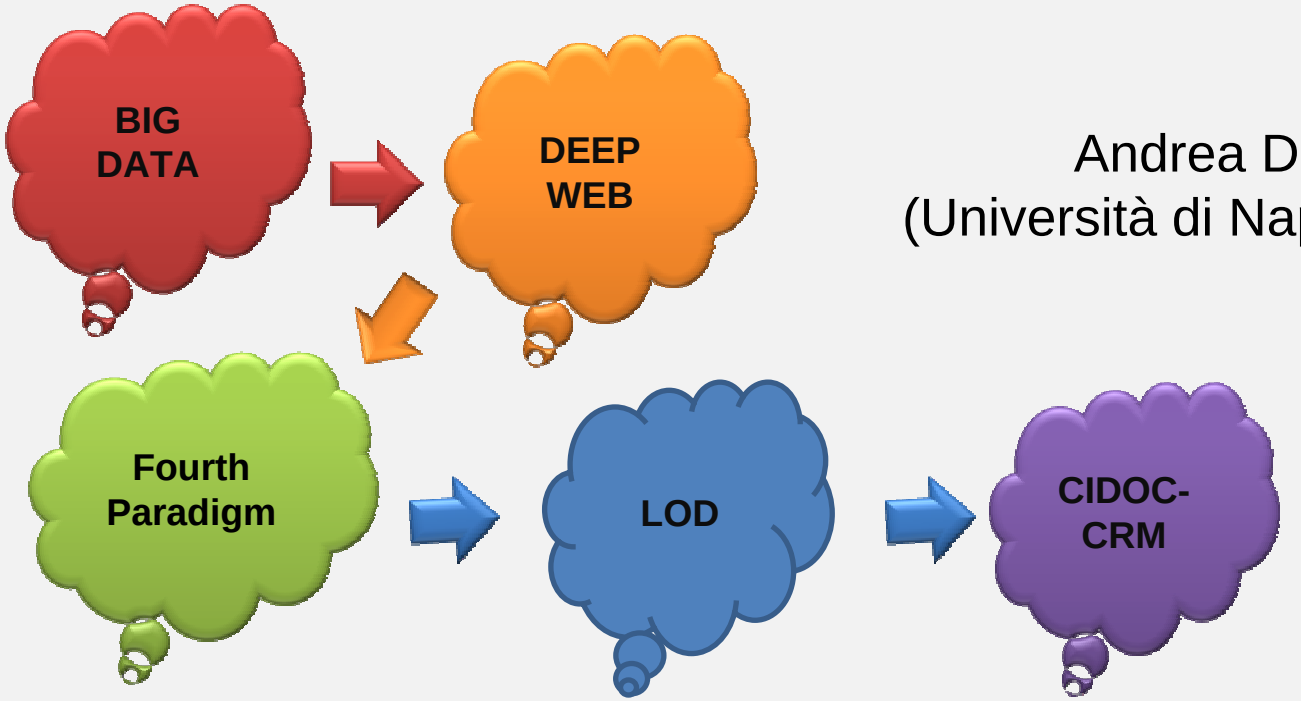
CIDOC-CRM
&
Linked Open Data

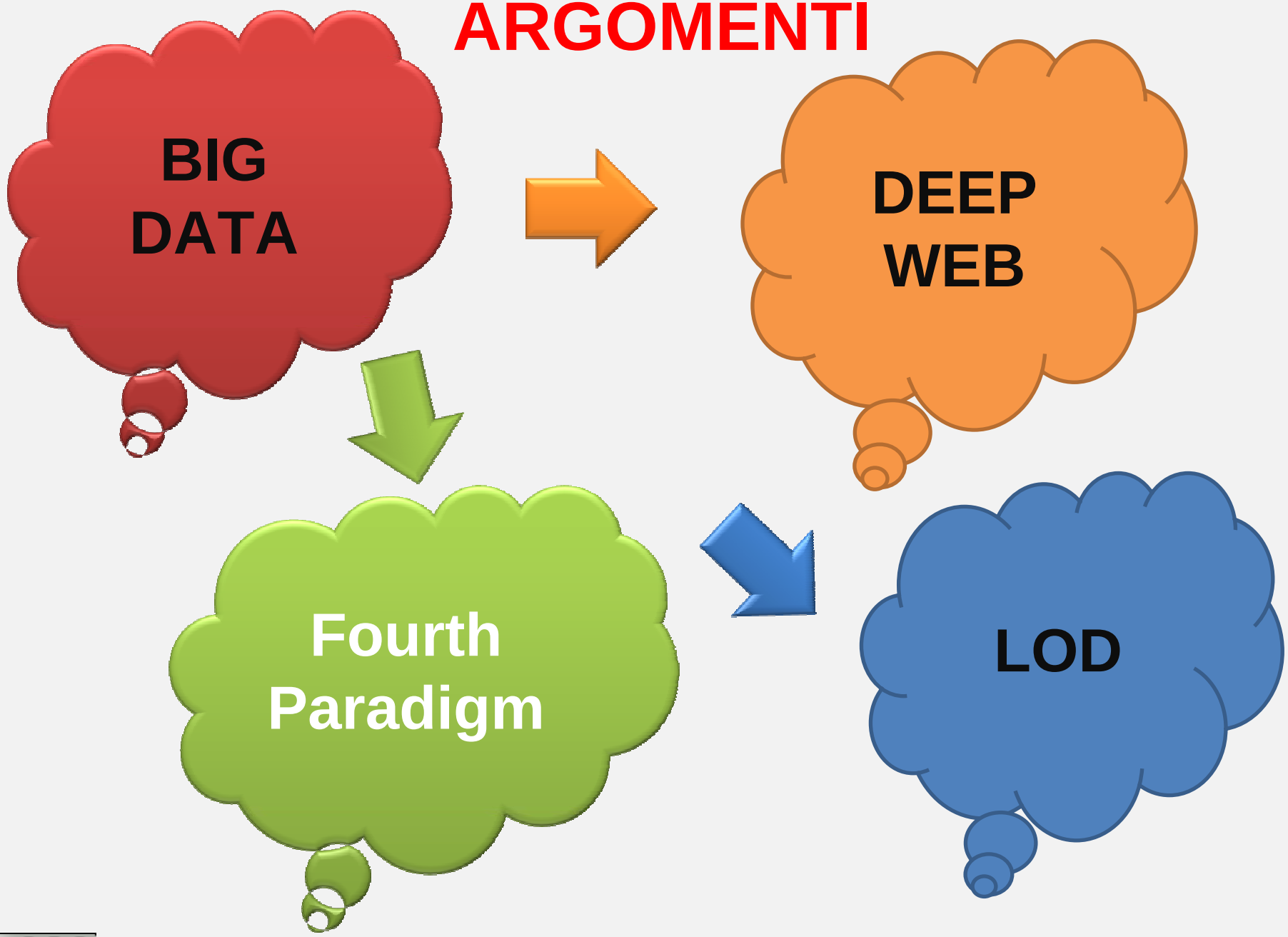


Andrea D'ANDREA
(Università di Napoli "L'Orientale")

BY NC SA

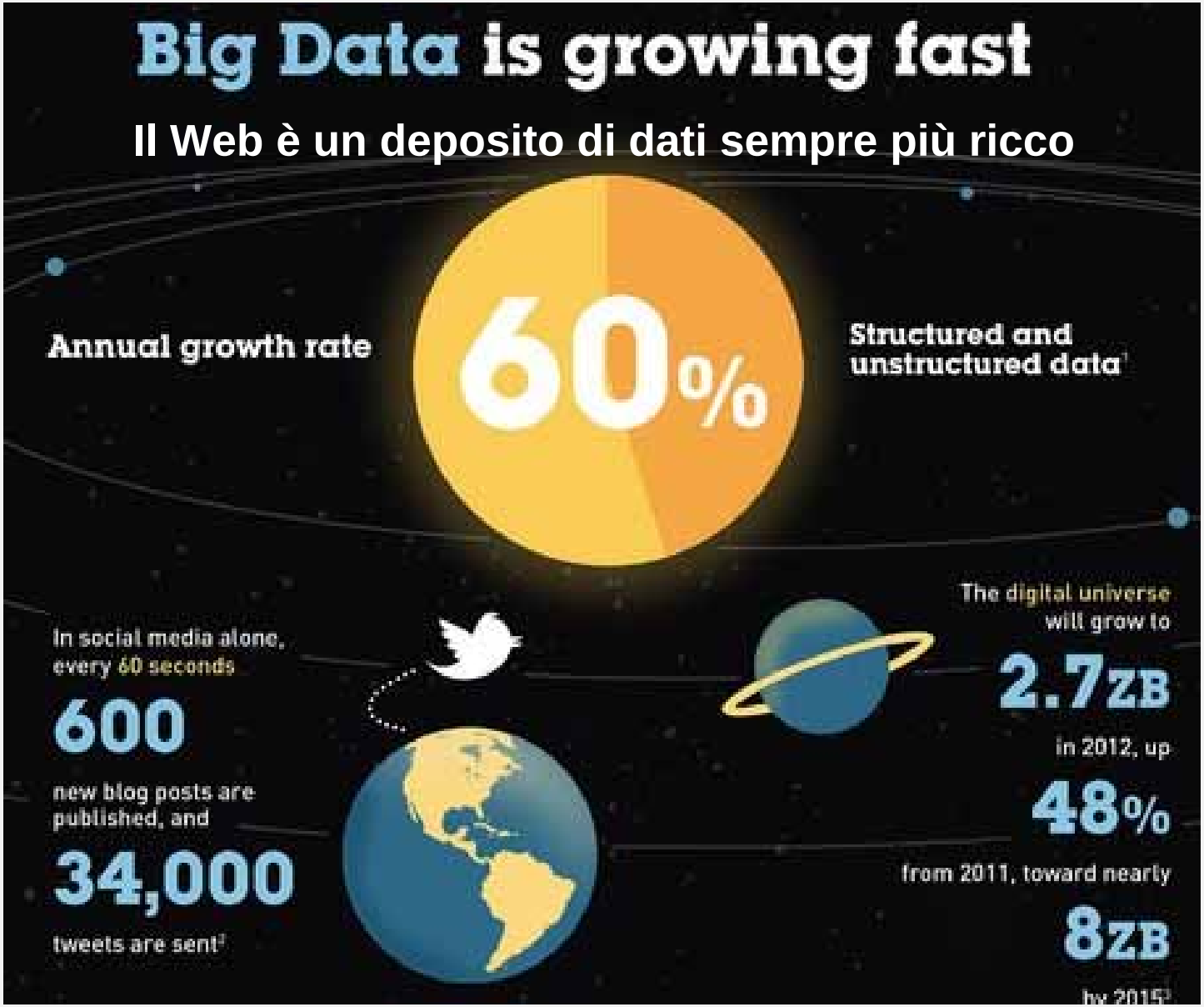
dandrea@unior.it

ARGOMENTI



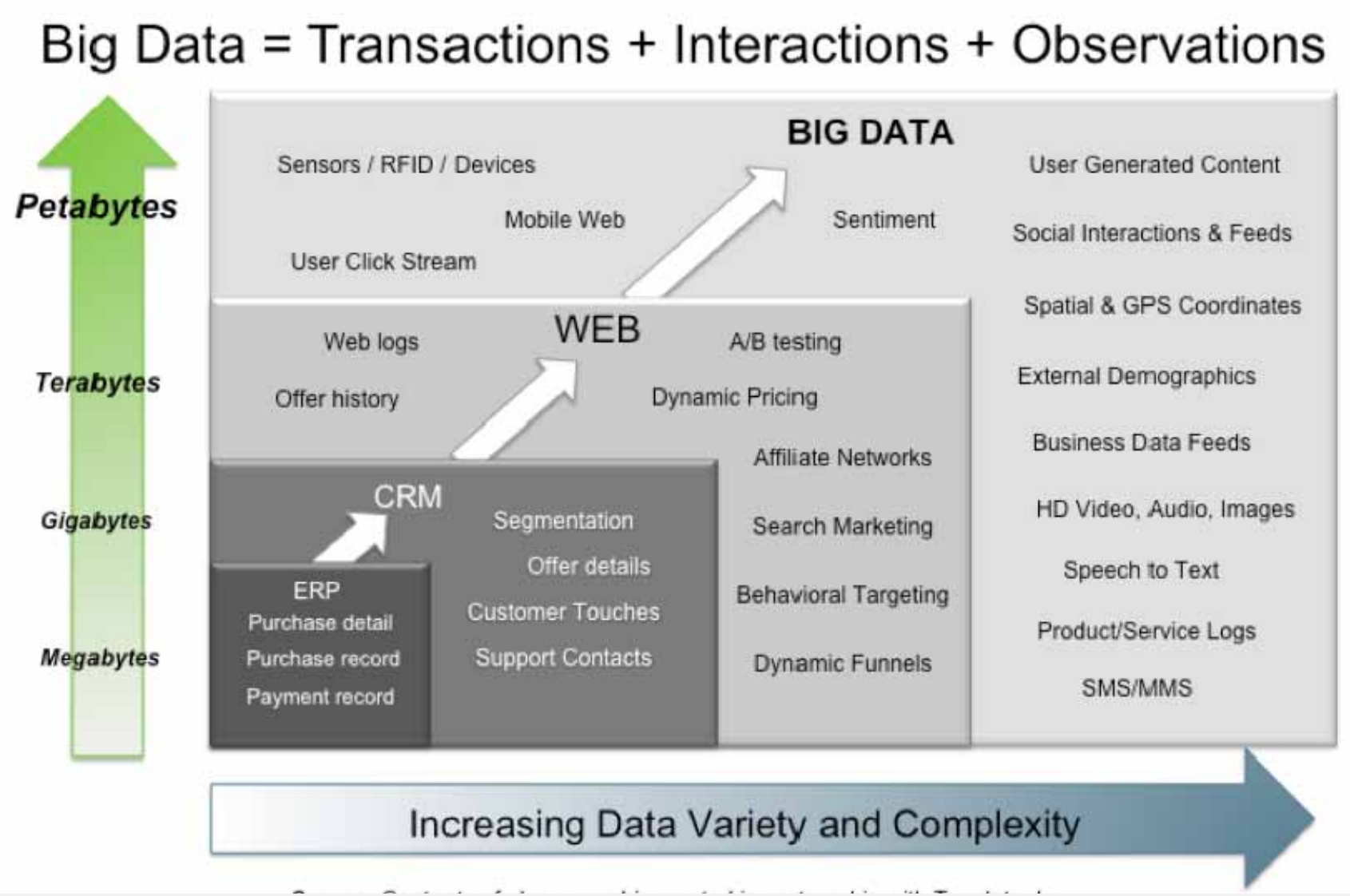
BY NC SA

BIG DATA



BY NC SA

BIG DATA



BY NC SA

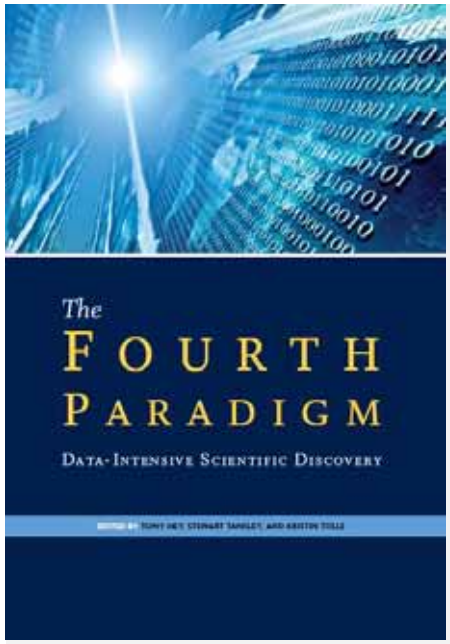
BIG DATA

Big data is a collection of data sets large and complex that it becomes difficult to process using on-hand database management tools. The challenges include capture, storage, search, sharing, analysis, and visualization.

Data sets grow in size in part because they are increasingly being gathered by ubiquitous information-sensing mobile devices

Big data is difficult to work with using relational databases and desktop statistics and visualization packages, requiring instead "massively parallel software running on tens, hundreds, or even thousands of servers“.

What is considered "big data" varies depending on the capabilities of the organization managing the set. "For some organizations, facing hundreds of gigabytes of data for the first time may trigger a need to reconsider data management options. For others, it may take tens or hundreds of terabytes before data size becomes a significant consideration.



FOURTH PARADIGM

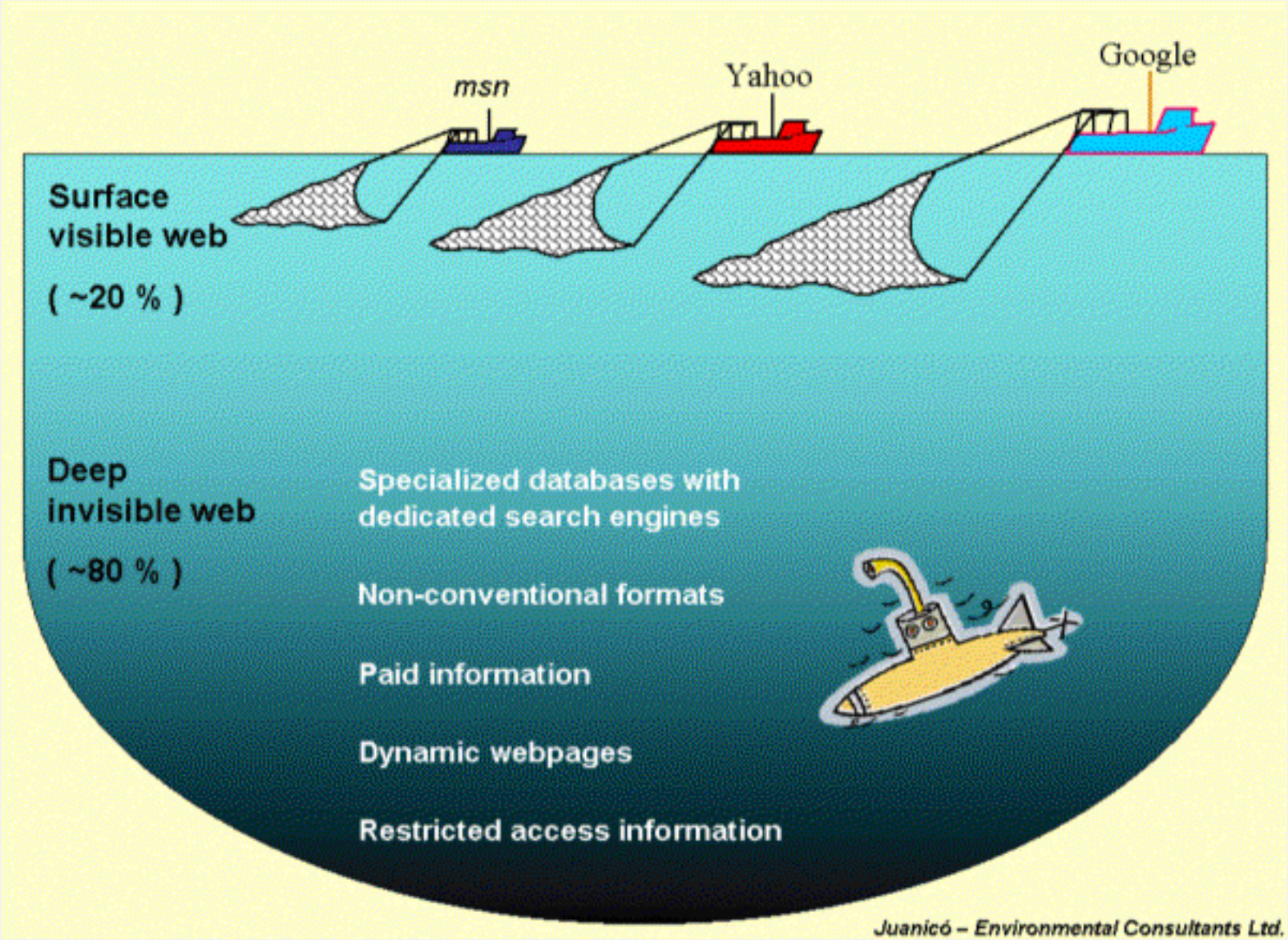
"*Science is becoming increasingly dependent on data*."

Gathering data is so easy and quick that it exceeds our capacity to validate, analyze, visualize, store, and curate the information.

Most consider how research practices will be transformed by computational thinking—scientific instruments powered by computers in ways that make them a universal intellectual amplifier. **In what if any sense will computational thinking transform the process of scientific discovery?** Will data-intensive science cause an incremental change in how science is done? Or will the changes be a true step function?
"... a world of scholarly resources - text, databases, and any other associated materials - that were seamlessly navigable and interoperable."

BY NC SA

DEEP WEB

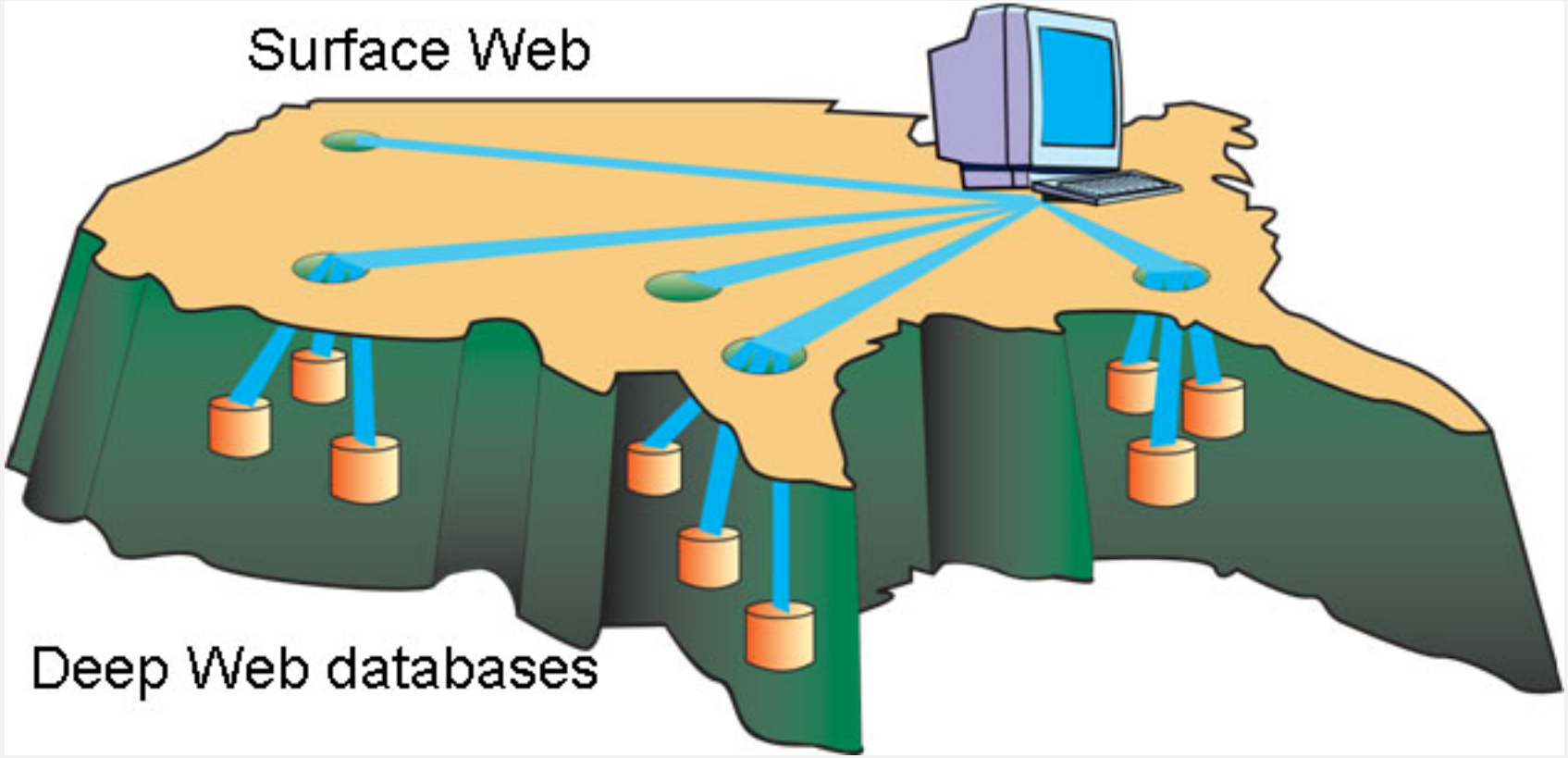


BY NC SA

DEEP WEB

Il 99% dei documenti scientifici è nascosto in Database

Il Deep-WEB non è accessibile dai motori di ricerca



BY NC SA

Dal WEB 1.0 al WEB Semantico

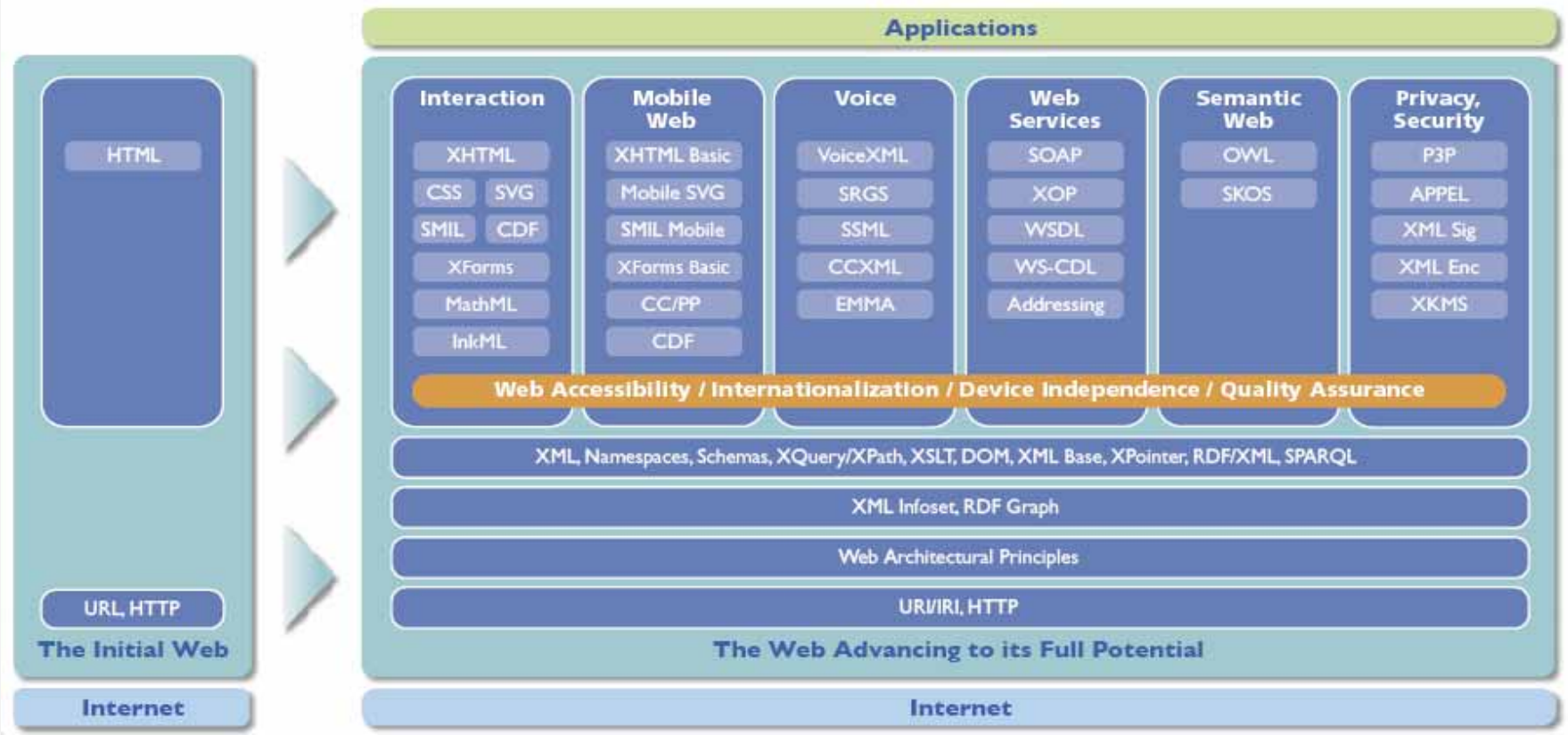
✓Il Web 1.0 era un sistema di **documenti interconnessi (interlinked) accessibili attraverso Internet.** I **Web browsers usavano il protocollo HTTP per comunicare con Web Servers.** L'utente accedeva a pagine **HTML e usava gli hyperlinks per navigare tra le pagine.** L'utente poteva così accedere a qualsiasi **documento raggiungibile sul Web**

✓Nel Web 2.0 gli utenti **condividono (share) foto, video e pensieri, interagiscono sui social networks, pubblicano contenuti nei blogs, usano tags per dare significato ai documenti.** L'utente ha piena libertà di interagire con i siti Web, che diventano dinamici

✓ Se una macchina potesse comprendere il significato dell'informazione, potrebbe aiutare l'utente a trovare ciò di cui ha veramente bisogno. Mentre il Web attuale si basa su documenti interconnessi, il **Web semantico** sarà basato su **oggetti e sulle relazioni tra di essi**

BY NC SA

Dal WEB 1.0 al WEB Semantico

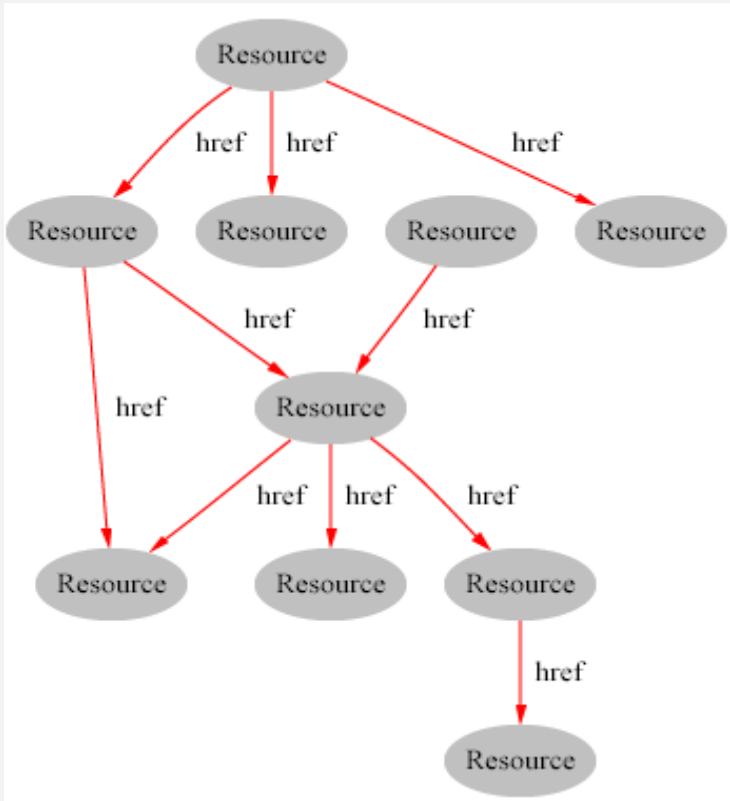


WEB
di Documenti

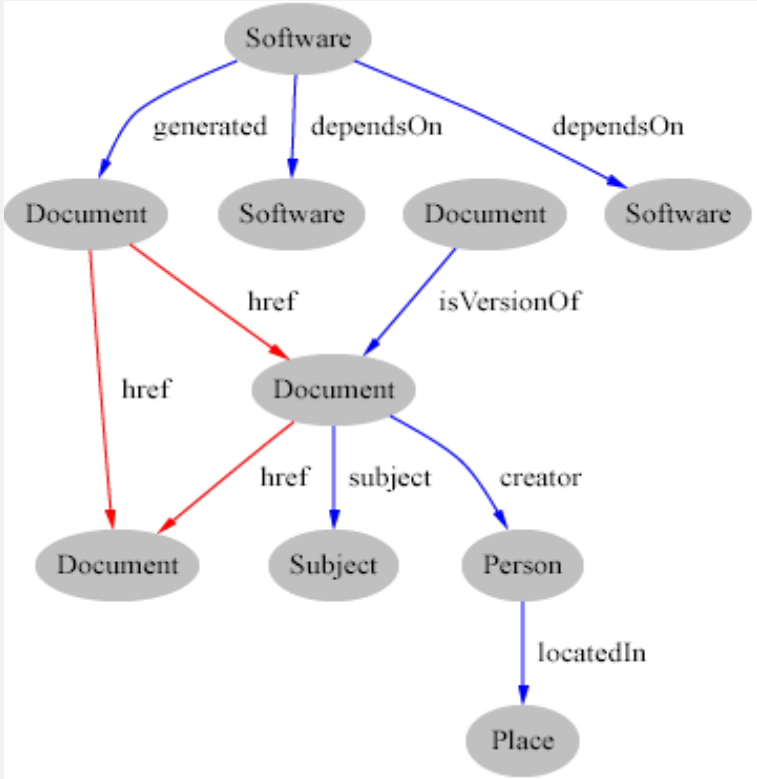
WEB
di Dati

BY NC SA

Dal WEB 1.0 al WEB Semantico



Human understandable but *"only"* machine-readable



Human and machine *"understandable"*

BY NC SA

Il WEB Semantico

Domande per il Web Semantico

✓Quali sono gli archeologi che hanno studiato all'Università di Roma e hanno iniziato la loro carriera come Funzionari del MIBAC?

✓Quali Funzionari hanno iniziato la loro carriera nello stesso anno di …… ?

✓Quali Funzionari sono andati all'Università di Roma e sono stati colleghi di …… ?

… Qualsiasi tipo di domanda implica una diversa formalizzazione dei dati. **Google è insufficiente.**

BY NC SA

Il WEB Semantico

Google

Quali sono gli archeologi che hanno studiato all'Università di Roma e hanno ini

Ricerca

Circa 28.800 risultati (0,66 secondi)

Web
Immagini
Maps
Video
Notizie
Shopping
Più contenuti

Nel Web
Pagine in italiano
Pagine da: Italia
Pagine straniere tradotte

Qualsiasi data
Ultima ora
Ultime 24 ore
Ultima settimana
Ultimo mese
Ultimo anno
Intervallo di date

Tutti i risultati

Pagine in italiano

[PDF] 323 VED EMILIA
www.ilgiornaledellarte.com/immagini/IMG20120831162926635.pdf
Formato file: PDF/Adobe Acrobat - Visualizzazione rapida
31 ago 2012 – Inaugurato il 2 ottobre **del** 1841, il Teatro **è** stato intitolato al grande tenore modenese **...** re Classiche **all'Università di** Bologna si **è** specializzato in **...** Successivamente **ha** continuato la **carriera** accademica a Ve- **... che sono** anche grandi collezionisti, **come** Gaetano Maccafer- **.....** dei **funzionari del Mibac**.

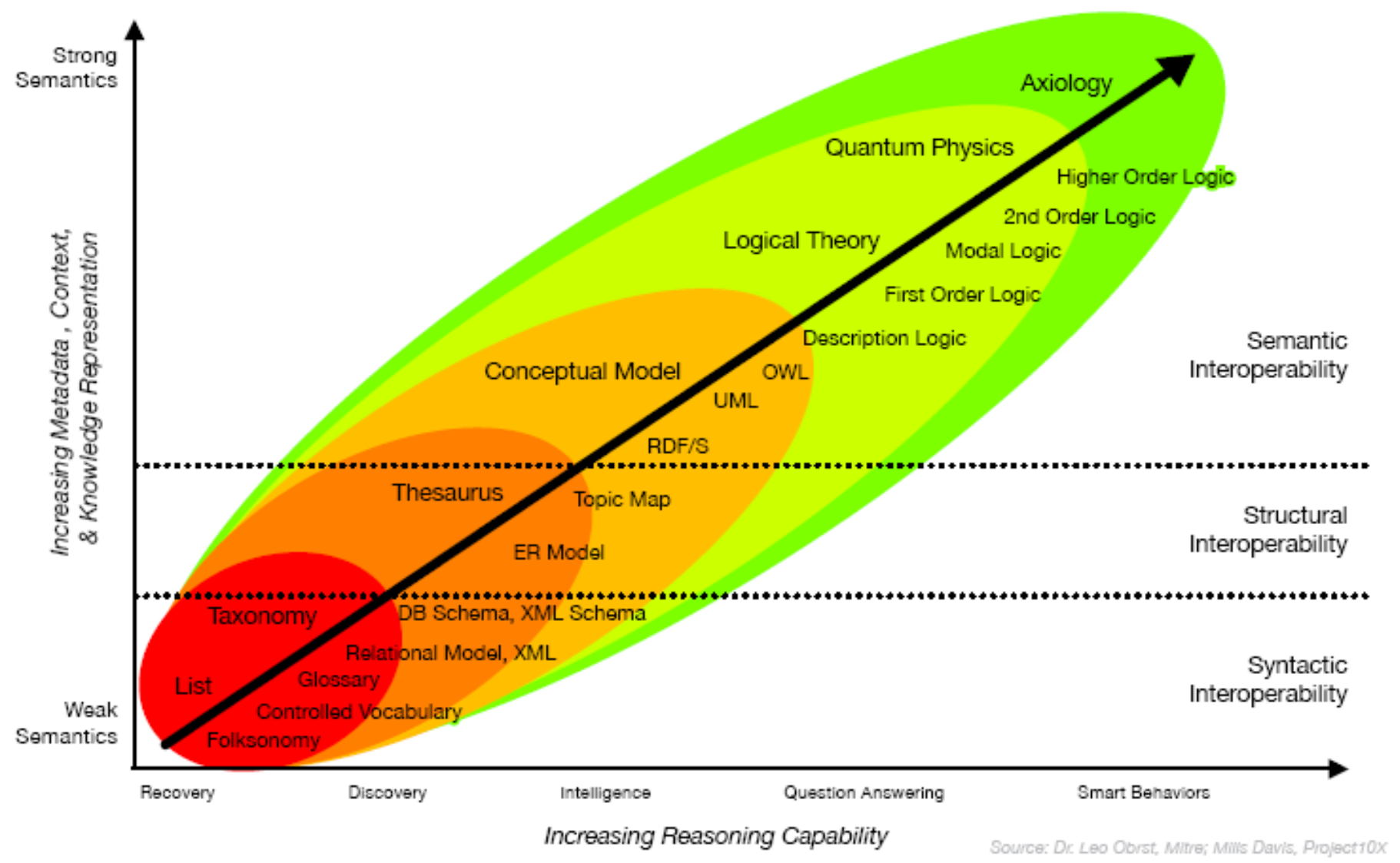
[PDF] Museo Nazionale Preistorico Etnografico "Luigi Pigorini"
www.pigorini.beniculturali.it/doc/RIME_IT.pdf
Formato file: PDF/Adobe Acrobat - Visualizzazione rapida
no anche i musei etnografici a rinnovare la **loro** missione propo- **...** parte **di** una crisi più generale **che** investe tutti i musei, i **quali**, mentre **sono ...** ta **all**'attuale congiuntura economica sfavorevole **e** alla mancanza **di ... come** protagonista della vita **del** museo **e hanno** imposto una tra- **..... studiato archeologia e** storia a Parigi.

[PDF] FRAMMENTI DEL PASSATO TESORI DALL'AGER TIBURTINUS
www.beniculturali.it/mibac/multimedia/MiBAC/.../1240912583379_...
Formato file: PDF/Adobe Acrobat - Visualizzazione rapida
largo seguito **del** culto **di** Hercules Victor, **...** quotidianamente per catalogare **e studiare ... come** un sito chiave **del** Lazio, vitale asse **di ... all**'Urbe. Naturalmente, la presenza della villa dell'imperatore Adriano non farà **che ...** monumenti erano espressione nel **loro** tempo **....** commerciali **e** culturali **che avevano** in **Roma** lo **...**

[SIAENews]
www.siae.it/Edicola.asp
La mostra dedicata **all**'attrice **e** regista Tatiana Pavlova, **che** si tiene a **Roma**, presso il Museo Siae **...** La lunga **carriera** artistica **di** Tatiana Pavlova (1890-1975) **come** attrice, regista **e** docente **...** Nei lidi **di** Salerno **e** Pontecagnano i militari dell'Arma dei Carabinieri **hanno** condotto una vasta operazione con **funzionari del** .

CC BY NC SA

Il WEB Semantico

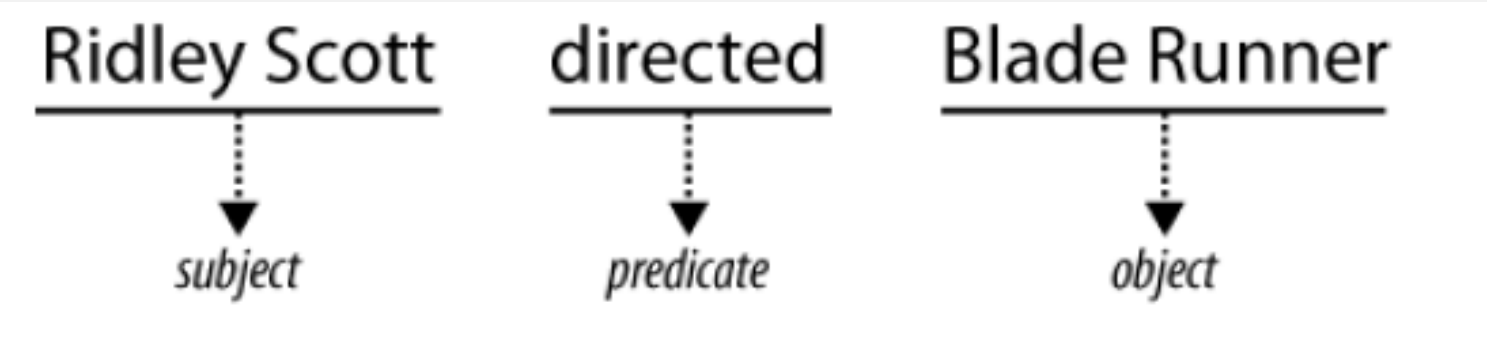


BY NC SA

Il WEB Semantico

Bisogna esporre i dati in un linguaggio che ne evidenzi la semantica disambiguando i significati e che sia comprensibile dalle macchine.

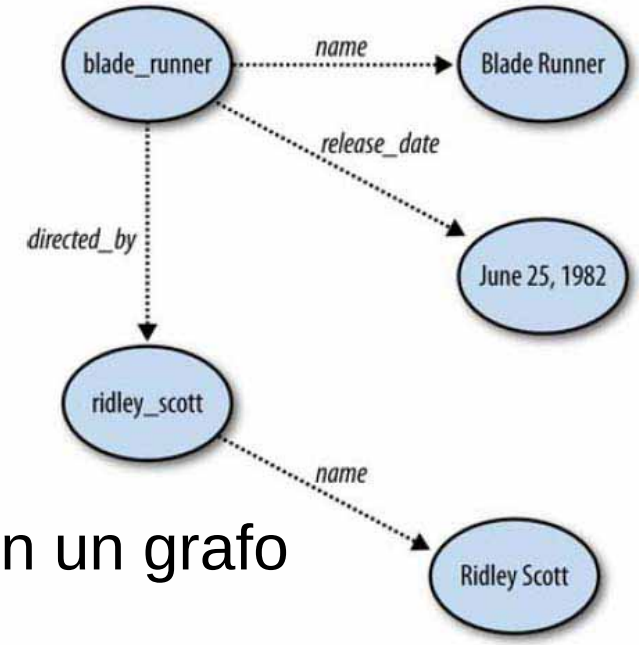
La tripla è il perno di una rappresentazione semantica. Essa è composta da un soggetto, un predicato e un oggetto (asserzioni della logica dei predicati)



BY NC SA

Il WEB Semantico

Il **soggetto** corrisponde in genere ad un concetto univoco, o entità, che può essere una persona, un luogo, un'idea.
I **predicati** rappresentano proprietà dell'entità a cui sono collegati
Gli **oggetti** possono essere: Soggetti di altre triple
Valori letterali, come stringhe o numeri



La **tripla** si rappresenta con un grafo

BY NC SA

LINKED DATA

4 regole per la creazione di LD sul WEB

✓Usare **URI (Universal Resource Identifier**) per identificare ogni risorsa (concetto/documento). pleiades.stoa.org/places/579885

✓Usare **HTTP URI** per individuare la risorsa sulla rete: http:\\pleiades.stoa.org/places/579885

✓Fornire informazioni utili sull'oggetto usando formati standard (es. **RDF**)

✓Includere **link** ad altri URI. Le interconnessioni rendono più ricchi e visibili i nostri dati

BY NC SA

LINKED DATA

Un altro requisito fondamentale è rappresentato dall'accessibilità dei dati che devono essere aperti e connessi in modo semplice tramite ad esempio RDF (Resource Description Framework) che è un modello di **rappresentazione dei dati** e non un linguaggio

BY NC SA

WEB Semantico + LD

Comprende asserzioni (o frasi) collegate tramite link qualificati e autoparlanti che diventano una collezione di conoscenza leggibile da una macchina.

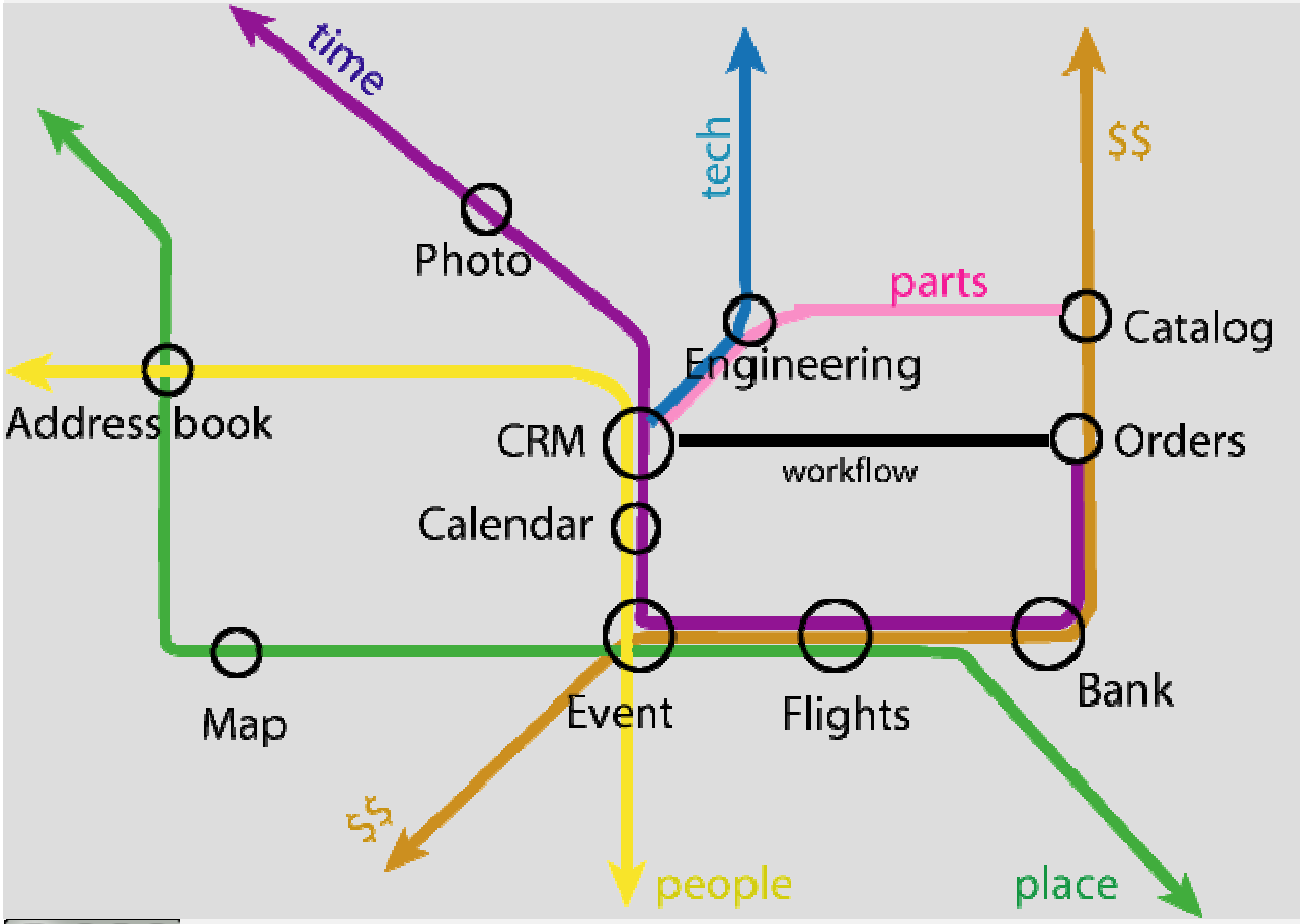
Il modello dati è connesso col modello dei **database relazionali** poiché il perno dei linked data sono i "***predicati***" che esprimono tipi di relazione.

Le entità, rappresentate in strutture, sono articolate in un insieme di dati, ciascuno in sé autonomo, combinabile logicamente con altri dati per produrre nuove entità.

BY NC SA

WEB Semantico + LD

Tim Berners-Lee ha proposto l'analogia con una **rete di metropolitane**



Le **ontologie** sono le **linee** che connettono tra loro le **applicazioni** (le **stazioni**).

I **metadati** sono i **convogli** che trasportano **l'informazione** (i **passeggeri**) tra una stazione e l'altra.

cc BY NC SA

CIDOC-CRM

CIDOC-ICOM **recommendation** on Linked Open Data for museums –Draft

When museum objects are referred to in Internet applications, it is necessary that the objects are uniquely identified by suitable **URIs**

This URI should be derived in a simple way from the **inventory numbers** published in exhibition catalogues, through on-line museum catalogue access or by asking museum staff. This will avoid the need go through a matching process that could generate errors.

http://www.cidoc-crm.org/URIs_and_Linked_Open_Data.html

BY NC SA

CIDOC-CRM + LD (1)

Per essere parte del WEB di data non si deve lavorare sulla progettazione di interfacce, ma pubblicare i propri dati in base ai principi del LD adoperando strutture formali e standard (modelli ontologici e metadati) che semplifichino il raggiungimento della interoperabilità

```
<CRMset>
  <CRM_Entity>Epitaphios GE34604
    <in_class>E23: Information Carrier</in_class>
    <is_identified_by>GE 34604
      <in_class>E42: Object Identifier</in_class>
    </is_identified_by>
    <is_documented_in>
        http://cidoc.ics.forth.gr/graphics/epit1.jpg
      <in_class>E31: Document</in_class>
    </is_documented_in>
    <current_owner>Museum Benaki !!!
      <in_class>E40: Legal Body</in_class>
    </current_owner>
  </CRM_Entity>
</CRMset>
```

BY NC SA

CIDOC-CRM + LD (1)

```
<CRMset>
  <CRM_Entity>Epitaphios GE34604
    <in_class>E23: Information Carrier</in_class>
    <is_identified_by>GE 34604
      <in_class>E42: Object Identifier</in_class>
    </is_identified_by>
    <is_documented_in>
        http://cidoc.ics.forth.gr/graphics/epit1.jpg
      <in_class>E31: Document</in_class>
    </is_documented_in>
    <current_owner>Museum Benaki !!!
      <in_class>E40: Legal Body</in_class>
    </current_owner>
  </CRM_Entity>
</CRMset>
```

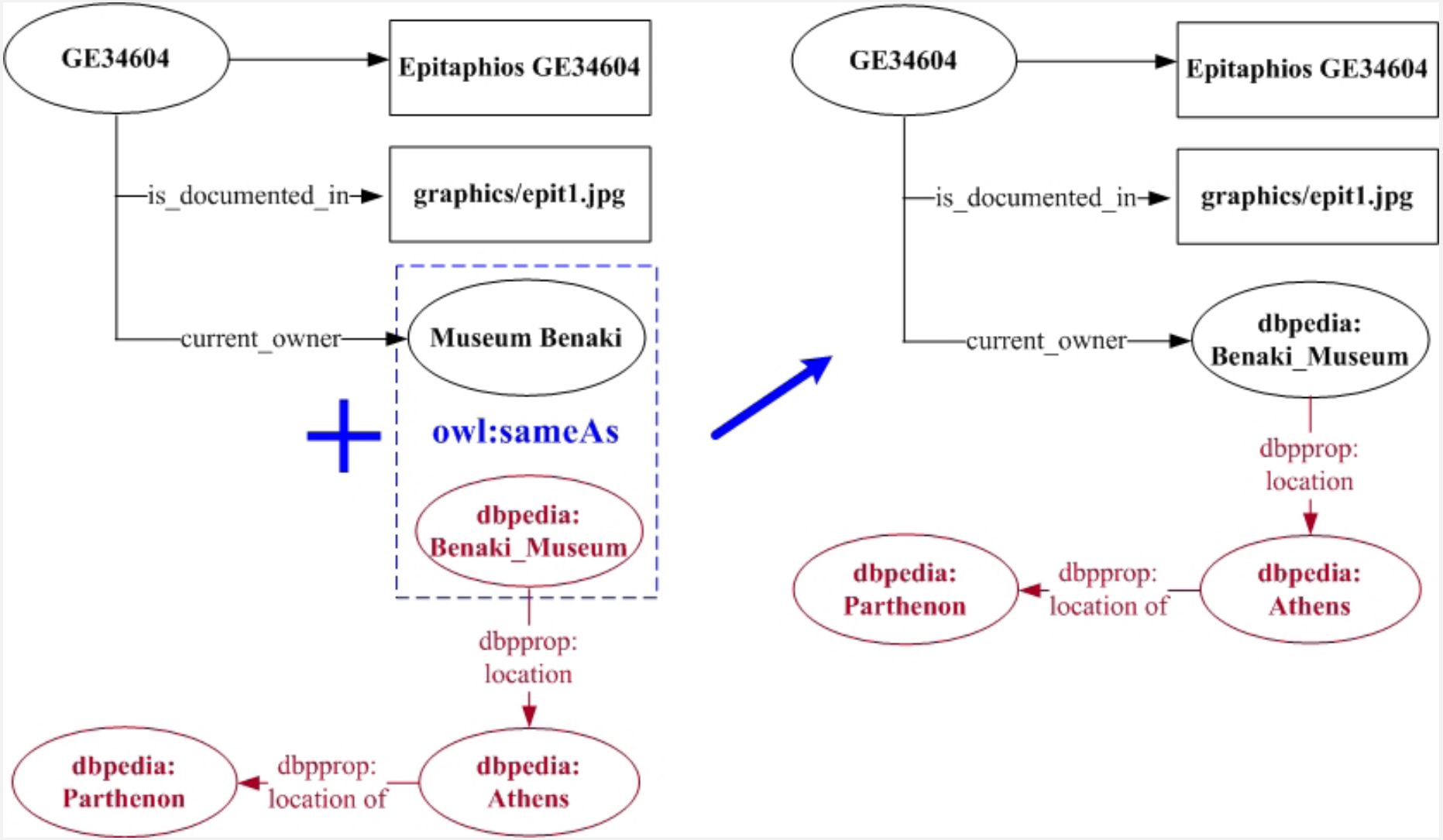
<http://CHInstitution.org/MuseumBenaki>

owl:sameAs

<http://dbpedia.org/resource/Benaki_Museum>

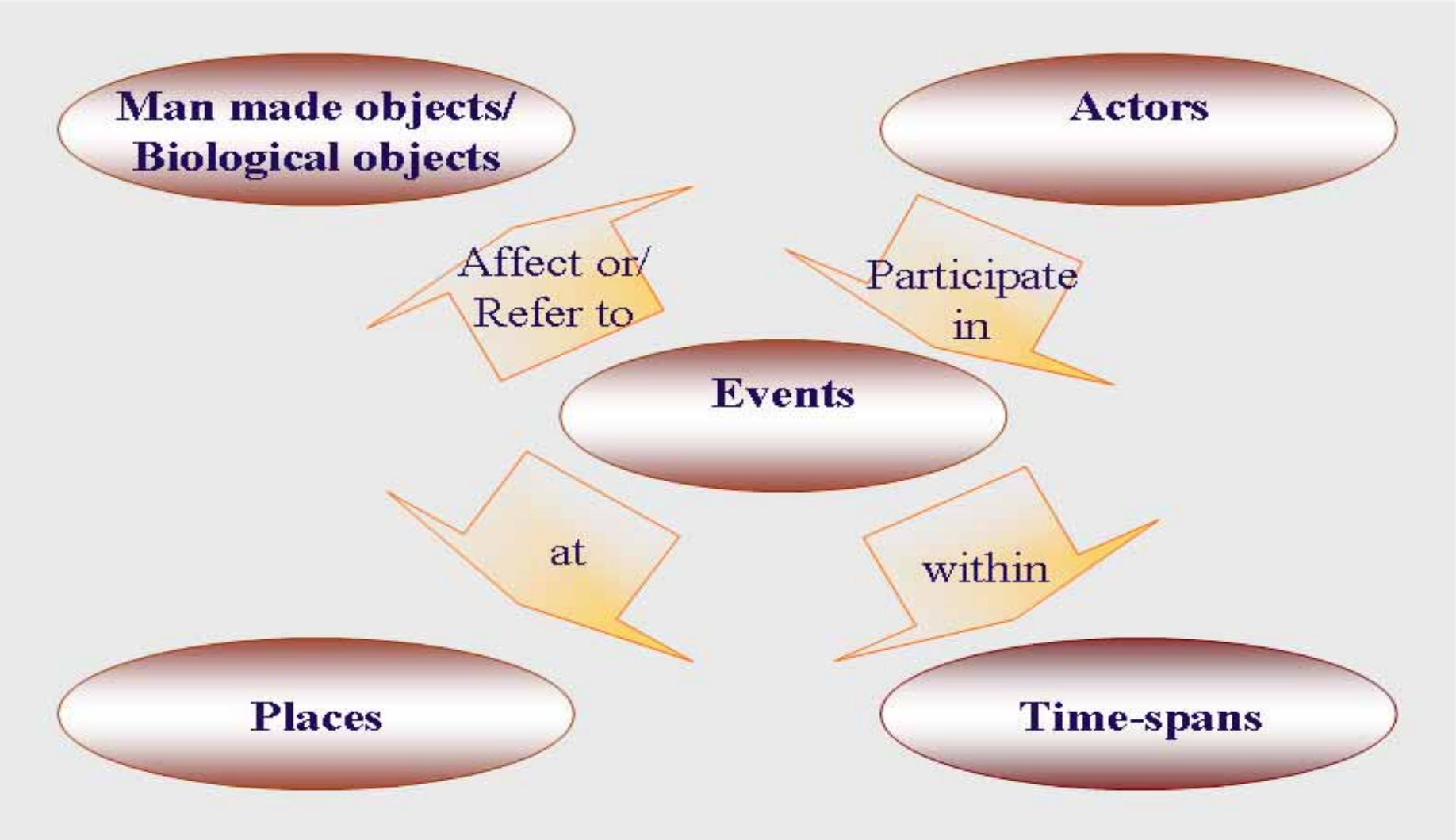
BY NC SA

CIDOC-CRM + LD (1)



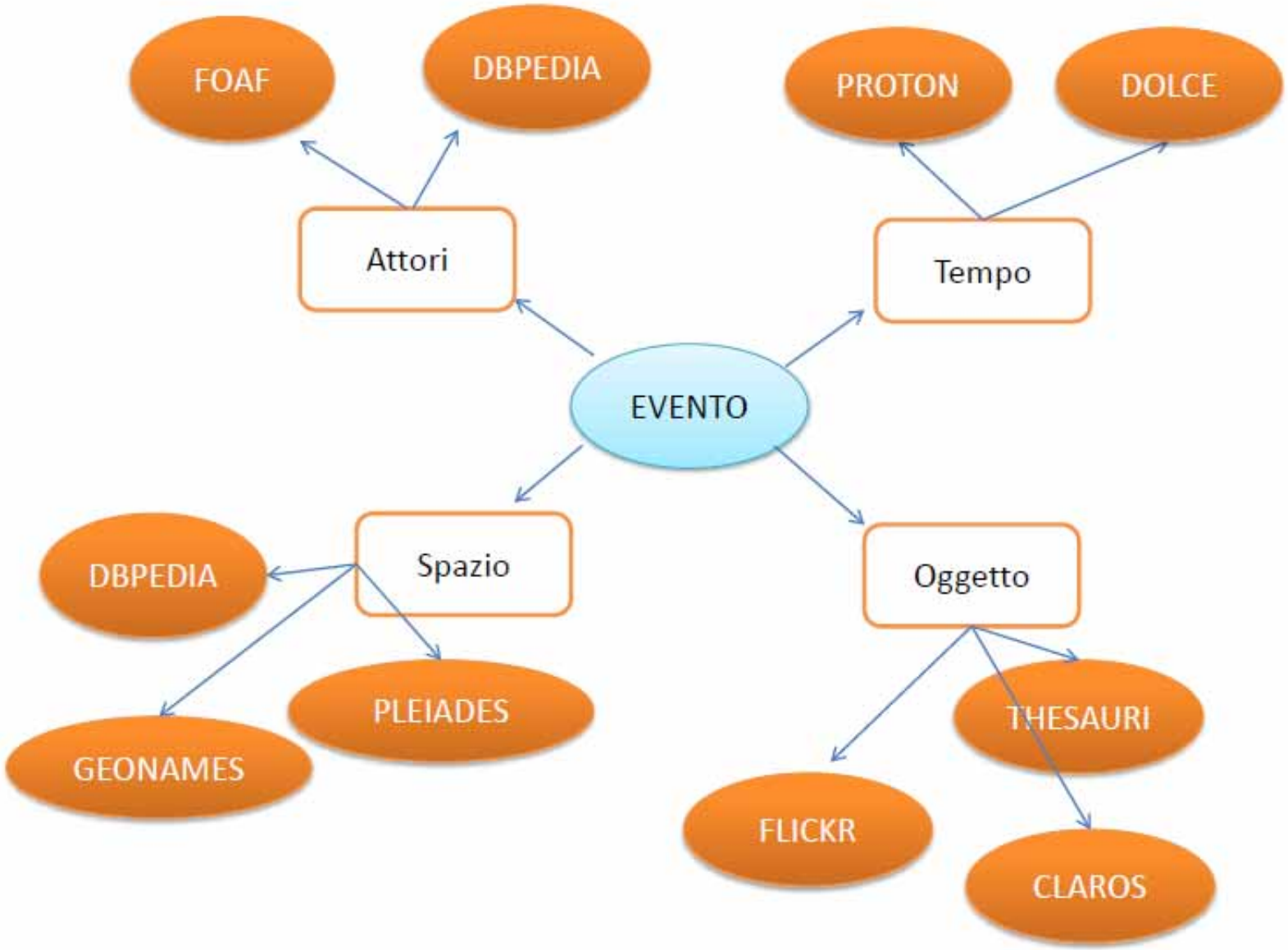
BY NC SA

CIDOC-CRM + LD (2)



BY NC SA

CIDOC-CRM + LD (2)



BY NC SA

CIDOC-CRM + LD (3)

```
<http://lod.ac/ref/18731>
<http://lod.ac/ns/lodac#exhibitionHistory> "個展(東京、東京画廊1969)";
<http://lod.ac/ns/lodac#genre> "Prints:", "版画:";
<http://purl.org/NET/cidoc-crm/core#P62I_is_depicted_by> "右下に署名(刷)";
<http://purl.org/dc/elements/1.1/creator> "横尾忠則";
<http://purl.org/dc/terms/created> "1969", "昭和44";
<http://purl.org/dc/terms/extent> "90.0×90.0", "on paper, acrylic films and
acrylic sheet90.0×90.0";
<http://purl.org/dc/terms/identifier> "P01847";
<http://purl.org/dc/terms/isReferencedBy> <http://lod.ac/id/18731>;
<http://purl.org/dc/terms/medium> "silkscreen", "シルクスクリーン・紙、アクリルフィルム、アクリル
板・1";
<http://purl.org/dc/terms/provenance> "平成17年度購入P01847";
<http://purl.org/dc/terms/source> <http://search.artmuseums.go.jp>;
<http://purl.org/dc/terms/title> "Landscape No.1 Girl","風景 No.1 女の子";
a <http://lod.ac/ns/lodac#WorkReference>;
<http://www.w3.org/2004/02/skos/core#prefLabel> "Landscape No.1 Girl", "風景 No.1
女の子" .
```

```
<http://lod.ac/id/18731>
<http://purl.org/NET/cidoc-crm/core#P55_has_current_location> <http://lod.ac
/id/912>;
<http://purl.org/dc/terms/creator> <http://lod.ac/id/874>;
<http://purl.org/dc/terms/references> <http://lod.ac/ref/18731>;
<http://purl.org/dc/terms/title> "Landscape No.1 Girl","風景 No.1 女の子";
a <http://lod.ac/ns/lodac#Work>;
<http://www.w3.org/2004/02/skos/core#prefLabel> "Landscape No.1 Girl",
"風景 No.1 女の子" .
```

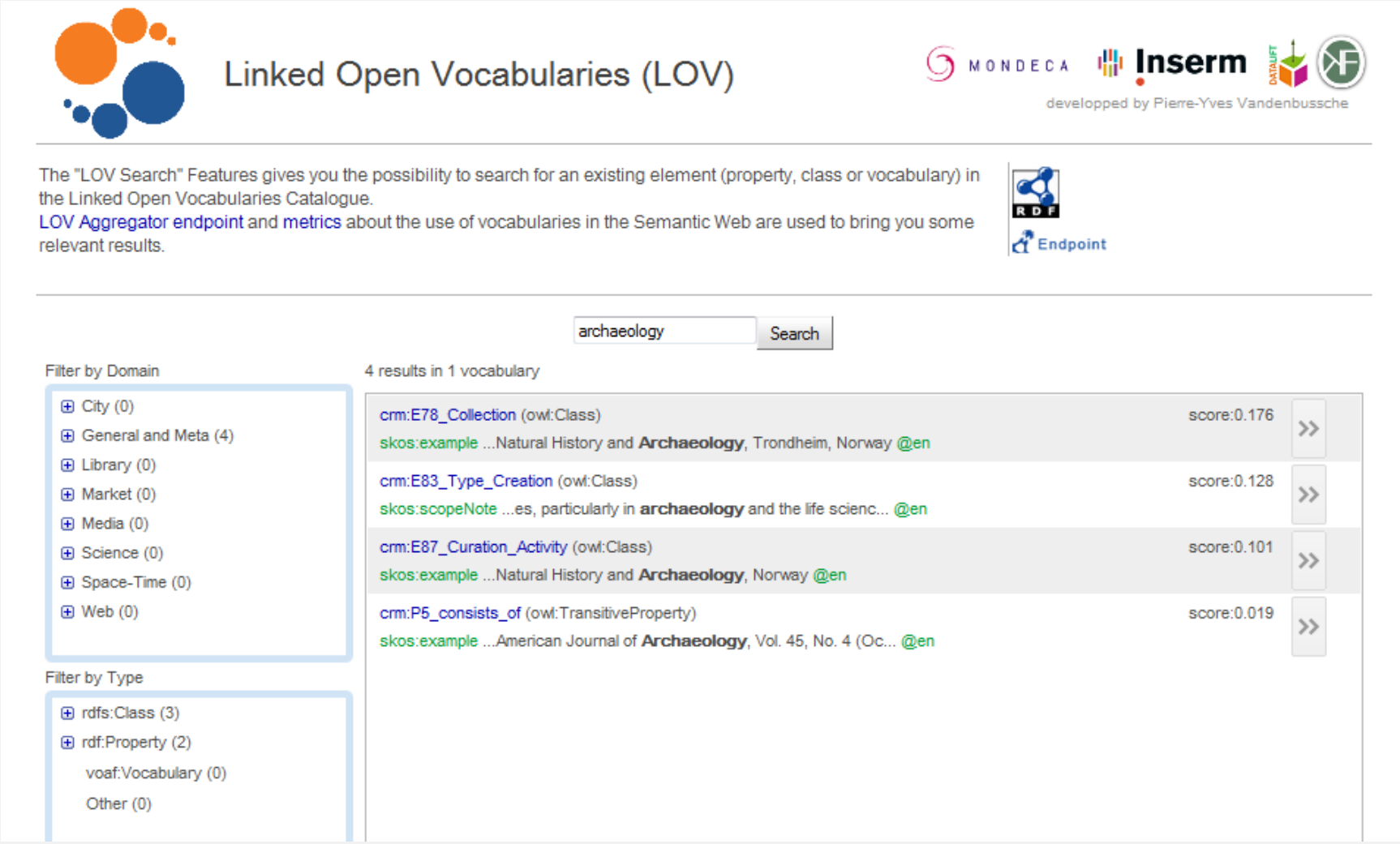
BY NC SA

CIDOC-CRM + LD (3)

BY NC SA

CIDOC-CRM + LD

http://lov.okfn.org/dataset/lov/search/#s=archaeology



BY NC SA

CIDOC-CRM + LD (4)

ARCHAEOLOGY DATA SERVICE

http://data.archaeologydataservice.ac.uk/

Archaeology Data Service Linked Data

These Linked Data are made available as a direct result of the STELLAR project, a joint project between the University of Glamorgan, the ADS and English Heritage. Linked Data is a growing part of the Web where data is published on the Web and then linked to other published data in much the same way that web pages are interlinked using hypertext. The term Linked Data is used to describe a method of exposing, sharing, and connecting data via URIs on the Web. For more information go to http://linkeddata.org/. The ADS cannot give advice or support on using RDF, SPARQL or SPARQL Endpoints. For a full description of the ADS datasets used in this initial evaluation version of an ADS Linked Data store please see the ADS STELLAR Research Page. The STELLAR tools and templates are available from the University of Glamorgan.

Property	Value
?:coverage	▪ <http://data.ordnancesurvey.co.uk/id/country/england>
?:description	▪ The Archaeology Data Service (ADS) supports research, learning and teaching with high quality and dependable digital resources. It does this by preserving ... »more»
?:exampleResource	▪ <http://data.archaeologydataservice.ac.uk/10.5284/1000230/cobham/> ▪ <http://data.archaeologydataservice.ac.uk/10.5284/1000230/cuxton/> ▪ <http://data.archaeologydataservice.ac.uk/10.5284/1000230/eyhorne/> ▪ <http://data.archaeologydataservice.ac.uk/10.5284/1000230/saltwood/> ▪ <http://data.archaeologydataservice.ac.uk/10.5284/1000230/sittingbourne/> ▪ <http://data.archaeologydataservice.ac.uk/10.5284/1000230/thurnham/> ▪ <http://data.archaeologydataservice.ac.uk/10.5284/1000230/tutt/> ▪ <http://data.archaeologydataservice.ac.uk/10.5284/1000365/> ▪ <http://data.archaeologydataservice.ac.uk/10.5284/1000389/> ▪ <http://data.archaeologydataservice.ac.uk/10.5284/1000392/>
?:issued	▪ 2011-06-30
?:label	▪ Archaeology Data Service Linked Data
?:license	▪ Subject to ADS Terms of Use and Access
?:publisher	▪ Archaeology Data Service
?:rights	▪ All rights reserved, copyright retained by original depositors
?:title	▪ Archaeology Data Service Linked Data
?:uriRegexPattern	▪ http://data.archaeologydataservice.ac.uk/.+

cc BY NC SA

CIDOC-CRM + LD (4)

ARCHAEOLOGY DATA SERVICE

SEARCH

HOME | ARCHSEARCH | ARCHIVES | LEARNING | ADVICE | OUR RESEARCH | BLOG | ABOUT US | LOGIN

ADS Linked Data Information Page

An example query for the ADS Linked Data Repository in URL format:

This query will return results from the SPARQL endpoint in XML format.
Any query can be sent to the endpoint if it is converted to URL safe encoding as below:

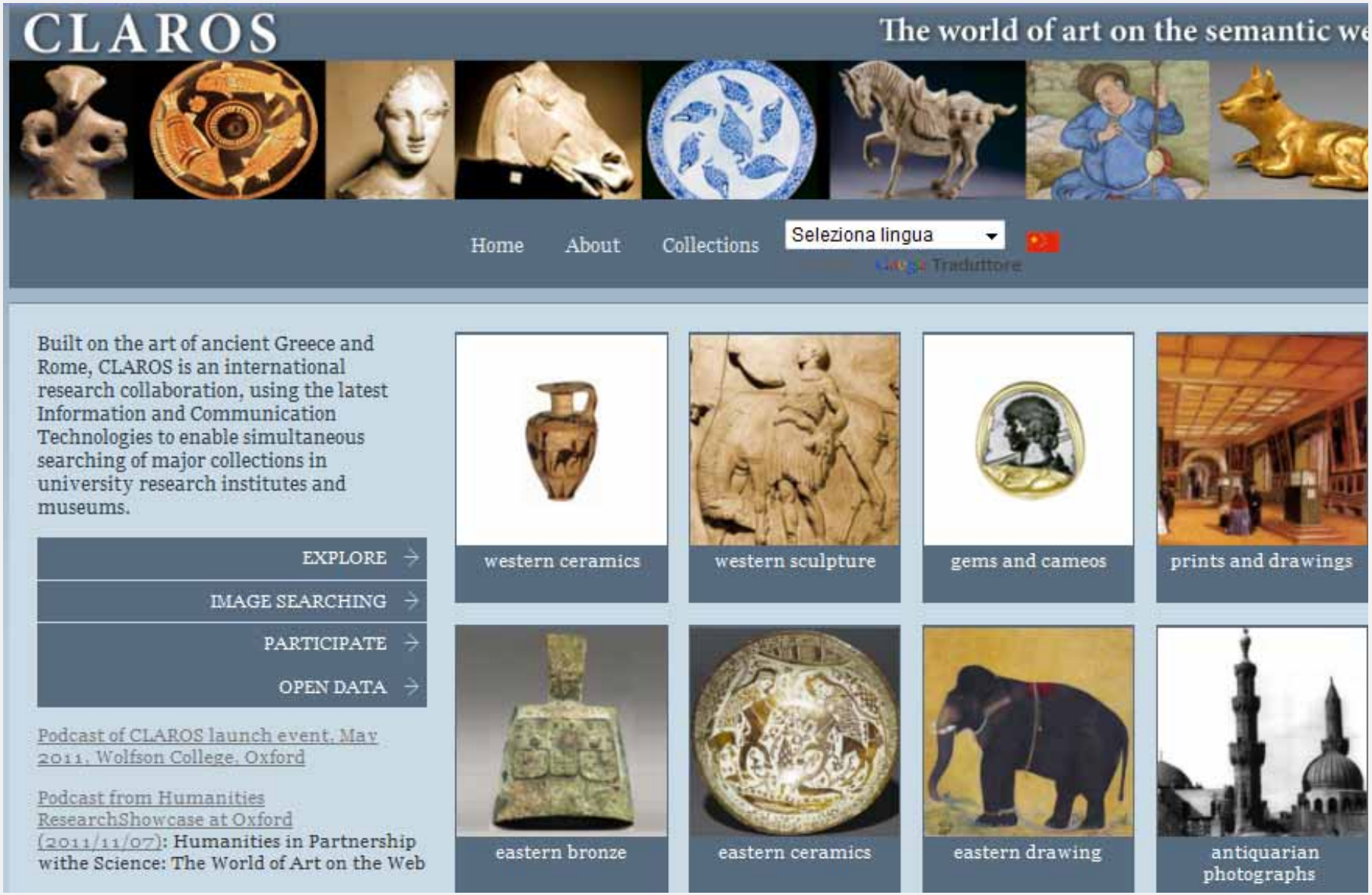
This example query returns the originating context of all 'glass' finds:
Here is the SPARQL notation:

```
PREFIX crm:<http://erlangen-crm.org/101001/>
PREFIX crmeh:<http://purl.org/crmeh#>
SELECT DISTINCT ?contextType ?findType
WHERE {
  ?ehe0007 a crmeh:EHE0007_Context;
     crm:P2_has_type [rdfs:label ?contextType];
     crm:P26i_was_destination_of [crm:P25_moved ?ehe0009] .
  ?ehe0009 a crmeh:EHE0009_ContextFind ;
     crm:P2_has_type [rdfs:label ?findType];
     crm:P45_consists_of [rdfs:label ?material].
   FILTER(REGEX(?material, "(?i)glass")) .
} LIMIT 100
```

Here is the URL link for this query: URL

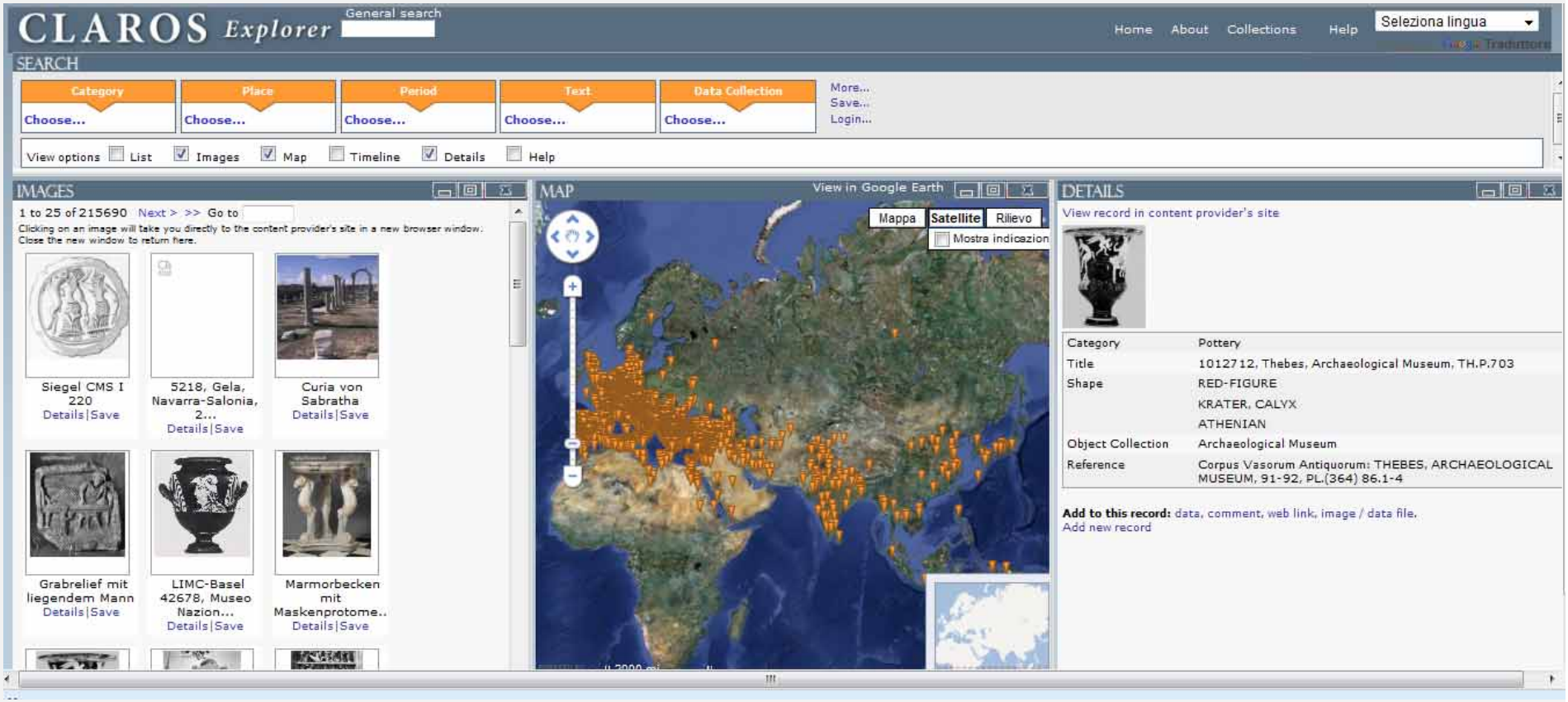
CC BY NC SA

CIDOC-CRM + LD (5)



BY NC SA

CIDOC-CRM + LD (5)



BY NC SA

CIDOC-CRM + LD (6)

CIDOC–CRM

This site lists the ontology development work carried out during Stage 2 of the ResearchSpace Project by Seme4 Ltd. We present the CIDOC-CRM (RDFS, v5.0.2) with a modified namespace such that it can be served as Linked Data and browsed on this site. In addition a number of extensions have been created to encapsulate concepts and relationships within the BM model that are not natively expressed in the CRM. You can download the cidoc-crm and bm-extensions.

CRM Classes

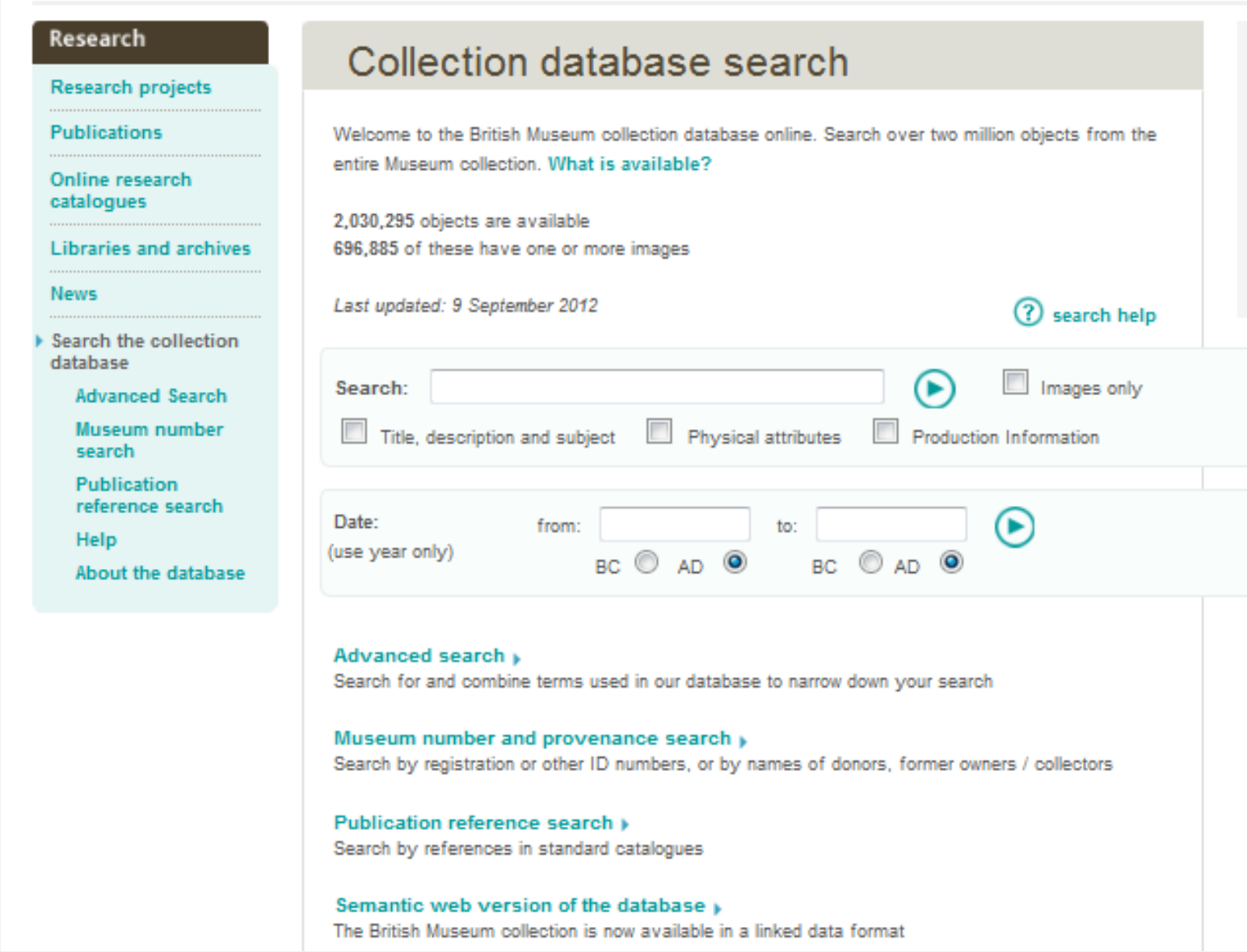
E1.CRM_Entity
E2.Temporal_Entity
E3.Condition_State
E4.Period
E5.Event
E6.Destruction
E7.Activity
E8.Acquisition
E9.Move
E10.Transfer_of_Custody
E11.Modification
E12.Production
E13.Attribute_Assignment
E14.Condition_Assessment
E15.Identifier_Assignment
E16.Measurement
E17.Type_Assignment

CRM Predicates

P1F.is_identified_by
P1B.identifies
P2F.has_type
P2B.is_type_of
P3F.has_note
P4F.has_time-span
P4B.is_time-span_of
P5F.consists_of
P5B.forms_part_of
P7F.took_place_at
P7B.witnessed
P8F.took_place_on_or_within
P8B.witnessed
P9F.consists_of
P9B.forms_part_of
P10F.falls_within
P10B.contains

BY NC SA

CIDOC-CRM + LD (6)



BY NC SA

CIDOC-CRM + LD (6)

The British Museum

Home | Sparql | Help | Licensing | About

British Museum Semantic Web Collection Online

Welcome

Welcome to this Linked Data and SPARQL service. It provides access to the same collection data 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 current version is beta and development work continues to improve the service. We hope that the service will be used by the community to develop friendly web applications that are freely available to the community.

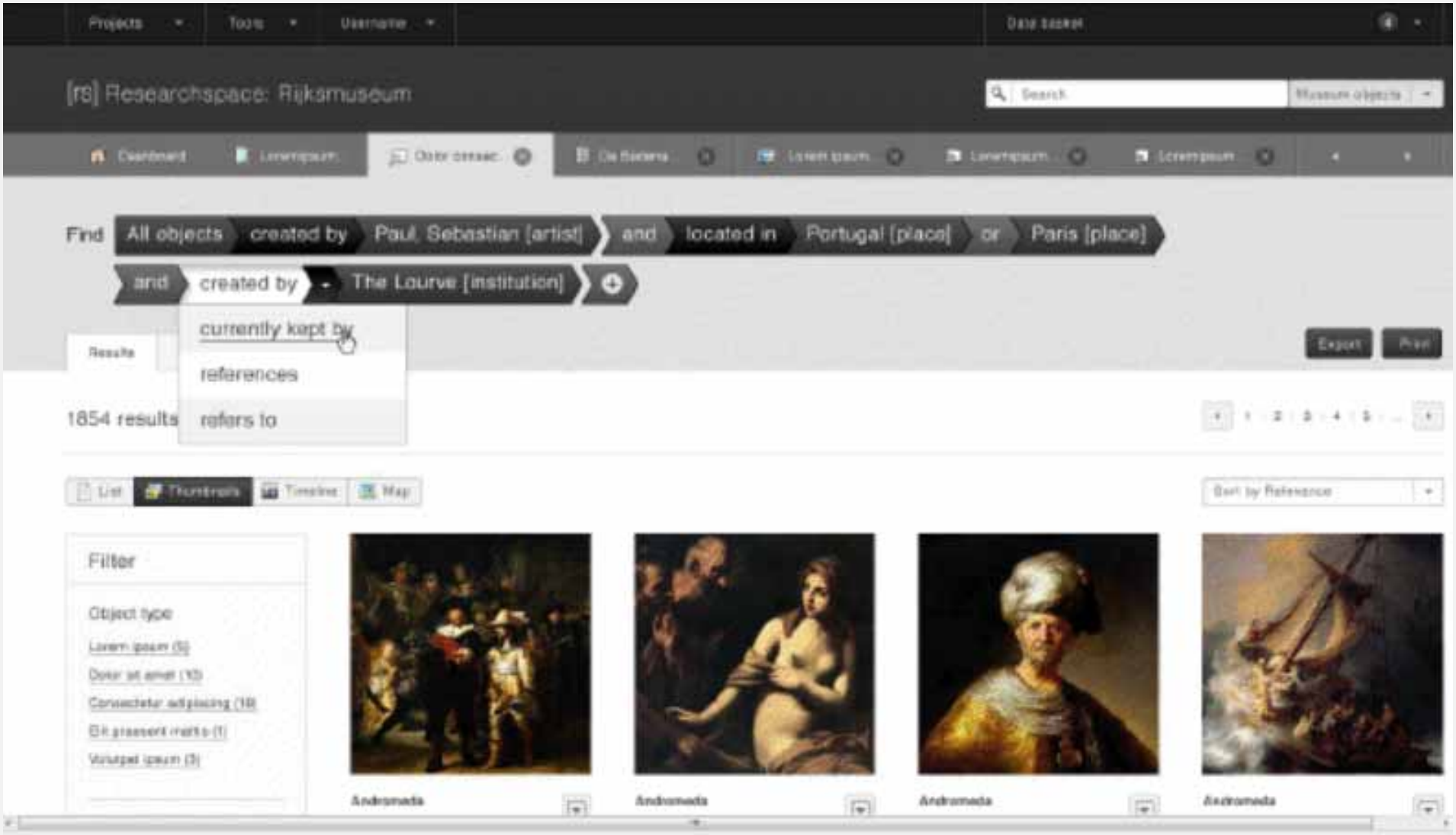
Please use the SPARQL menu item to use the SPARQL user interface or click here.

Date: 5th September 2011

http://collection.britishmuseum.org/description/object/YCA62958.html

CC BY NC SA

CIDOC-CRM + LD



BY NC SA

CIDOC-CRM + LD

annotate.oldman.me.uk/federate.html

researchspace riijksmuseum

Napoli Corriere Mezzogiorno Mattino Sole24ORE PostaUNO Facebook FLC CGIL RegioneCampania L'espresso PostaCertificat@ IlFatto Docente_UNIOR

Semantic Federated Search - Test developed by Yale Center for British Art, the British Museum and Ontotext

This web page retrieves results from two different decentralised databases with a single federated search query. There is no middleware or program code to connect the two data sources. One database contains data provided from the RKD's Rembrandt Project and is located in Sofia, Bulgaria. The other store belongs to the Yale Center for British Art at Yale University (YCBA), New Haven in the United States.This web site is located in Amsterdam. This test is part of work being undertaken by the British Museum and YCBA. Additional RDF stores will be added and functionality developed over time.The main features of the test are;

1. No middleware is used for the search. There is one query searching two separately located data stores. 2. The data is semantically harmonised simply by conforming to the CIDOC-CRM ontology. Both sets of data agree about the meaning of certain concepts. 3. There has been no hardcoded links between the datasets. They have been created separately and no RDF triples have been inserted to assert any direct connections. 4. The search illustrates that both sets of data agree about the meaning of a painting title, even though these concepts may have been implemented differently within the respective organisations. 5. The web page has been tested with Firefox but may work with later versons of Internet Explorer. 6. British Museum data will be included in short course. 7. There are no database optimisations. The stores have only recently be installed.

Dominic Oldman - April 2012

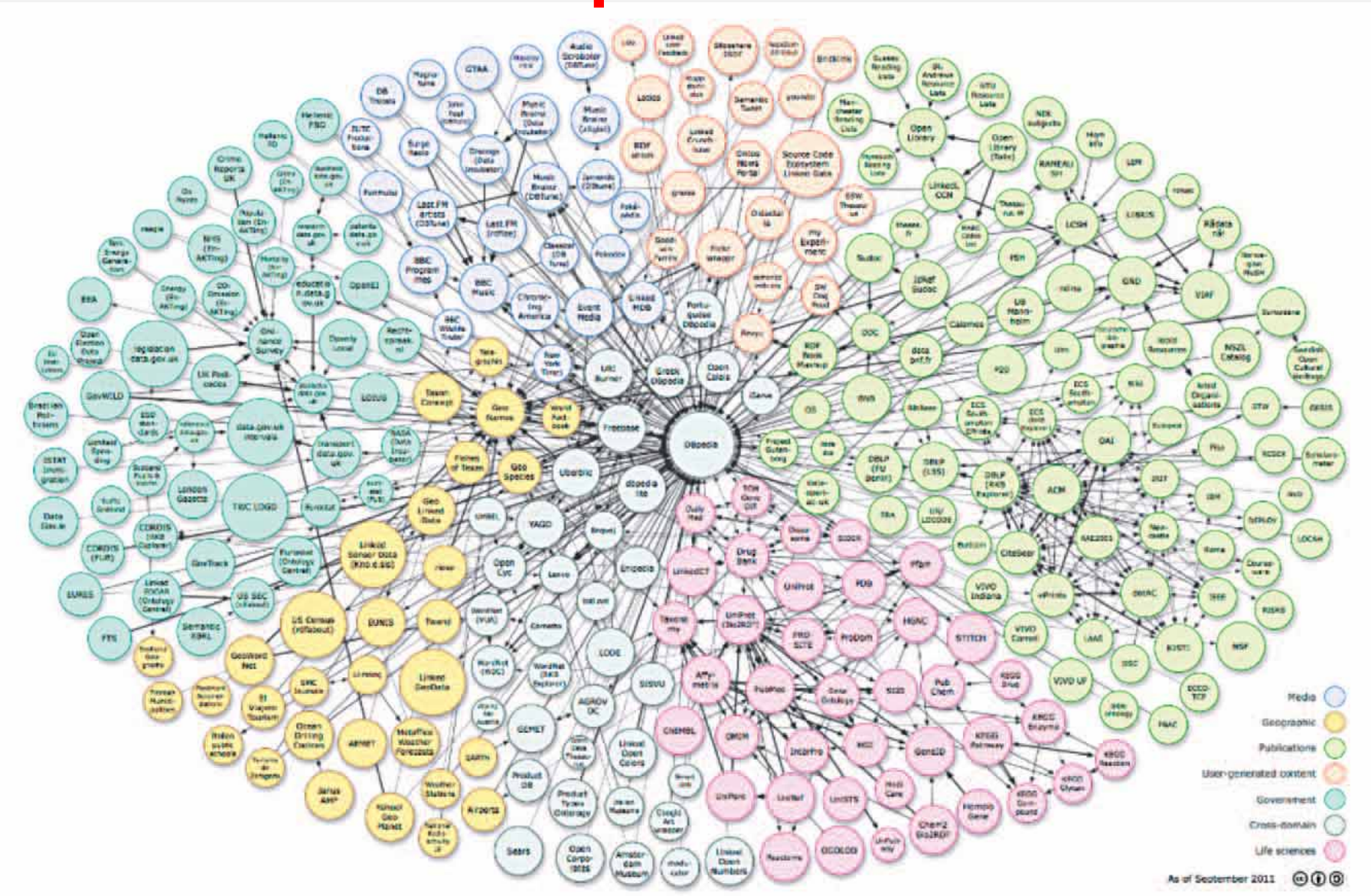
noble Search

Rembrandt Federated Search Results (Sofia - OWLIM)

Yale Federated Search Results (New Haven - OWLIM)

BY NC SA

Linked Open Data Cloud



BY NC SA

Linked Open Data

http://linkeddata.org/

http://www.linkedopendata.it/

http://thedatahub.org/

BY NC SA