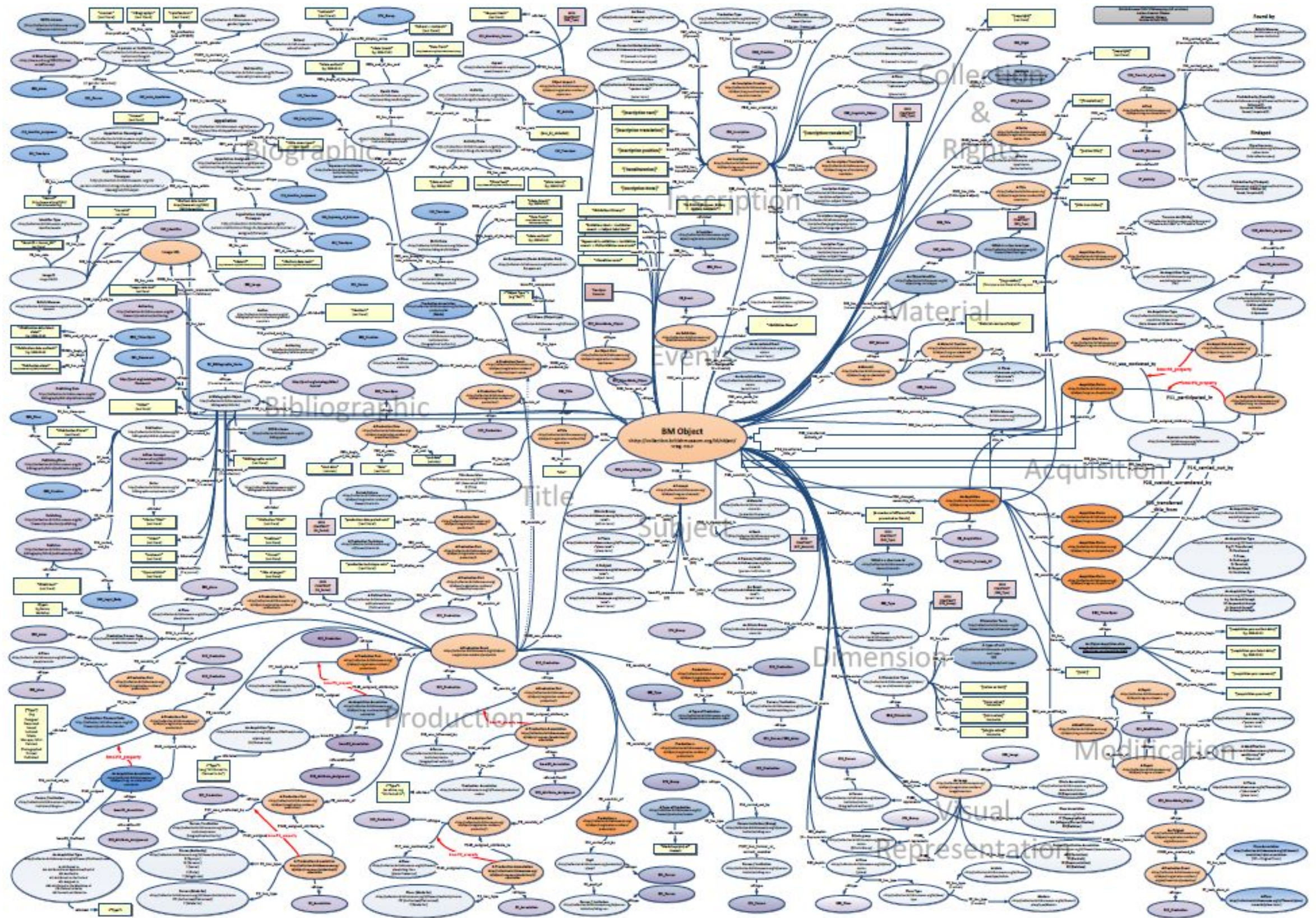




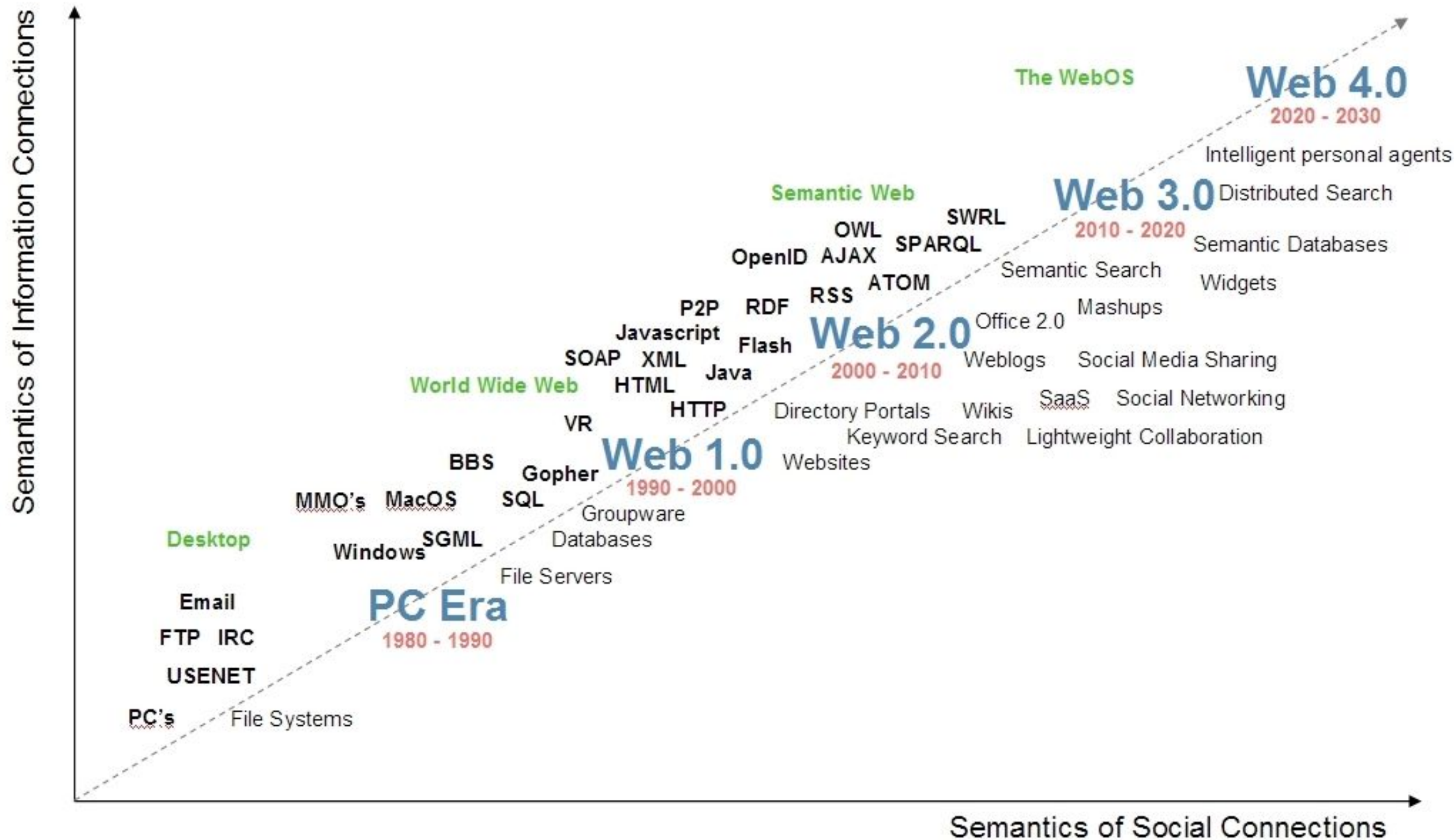
Antoine Courtin

Web.2.0 et structuration des données

avec Olivier Bonfait
à partir d'une présentation sur humanités
numériques et histoire de l'art
à l'Université de Bourgogne

Ce que font (ou non) les données à la pratique de l'histoire de l'art (2)

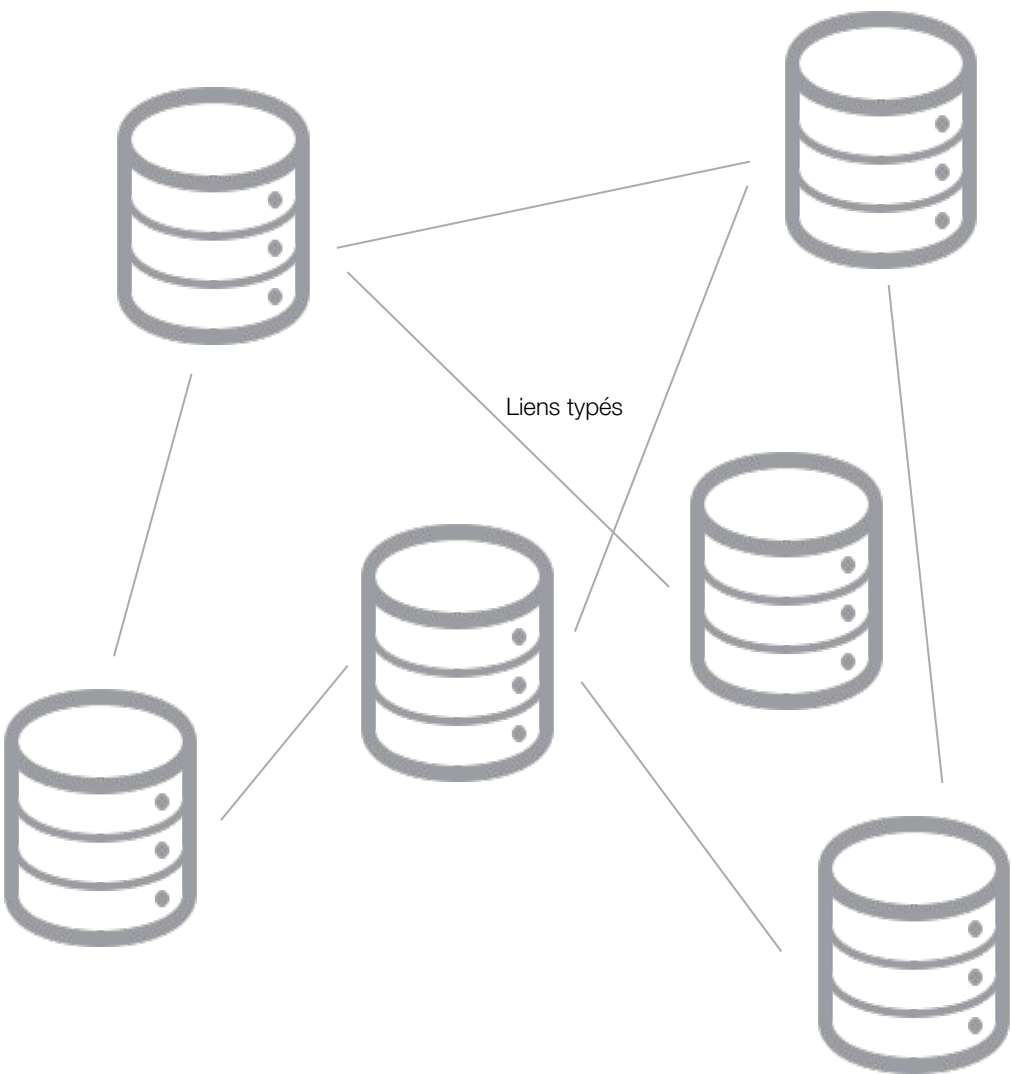
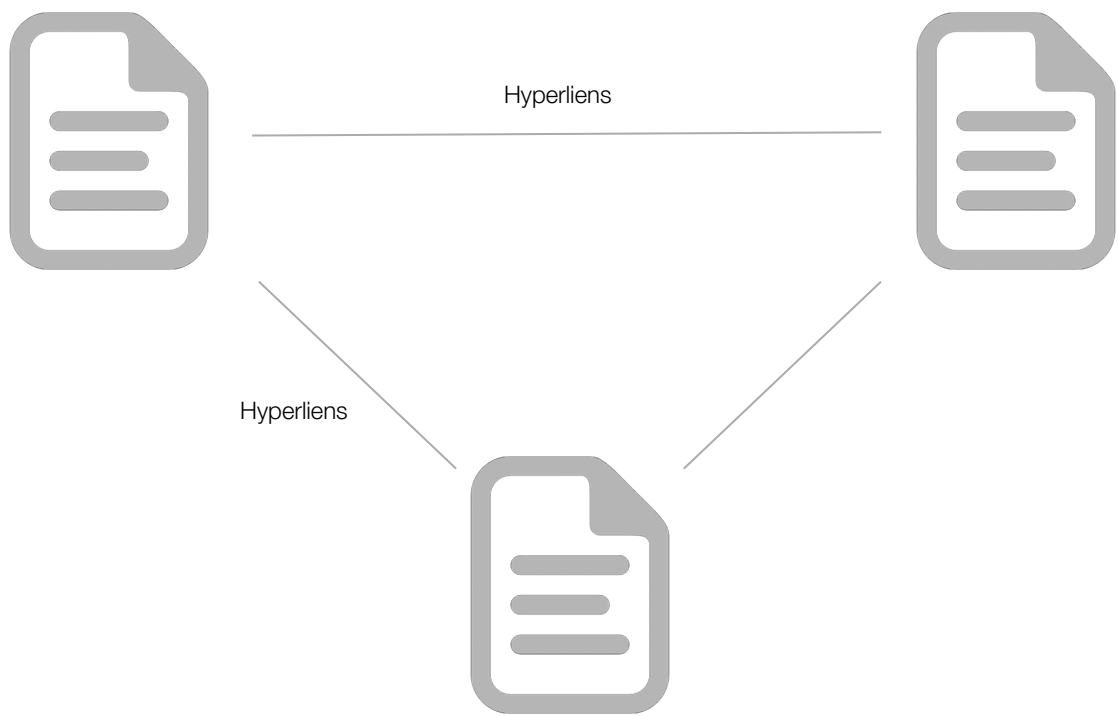
Le WWW a déjà fêté ses 25 ans



Un web de données liées (1) ←

Web de document

Web de données



Une transition pour passer d'un web :

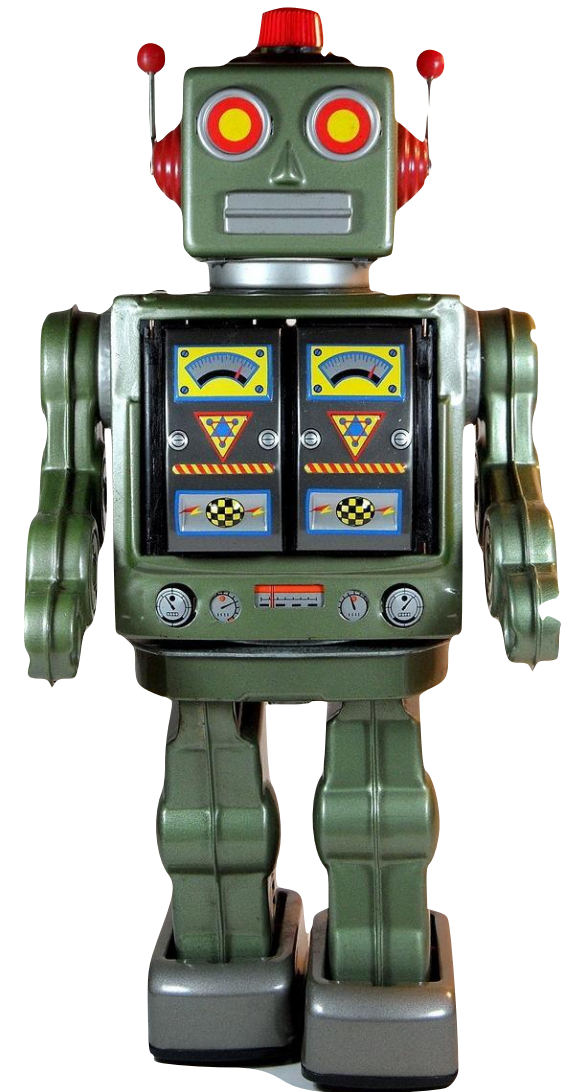
compréhensible par l'humain



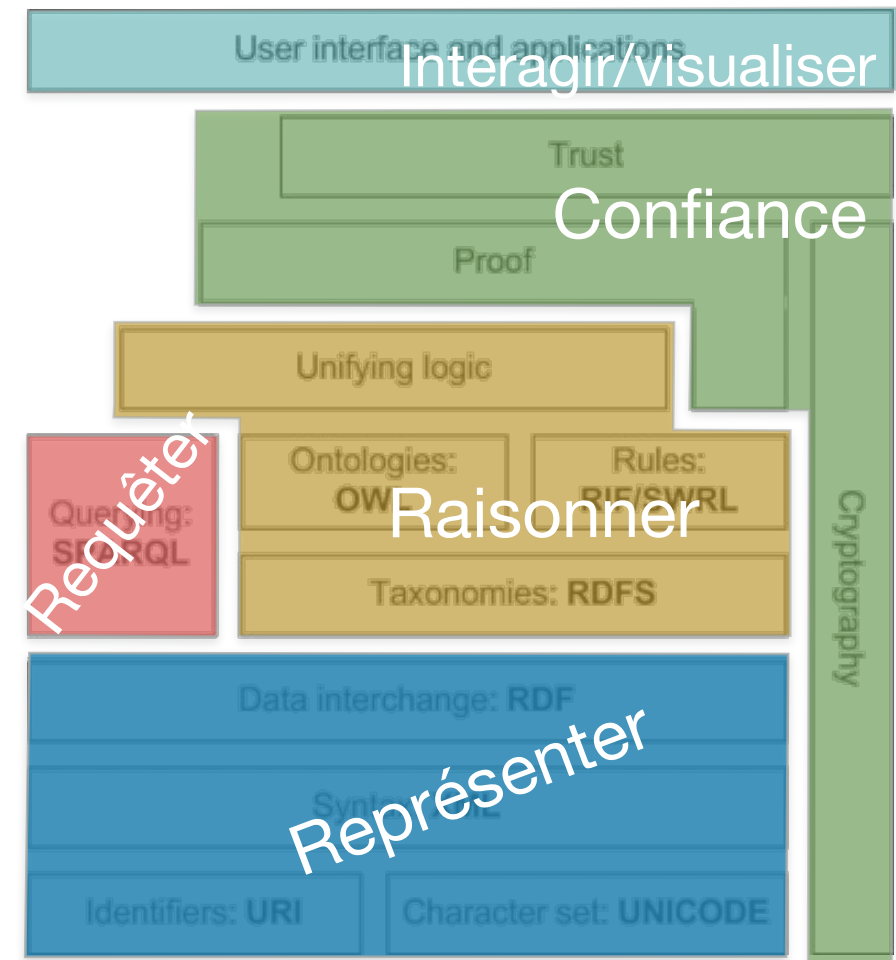
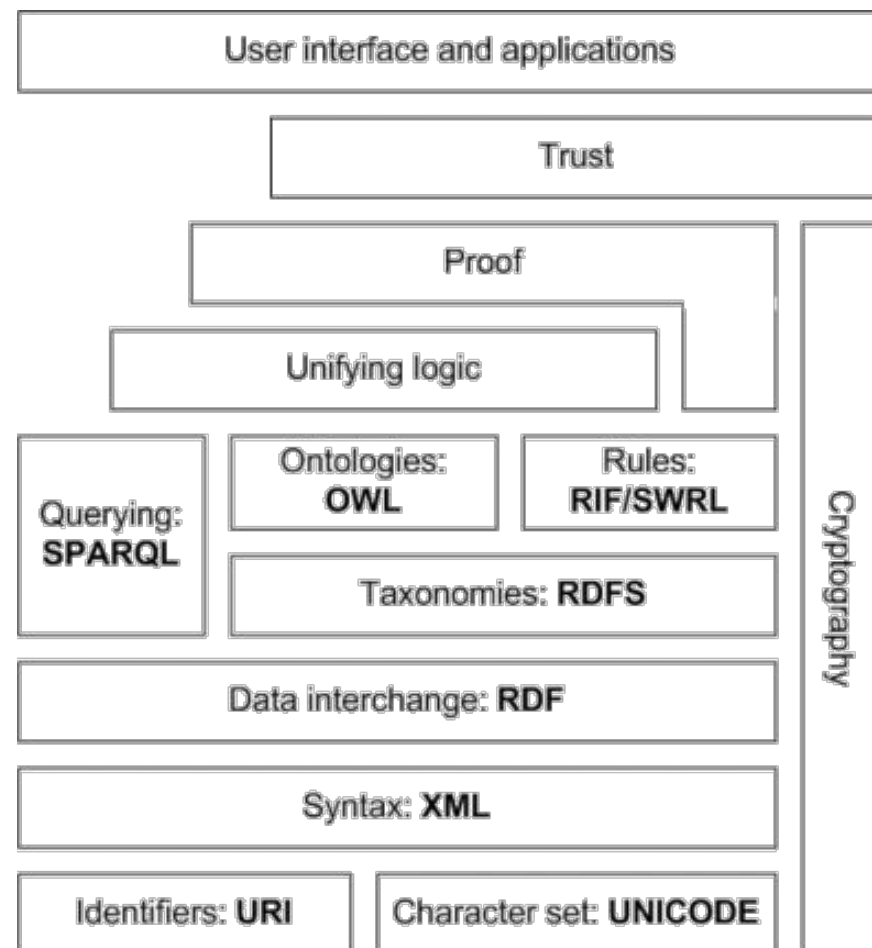
« Devinable » pour les machines

« affichable » pour les machines

à un web manipulable pour les machines



Un web de données liées (2) ←



d'après Fabien Gandon <http://goo.gl/vmYqDA>

Le Web de données souvent dénommé linked data vise à favoriser la publication de données structurées, en reliant ces données entre elles pour constituer une gigantesque base de données (Bizer et Heat, 2001) et non sous la forme de silos de données isolés. Pour ce faire, plusieurs briques technologiques ont été mises en place.

4 éléments fondamentaux pour le web de données

Système d'identifiants fiables et pérennes:

URI

Grammaire RDF et ses syntaxes

RDF/XML, N3, Turtle, N-triple, RDFa

Ontologies exprimées

RDFs, OWL

Protocole, langage de requête pour le RDF

SPARQL

URI Permet d'établir un identifiant pérenne pour chaque ressource sur le web, c'est-à-dire d'identifier sur le web tout ce qui existe !



<http://.../stylobille>



http://.../Harry_Potter



<http://.../tomate>

« Une URI par chose, une chose par URI »

Un principe fort **AAA**

Anycne can say **A**nycnng about **A**nycnng



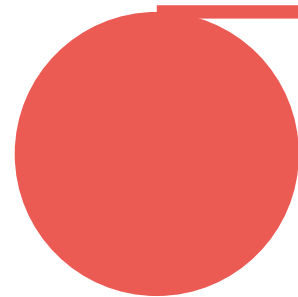
Grammaire RDF (Ressource Description Format)

Modèle où toute connaissance respecte la forme d'un triplet selon la syntaxe suivante

sujet

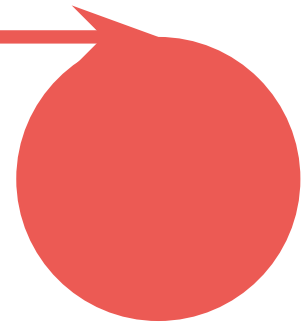
prédicat

objet



Bruxelles

LieuDeNaissanceDe



ClaudeLevyStrauss

<http://dbpedia.org/ressource/Bruxelles>

<http://dbpedia.org/ontology/birthPlace>

<http://dbpedia.org/ressource/ClaudeLevyStrauss>

Un web de données liées



est
une

sculpture

date
de

1565-1585

représente



s'appelle

Marc Aurèle

a sa biographie ici

est
un

site web

Un web de données liées

Il faut pouvoir exprimer ces descriptions sans ambiguïtés et surtout de manière commune à d'autres producteurs de données

Ontologies

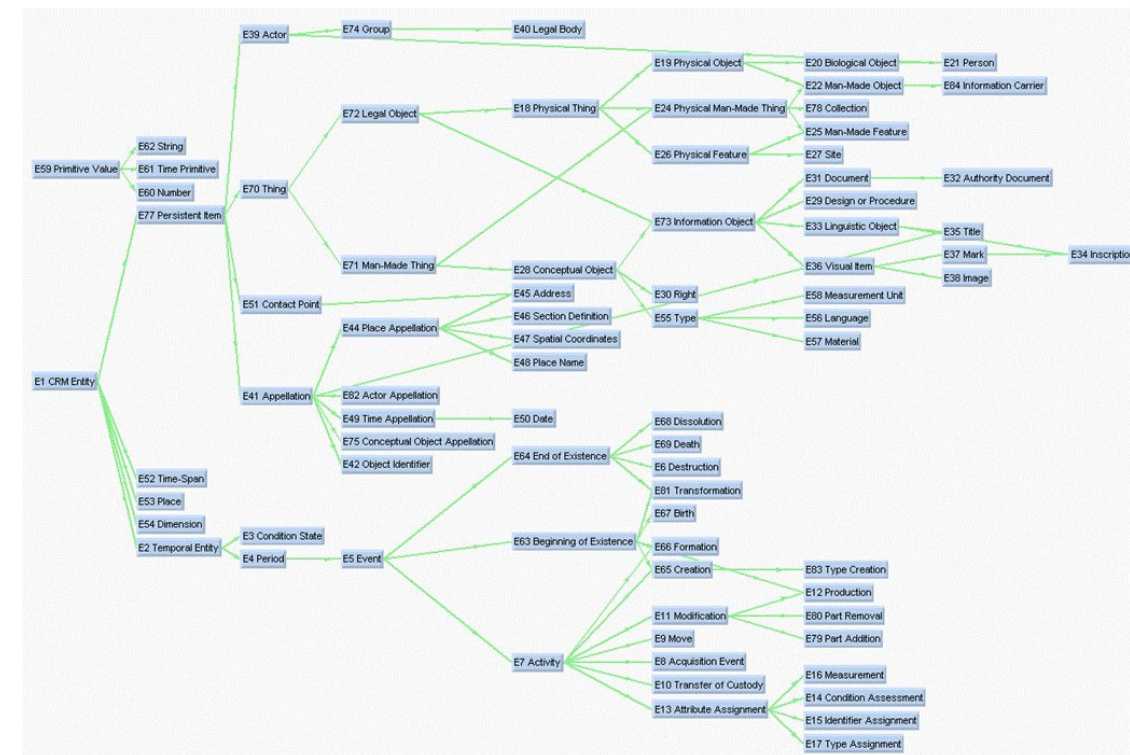
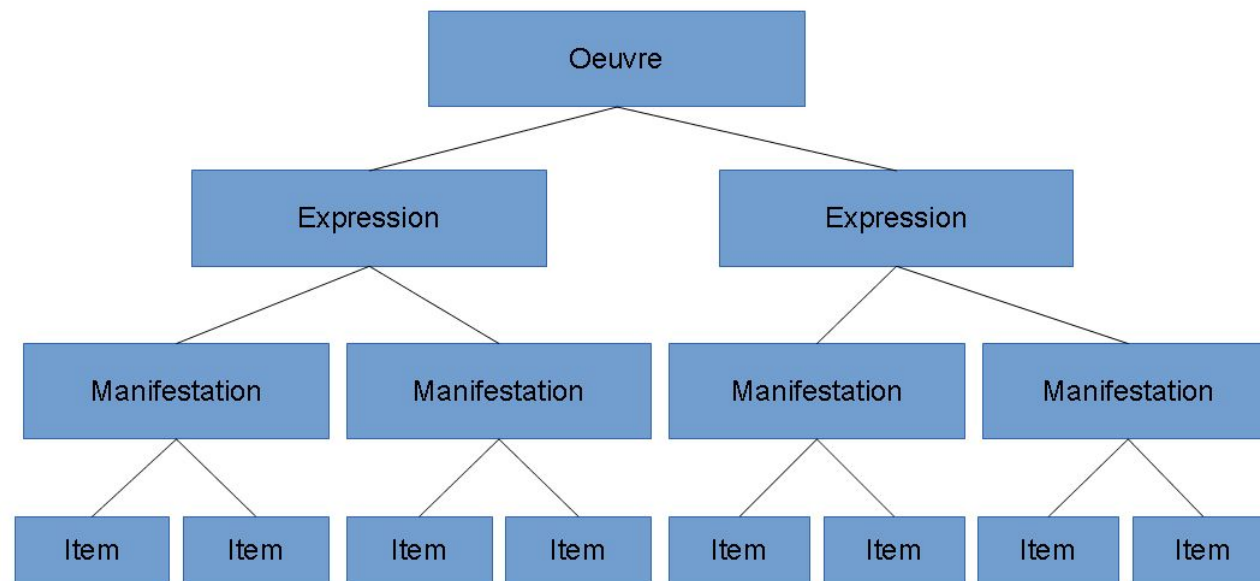
Dans le milieu culturel, plusieurs ontologies co-existent

FRBR

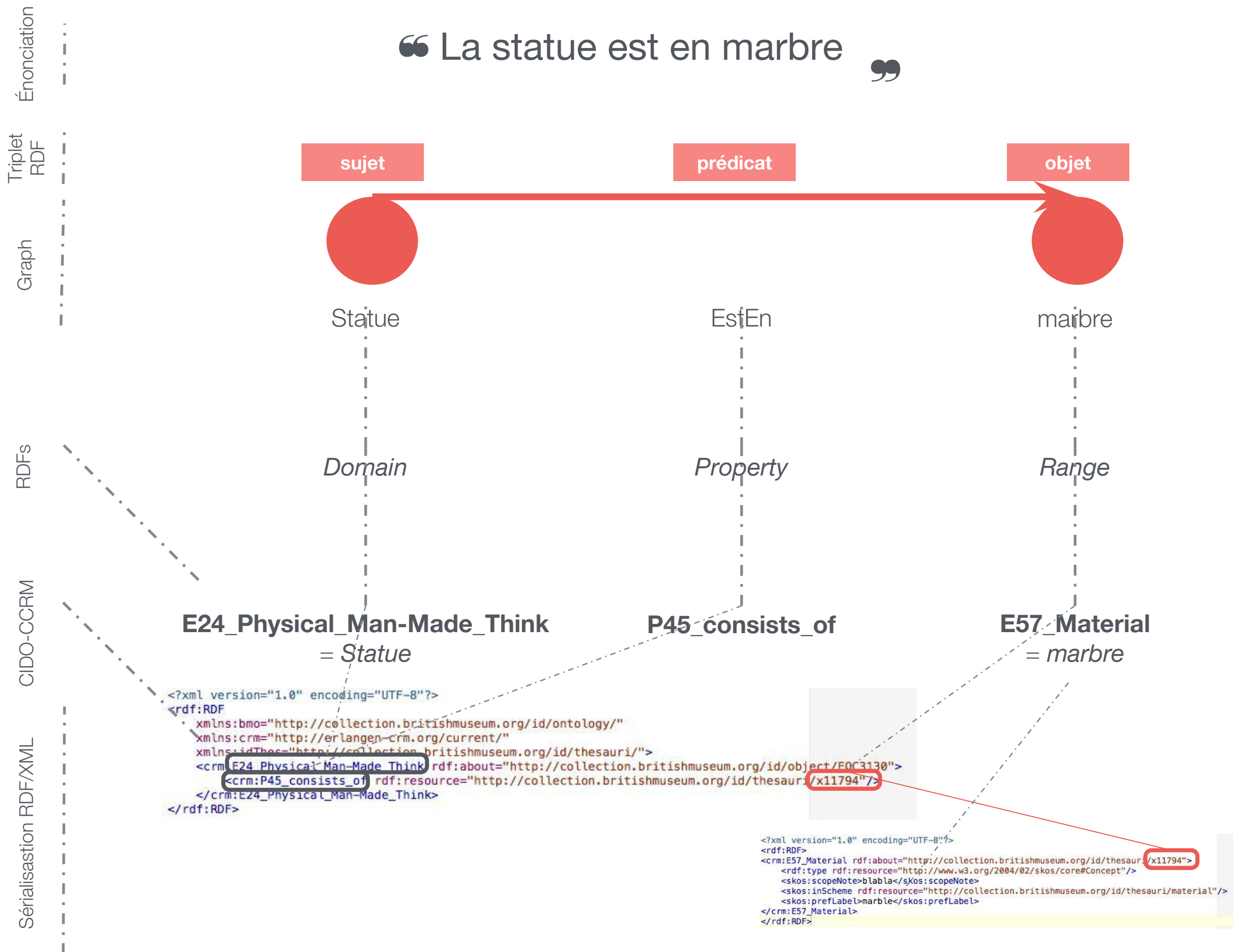
FRBR_{er}

FRBR_{oo}

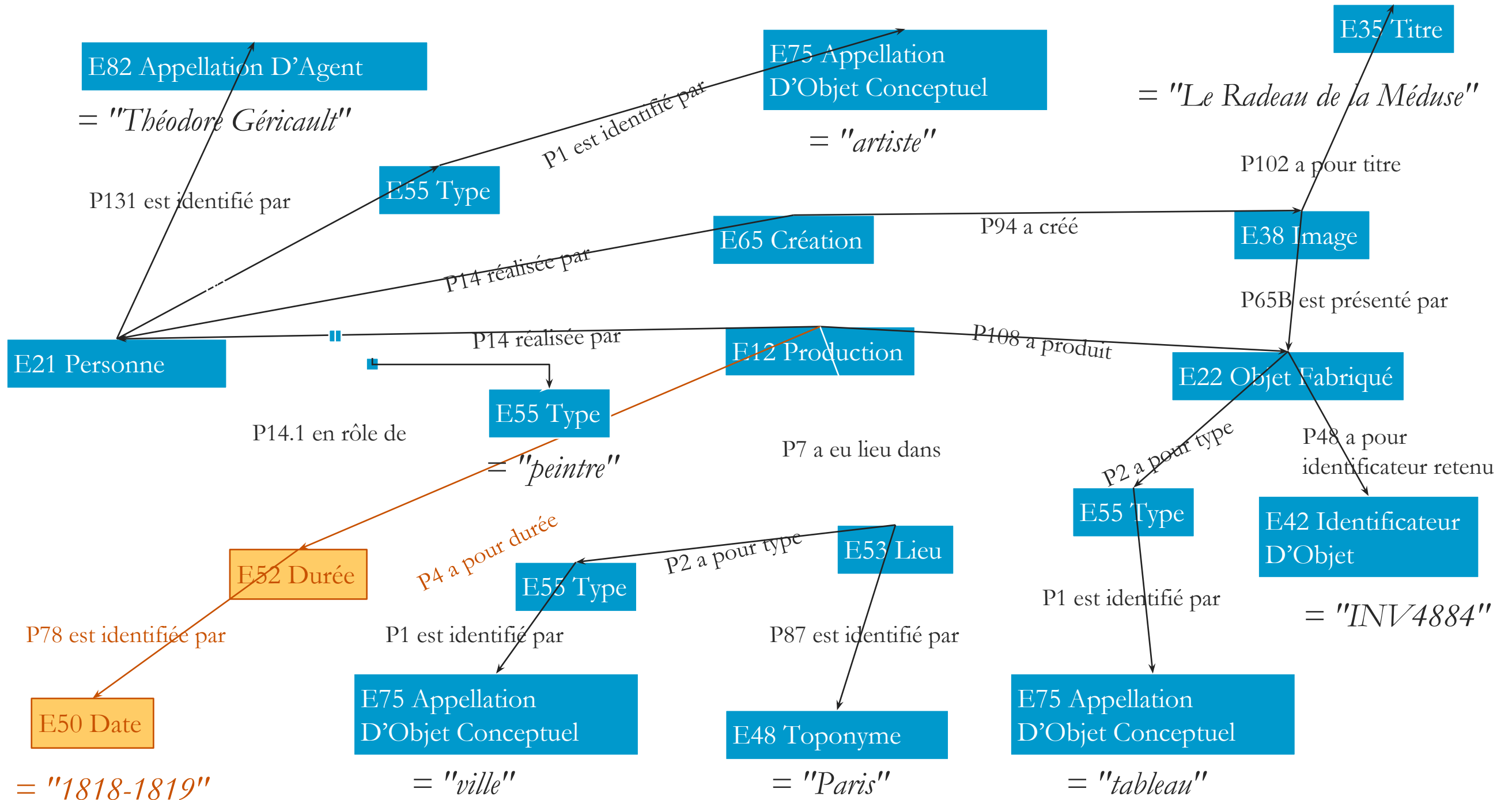
CIDOC-CRM



“ La statue est en marbre ”



- « L'artiste Théodore Géricault a peint son tableau *Le Radeau de la Méduse* (INV4884) dans la ville de Paris en 1818-1819. »





Données accessibles sur le web (sans condition de formats)



Données accessibles structurées (ex: fichier Excel plutôt que le PDF d'un tableur)



Données structurées dans des formats non-propriétaires (ex: CSV plutôt qu'Excel)

Open Data

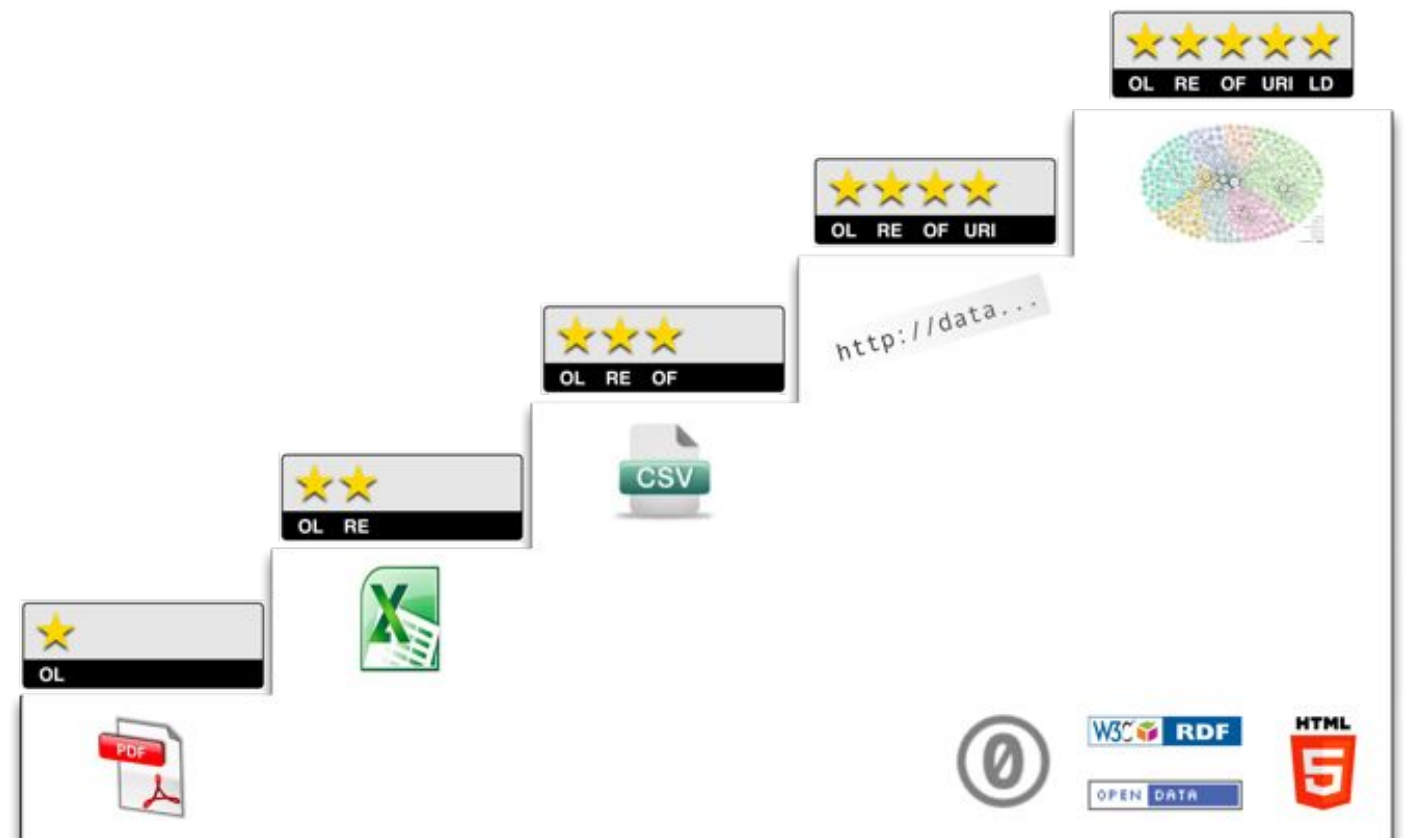


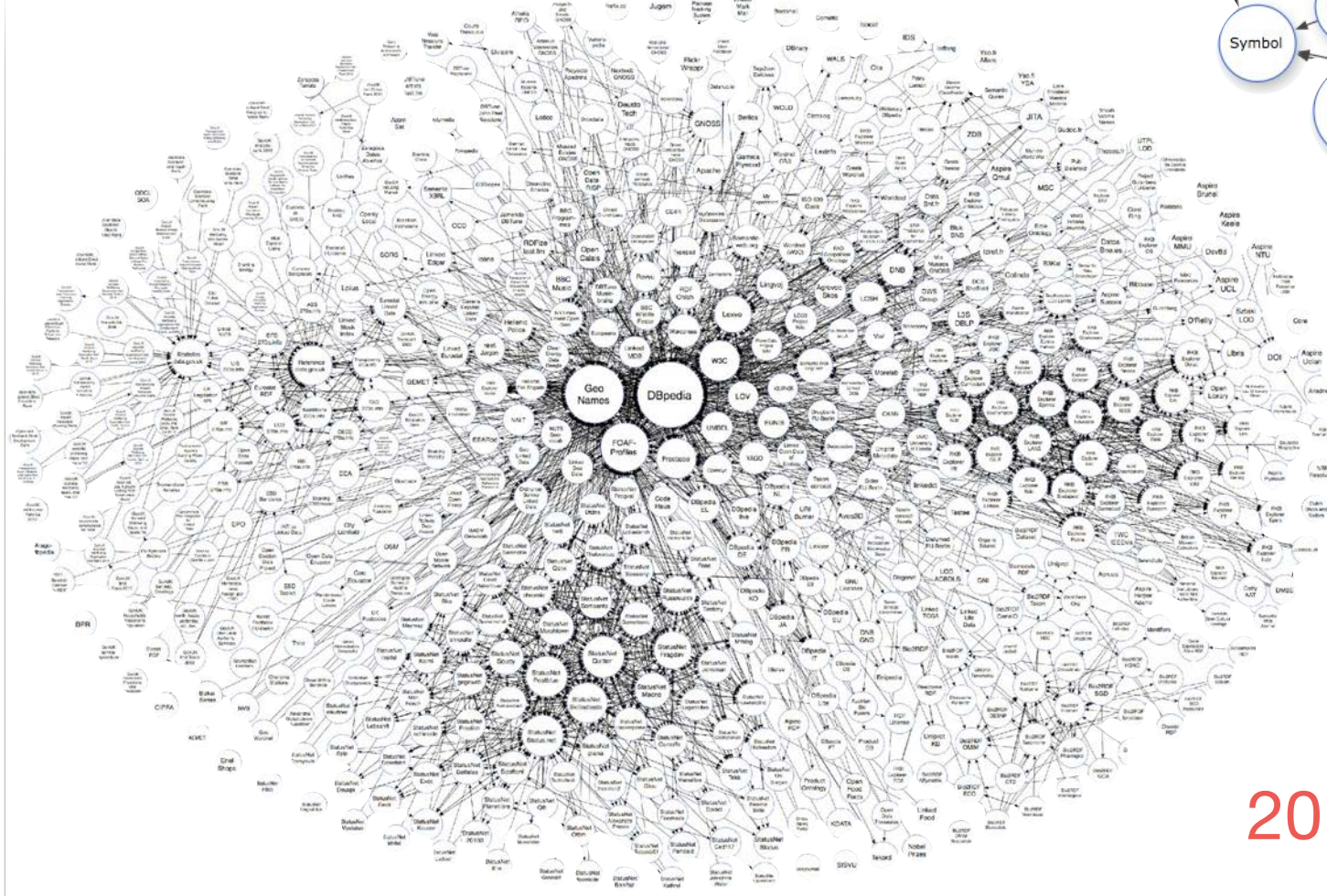
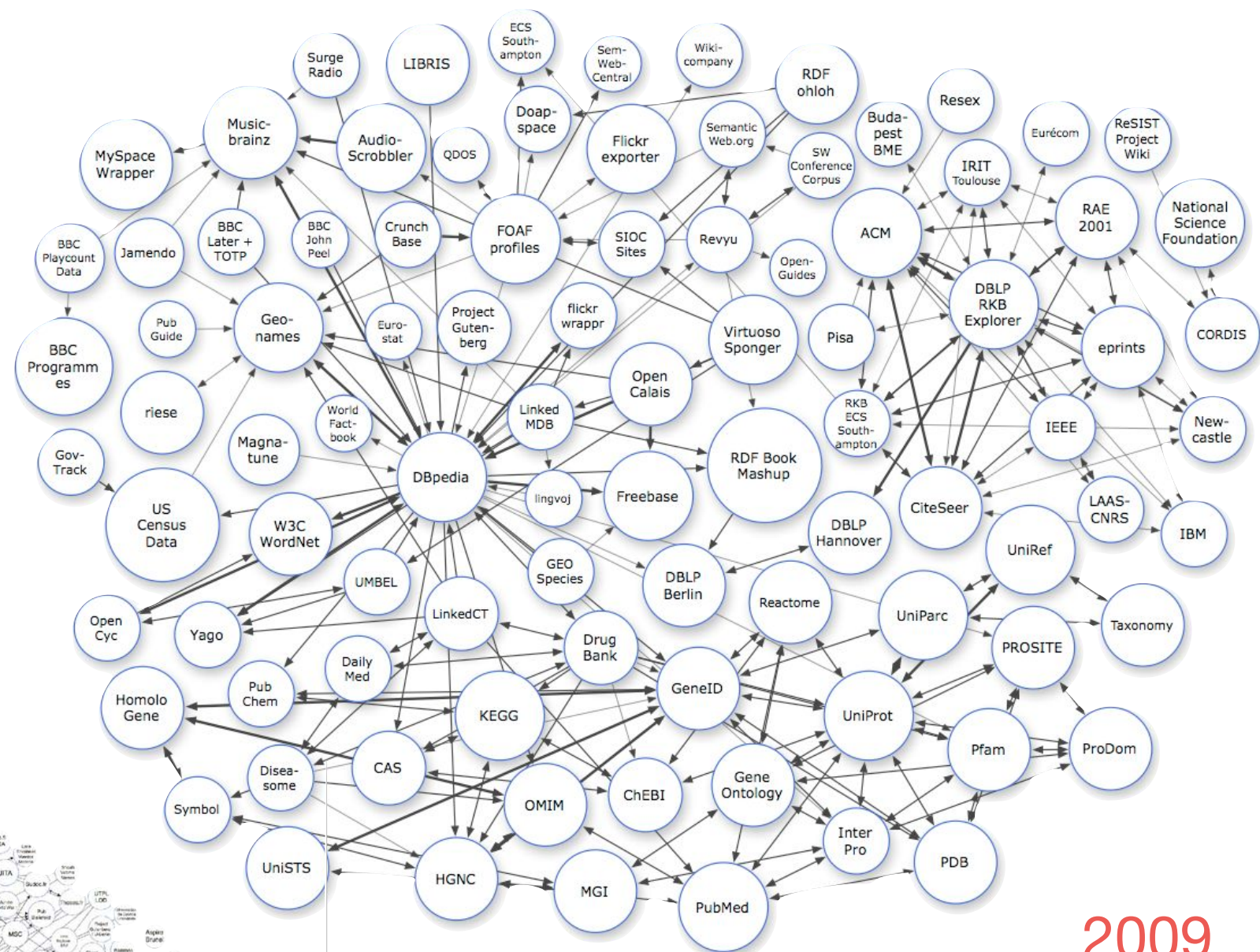
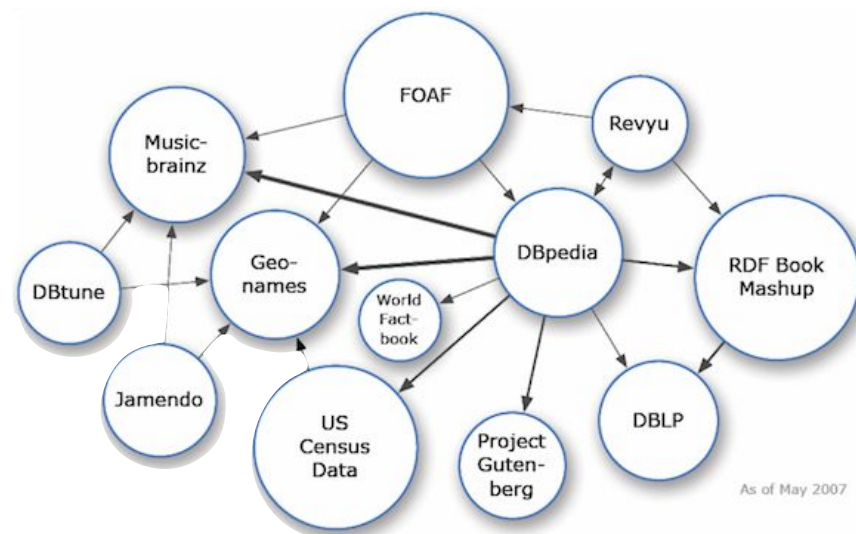
Utilisation des URIs pour identifier les ressources

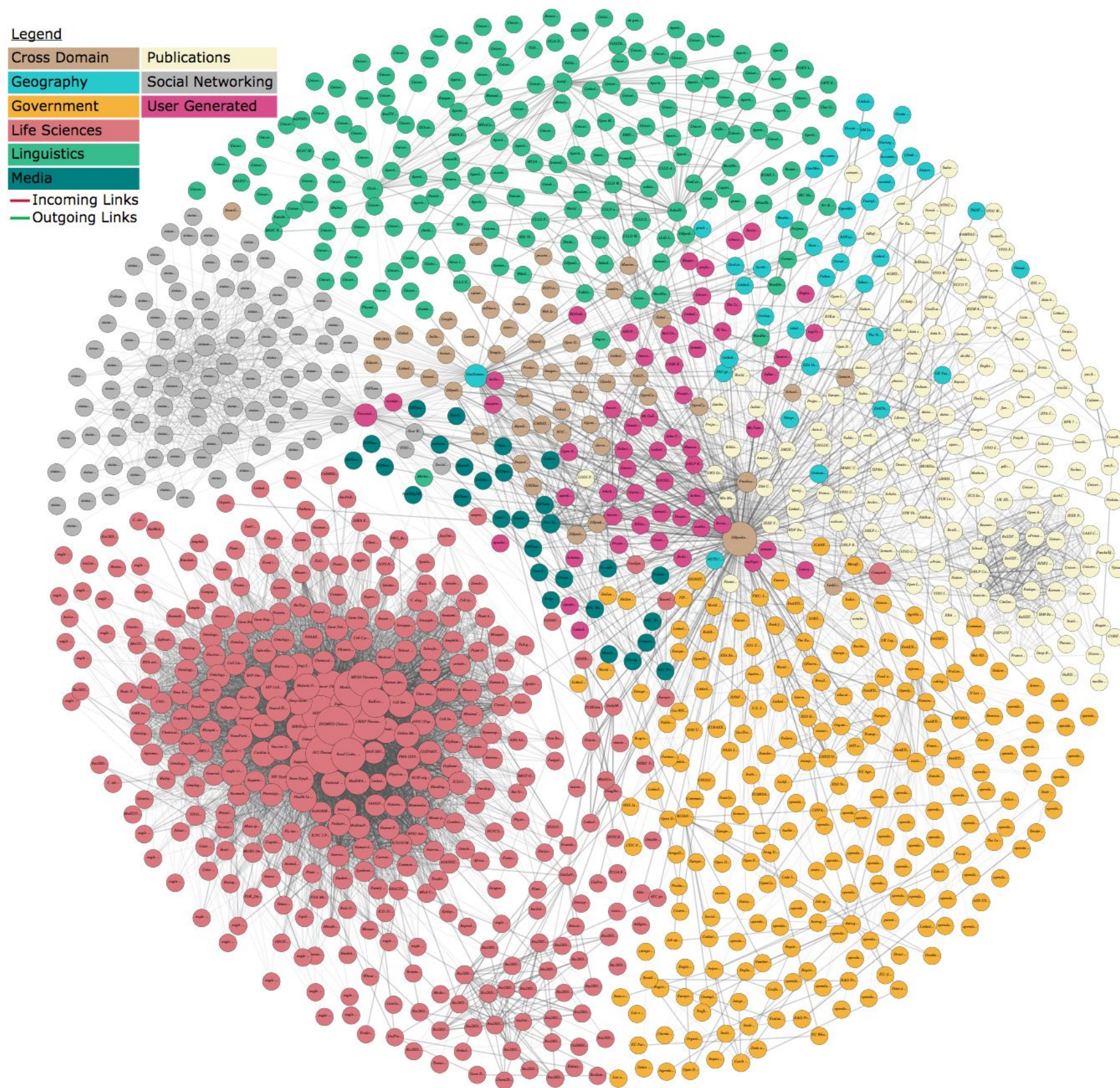
Linked Open Data



Les données sont reliées à d'autres données







2017

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

Ce n'est **pas une rupture**

Mais le prolongement du web qui développe des données structurées **compréhensibles** par les **machines...**

British Museum



Home | Sparql | Help | Licensing | About

British Museum Semantic Web Collection Online



Powered by the CIDOC Conceptual Reference Model and SPARQL 1.1



Welcome to this Linked Data and SPARQL service. It provides access to the same collection records available through the Museum's web presented **Collection Online**, but in a computer readable format. The use of the W3C open data standard, RDF, allows the Museum's collection data to join and relate to a growing body of linked data published by other organisations around the world interested in promoting accessibility and collaboration.

The data has also been organised using the **CIDOC CRM** (Conceptual Reference Model) crucial for harmonising with other cultural heritage data. The CIDOC CRM represents British Museum's data completely and, unlike other standards that fit data into a common set of data fields, all of the meaning contained in the Museum's source data is retained. This represents the only way in which museums can represent all their specialist and localised knowledge, essential for serious research, but also support highly precise cross organisational data discovery and semantic relationships.

We hope that the service will be used by the community to develop web applications for a wide range of audiences from general users to academic researchers – and that are freely available to the community:

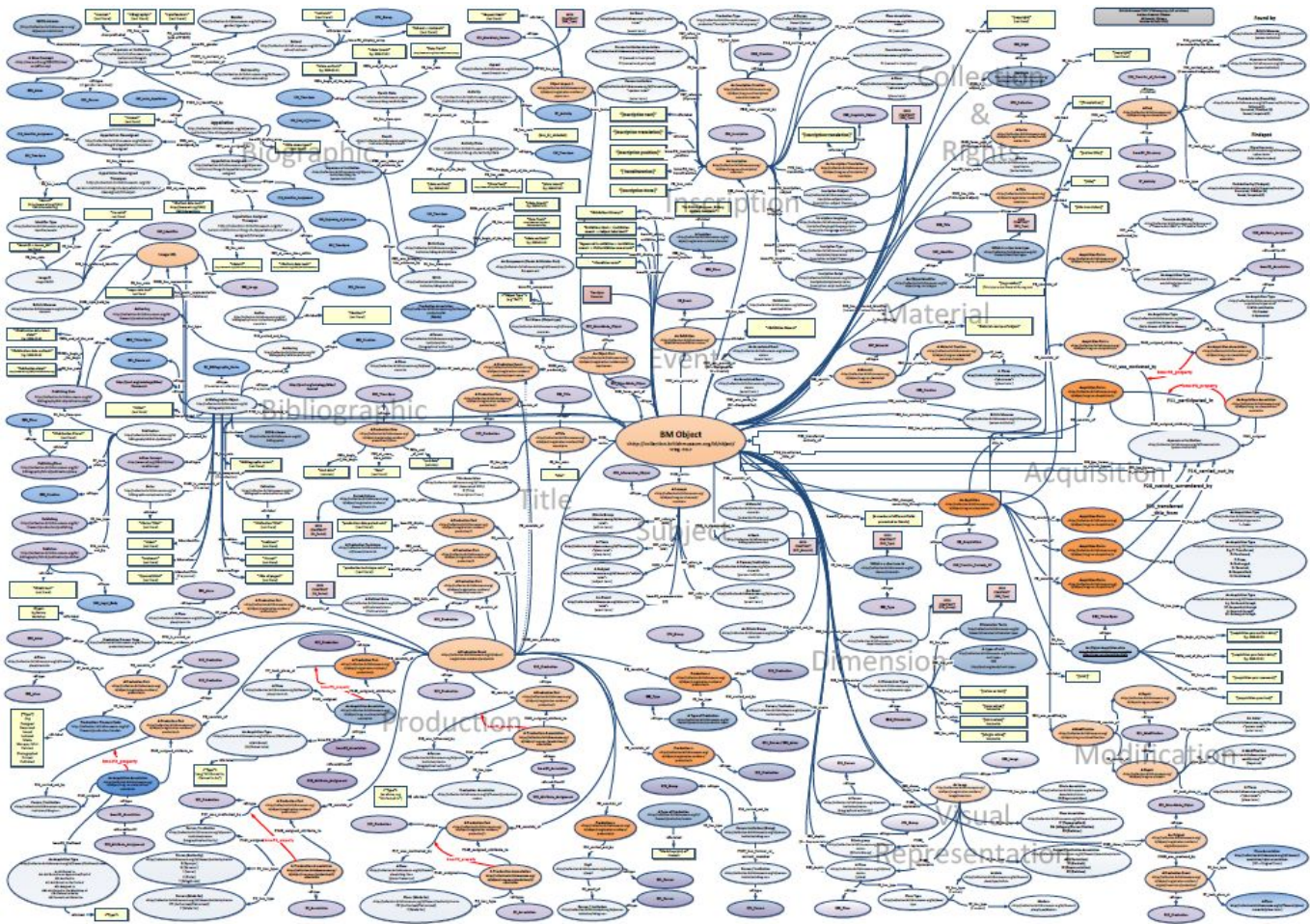
RDF and HTML-resolvable resources are available for every object, for example [here](#)

A SPARQL endpoint (and HTML user interface) is available [here](#)

An example web application built using the SPARQL endpoint using AngularJS (and Django) is available [here](#).

RDF (NQuads) dumps are available for the entire collection [here](#).

The collections data is also used in the [ResearchSpace](#) project:



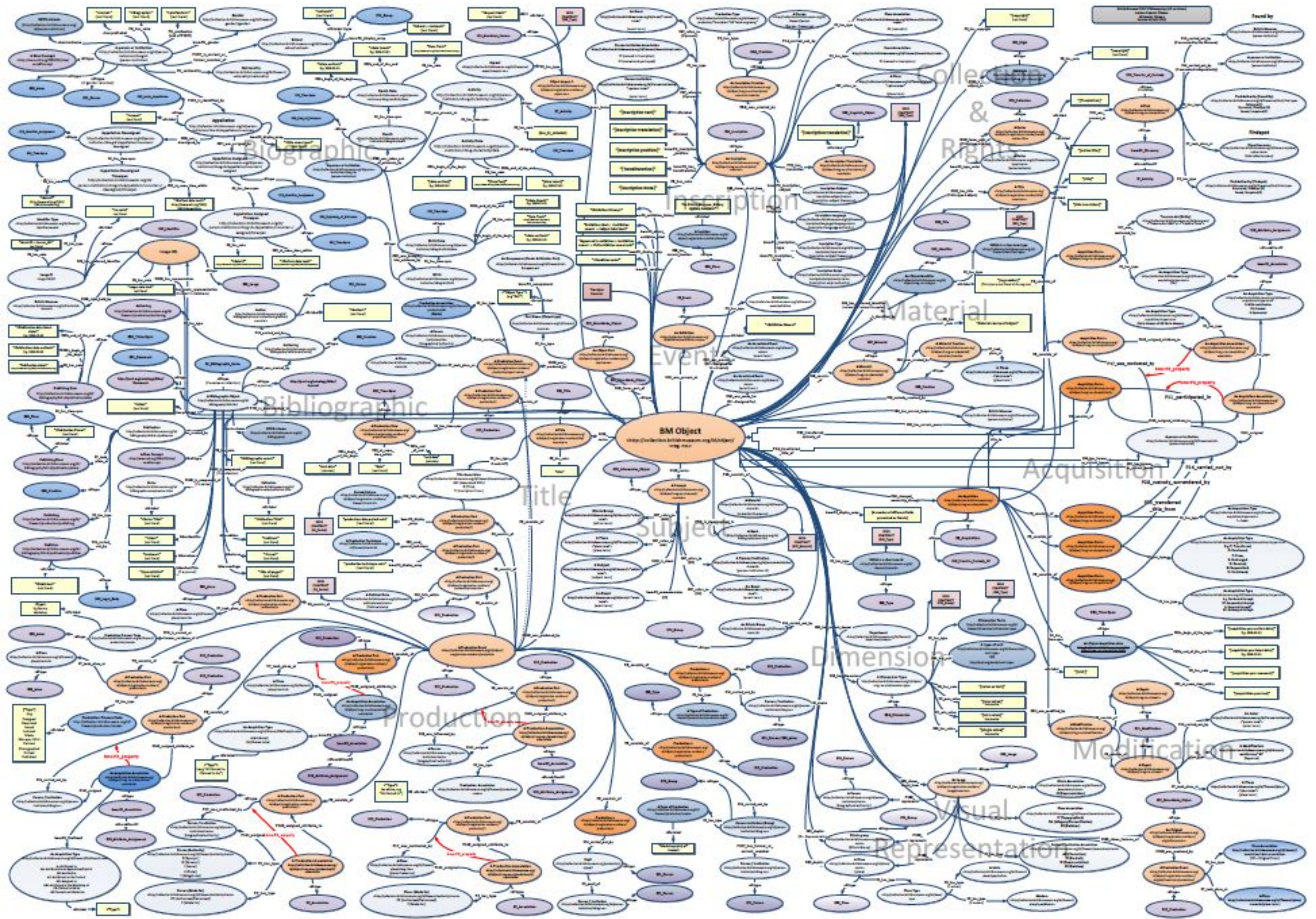


Schéma d'analyse des éléments de description d'une oeuvre réalisé par le British Museum à partir du modèle de données CIDOC-CRM

Find: **Things** FROM **Salamis Cyprus** and HAS TYPE **terracotta** and HAS TYPE **statue** and CREATED ON **Year 800 BC - Year 300 BC**

Find:

●

CREATED ON

Year 800 BC - Year 300 BC

 or

+

and

+

remove

●

FROM

Salamis Cyprus

 AND

HAS TYPE

terracotta

 AND

HAS TYPE

statue

 AND

CREATED ON

Year 800 BC - Year 300 BC



- Hide Filter
- ▶ acquired by (15)

▶ found by (4)

▶ from (15)

▼ has type (15)

☐ Proto-Cypriot (3)

☐ figure (1)

☐ hand-modelled (4)

☐ handmade (9)

☐ impressed (3)

☐ incised (3)

☐ mould-made (2)

☐ moulded (3)

☐ painted (15)

☐ slipped (1)

▶ created on (15)

▶ created in (15)

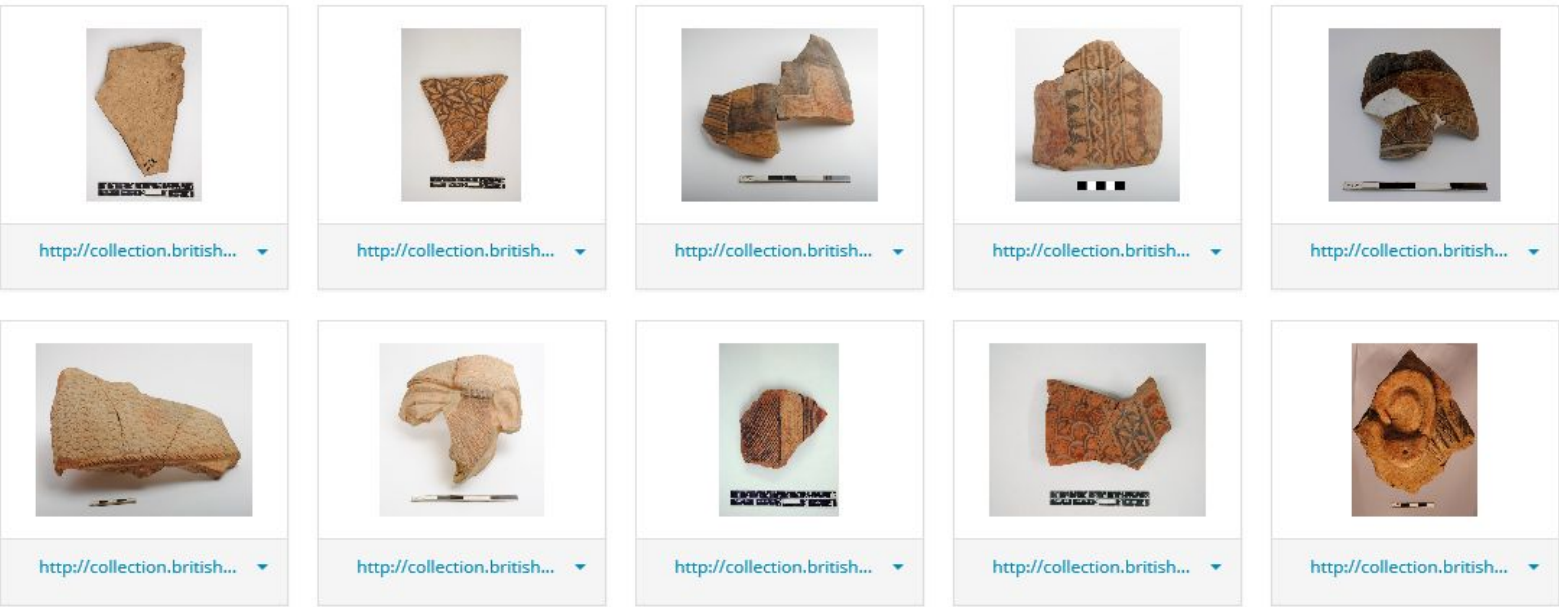
▶ from (15)

▶ modified in (15)

▶ found or acquired at (15)

▶ from (15)

▶ is similar or same with (9)



« 1 2 »

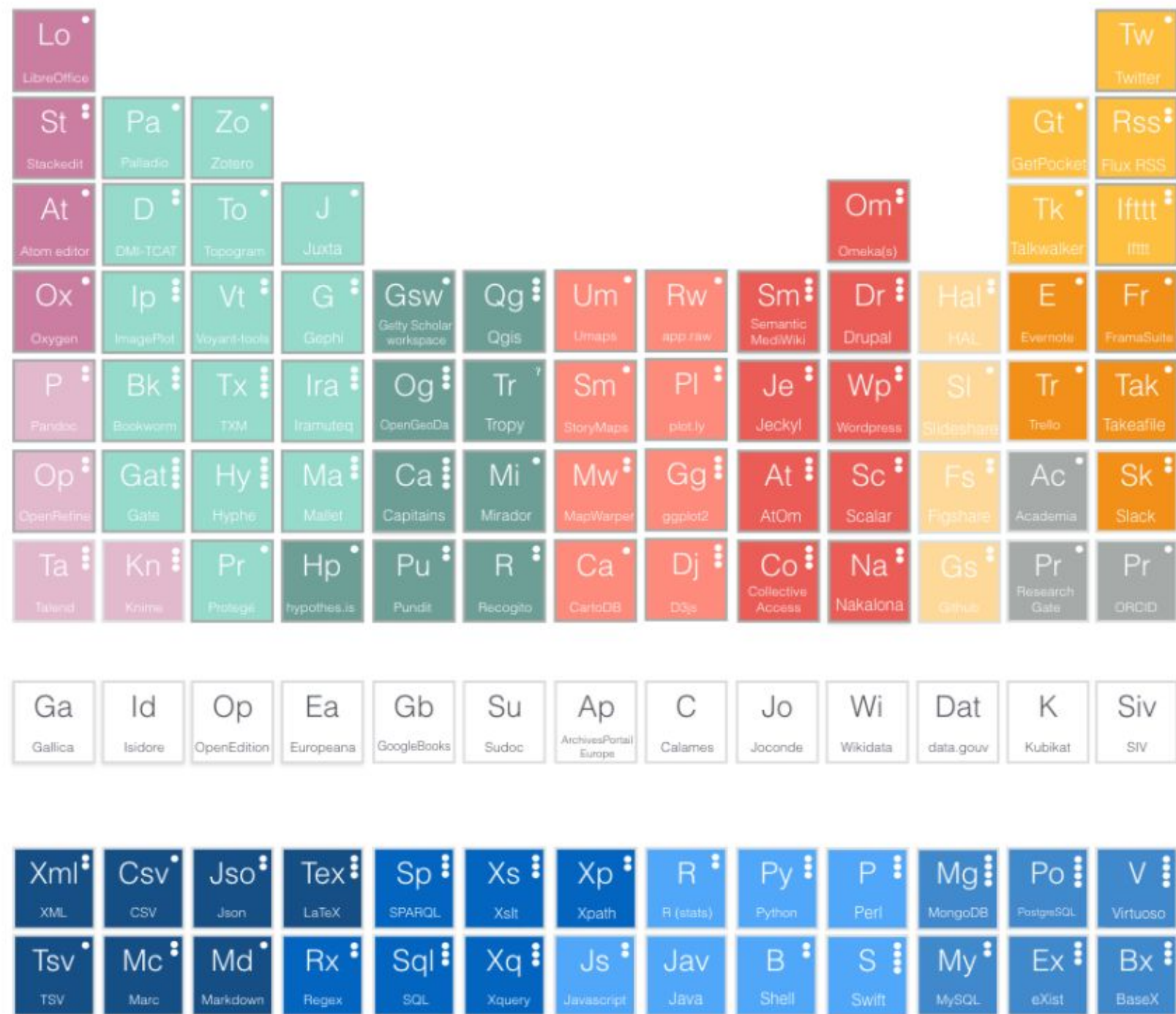
Save as New Set

Add to Existing Set

Semantic Search exploring Salamis, Cyprus for the European GRAVITATE project. <http://tinyurl.com/q9br83l> <https://public.researchspace.org/resource/Start>

Interface actuelle (2019) du projet Research Space qui contient l'ensemble de la base collection du British Museum après avoir été modélisé sous les principes du CIDOC-CRM

OUTILS



<https://goo.gl/LI9nVi>

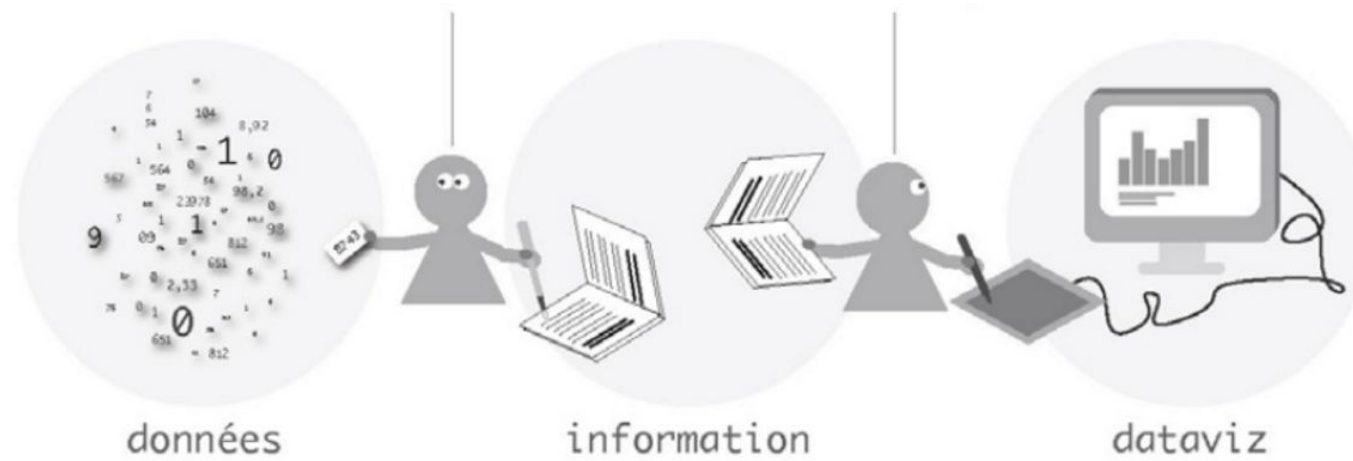


grille typologique des différents outils numériques
(Libre Office, Palladio, Voyant Tools)

- l'émergence d'outils *DH* : *une chaîne de traitement de corpus à la portée de tous ?*

> concilier l'exigence documentaire et l'exigence scientifique

1. acquisition des données
2. traitement des données
3. exploitation des données
4. diffusion des données
 1. sous forme de données structurées
 2. une éditorialisation nécessaire pour contextualiser
 3. diffusion des contenus numérisés
5. encourager la réutilisation



Acquisition des données

- saisie des données
- collecte de données
- enrichissement des données tierces



Structurer les données

- nettoyer les données
- optimiser les formats
- compléter les données



Traiter et visualiser l'information

- choisir des informations à révéler
- tester les formes de visualisation



Editorialiser la visualisation de données

- apporter un discours
- légender



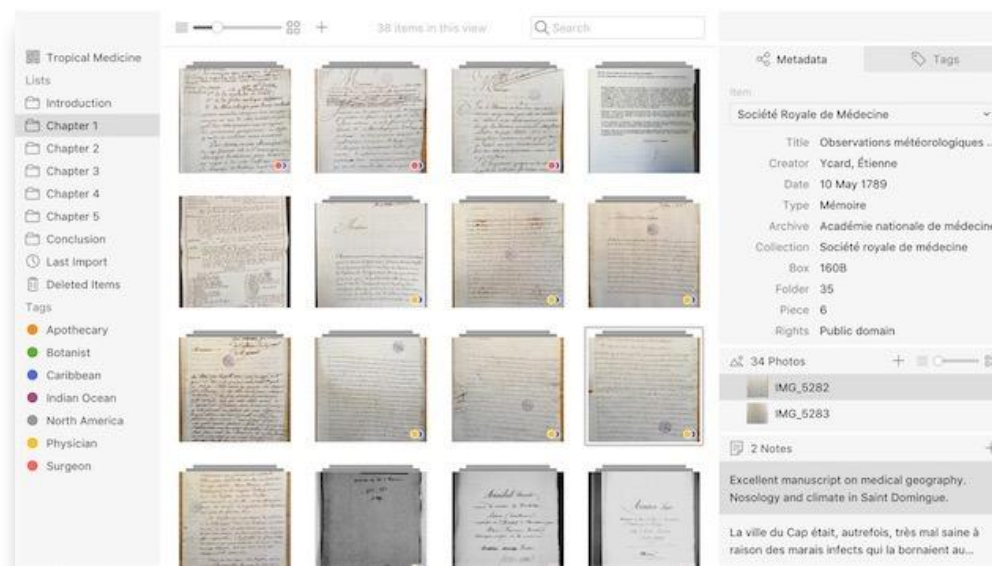
- Les outils ne sont pas un but ultime : il faut surtout favoriser la compréhension des enjeux numériques et le rôle des données structurées par les chercheurs
- « apprivoiser le logiciel et le plier à ses usages » Caroline Muller à propos de zotero (<https://consciences.hypotheses.org/1184>)

The screenshot shows the Pelagios 3 Project web interface. The top navigation bar includes the project name, a search icon, and a user login status 'Logged in as Rainer Simon'. The main content area displays a text snippet from 'Arrian: Periplus of the Euxine Sea' (130-131 CE). The text is annotated with green highlights around 'Trapezus' and 'Sinope'. A map overlay is visible, showing the location of Trapezus (modern Trabzon) on the Black Sea coast. The map includes a location pin and a text box with the name 'Trapezus' and coordinates 'pleiades:857359'. Below the map, there are buttons for 'Change' and 'OK', and a text input field for 'Add a comment...'. The interface also shows '427 Annotations' and 'No Other Contributors' for this text.

Créez des cartes à partir de données géographiques issues de textes. Annotez des images. Mettez vos données de recherche en relation avec d'autres données présentes sur le web, augmentez leur visibilité en les rendant partageables, sans avoir besoin d'être un expert des technologies du Web sémantique.

<https://recogito.pelagios.org/>

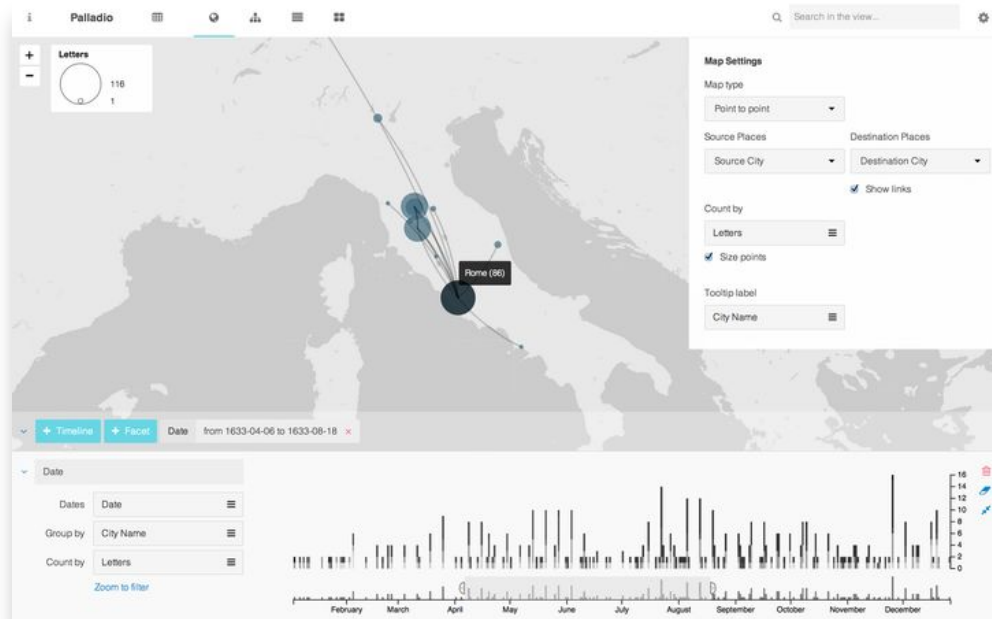
<https://tropy.org>



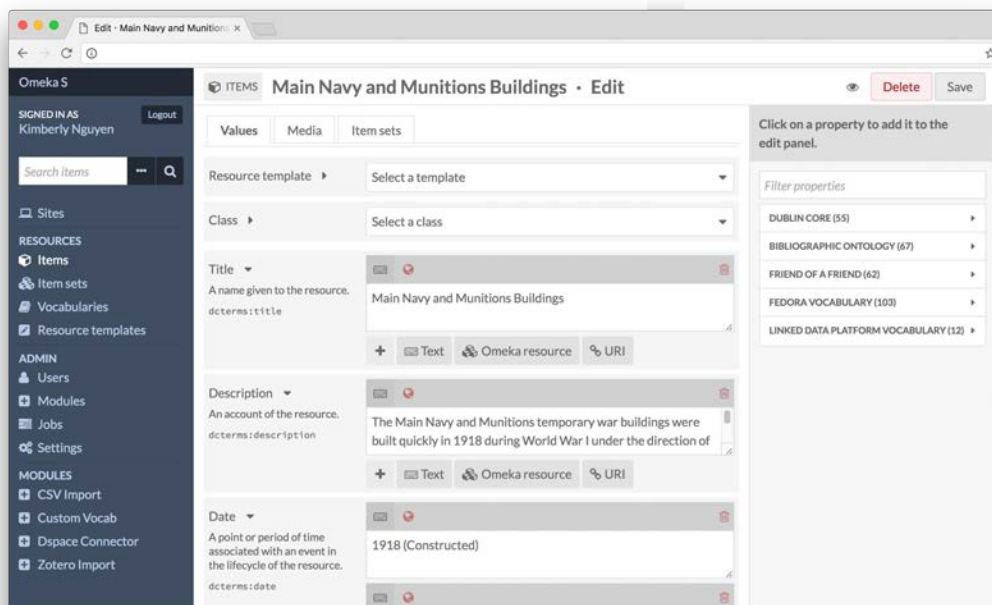
Trois exemples de programme utile à différentes étapes d'un projet

Tropy pour gérer ses photos de dépouillements d'archives notamment est disponible en v1. *Gérer la documentation*

<http://hdlab.stanford.edu/palladio-app/>



Palladio pour visualiser/explorer un corpus via différentes approches pouvant être combinées (galerie d'images, carte, réseaux, chrono, etc.) *Explorer les données*



<https://omeka.org/s/>

OmekaS (Plate-forme de publication Web orienté web sémantique dédiées aux collections numériques du patrimoine culturel) *Publier les données*