



Vers les données liées : conséquences théoriques et pratiques pour les sciences humaines

*Towards Linked Data:
Theoretical and Practical Consequences
for Humanities Research
(slides are in English!)*

Lyne Da Sylva, professeure agrégée
École de bibliothéconomie et des sciences de l'information
Université de Montréal

Digital Humanities, Montréal

11 août 2017

Web of documents vs Semantic Web



or Web of data

Current Web (« Document Web »)	Semantic Web
Set of documents	Set of knowledge
Based essentially on HTML	Based on XML and RDF(S)
Search by keywords	Search by concept
For human consumption	For machine consumption

Source :

- Gagnon, Michel. 2004. *Introduction au web sémantique. Tutoriel*, École Polytechnique de Montréal.

<http://www.professeurs.polymtl.ca/michel.gagnon/Publications/tutorielSWIG04.pdf>

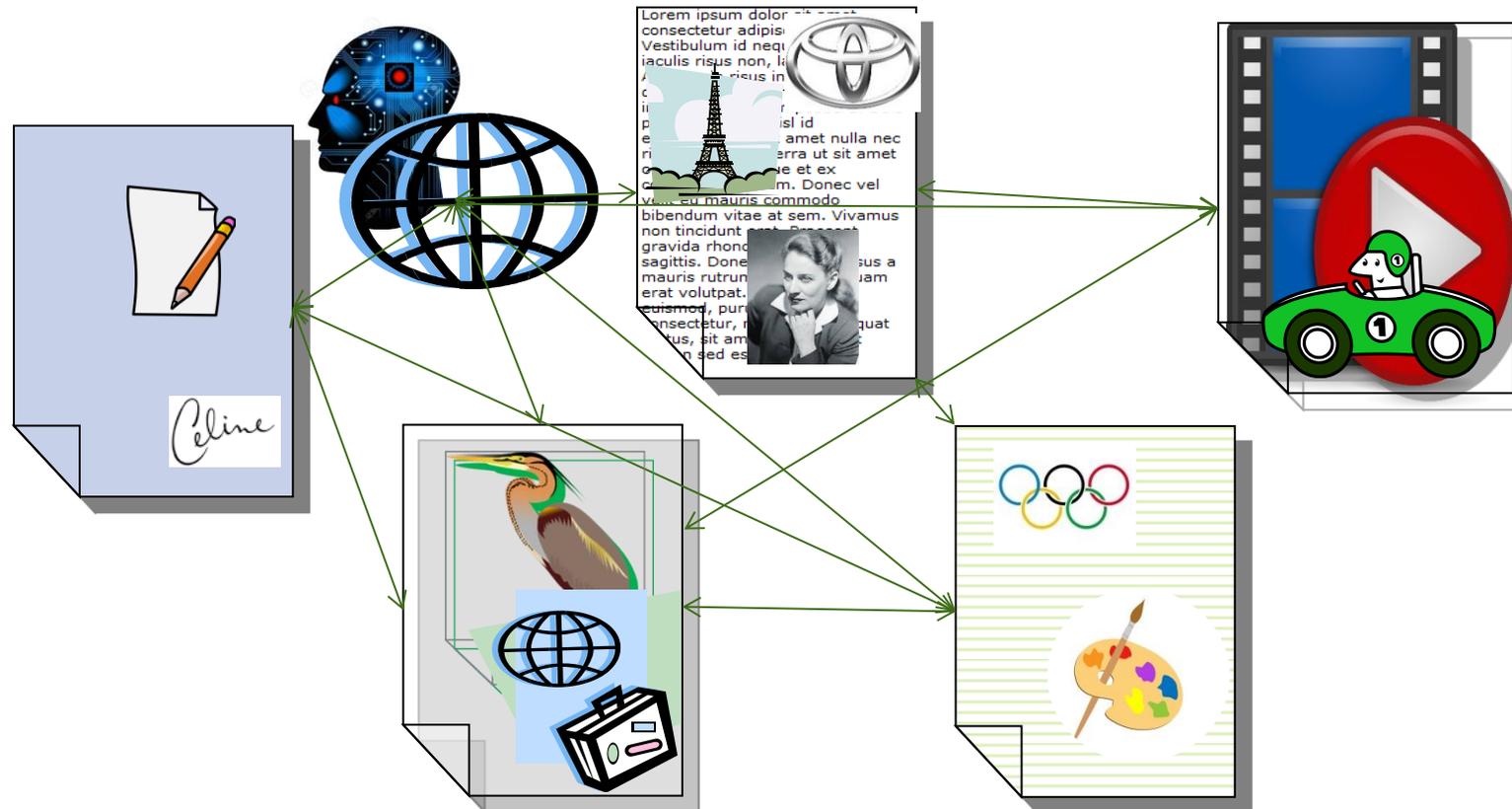


Objectives of this talk

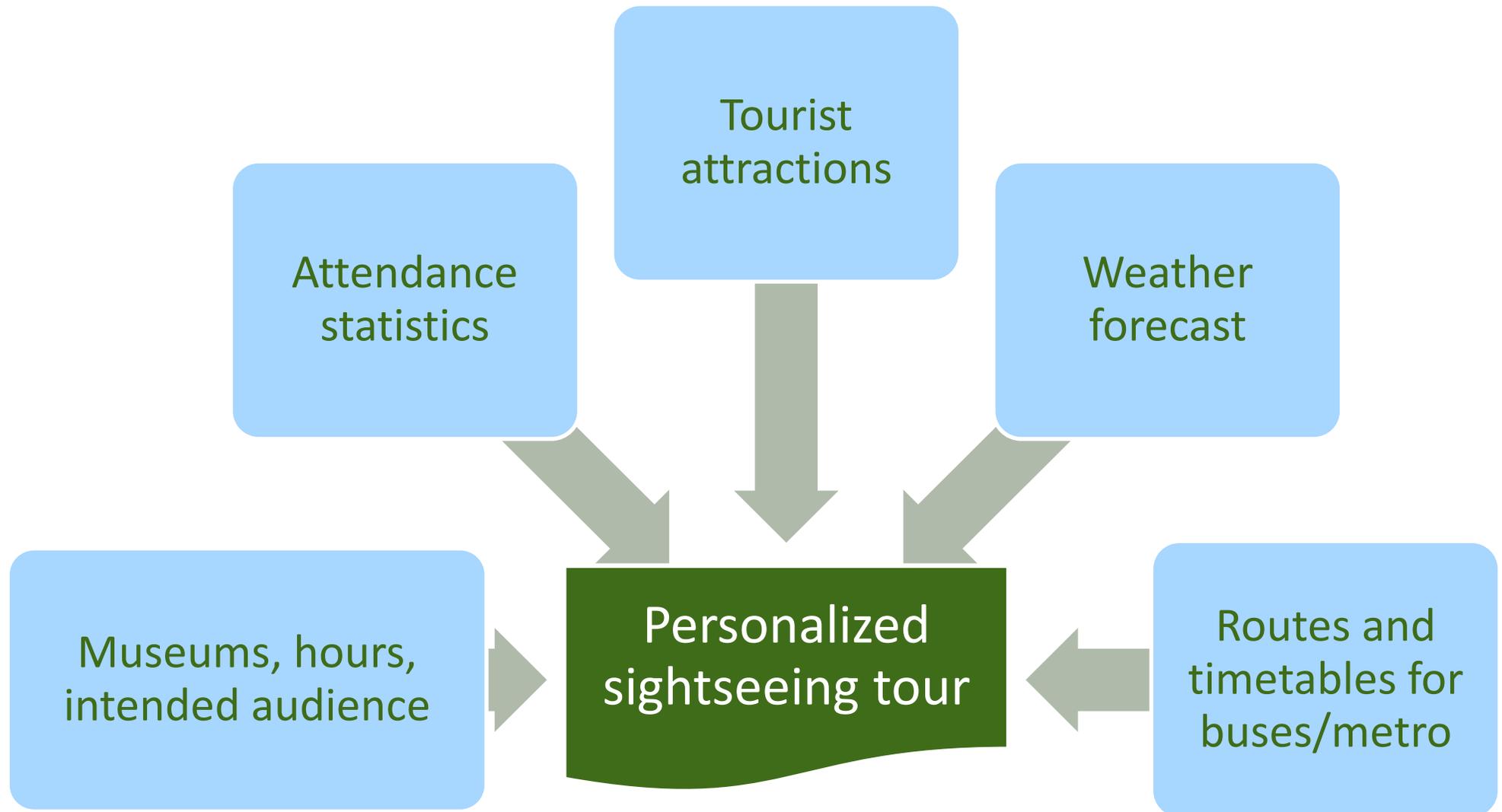
- Summarize basic notions of the Semantic Web (specifically Linked Data)
- Discuss certain implications of the Semantic Web (specifically Linked Data) for researchers in the Humanities and Social Sciences (HSS)

Goals of the Semantic Web

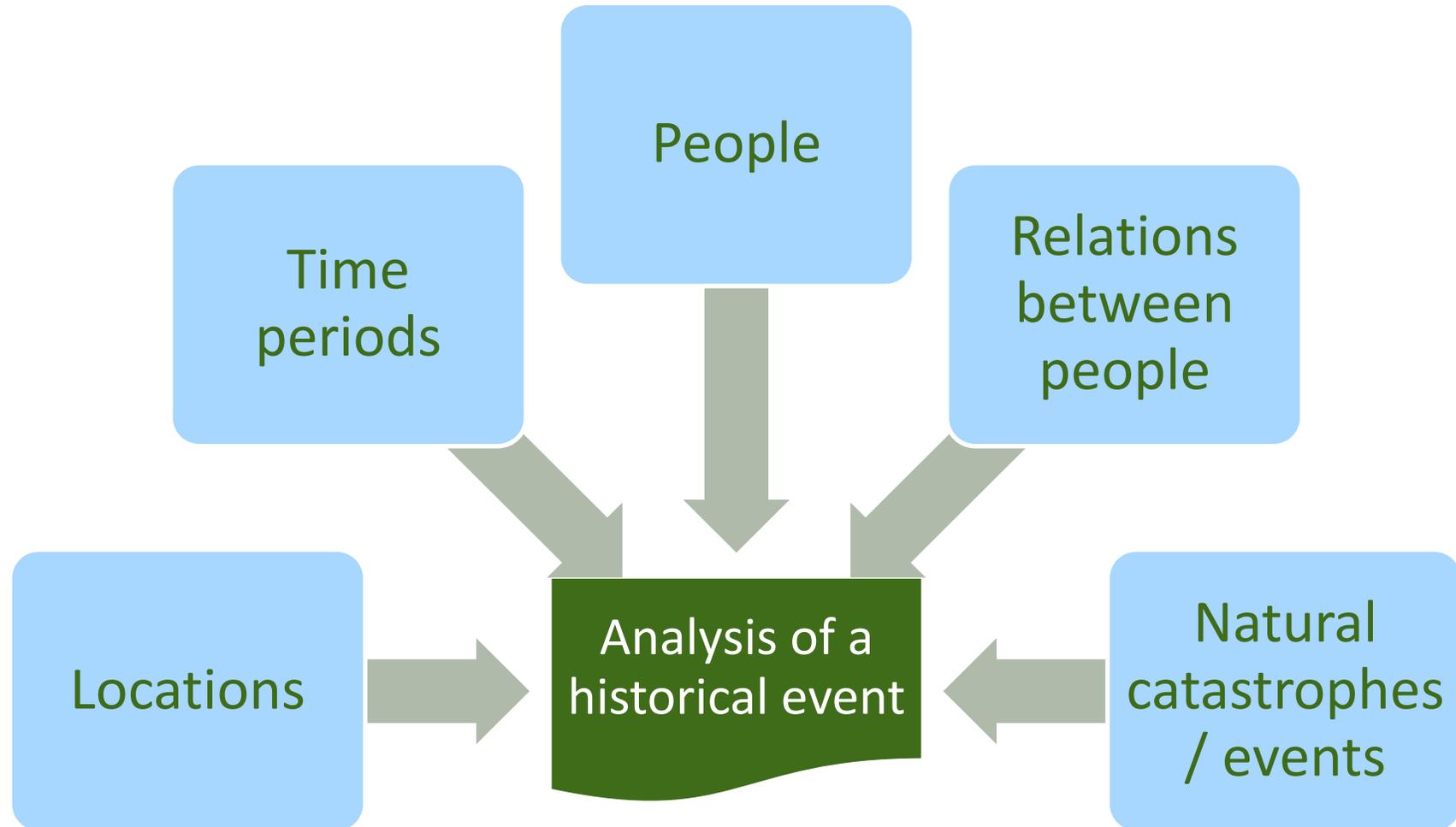
- To make explicit the meaning of information on the Web
 - to allow complex automated processing of the data



Complex automated processing



Complex automated processing (cont'd)





Outline

- Brief description of Linked Data technology
- Possibilities of Linked Data for HSS research
- Implications for research in HSS
 - Practical
 - Theoretical
- Some additional thoughts



BRIEF DESCRIPTION OF LINKED DATA



Basic notions of Linked Data

1. Identifiers
2. Relations
3. RDF triples
4. Vocabularies and ontologies
5. Triplestores

1. Identifiers

- To identify uniquely (and explicitly)

Entity

Identifier

Guerre et paix

<http://catalogue.bnf.fr/ark:/12148/cb31478402c>

Léon Tolstoï

<urn:ISNI:0000-0001-2242-4494>

Lyne Da Sylva

<http://dasylva.ebsi.umontreal.ca>

Lyne Da Sylva

<http://orcid.org/0000-0003-2530-1048>

Michèle Hudon

<http://catalogue.bnf.fr/ark:/12148/cb12534437b>

- Usually associated with names (labels that may vary with language)

War and Peace,
Война́ и мир

Leo Tolstoy,
Толстой, Лев
Николаевич,...

From many
sources –
standardized
or not

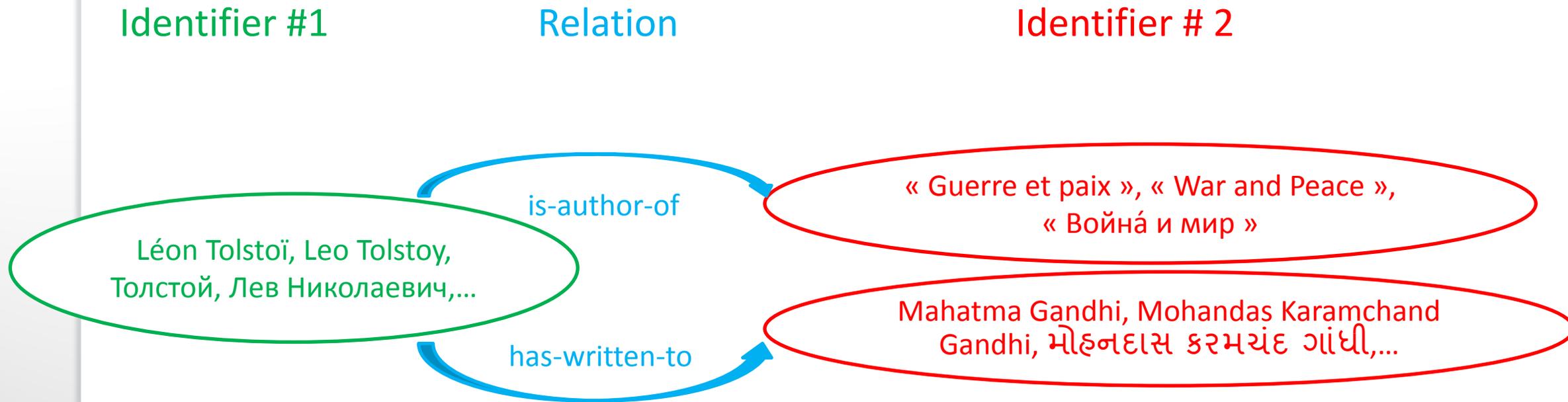
2. Relations

- To allow automatic processing
 - explicitly encode relations as well as entities

Entity 1	Relation	Entity 2
Leo Tolstoy	is-author-of	War and Peace
Leo Tolstoy	has-written-to	Mahatma Gandhi
http://dasylva.ebsi.umontreal.ca	firstName	Lyne
http://dasylva.ebsi.umontreal.ca	familyName	Da Sylva
http://dasylva.ebsi.umontreal.ca	knows	Michèle Hudon

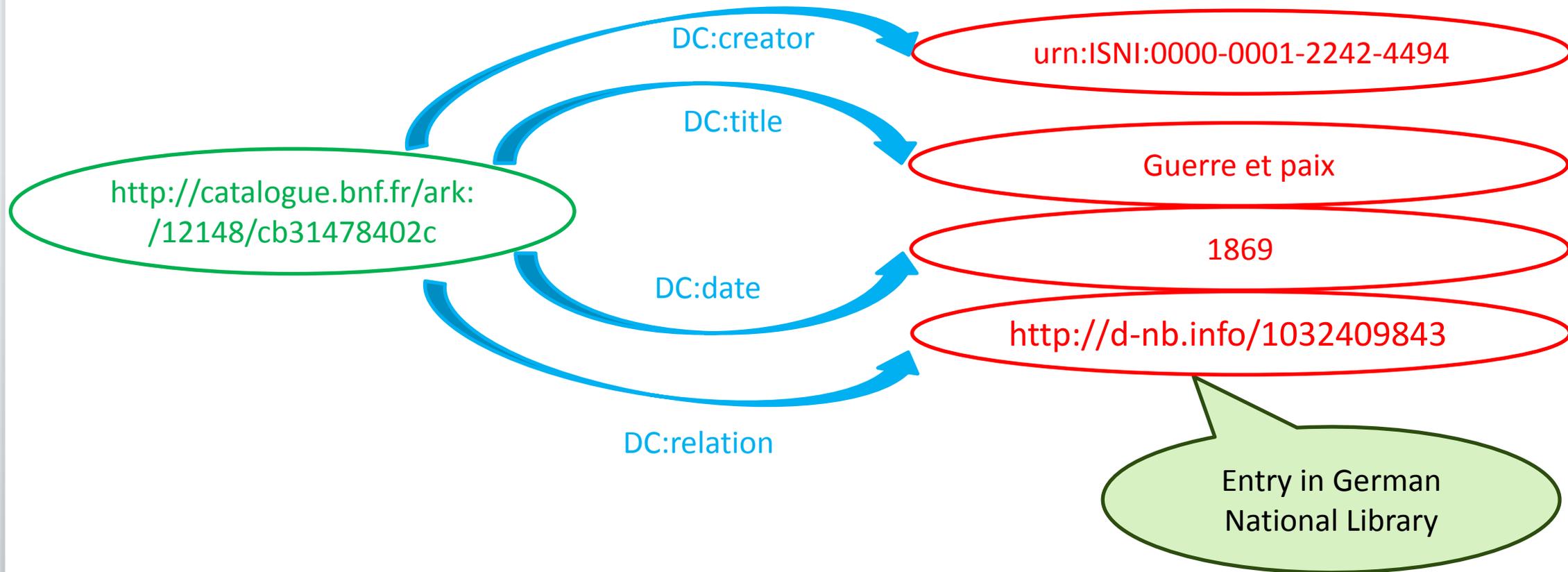
3. RDF Triples

- Uniform encoding: simple binary **relation** between **subject** and **object**



3. RDF triples (cont'd)

- With identifiers and Dublin Core (DC) vocabulary



3. RDF triples (cont'd)

- In XML format...

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dcterms="http://purl.org/dc/terms/">

  <rdf:Description rdf:about="http://catalogue.bnf.fr/ark:/12148/cb31478402c">
    <dcterms:title>Guerre et paix</dcterms:title>
    <dcterms:created>1869</dcterms:created>
    <dcterms:creator rdf:resource="urn:ISNI:0000-0001-2242-4494"/>
    <dcterms:relation rdf:resource="http://d-nb.info/1032409843"/>
  </rdf:Description>

</rdf:RDF>
```

4. Vocabularies and ontologies

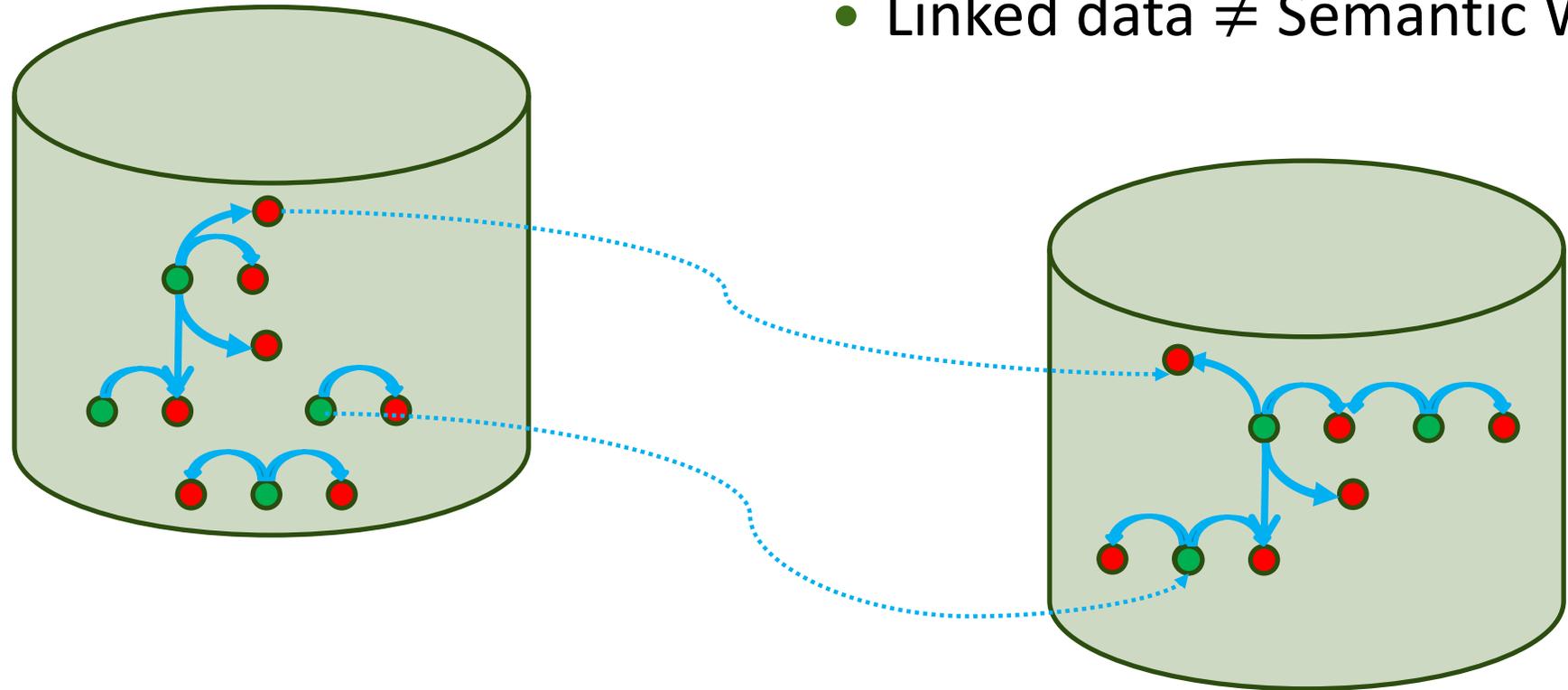
- For uniform encoding of relations
- A small sample:

Vocabulary	Main use	Sample relations
Dublin Core	Documents	Creator Date Rights
FOAF	People and relationships	firstName familyName knows
SKOS	Thesauri	broader narrower prefLabel
OWL	Ontologies	sameAs

5. Triplestores

- To store triples

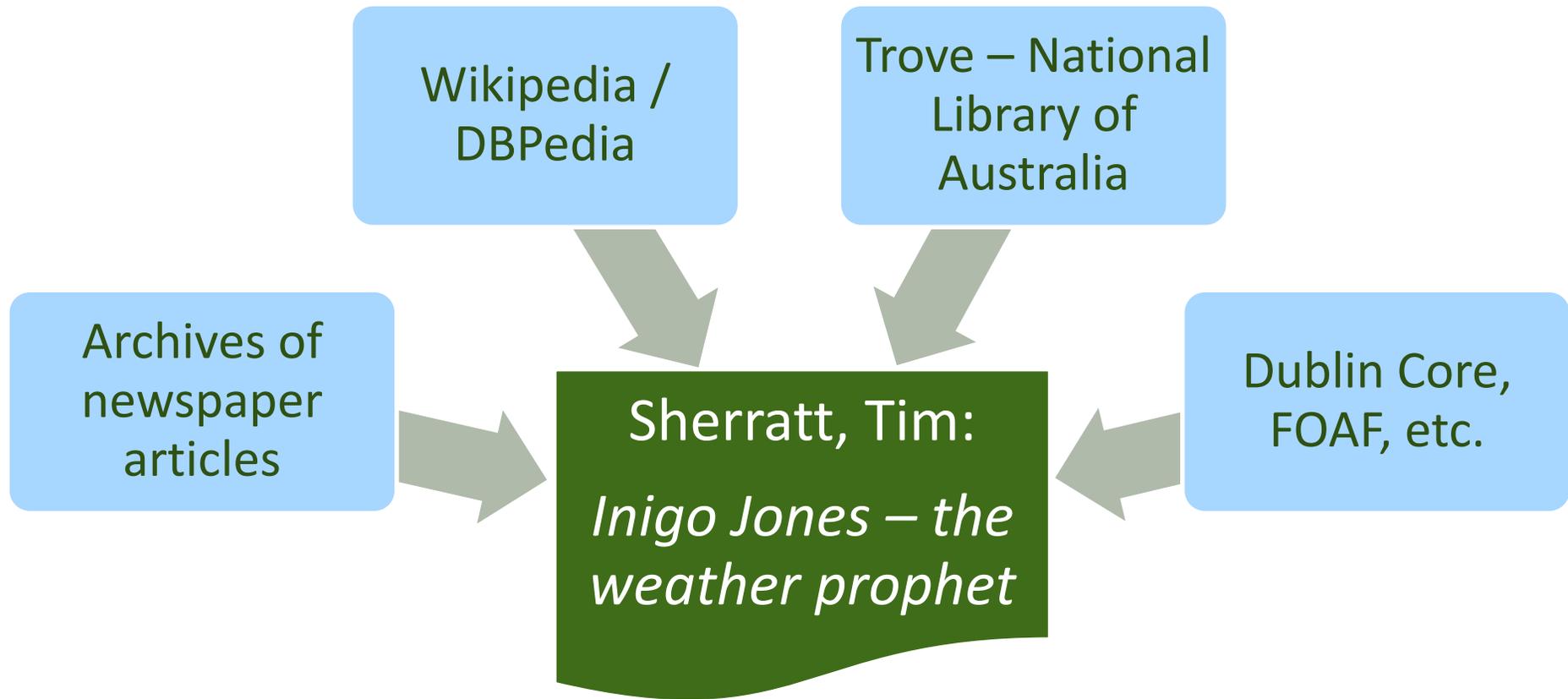
- Hopefully: open data to ensure accessibility / access
- Linked data \neq Semantic Web





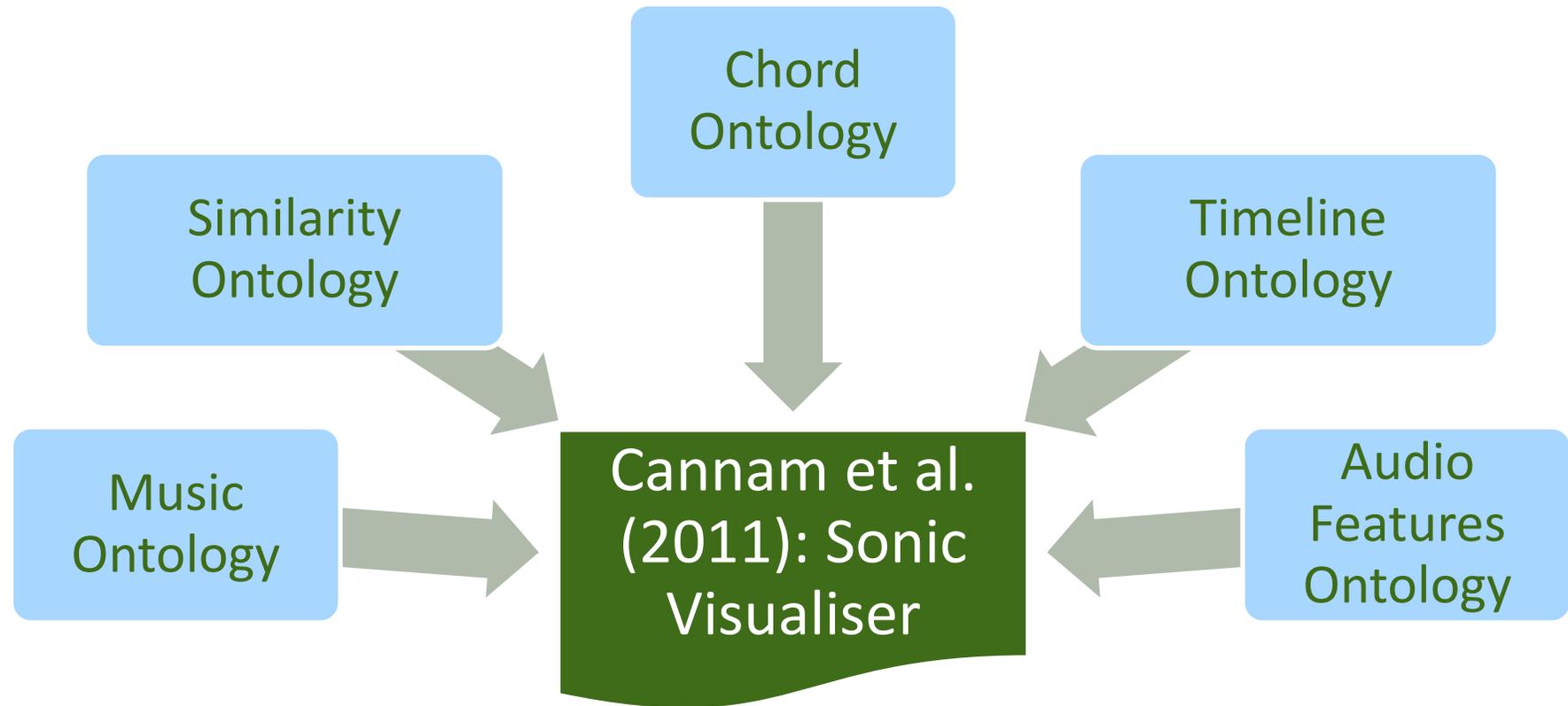
POSSIBILITIES OF THE SEMANTIC WEB FOR HSS RESEARCH

Examples of implementations



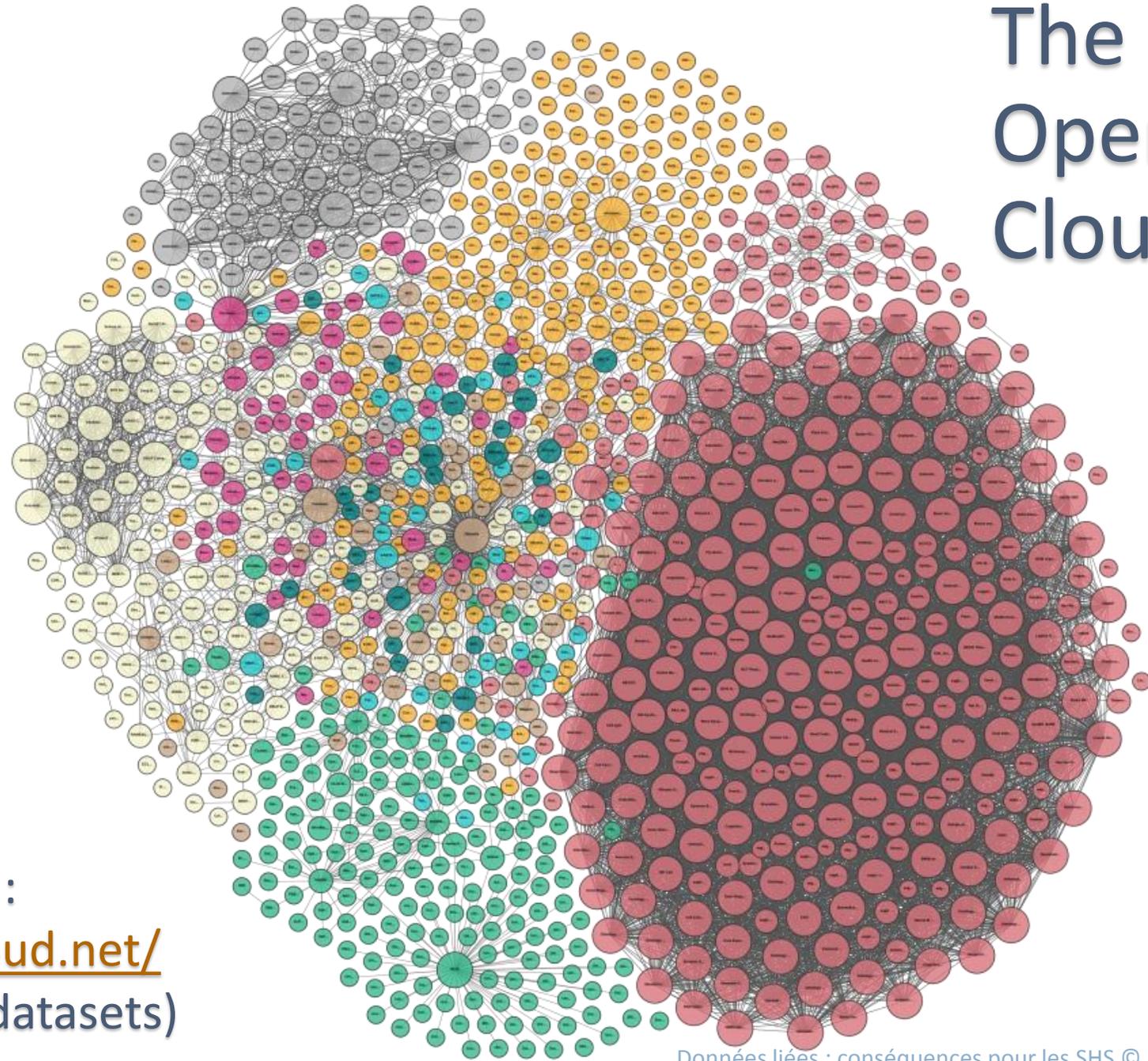
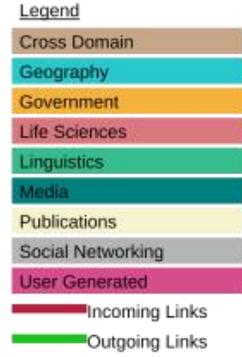
- Document, augmented with data: <http://lodbookdev.herokuapp.com/#!/text/1/>

Examples of implementations



- OMRAS2 Music Information Retrieval project
- music informatics research
- Sonic Visualiser: application to visualize, annotate, and automate analysis of audio recordings

The Linked Open Data Cloud



State in 2017 :
<http://lod-cloud.net/>
(about 1139 datasets)



**PRACTICAL IMPLICATIONS FOR
RESEARCH IN HUMANITIES AND SOCIAL
SCIENCES (HSS)**

Practical implications

Focus of
research

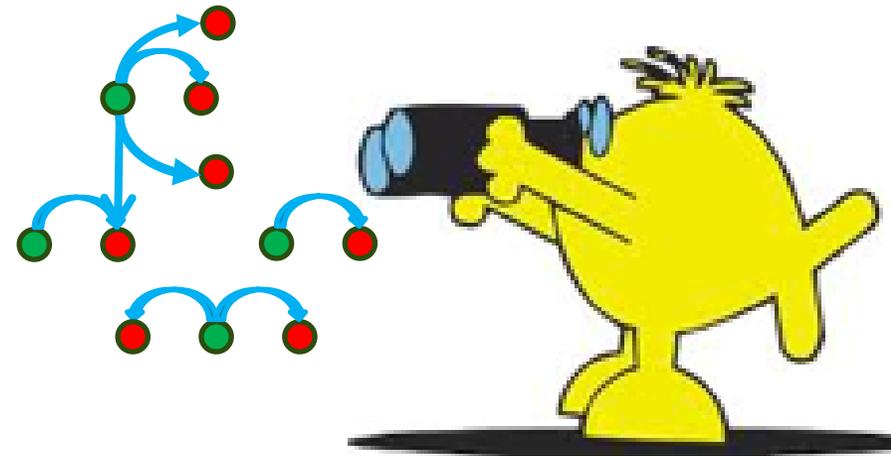
« Redocumen-
tarisation »
- « redocumenting »

Effort needed
to extract
information

Technical
implications

Focus of
research

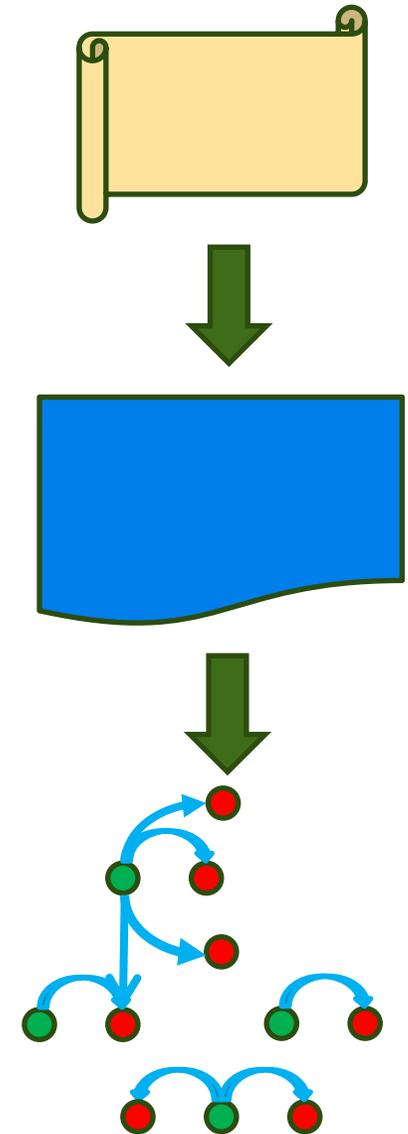
- Piecemeal information – not documents nor overview



« Redocumentarisation »

- « redocumenting »

- see Pédaque (2007):
i.e. digitizing existing documents
- now: transcoding of
extracted information
- see also atomisation,
below



Effort needed to extract information

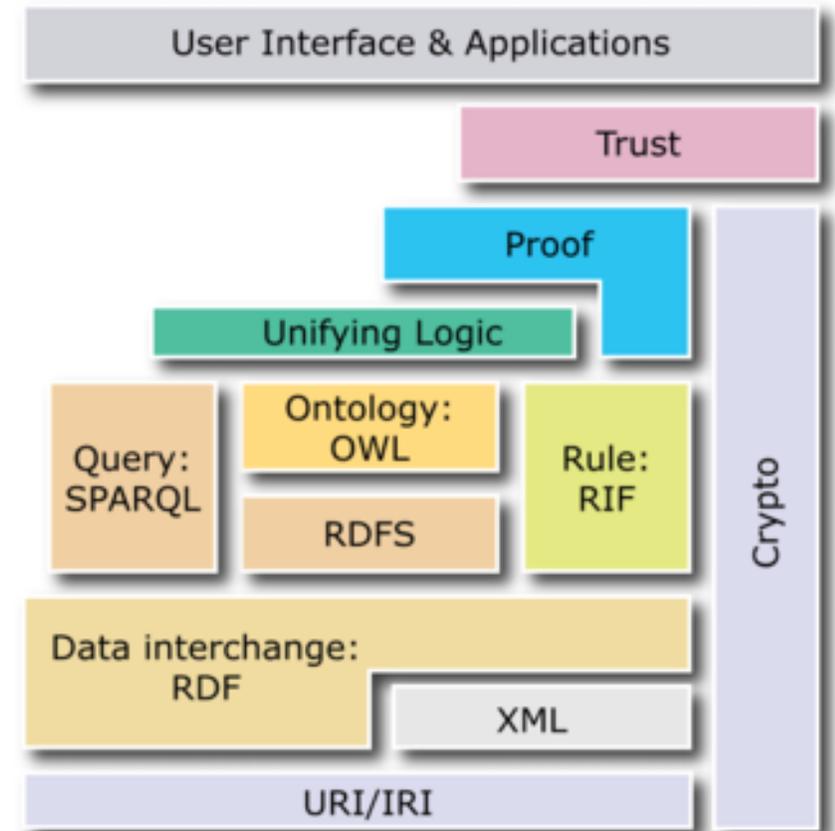
Compare

- HSS – manual and intellectual (reading, interpretation, observations)
- pure sciences (STEM) – automatic (probes, captors, sensors, databases, computer generation ...)



Technical implications

- Skills needed
 - Formal and analytical approach
- Tools
 - Semantic Web technologies



« L'historien doit s'intéresser à la structure des données et non plus simplement à la livraison de celles-ci. » (Michon, 2016:17)



THEORETICAL IMPLICATIONS FOR RESEARCH IN HSS

Theoretical implications

Atomisation of
research objects

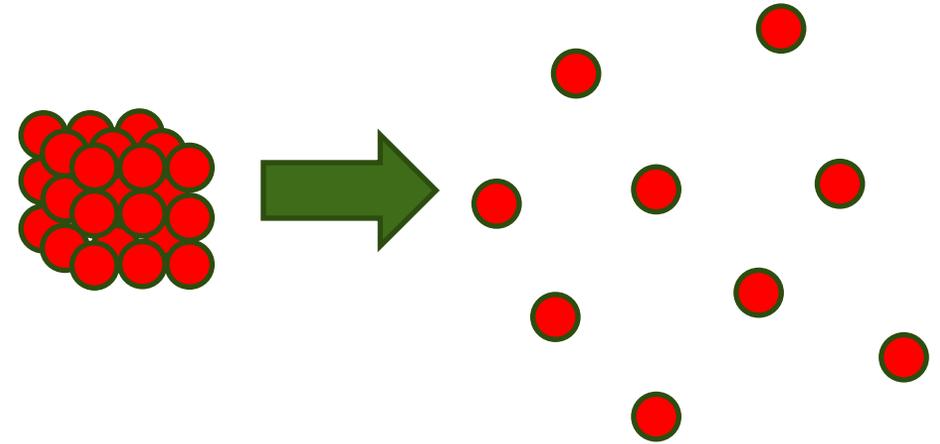
Reification of
research objects

Blurred distinction
between data and
metadata



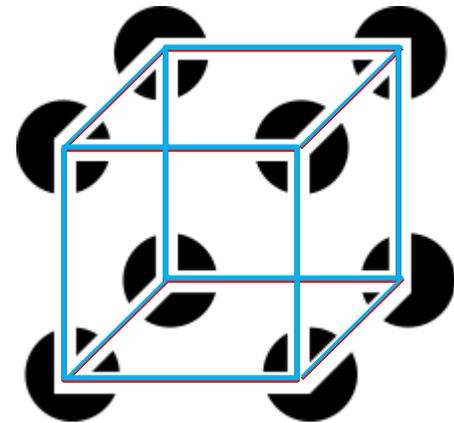
Atomisation of research objects

- Contrary to usual
synthetical
approach of
researchers



Reification of research objects

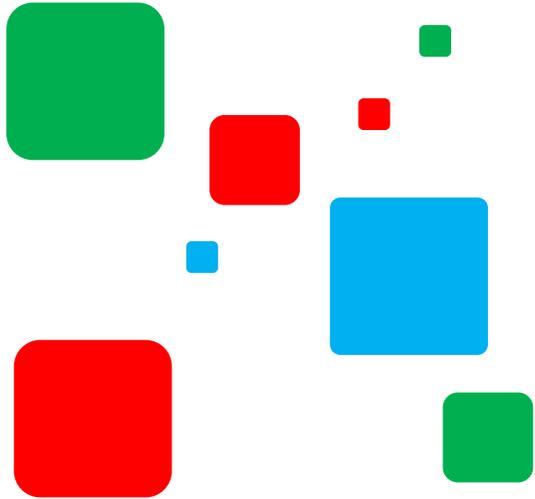
- Entity#1 and Entity#2
of RDF triples – OK
- but relations???



Blurred distinction between data and metadata



- Some triples are clearly metadata
- Others introduce true relationships
- In general, what data are the triples metadata for?





Some additional thoughts

- Implications vary for different disciplines
 - Data-oriented disciplines
 - economics, demographics, etc.
 - Metadata-oriented disciplines
 - library science, museum studies, etc.
 - Document-oriented disciplines
 - history, literature, etc.



And yet...

Let's think for a moment about the work of a historian — identifying actors, defining relationships, documenting the complex networks that bring together people, places and events over time. It's painstaking, exhilarating [sic] and potentially soul-destroying work. It's also an exercise in data modelling. Whether the results are preserved in a triplestore, a spreadsheet, or on a drawer full of index cards — it's nodes and edges, it's entities and relationships, it's data.

(Sherratt, 2015)



Conclusion

- Transformations on the conduct of research
 - inevitable
- Needed: new solutions for new challenges
 - *Le texte exprimant avec éloquence les faits historiques, comment peut-on en optimiser la réutilisation dans le cadre des données liées sans en morceler la narration? (Michon, 2016:26)*



THANK YOU!
MERCI!

Lyne Da Sylva

lyne.da.sylva@umontreal.ca

<http://dasylva.ebsi.umontreal.ca>

References

- Cannam, Chris; Sandler, Mark; Jewell, Michael O; Rhodes, Christophe; d'Inverno, Mark. 2011. Linked Data and You: Bringing Music Research Software into the Semantic Web. *Journal of New Music Research*, 39(4): Music Informatics and the OMRAS2 Project.
- Michon, Philippe. (2016). *Vers une nouvelle architecture de l'information historique : L'impact du Web sémantique sur l'organisation du Répertoire du patrimoine culturel du Québec*. Mémoire de maîtrise, Université de Sherbrooke.
http://savoirs.usherbrooke.ca/bitstream/handle/11143/8776/Michon_Philippe_MA_2016.pdf?sequence=8&isAllowed=y
- Pédauque, Roger T. (collectif), (2007), *La Redocumentarisation du Monde*, Paris : Éditions Cepadues.
- Sherratt, Tim. (2015). *Stories for machines, data for humans*, <http://discontents.com.au/data-for-humans/>, 10 avril 2015.