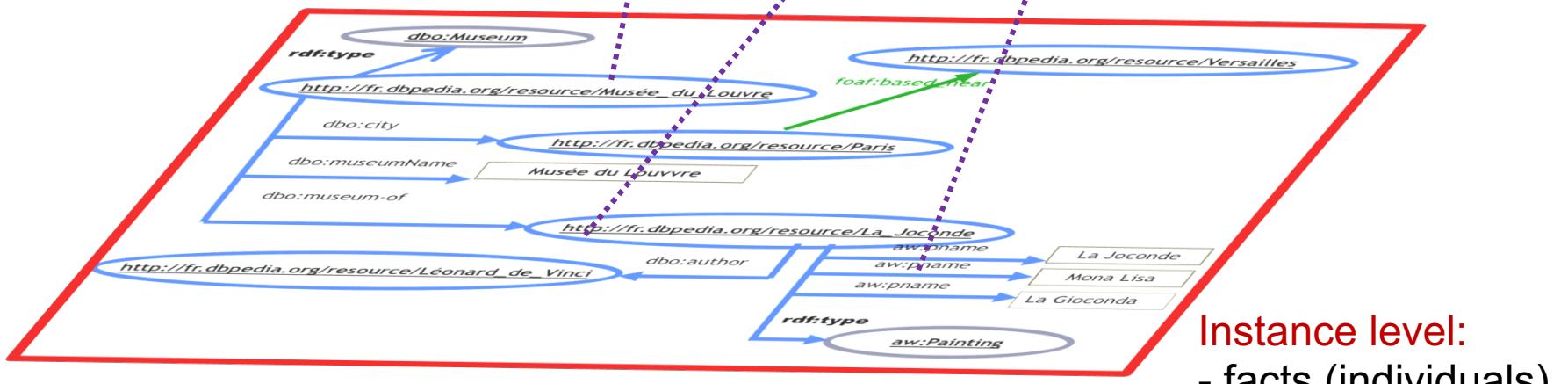
ONTOLOGY LEVELS

Conceptual level:
- classes, properties
(relations)

Instance level:
- facts (individuals)

OWL ONTOLOGY - AXIOMS

- **Axioms:** knowledge definitions in the ontology that were **explicitly defined** and have **not been proven true**.

    - **Reasoning over an ontology**

        → Implicit knowledge can be made explicit by logical reasoning

- **Example:**

    Pompidou museum is an **Art Museum**

    < Pompidou_museum rdf:type ArtMuseum> .

    Pompidou museum contains *Hallucination partielle*

    < Pompidou_museum ao:contains *Hallucination_partielle> .*

- **Infer that:**

➔ Pompidou museum is a CulturalPlace

    < Pompidou_museum rdf:type CulturalPlace> .

Because: **Museum** subsumes **ArtMuseum** and **CulturalPlace** subsumes **Museum**

➔ *Hallucination partielle* is a Work

    <Hallucination_partielle rdf:type ao:Work> .
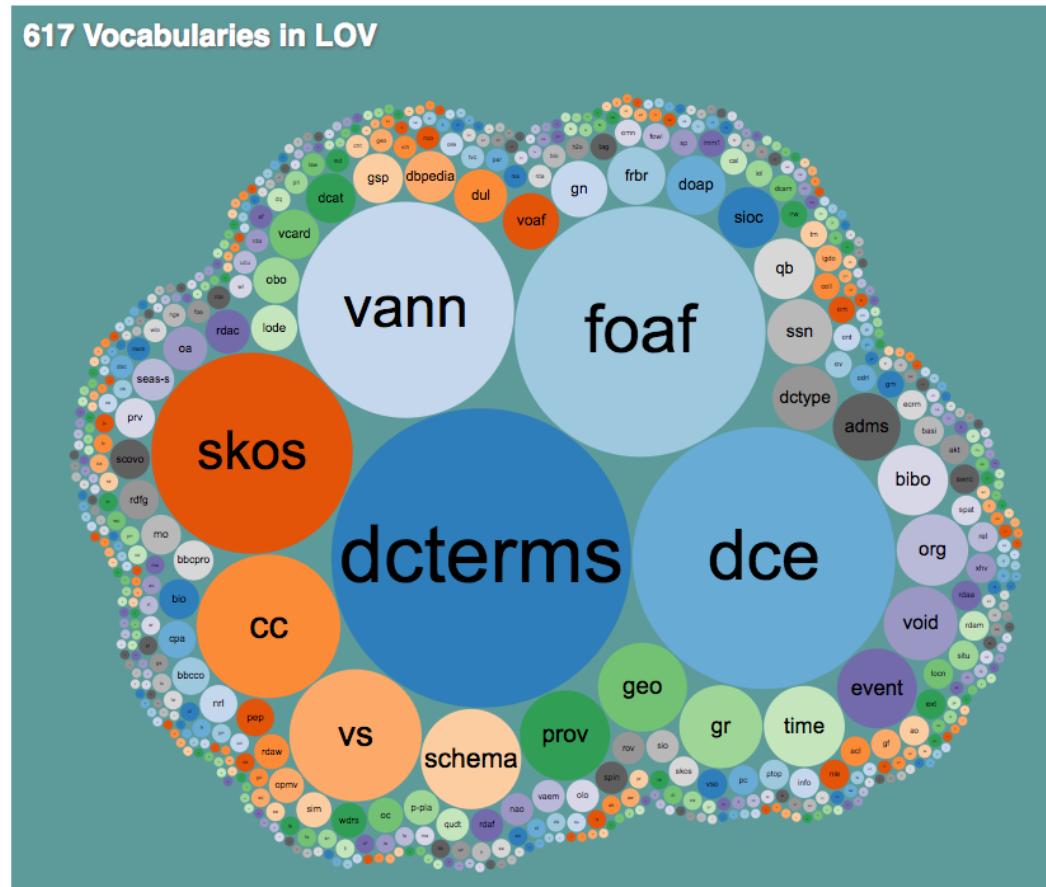
Because: the **range** of the object property **contains** is the class **Work**.

LINKED OPEN VOCABULARIES (LOV)

Linked Open Vocabularies

- Keeps track of available open ontologies and provides them as a graph

- Search for available ontologies, open for reuse

- Example:

http://lov.okfn.org/dataset/lov/vocabs/foaf

617 Vocabularies in LOV
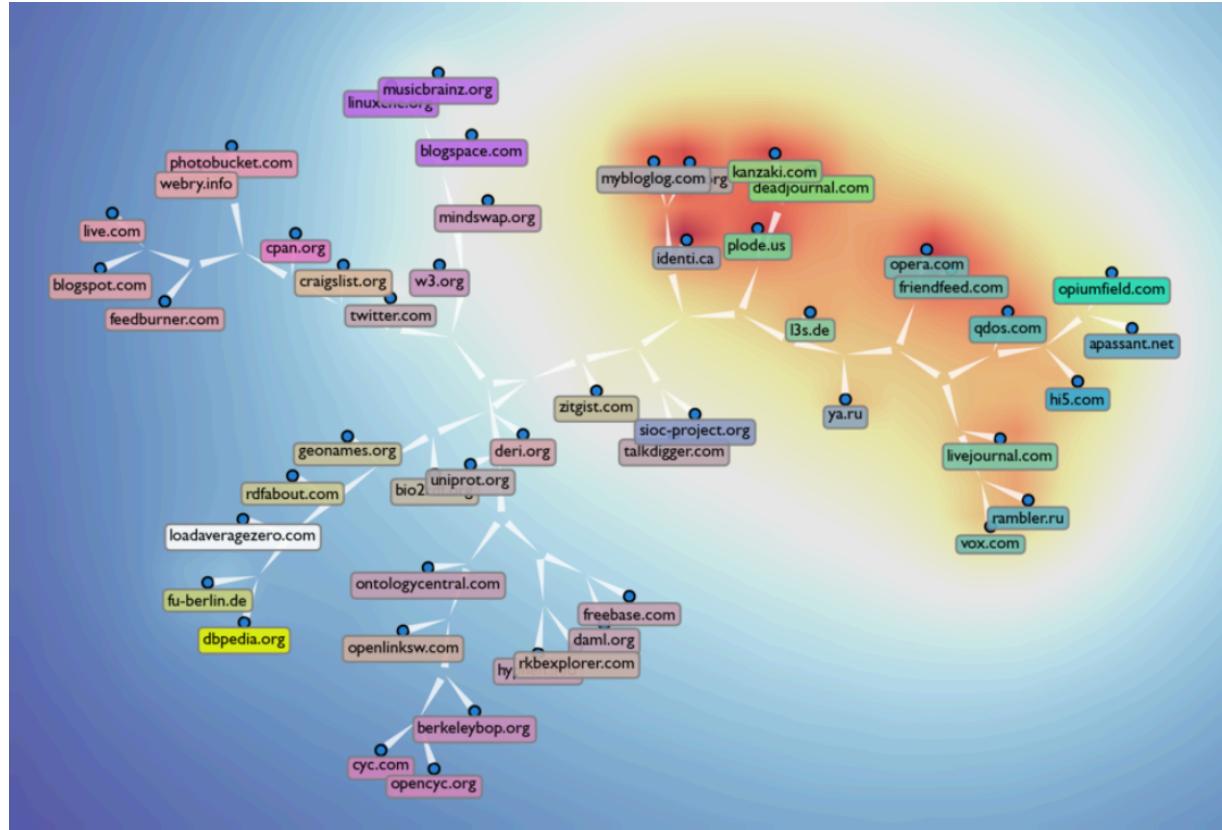
ONTOLOGY PREDICATES SPREAD ON THE SEMANTIC WEB

- **RDFa (or [Resource Description Framework](#) in Attributes)**

- Top 50 web sites publishing Semantic Web data, clustered by predicates used.
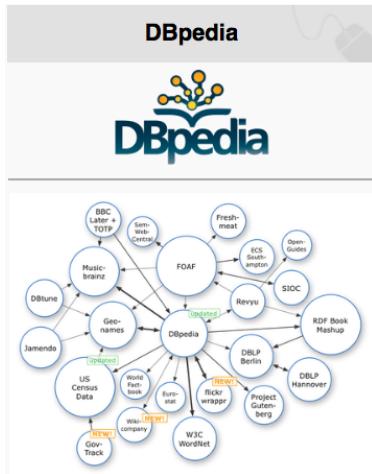
FOAF

KNOWLEDGE GRAPHS

WHO IS DEVELOPING KNOWLEDGE GRAPHS?

2013

2012

2007

2008

2012

2015

2016

2007

2013

Academic side
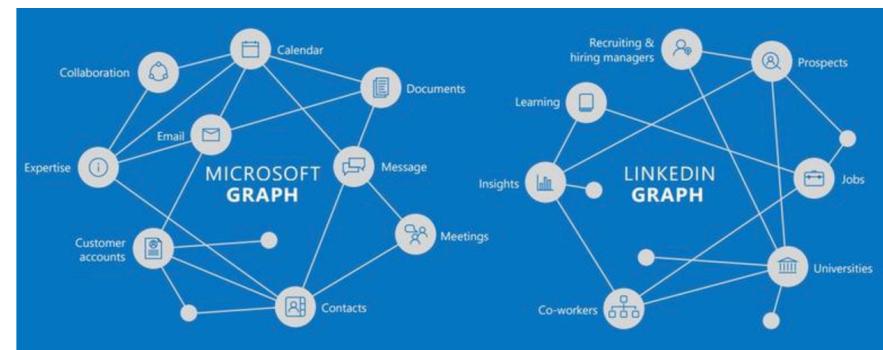
Commercial side

WEB SEARCH WITHOUT KNOWLEDGE GRAPHS

WEB SEARCH WITH KNOWLEDGE GRAPHS

QUESTION ANSWERING WITH KNOWLEDGE GRAPHS

CONNECTING EVENTS AND PEOPLE WITH KNOWLEDGE GRAPHS

TOWARDS A KNOWLEDGE-POWERED DIGITAL ASSISTANT

Cortana (Microsoft)

- Natural access and storage of knowledge
- Chat bots
- Personalization
- Emotion

KNOWLEDGE GRAPH: A DEFINITION …

The **Knowledge Graph** is a knowledge base used by Google to enhance its search engine's search results with semantic-search information gathered from a wide variety of sources. Knowledge Graph display was added to Google's search engine in 2012, starting in the United States, having been announced on May 16, 2012.[1] It uses a graph database to provide structured and detailed information about the topic in addition to a list of links to other sites. The goal is that users would be able to use this information to resolve their query without having to navigate to other sites and assemble the information themselves.[2] The short summary provided in the knowledge graph is often used as a spoken answer in Google Assistant searches.[3]

Wikipedia (en)

This is not a formal definition!

KNOWLEDGE GRAPH: A DEFINITION …

[L. Ehrlinger and W. Wöß, SEMANTiCS'2016]

Definition	Source
"A knowledge graph (i) mainly describes real world entities and their interrelations, organized in a graph, (ii) defines possible classes and relations of entities in a schema, (iii) allows for potentially interrelating arbitrary entities with each other and (iv) covers various topical domains."	Paulheim [16]
"Knowledge graphs are large networks of entities, their semantic types, properties, and relationships between entities."	Journal of Web Semantics [12]
"Knowledge graphs could be envisaged as a network of all kind things which are relevant to a specific domain or to an organization. They are not limited to abstract concepts and relations but can also contain instances of things like documents and datasets."	Semantic Web Company [3]
"We define a Knowledge Graph as an RDF graph. An RDF graph consists of a set of RDF triples where each RDF triple (s, p, o) is an ordered set of the following RDF terms: a subject $s \in U \cup B$, a predicate $p \in U$, and an object $U \cup B \cup L$. An RDF term is either a URI $u \in U$, a blank node $b \in B$, or a literal $l \in L$."	Färber et al. [7]
"[…] systems exist, […], which use a variety of techniques to extract new knowledge, in the form of facts, from the web. These facts are interrelated, and hence, recently this extracted knowledge has been referred to as a knowledge graph."	Pujara et al. [17]

→ Populated Ontology

→ RDF Graph

→ Populated Ontology

→ RDF Graph

→ Extracted RDF Graph

[3] A. Blumauer. From Taxonomies over Ontologies to Knowledge Graphs, July 2014. https://blog.semanticweb.at/2014/07/15/from-taxonomies-over-ontologiesto-knowledge-graphs [August, 2016].

[7] M. Farber, B. Ell, C. Menne, A. Rettinger, and F. Bartscherer. Linked Data Quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. Semantic Web Journal, 2016. http://www.semantic-web-journal. net/content/linked-data-quality-dbpedia-freebaseopencyc-wikidata-and-yago [August, 2016] (revised version, under review).
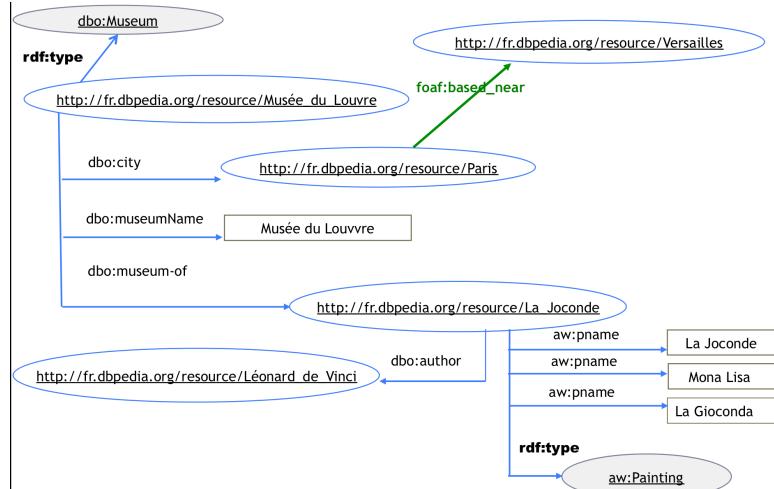
[12] M. Kroetsch and G. Weikum. Journal of Web Semantics: Special Issue on Knowledge Graphs. http://www.websemanticsjournal.org/index.php/ps/announcement/view/19 [August, 2016].

[16] H. Paulheim. Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods. Semantic Web Journal, (Preprint):1–20, 2016.
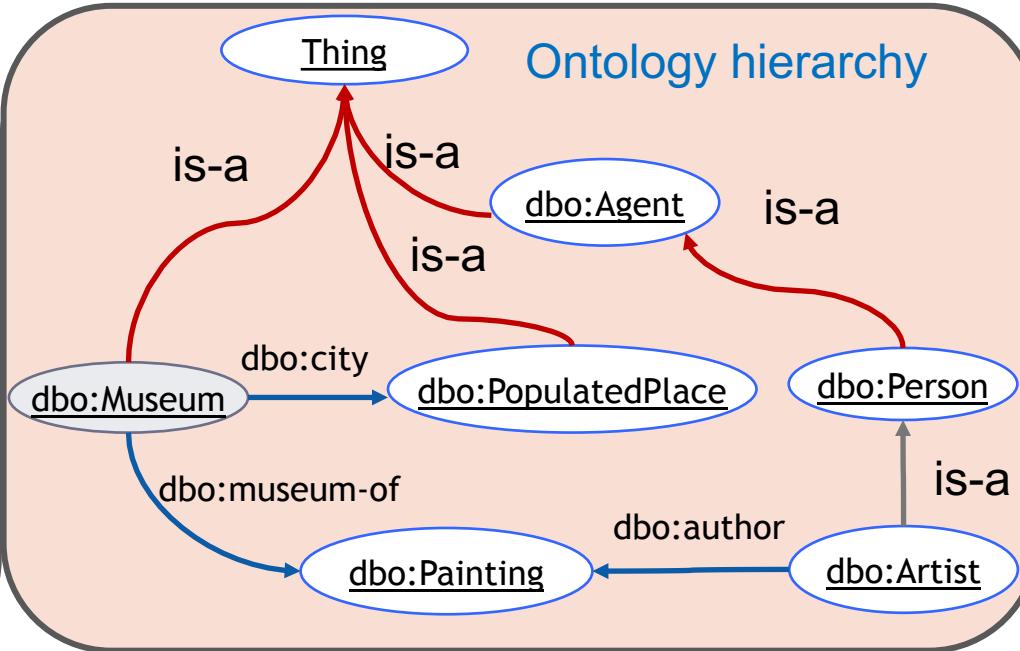
[17] J. Pujara, H. Miao, L. Getoor, and W. Cohen. Knowledge Graph Identification. In Proceedings of the 12th International Semantic Web Conference - Part I, ISWC '13, pages 542–557, New York, USA, 2013. Springer.

KNOWLEDGE GRAPH (KG)

RDF Graphs

Ontology hierarchy

Querying (SPARQL)

PREFIX dbo: <http://dbpedia.org/ontology#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?m ?p
WHERE { ?m rdf:type dbo:Museum . ?m dbo:musuem-of ?p .}

Reasoners: (Pellet, Fact++, Hermit, etc.)

- KG saturation: infer whatever can be inferred from the KG.
- KG consistency checking: no contradictions
- KG repairing
- …

Ontology axioms and rules

owl:equivalentClass(dbo:Municipality, dbo:Place)
owl:equivalentClass(dbo:Place, dbo:Wikidata:Q532)
owl:equivalentClass(dbo:Village, dbo:PopulatedPlace)
owl:equivalentClass(dbo:PopulatedPlace,dbo:Municipality)
owl:disjointClass(dbo:PopulatedPlace, dbo:Artist)
owl:disjointClass(dbo:PopulatedPlace, dbo:Painting)
owl:FunctionalProperty(dbo:city)
owl:InverseFunctionalProperty(dbo:museum-of)

dbo:birthPlace(X, Y) => dbo:citizsenOf(X, Y)
dbo:parentOf(X, Y) => dbo:child(Y, X)

24

KNOWLEDGE GRAPH EXPANSION AND ENRICHMENT

- **Expansion: knowledge graphs are incomplete**

  - Data linking (entity resolution, reference reconciliation, …)
  - Link prediction: add relations
  - Ontology matching: connect graphs
  - Missing values prediction/inference

- **Enrichment: can new knowledge emerge from knowledge graphs?**

  - Knowledge discovery: Key discovery
  - Automatic reasoning and planning

- **Validation: knowledge graphs may contain errors**

  - Link Validation
  - Dealing with errors and ambiguities

KNOWLEDGE GRAPH EXPANSION AND ENRICHMENT

- **Expansion: knowledge graphs are incomplete**

  - **Data linking (entity resolution, reference reconciliation, …)**
  - Link prediction: add relations
  - Ontology matching: connect graphs
  - Missing values prediction/inference

- **Enrichment: can new knowledge emerge from knowledge graphs?**

  - **Knowledge discovery: Key discovery**
  - Automatic reasoning and planning

- **Validation: knowledge graphs may contain errors**

  - **Link Validation**
  - Dealing with errors and ambiguities

OUTLINE

▪ **Introduction**   (F. Saïs)

▪ **Technical Part**

1.   Data Linking   (N. Pernelle)
2.   Key Discovery   (D. Symeonidou)
3.   Link Validation   (F. Saïs)

▪ **Future Challenges** (F. Saïs)