When it comes to Querying Semantic Cultural Heritage Data

Béatrice MARKHOFF[1], Thanh Binh NGUYEN[2], Cheikh NIANG[3]

[1] *University François Rabelais Tours, LI, Tours, France*
[2] *University of Orléans, LIFO, Orléans, France*
[3] *AUF, Paris, France*

SW4CH@ADBIS'17

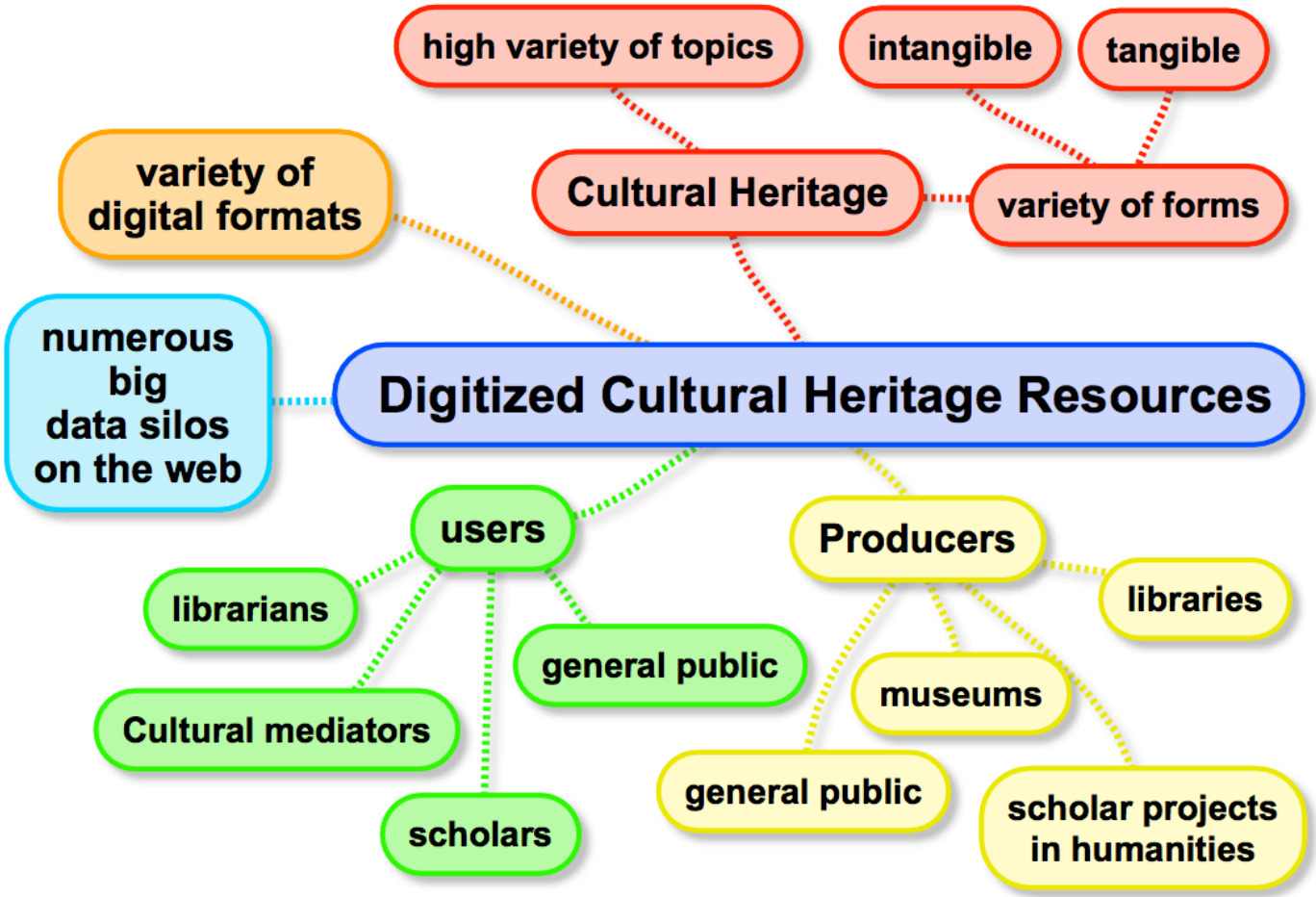
UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Outline

- Cultural Heritage Data and the Semantic Web
- Querying the Semantic Web
- Ontology-based Data Integration Systems
- Conclusion

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Digitized Cultural Heritage Ecosystem



UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

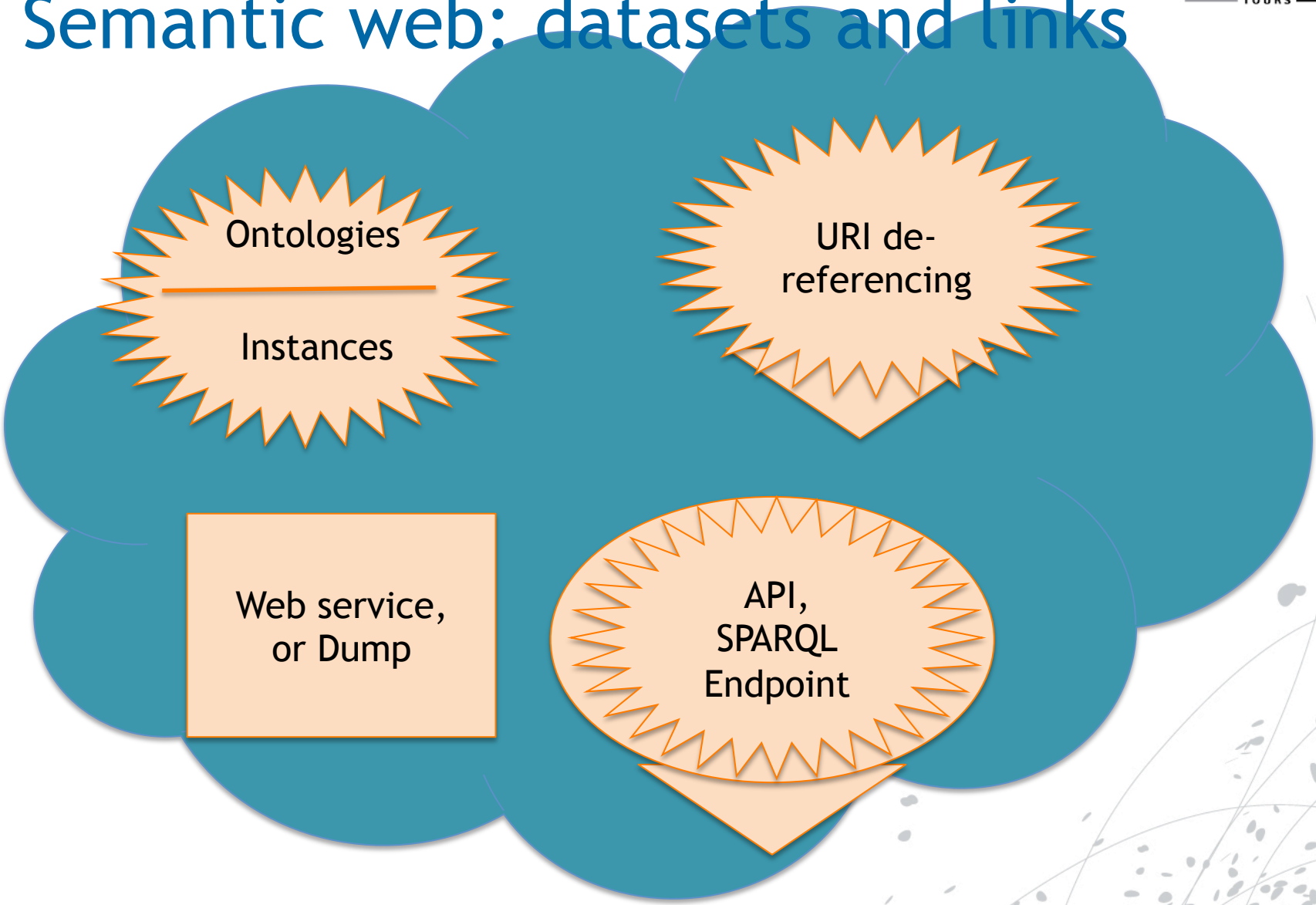
Cultural Heritage Representation

- High abstract level common points
  - Times (when…)
  - Places (where…)
  - Events, objects (what…)
  - People (who…)
- Search catalogue: for applications to find where it is stored
  - Vocabularies, terminologies, taxonomies, thesaurus (SKOS)
  - Support **human-human** and **human-machine** communication
- Representing: for applications to connect and reason with
  - Ontologies* (RDFS, OWL, rules…), s.t. **CIDOC CRM**
  - Bonus of **machine-machine** communication

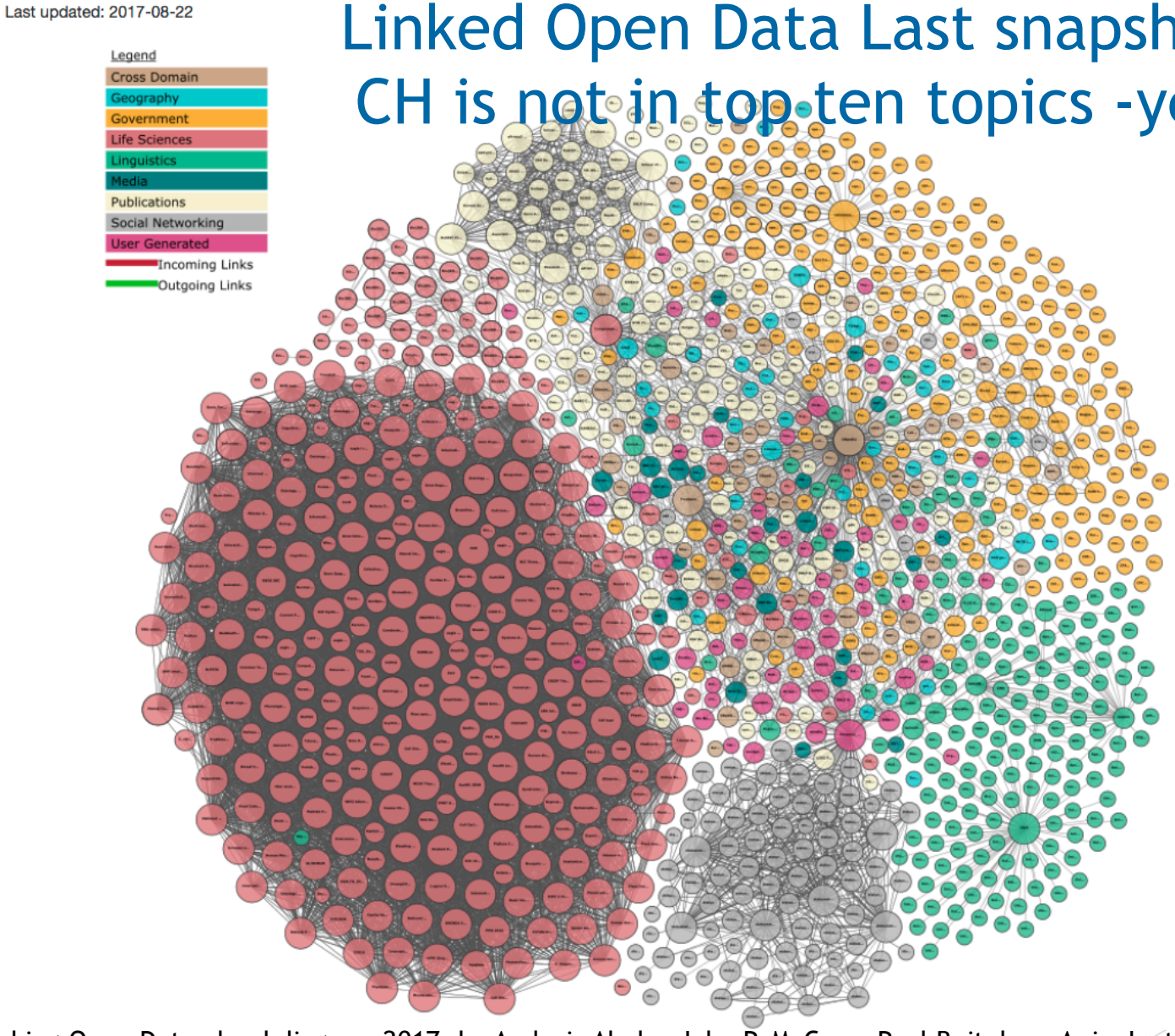
* Formal representations of a shared consensus about the concepts and their relationships that are necessary for describing a domain

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Semantic web: datasets and links



Linked Open Data Last snapshot: CH is not in top-ten topics -yet-



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

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

CH projects and the semantic web

- Building metadata or ontologies [CIDOC CRM] for representing cultural heritage data
- Web portals
  - Europeana: www.europeana.eu/
  - CultureSampo: http://www.kulttuurisampo.fi/?lang=en
  - Ariadne: www.ariadne-infrastructure.eu
  - …
- Migrating via ETL (extract-transform-load) processes
- Curating, mapping, connecting, enriching cultural heritage data

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Ways for Querying

- SPARQL
  - One knowledge graph
  - Several graphs, even in distant datasets - provided that you know them -
- LOD Querying
  - Link traversal, or « full web » querying systems
  - Fixed set of datasets: Federated Query systems
  - Through Big Centralized Knowledge Bases (Dbpedia, Yago, BabelNet, …)

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Querying one triplestore with ontology: entaiment regimes / OBDA query rewriting

- OBDA query rewriting (Ontology-Based Data Access)
- Data

  Immovable_Cultural_Object(œuvre_788), has_appellation(œuvre_788, « Adoration of Shepherds »), Diagnosis(act_157), Analysis(ana_11), required(act_157, ana_11), performed(act_157, œuvre_788)
- Query

  Arts on which some scientific study has been performed?
  q(n) ← has_appellation(x, n), Cultural_Object(x), performed(y,x), required(y,z), Scientific_Study(z)
- Ontology

  subClassOf(Analysis, Scientific_Study)
  range(performed, Cultural_Object)

  answer : Adoration of Shepherds

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Querying several semantic datasets

- Link traversal or « full-web » [O. Hartig, 2014]
  - software clients may retrieve more data by looking up the URI on the Web
  - challenges:
    - data source selection, data source ranking => Indexes, knowledge storing
    - integration of retrieved data and result construction
  - public implemented system: http://squin.sourceforge.net/index.shtml
- Federated Query systems [survey in M. Saleem, 2015]
  - unique interface for querying data from a fixed set of independent data sources
  - challenges: … data source selection, query optimization…
  - Several existing prototypes (cited in the paper)

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Querying several semantic datasets

- Link traversal or « full-web » [O. Hartig, 2014]
  - software clients may retrieve more data by looking up the URI on the Web

(http://chroniclingamerica.loc.gov/lccn/sn85038614#title, dct:title, "Richmond dispatch.")
(http://chroniclingamerica.loc.gov/lccn/sn85038614#title, rdf:type, http://purl.org/ontology/bibo/Newspaper)
(http://chroniclingamerica.loc.gov/lccn/sn85038614#title, frbr:successorOf, http://chroniclingamerica.loc.gov/lccn/sn84024738#title)
(http://chroniclingamerica.loc.gov/lccn/sn85038614#title, dct:coverage, http://sws.geonames.org/4781708/)

(http://sws.geonames.org/4781708/, gn:name, "Richmond")
(http://sws.geonames.org/4781708/, gn:parentCountry, http://sws.geonames.org/6252001/)
(http://sws.geonames.org/4781708/, geo:lat, "37.55376")
(http://sws.geonames.org/4781708/, geo:long, "-77.46026")

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Querying several semantic datasets

- Link traversal or « full-web » [O. Hartig, 2014]
  - software clients may retrieve more data by looking up the URI on the Web
  - challenges:
    - data source selection, data source ranking
      => Indexes, knowledge storing
    - integration of retrieved data and result construction
  - public implemented system: http://squin.sourceforge.net/index.shtml
- Federated Query systems [survey in M. Saleem, 2015]
  - unique interface for querying data from a fixed set of independent data sources
  - challenges: … data source selection, query optimization…
    => Indexes, knowledge storing
  - Several existing prototypes (cited in the paper)

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

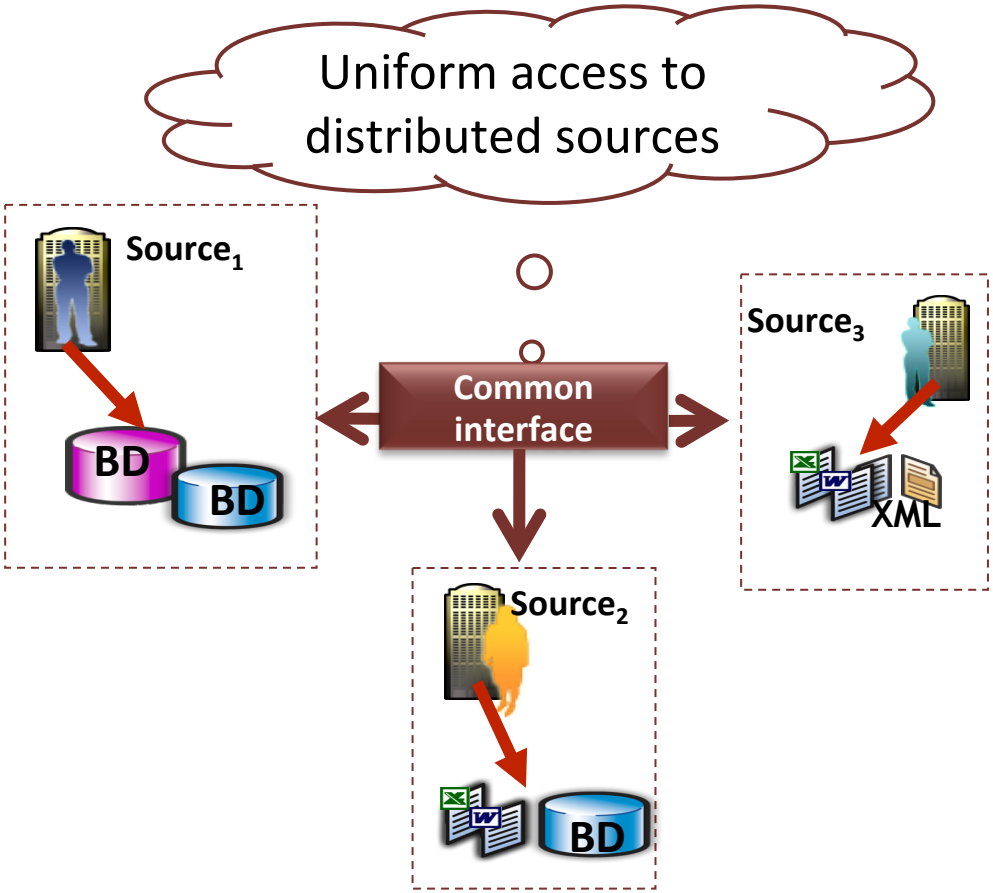
A network of semantic data sets: Just querying it? Integrating!

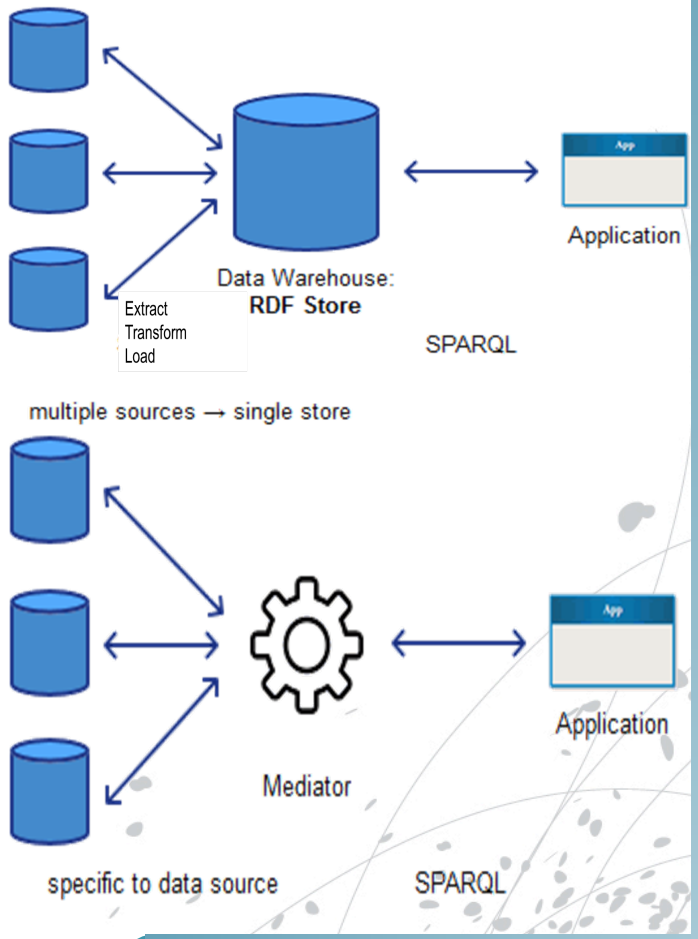
- Previously presented systems: no matter
  - what is queried
  - who queries
  - for which purposes

- Cultural Heritage projects:
  - Some determined datasets
  - Some specific users
  - Some explicit aims

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

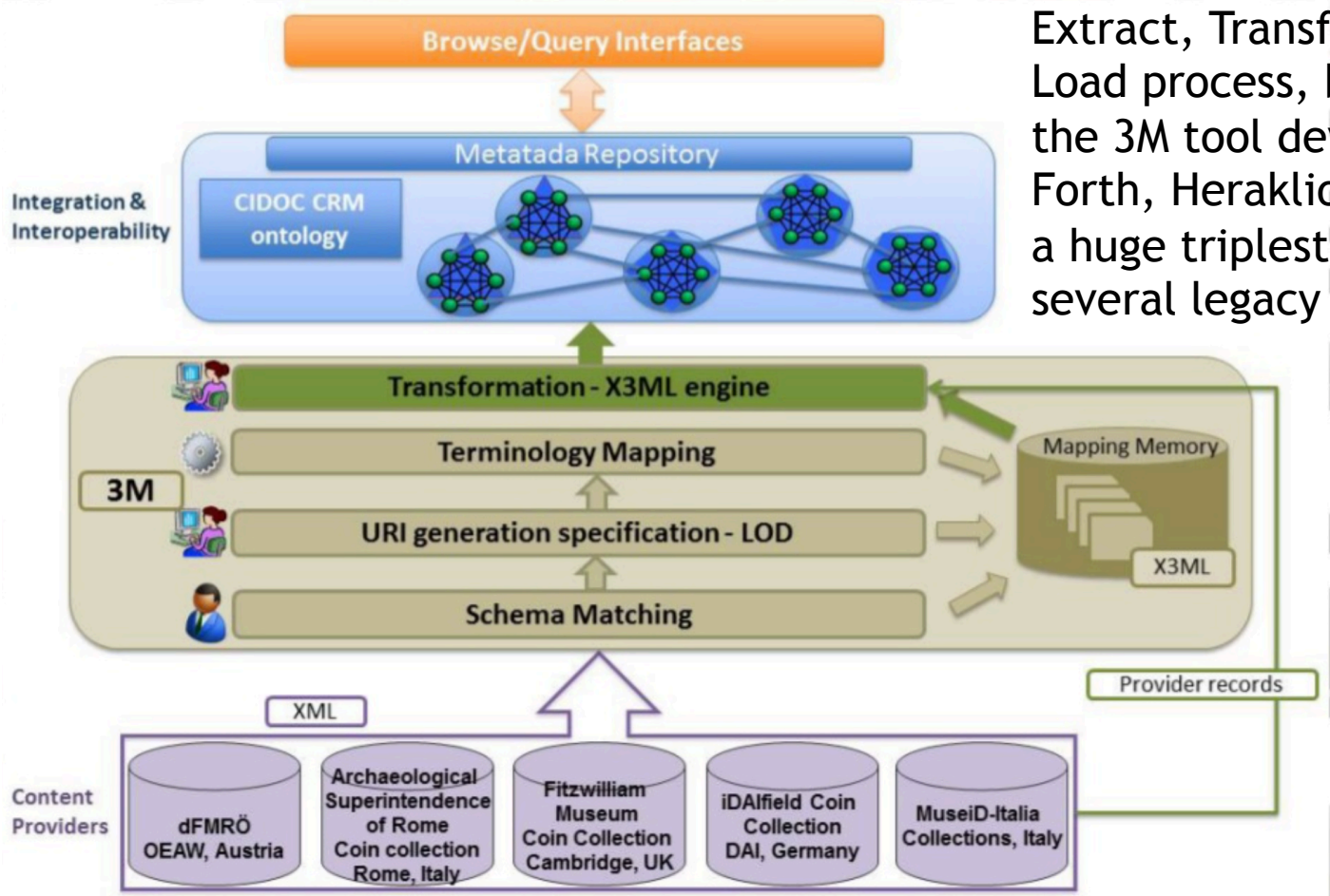
Ontology-based Data integration: ontology as « global schema »





UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Coin datasets [A. Felicetti et al., 2015]



Extract, Transform and Load process, based on the 3M tool developped at Forth, Heraklion, to build a huge triplestore from several legacy datasets

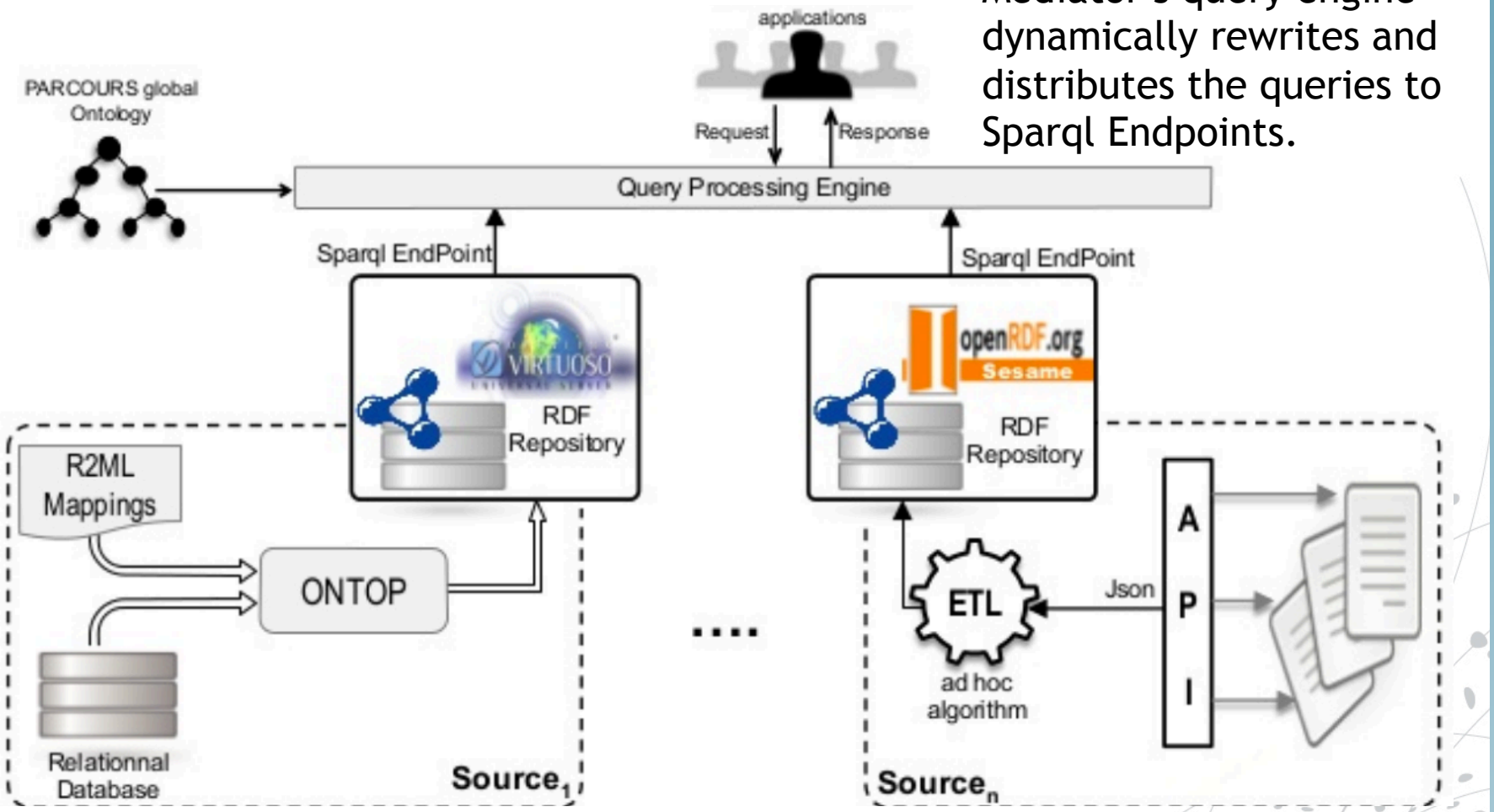
https://www.slideshare.net/ariadnenetwork/maria-theodoridou-semantic-integration-experiments

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Parcours: conservation-restoration

[C. Niang et al., 2017]

Mediator's query engine dynamically rewrites and distributes the queries to Sparql Endpoints.



UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

EP-Net [D. Calvanese et al., 2016]

Global ontology with ontop mappings to source datasets => answers from all datasets

- EP-Net Project: « Production and distribution of food during the Roman Empire: Economics and Political Dynamics »
- CIDOC CRM based Ontology plus OBDA mappings for **dynamically querying**:
  - EP-Net relational database

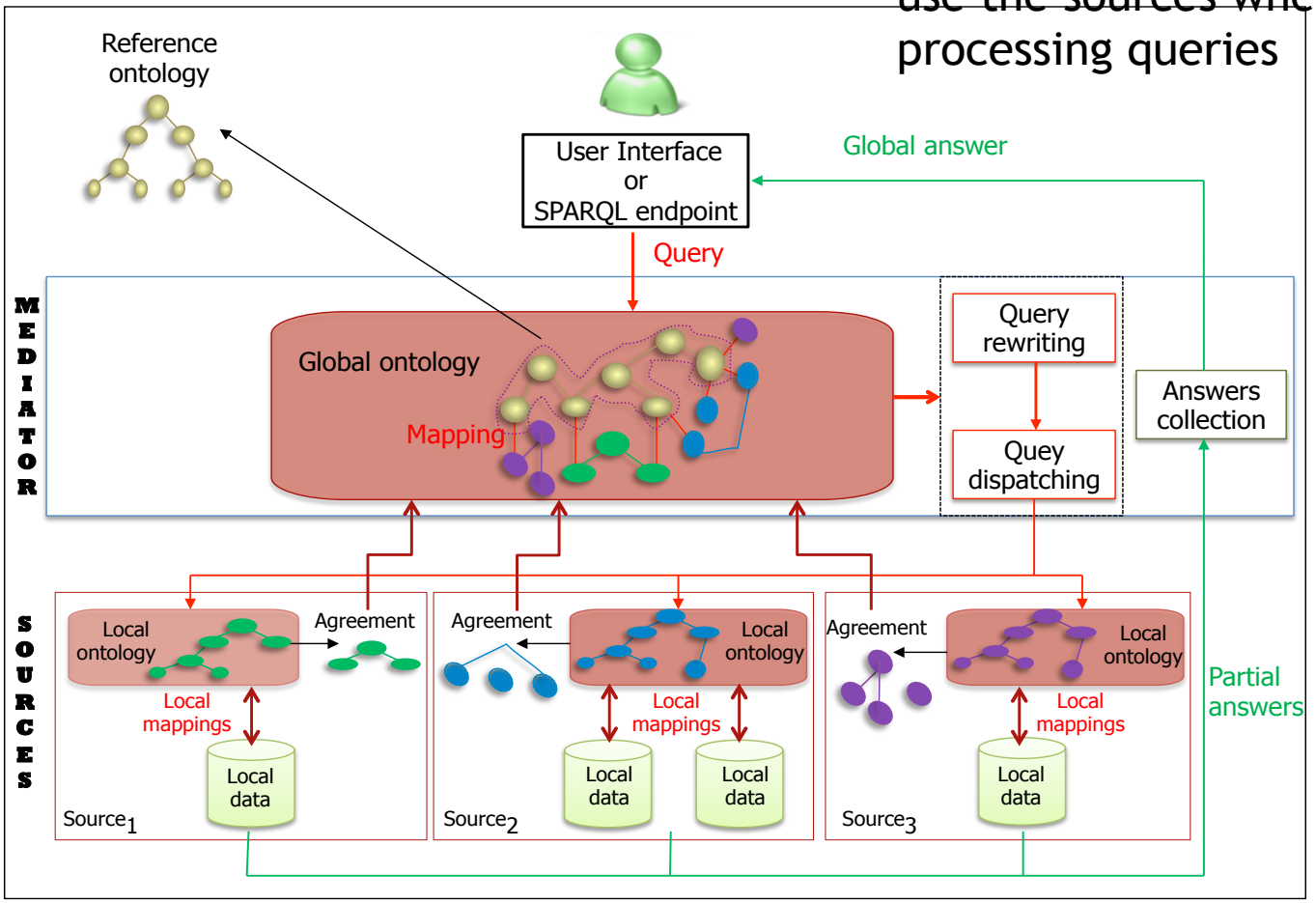
  - Epigraphic Database Heidelberg (EDH)

  - Pleiades dataset

Automatic Building of OBDI Systems

[B. Bouchou et al. 2014]

Reference Ontology provides the glue to assemble local concepts. Mappings = knowledge to use the sources when processing queries



UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Current uses in CH projects

Linked Open Data

Querying framework	CH
Quering one semantic dataset with SPARQL (may materialize several ones)	☑
Full web querying	☐
Federated Query systems	☐
Harvested centralized Knowledge Base (Yago, BabelNet…)	☐
Ontology-based mediation systems	☑

- ✓ **British Museum**
- ✓ **Coin datasets**
- ✓ …

- ✓ **PARCOURS**
- ✓ **EP-NET** [Calvanese et al., 2016]
- ✓ …

UNIVERSITÉ FRANÇOIS - RABELAIS TOURS

Conclusions

- Cultural Heritage stakeholders are opening their data to the semantic web
- OBDI systems are crucial for connecting such data
- Centralized materialized RDF data warehouses
  - efficient for operating complex computations on big data
- **Semantic mediators**
  - fit the **decentralized and highly evolutive** features of the web
  - rely on **mappings (= how to query the sources)**
- Other LOD querying systems should inspire automatic ways of building semantic mediators for Cultural Heritage data