Semantic Web approaches in digital history: an introduction

Michele Pasin
Kings College, London
November 2011



http://www.multiurl.com/g/bKQ

http://www.kcl.ac.uk/artshums/depts/ddh/

http://www.michelepasin.org

Outline

- the movements for open data
  - *what why who..*
- the semantic web initiative
  - *main principles and technologies; formal ontologies*
- semantic web approaches in digital history
  - *a few examples*
- hands on session
  - *design your own use-case for a semantic mash-up*

1. the movements for open data

What is the open data movement?

Numerous scientists have pointed out the irony that right at the historical moment when we have the technologies to permit worldwide availability and distributed process of scientific data, broadening collaboration and accelerating the pace and depth of discovery…..we are busy locking up that data and preventing the use of correspondingly advanced technologies on knowledge

John Wilbanks, Executive Director, Science Commons



http://creativecommons.org/science

Arguments pro and against...

- "Data belong to the human race".

Typical examples are genomes, data on organisms, medical science, environmental data.

- Facts cannot legally be copyrighted.

- It's the result of public money Public money was used to fund the work and so it should be universally available

- Helps scientific research In scientific research the rate of discovery is accelerated by better access to data.

- Intellectual property, copyright issues

especially with non-factual data

- Data is not information, nor knowledge

ie providing a 'data dump' doesn't produce transparency without experts interpreting it

- Revenue from publishing data can be used positively

eg permits non-profit organizations to recover costs or fund other activities

Open data: some big players

- Governmental data:

  U.S. government open-data *http://www.data.gov/*

  U.K. government open-data *http://data.gov.uk/*

  Financial information *http://openspending.org/*

- Science Data

  Biology: *http://www.biomedcentral.com/*

  Neuroscience: *http://openconnectomeproject.org/*

- Cultural Heritage Data

  British Library: *http://www.bl.uk/bibliographic/datafree.html*

  Europeana: *http://www.europeana.eu/portal/*

- News Data:

  The Guardian: *http://www.guardian.co.uk/data*

  BBC: *http://kasabi.com/browse/datasets/*

Examples of closed data

- Closed Databases: compilation in databases or websites to which only registered members or customers can have access.
- Closed Technologies: use of a proprietary or closed technology or encryption which creates a barrier for access.
- Copyright or License forbidding (or obfuscating) re-use of the data.
- Patent forbidding re-use of the data (for example the 3-dimensional coordinates of some experimental protein structures have been patented)
- Time-limited Access to resources such as e-journals (which on traditional print were available to the purchaser indefinitely)
- Webstacles, or the provision of single data points as opposed to tabular queries or bulk downloads of data sets.

A network of open data activities

- Open Access: making scholarly publications freely available on the internet.
- Open Content: making resources aimed at a human audience (such as prose, photos, or videos) freely available.
- Open Notebook Science: application of the Open Data concept to as much of the scientific process as possible, including failed experiments and raw experimental data.
- Open Knowledge: even broader perspective than Open Data. It covers (a) scientific, historical, geographic or otherwise (b) Content such as music, films, books (c) Government and other administrative information.
- Open Source (Software): licenses under which computer programs can be distributed and is not normally concerned primarily with data.

So, what can we do with open data?

- They allow programmatic access to resources
  - can use the power of computers to analyse the data
  - can draw inferences by ourselves, rather than relying on other applications/interfaces to the raw data

- Notion of ‘Mashup’
  - Def.: “Web page or application that uses and combines data, presentation or functionality from two or more sources to create new services” *http://en.wikipedia.org/wiki/Mashup_(web_application_hybrid)*
  - basic idea: generate new information by combining independent datasets
  - computational equivalent of an intellectual ‘synthesis’
    - *combination, visualization, and aggregation*

Mash-up: “BBC Dimensions”



“bring home the human scale of events and places in history”

http://howbigreally.com/

Mash-up: “England riots: was poverty a factor?”

David Cameron: "These riots were not about poverty"

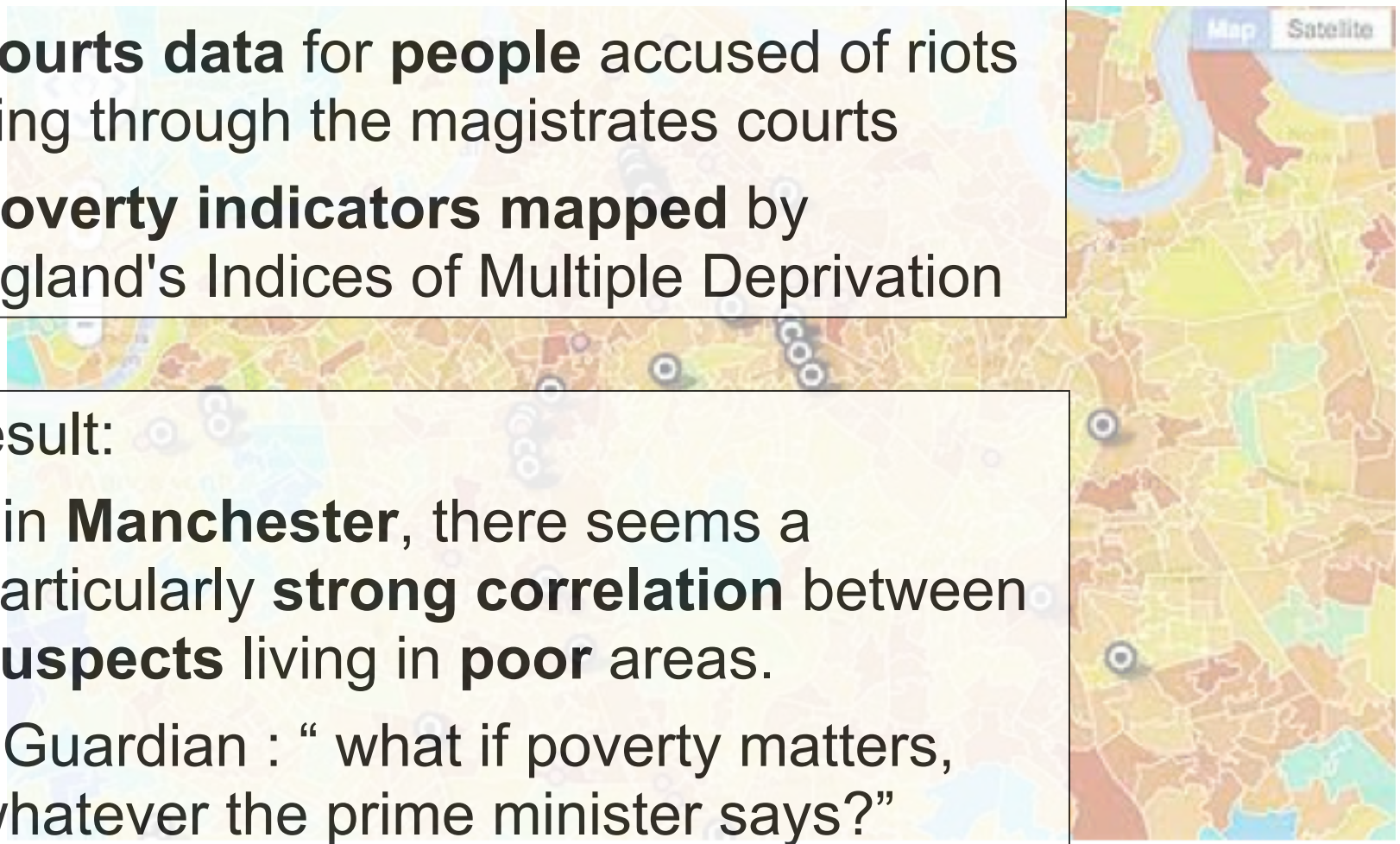


http://www.guardian.co.uk/news/datablog/2011/aug/16/riots-poverty-map-suspects

“..was poverty a factor?” behind the scenes

Two datasets:

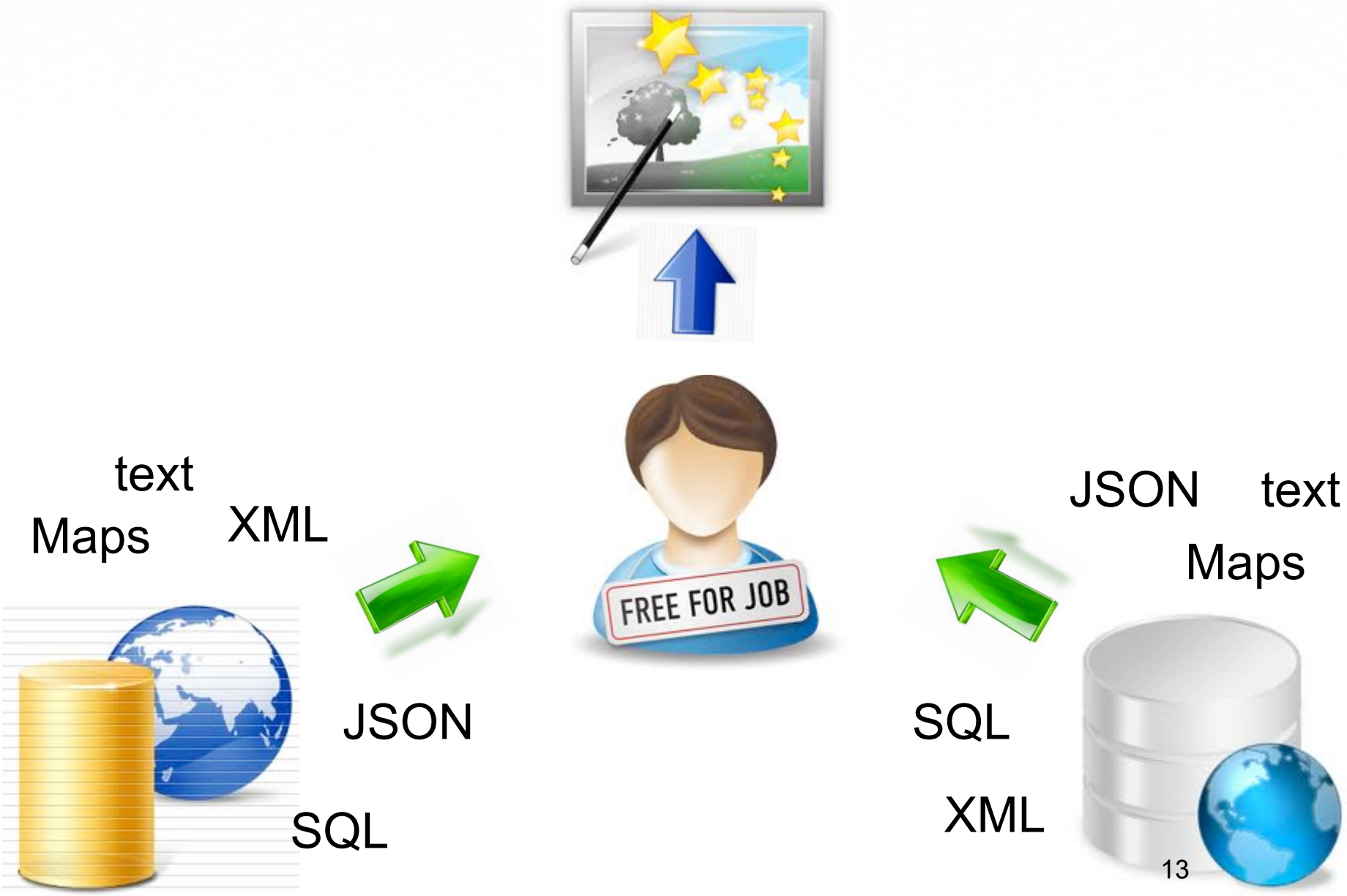
- **courts data** for **people** accused of riots going through the magistrates courts
- **poverty indicators mapped** by England's Indices of Multiple Deprivation



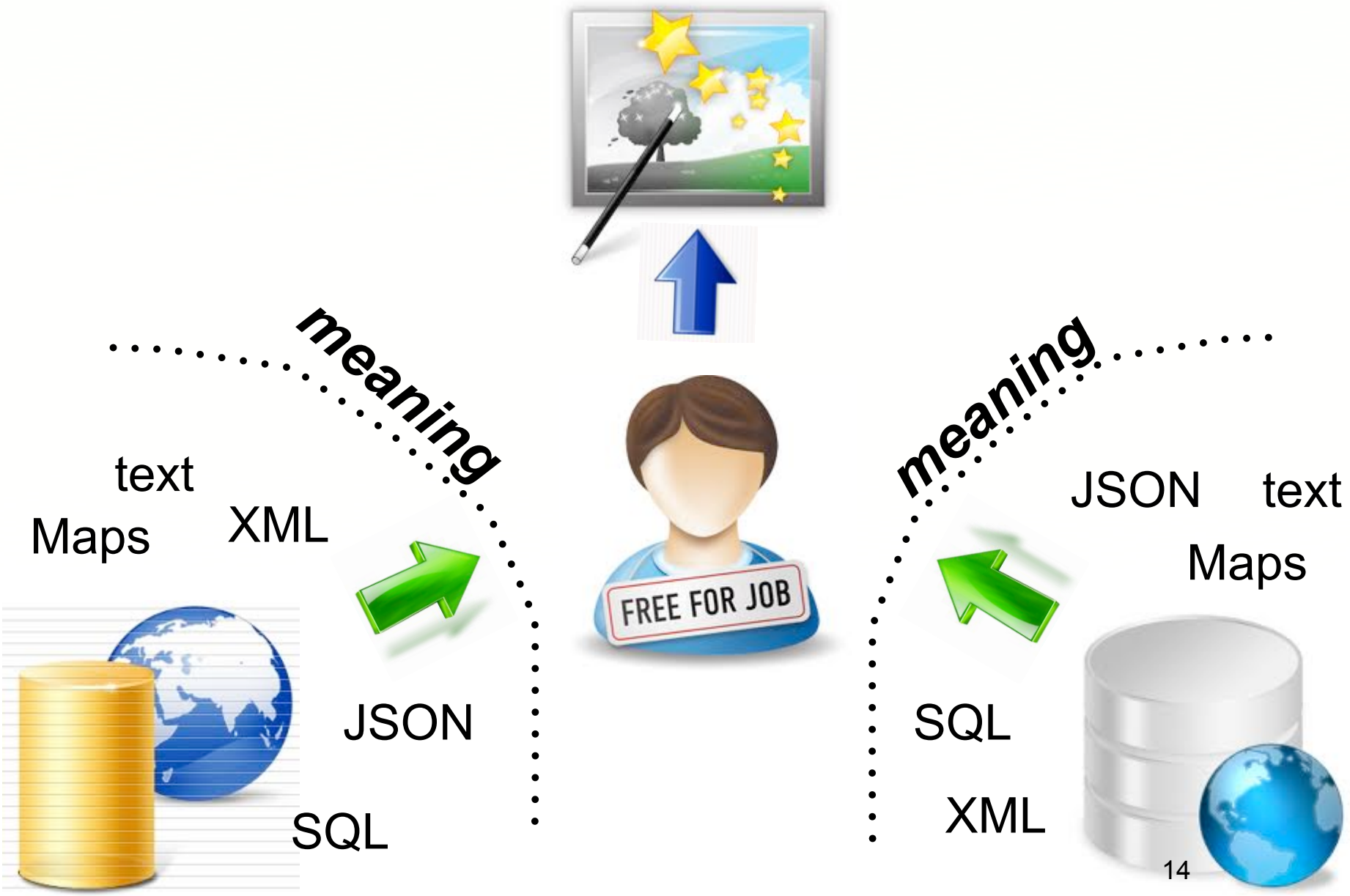
Result:

- in **Manchester**, there seems a particularly **strong correlation** between **suspects** living in **poor** areas.
- Guardian : “ what if poverty matters, whatever the prime minister says?”

What it takes to build a mash-up:



What it takes to build a mash-up:



Obstacles to creating mash-ups:

Text–data mismatch

A large portion of data is described in text, thus making it difficult for softwares to detect 'identity' of things. Eg ("World War 1", "The great War", "The first war of the 20th century")

Data format mismatch

Structured data is available in a plethora of formats. Different data providers use different computer languages eg XML, JSON, SQL, so the programmers needs to know how to operate with all of them.

Object identity and separate schema

Even if all data is available in a common format, in practice sources differ in how they state what essentially the same fact is. Eg two data providers refer to the same person, but one uses its NIN and the other the name+ surname+address to identify him/her.

Data quality

Data aggregators have little to no influence on the data publisher. Data is often erroneous, and combining data often aggravates the problem. Especially when performing reasoning (automatically inferring new data from existing data), erroneous data has potentially devastating impact on the overall quality of the resulting dataset.

Notion of Interoperability

Interoperability means the capability of different information systems to communicate some of their contents. In particular, it may mean that

1. two systems can exchange information, and/or

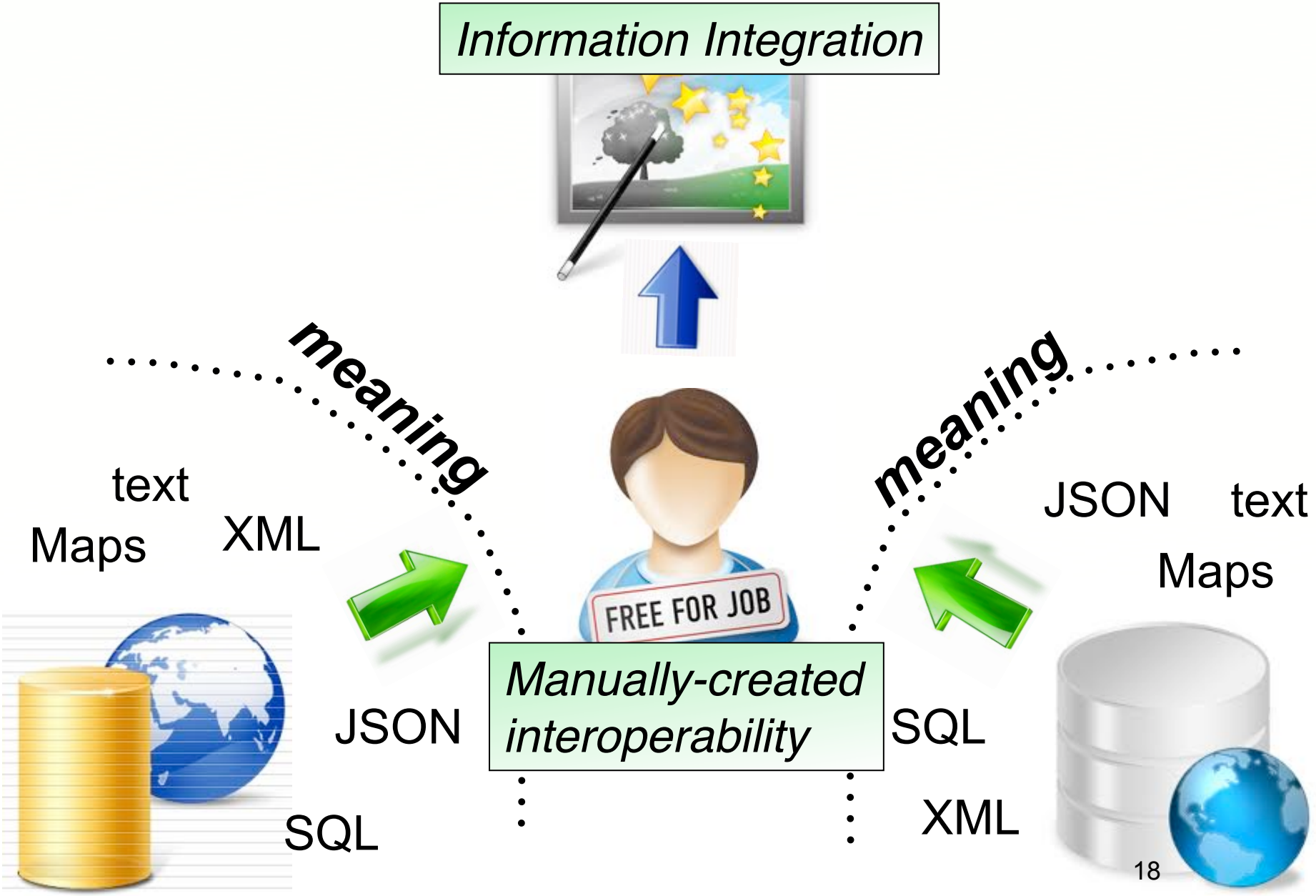
2. multiple systems can be accessed with a single method.

Notion of Information Integration

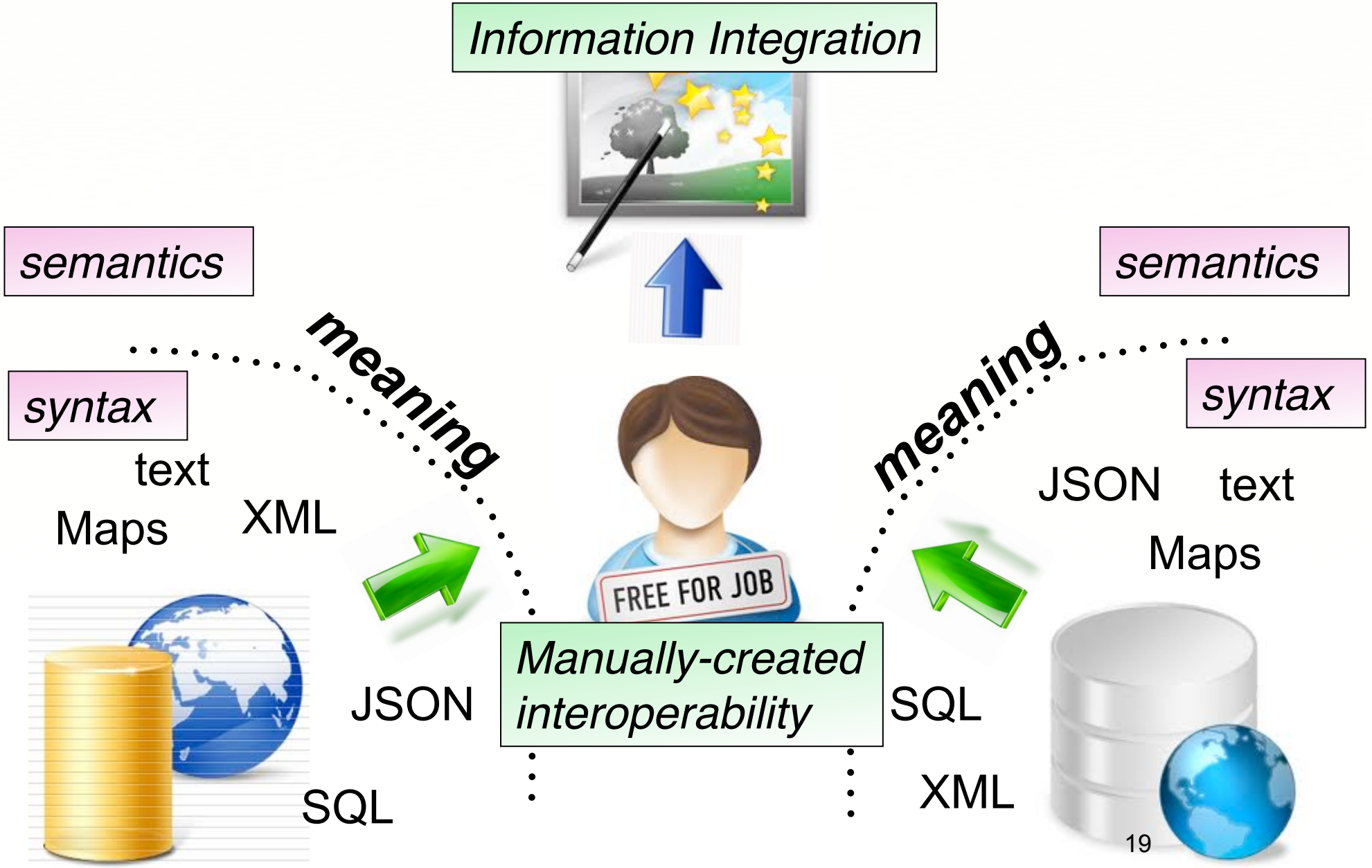
[...] information integration provides the basis for a rich "knowledge space" built on top of the basic web "data layer". This knowledge layer is composed of value-added services that process and offer abstracted information and knowledge, rather than returning documents (in the manner of most current web search engines).

Towards a Core Ontology for Information Integration, Doerr, 2003.

What it takes to build a mash-up:



What it takes to build a mash-up:



Notion of Syntactic Interoperability

*Syntactic interoperability means that the **information encoding** of the involved systems and the **access protocols** are compatible, so that information can be processed as described above without error. However, this does not mean that each system processes the data in a manner consistent with the intended meaning.*

For example, one system may use a table called "Actor" and another one called "Agent". With syntactic interoperability, data from both tables may only be retrieved as distinct, even though they may have exactly the same meaning.

Notion of Semantic Interoperability

*Semantic interoperability means the capability of different information systems to communicate information consistent with the **intended meaning**. In more detail, the intended meaning encompasses*

*1. the data **structure elements** involved,*

*2. the **terminology** appearing as data and*

*3. the **identifiers** used in the data for factual items such as places, people, objects etc.*

2. The Semantic Web vision

A little history

The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.

Berners-Lee, T., Hendler, J. and Lassila, O. The Semantic Web, Scientific American, 2001.

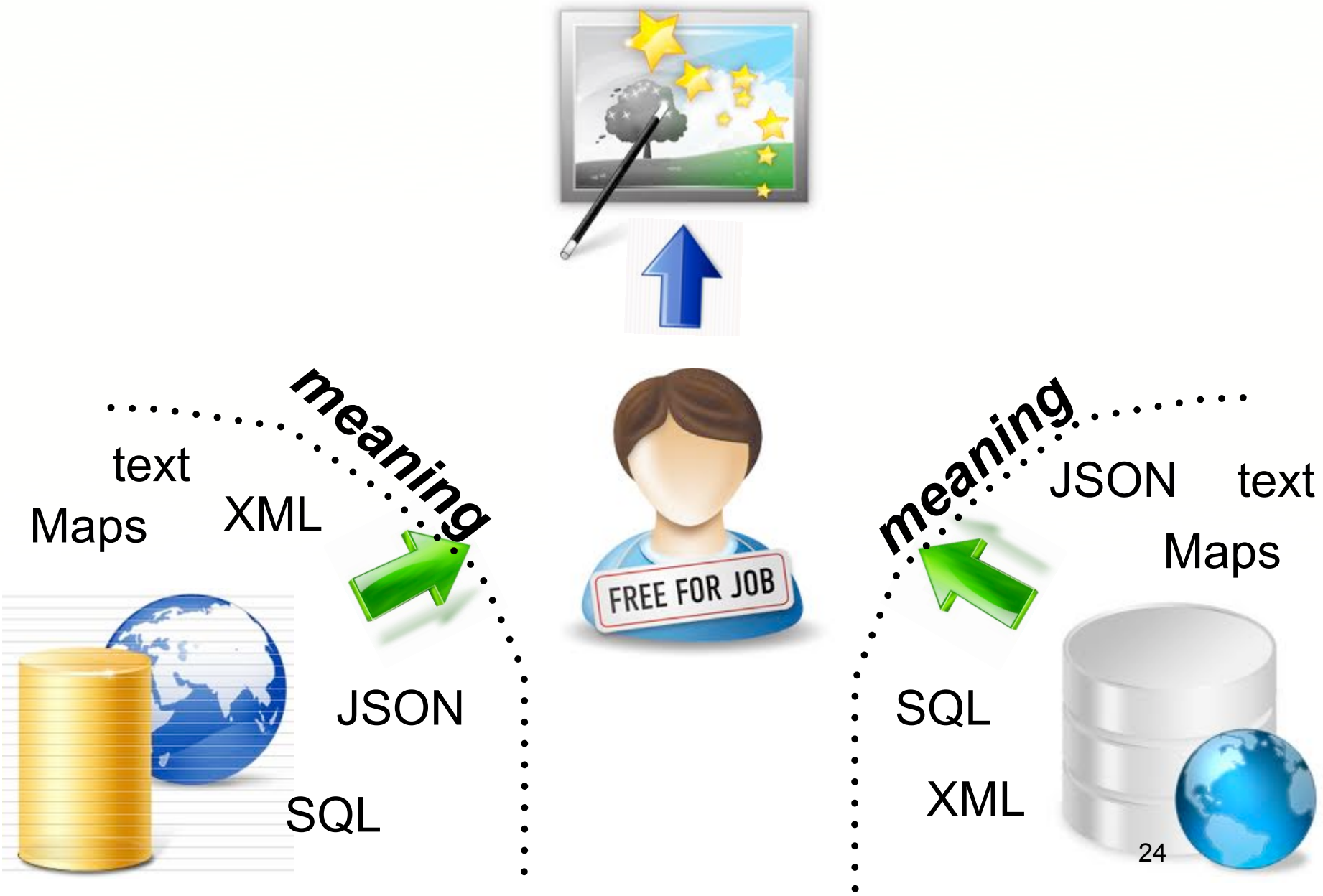
The Semantic Web is a vision: the idea of having data on the Web defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications.

World Wide Web Consortium, Semantic Web Activity Statement, 2001.



http://www.w3.org/2001/sw/Activity

Example: remember the mashup diagram..

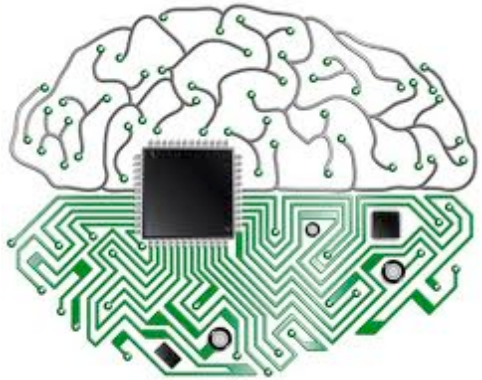


... spiced-up with some 'artificial' intelligence!



request





meaning

text

Maps

XML



JSON

SQL

meaning

JSON

text

Maps

SQL

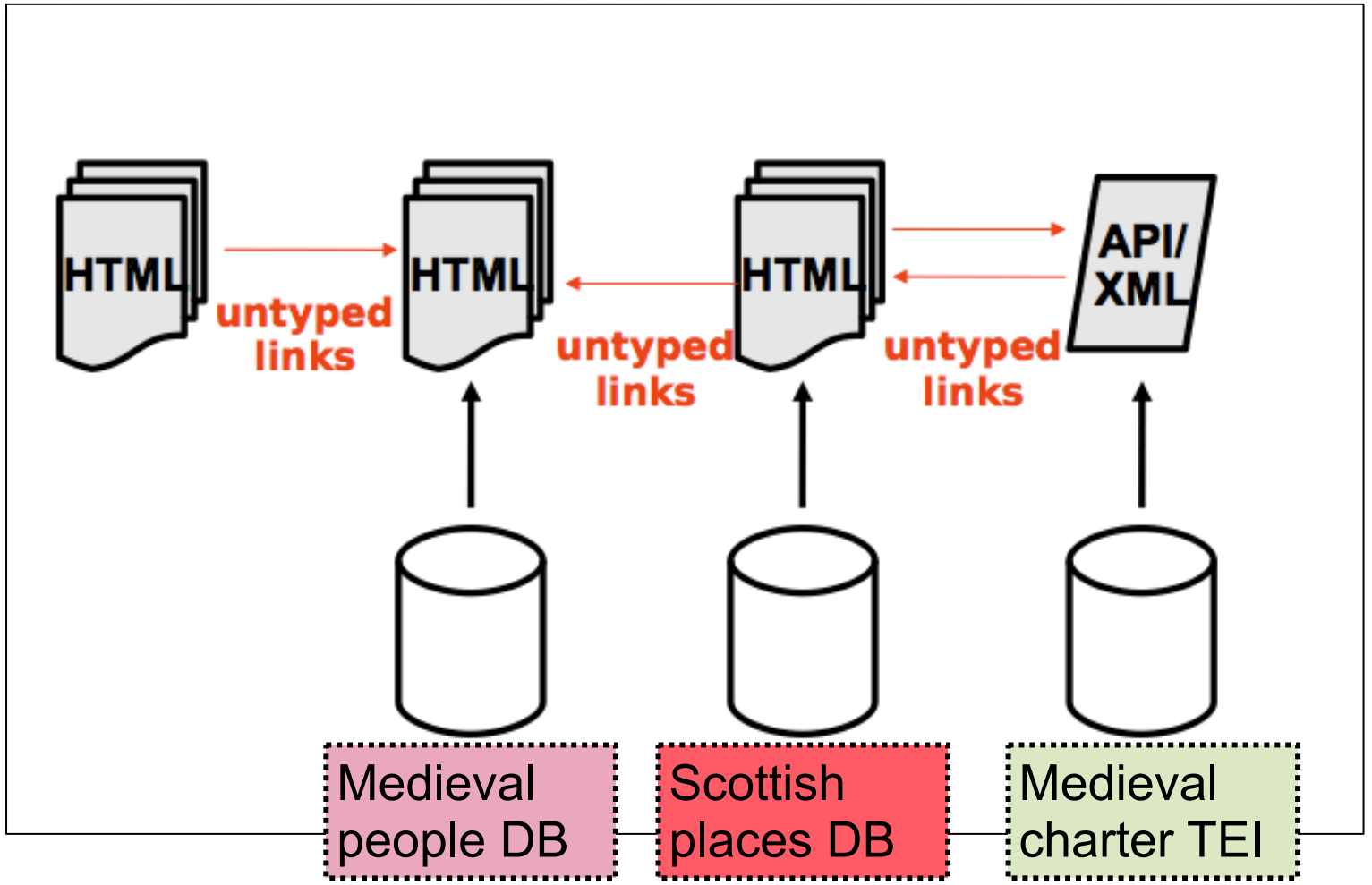
XML



Web vs Semantic web: overview of features

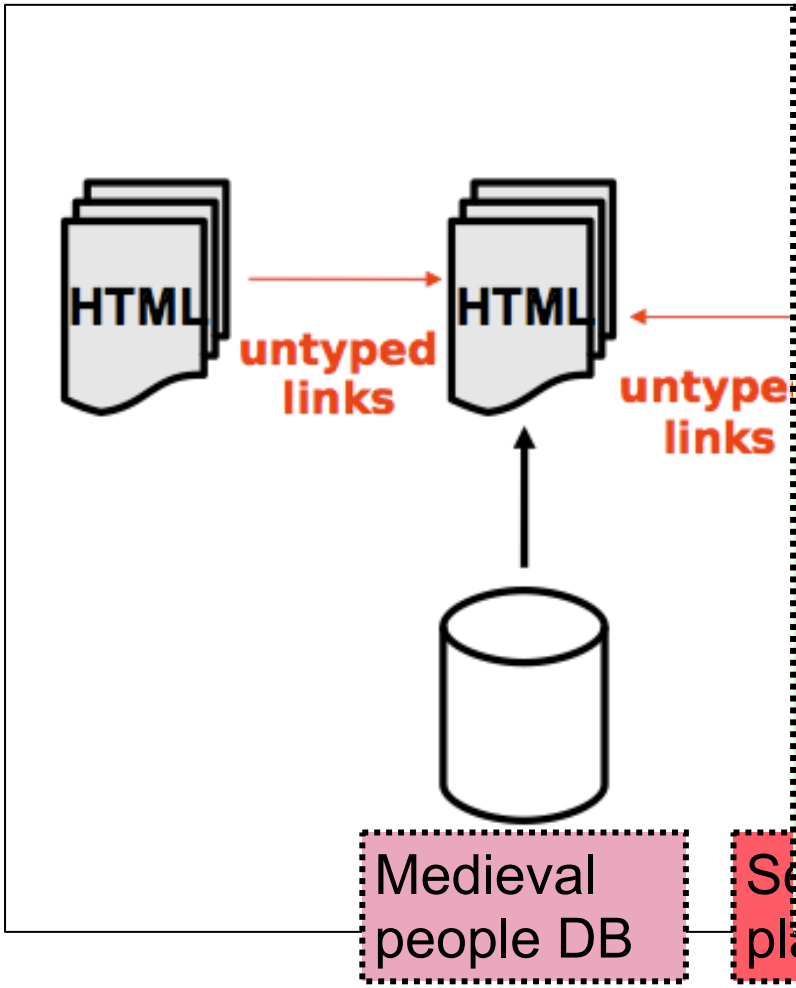
URL Uniform Resource Locator (=web pages)	**URI** Uniform Resource Identifier (=real things)
HTML, CSS etc. Technologies for the presentation of data	**RDF, RDFS, OWL** Technologies for encoding the meaning of data
Databases E.g., MySQL, Postgre, etc..	**TripleStores** Databases for semantic data (=RDF)
(Humans)	**Ontologies** 'knowledge charts' that let computers make sense of semantically-encoded information
(Humans)	**Reasoners** Softwares that apply logical deductions to semantic information so to derive new facts
(Humans)	**Agents** Web-bots: softwares that can carry out complex tasks by mediating between us and the SW

Standard web architecture: a simplified view



Adapted from Heath. An Introduction to Linked Data. (2007)

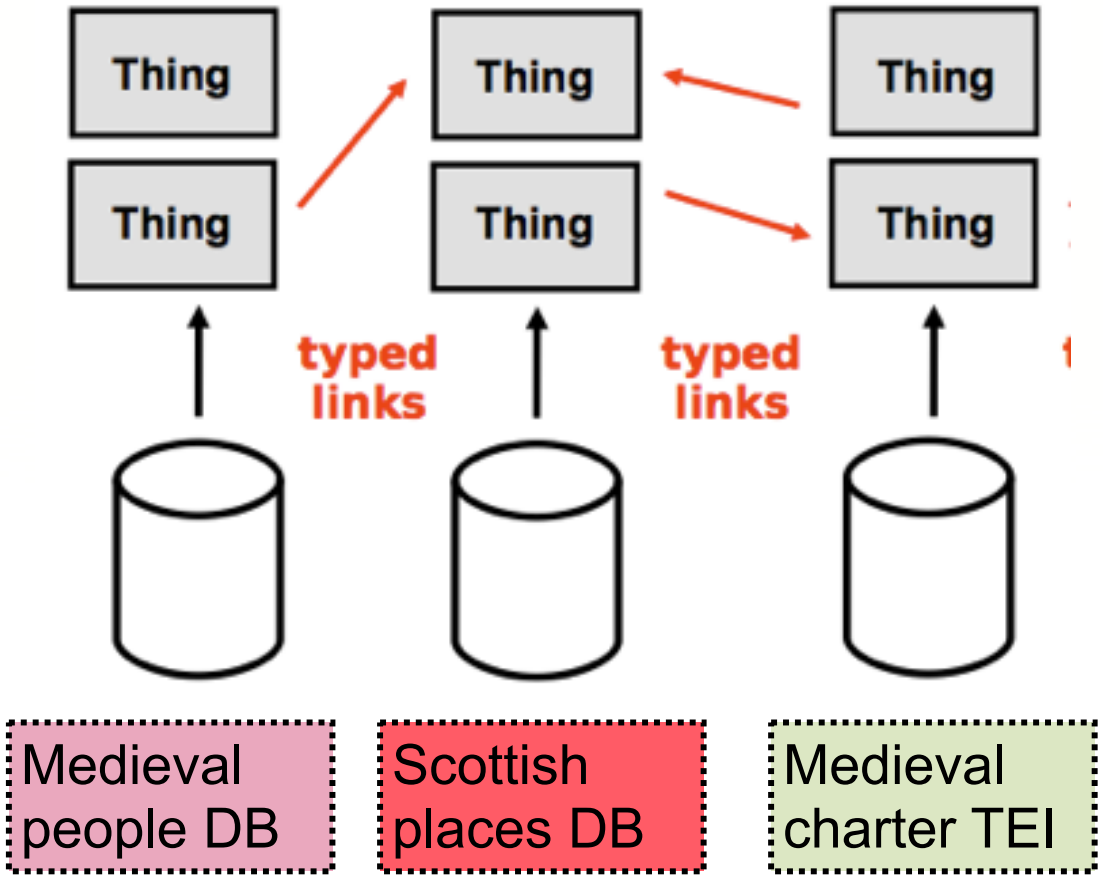
Standard web architecture: a simplified view



- Analogy
  - a global **filesystem**
- Designed for
  - **human** consumption
- Primary objects
  - **documents**
- Links between
  - **documents** (or sub-parts of)
- Degree of structure in objects
  - fairly **low**
- Semantics of content and links
  - **implicit**

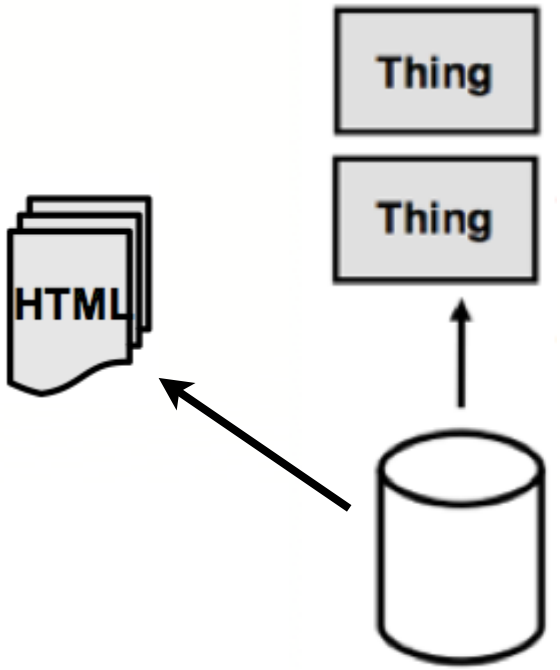
Adapted from Heath. An Introduction to Linked Data. (2007)

SW architecture: a simplified view



Adapted from Heath. An Introduction to Linked Data. (2007)

SW architecture: RDF triples



Medieval people DB

<http://www.medievaldb.uk/entity/person#Gustave-I>

<http://www.medievaldb.uk/entity/relation#lives-in>

<http://www.medievaldb.uk/entity/place#Glasgow>

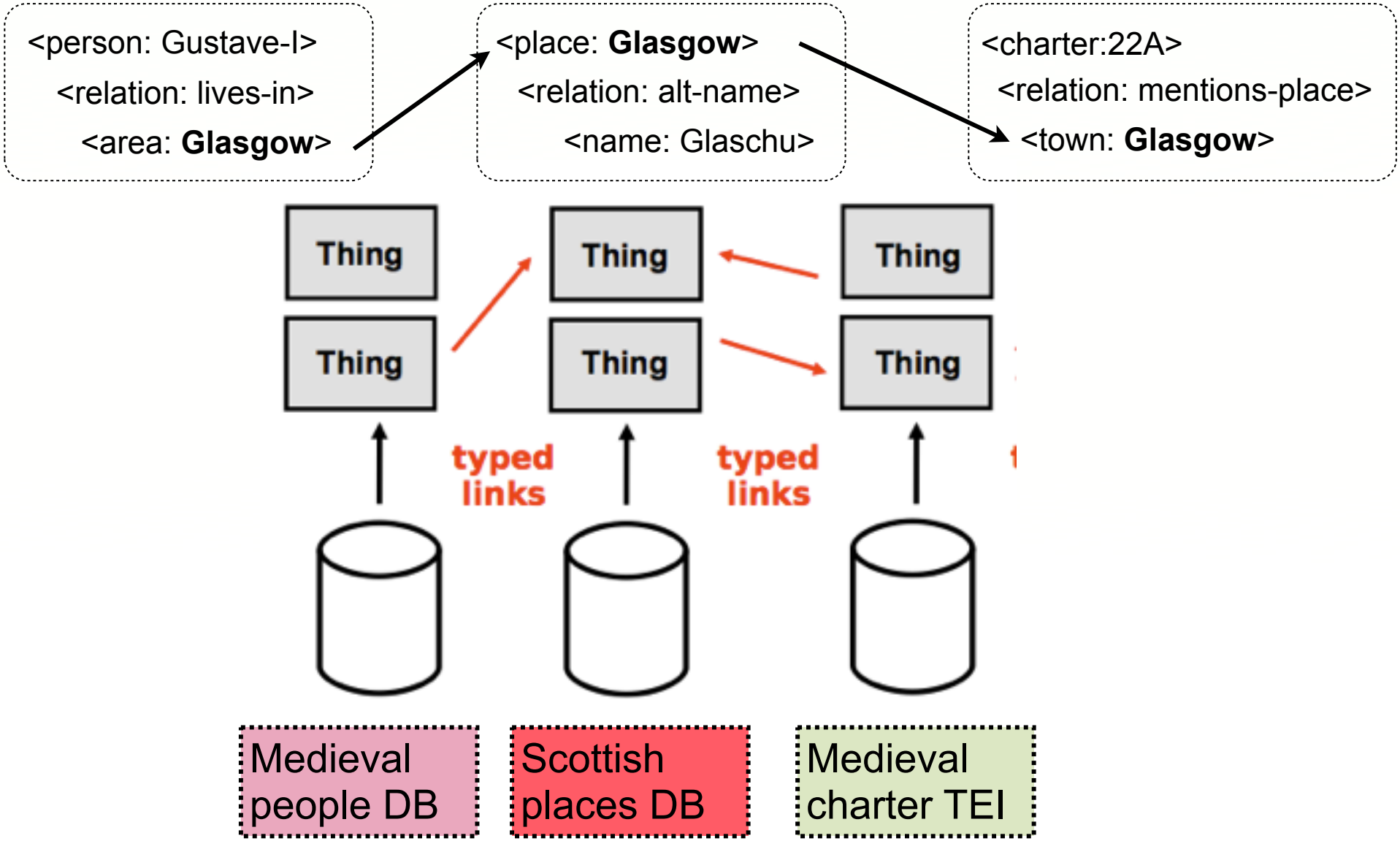
RDF

<Subject URI>

<Predicate URI>

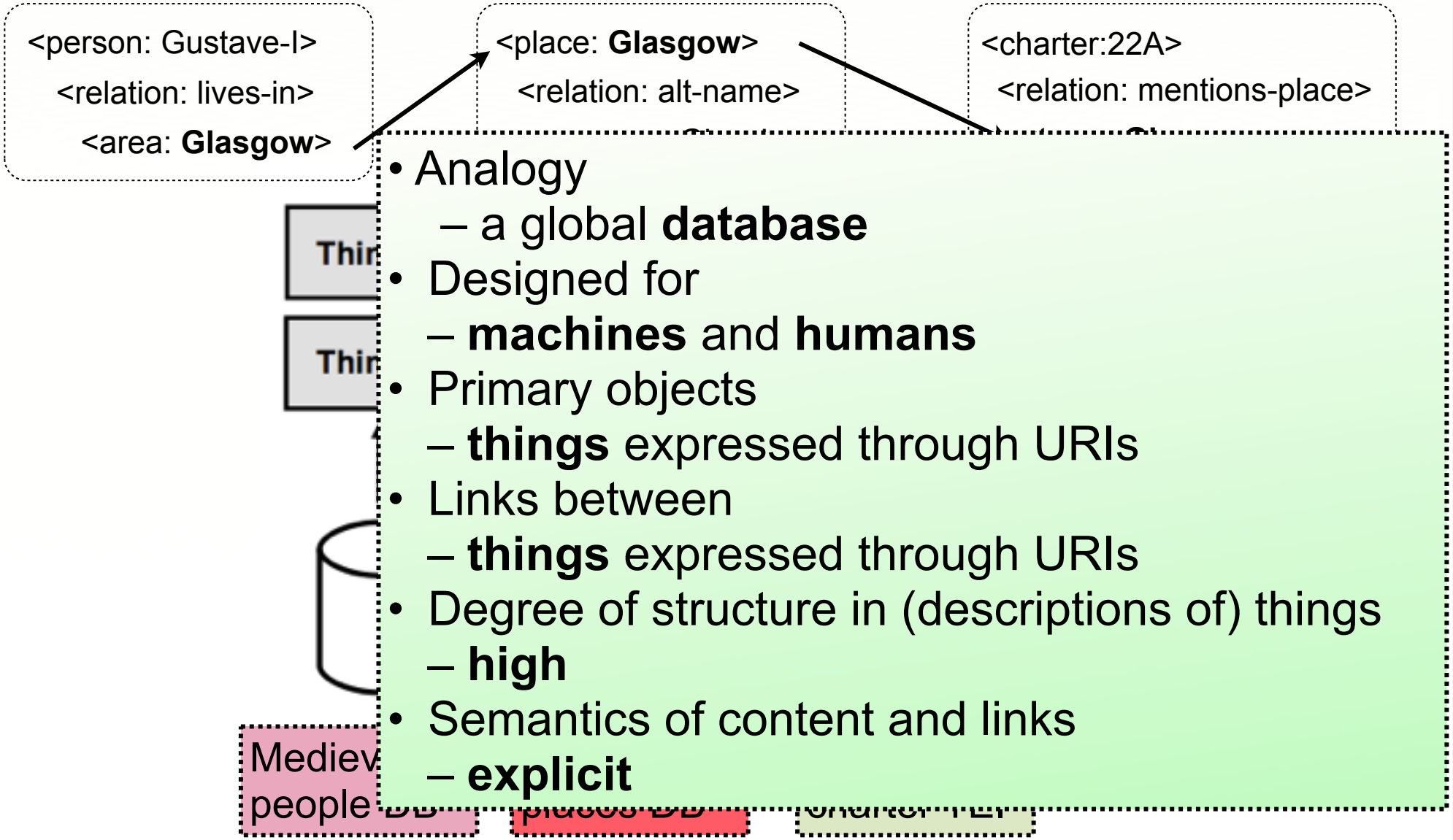
<Object URI>

SW architecture: a simplified view



Adapted from Heath. An Introduction to Linked Data. (2007)

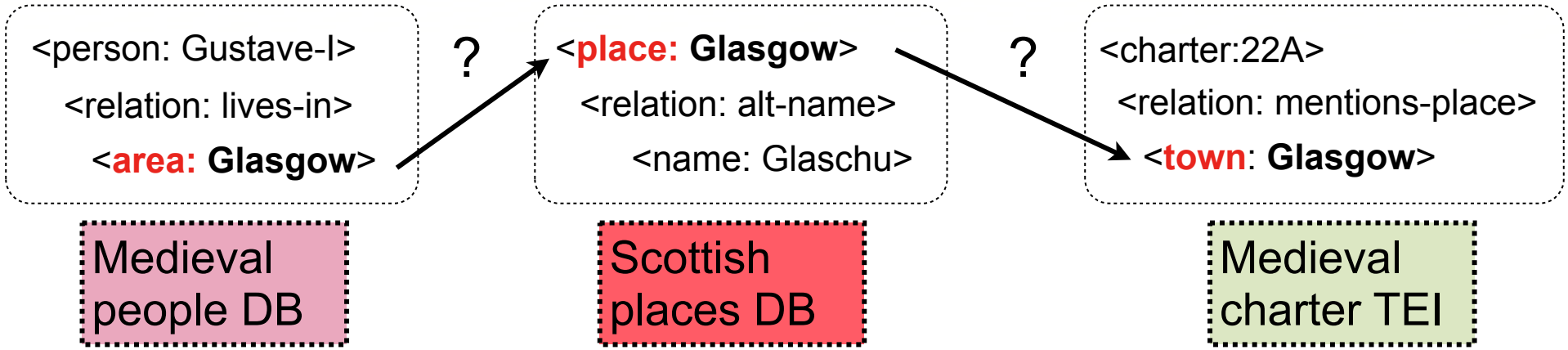
SW architecture: a simplified view



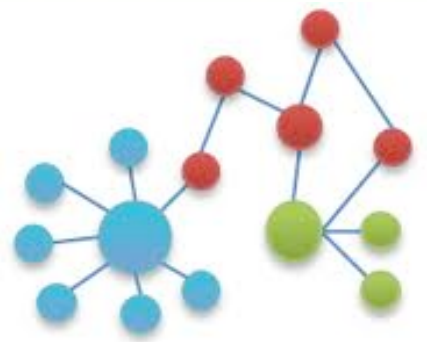
- Analogy
  - a global **database**
- Designed for
  - **machines** and **humans**
- Primary objects
  - **things** expressed through URIs
- Links between
  - **things** expressed through URIs
- Degree of structure in (descriptions of) things
  - **high**
- Semantics of content and links
  - **explicit**

Adapted from Heath. An Introduction to Linked Data. (2007)

Negotiating ‘meaning’ on the semantic web:



Negotiating ‘meaning’ on the semantic web:



Places Ontology:

MedievalDB:area

==

ScottishPlaces:place

==

MedievalCharter:town

then

<person: Gustave-I>
<relation: lives-in>
<area: **Glasgow**>
<relation: alt-name>
<name: Glaschu>

<person: Gustave-I>
<relation: lives-in>
<**area: Glasgow**>

=

Medieval people DB

<**place: Glasgow**>
<relation: alt-name>
<name: Glaschu>

Scottish places DB

=

<charter:22A>
<relation: mentions-place>
<**town**: **Glasgow**>

Medieval charter TEI

So what is an ontology?

- Philosophy:
  the inquiry into being in so much as it is being, or into beings insofar as they exist

- Digital world:
  the inquiry into being in so much as it can be represented (=modeled) with computers

- A definition:
  “a formal ontology is essentially a **formal model** which **represents** a target **domain**, and usually is constituted by a **hierarchy** of **concepts** which are **interlinked** by defined **relations**”.

Pitfall: Ontologies and data models

- Data schemas are not ontologies!

  - Writing something in XML/RDF/OWL does not make it an ontology! The key difference is not the language the **intended use**

  - making representational choices at the ***highest*** **level of** ***abstraction***, **while still being as clear as possible about the meaning of terms**

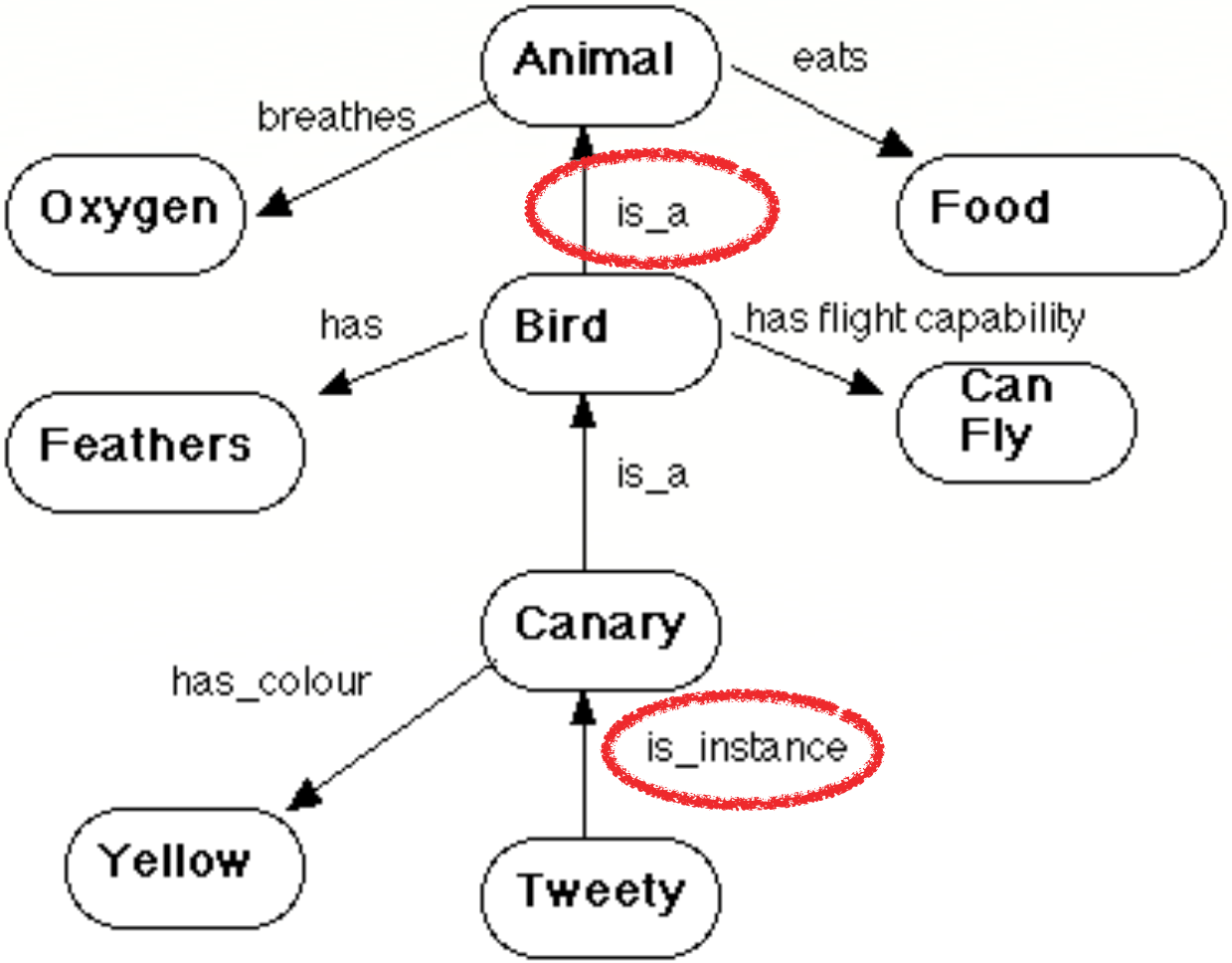
- Main difference with data models is not the content, but the purpose (= data sharing, interoperability)

  - Clarity: context dependent vs context independent design

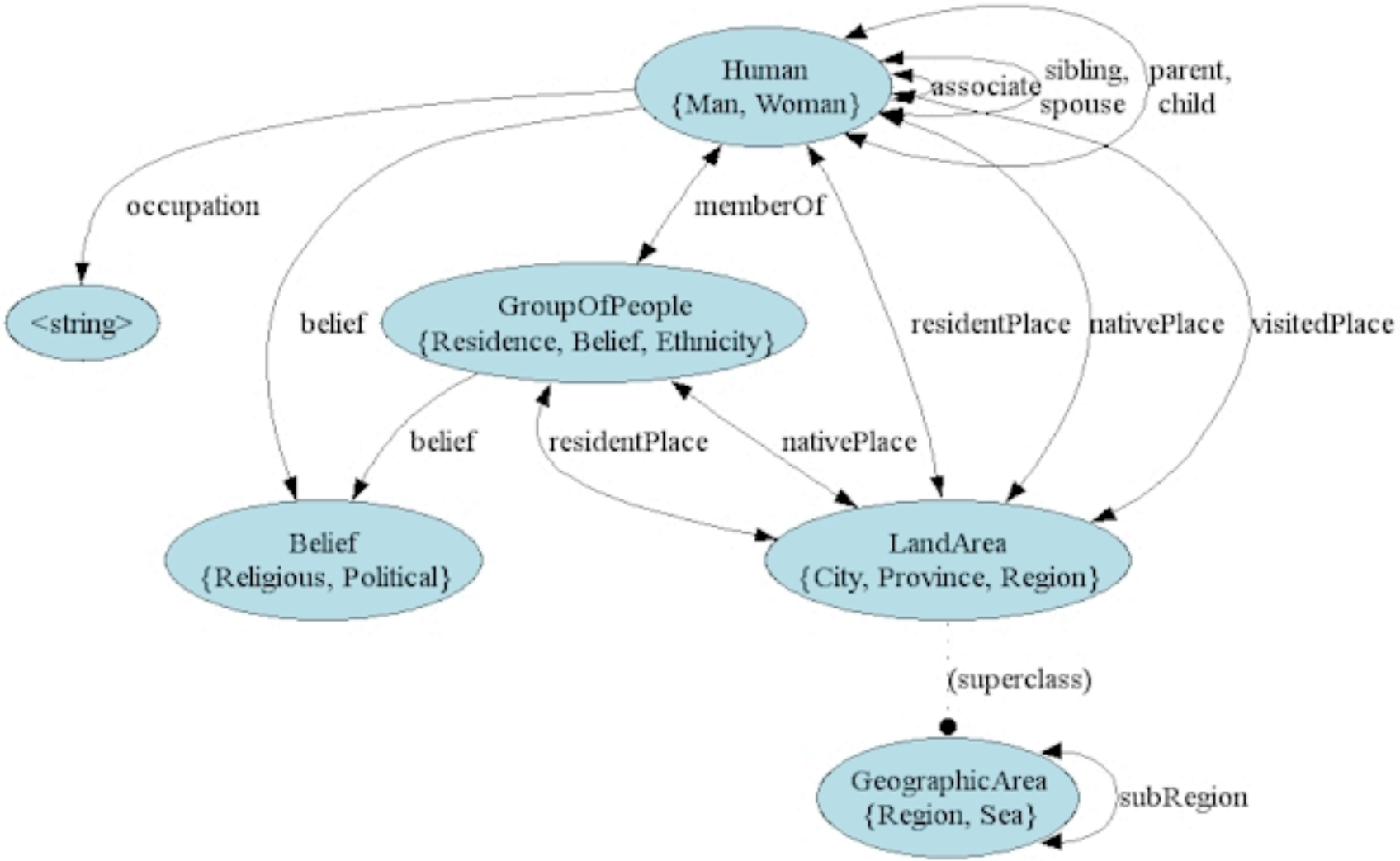
  - Extendibility: application oriented vs design for future reuse

  - Minimal Encoding Bias - avoid representational choice for benefit of implementation

A simple formal ontology for birds



A fragment of the 'Bible' ontology

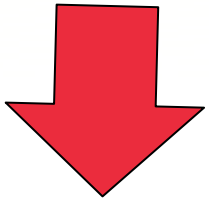


http://semanticbible.com/

Logic provides the ‘reasoning’ ...

- formal language for expressing the structures used in our inference processes

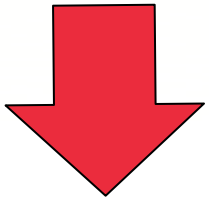
All x is b.	(Universal Affirmative)
There is a Y that is x.	(Particular Affirmative)
Therefore, y is b.	(Particular Affirmative)



All Roman tribunes have immunity	(Universal Affirmative)
Valerianus is a tribune.	(Particular Affirmative)
Therefore, Valerianus has immunity.	(Particular Affirmative)

.. and ontology provides the ‘meanings’ !

Tribune (from the Latin: tribunus; Byzantine Greek form τριβούνος) was a title shared by **10 elected officials** in the **Roman Republic**. Tribunes had the power to convene the **Plebeian Council** and to act as its **president**, which also gave them the **right to** propose **legislation** before it. They were sacrosanct, in the sense that any **assault** on their person **was prohibited**. They had the power to **veto** actions taken by **magistrates**, and specifically to **intervene legally** on behalf of **plebeians**. The tribune could also summon the **Senate** and **lay proposals** before it. [....]



For every x, if (x **isTribune**) ==> exists y such that (y **isCity**) and (y **hasName** Rome) and (**lives_in** x, y)

Making inferences by using ontologies:

<person: Gustave-I>
<relation: lives-in>
<**area:** **Glasgow**>

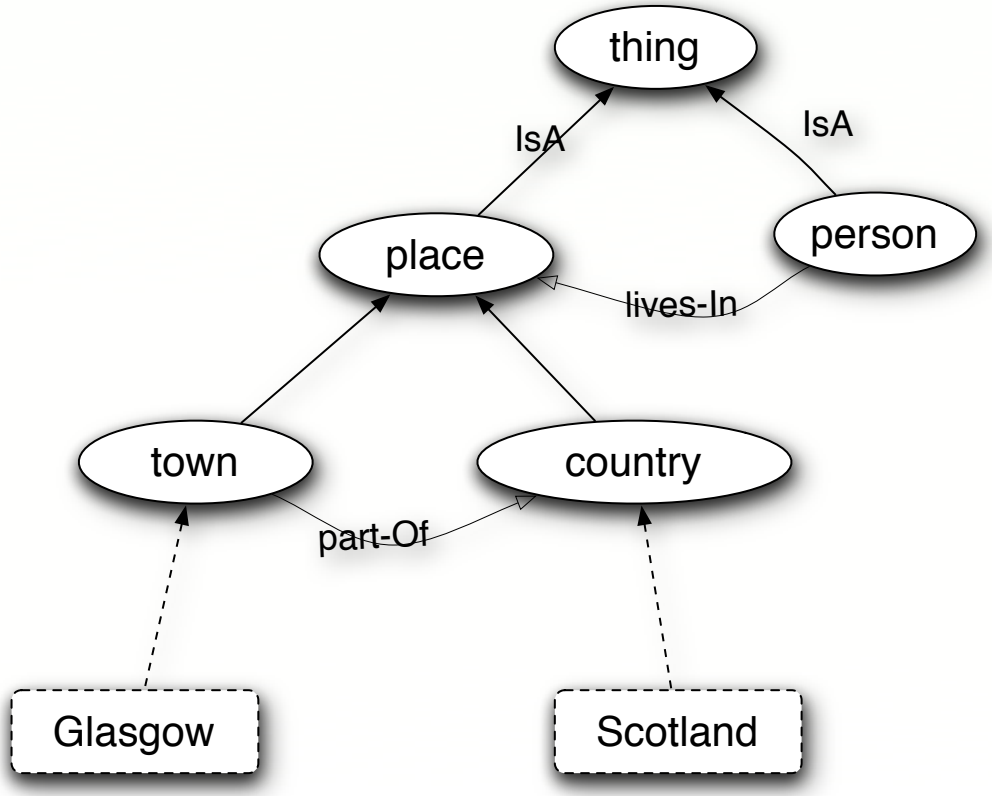
?

<**group**: **ScottishPeople**>
<relation: speak-language>
<langauge: gaelic>

Medieval people DB

Scottish places DB

Making inferences by using ontologies:



RULE:
If
P lives-in X
And
X part-Of Y
Then
X lives-in Y

<person: Gustave-I>
<relation: lives-in>
<**area:** **Glasgow**>

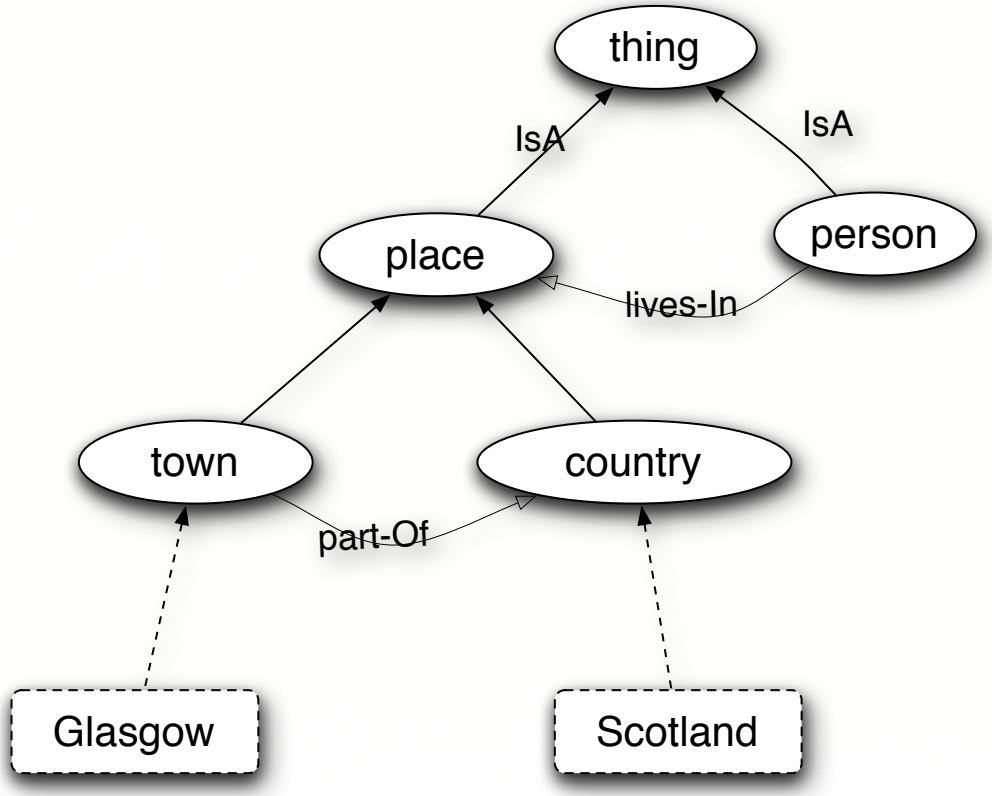
?

<**group**: **ScottishPeople**>
<relation: speak-language>
<langauge: gaelic>

Medieval people DB

Scottish places DB

Making inferences by using ontologies:



RULE:
If
P lives-in X
And
X part-Of Y
Then
X lives-in Y

then

<person: Gustave-I>
<relation: speak-language>
<language:gaelic>

<person: Gustave-I>
<relation: lives-in>
<**area: Glasgow**>

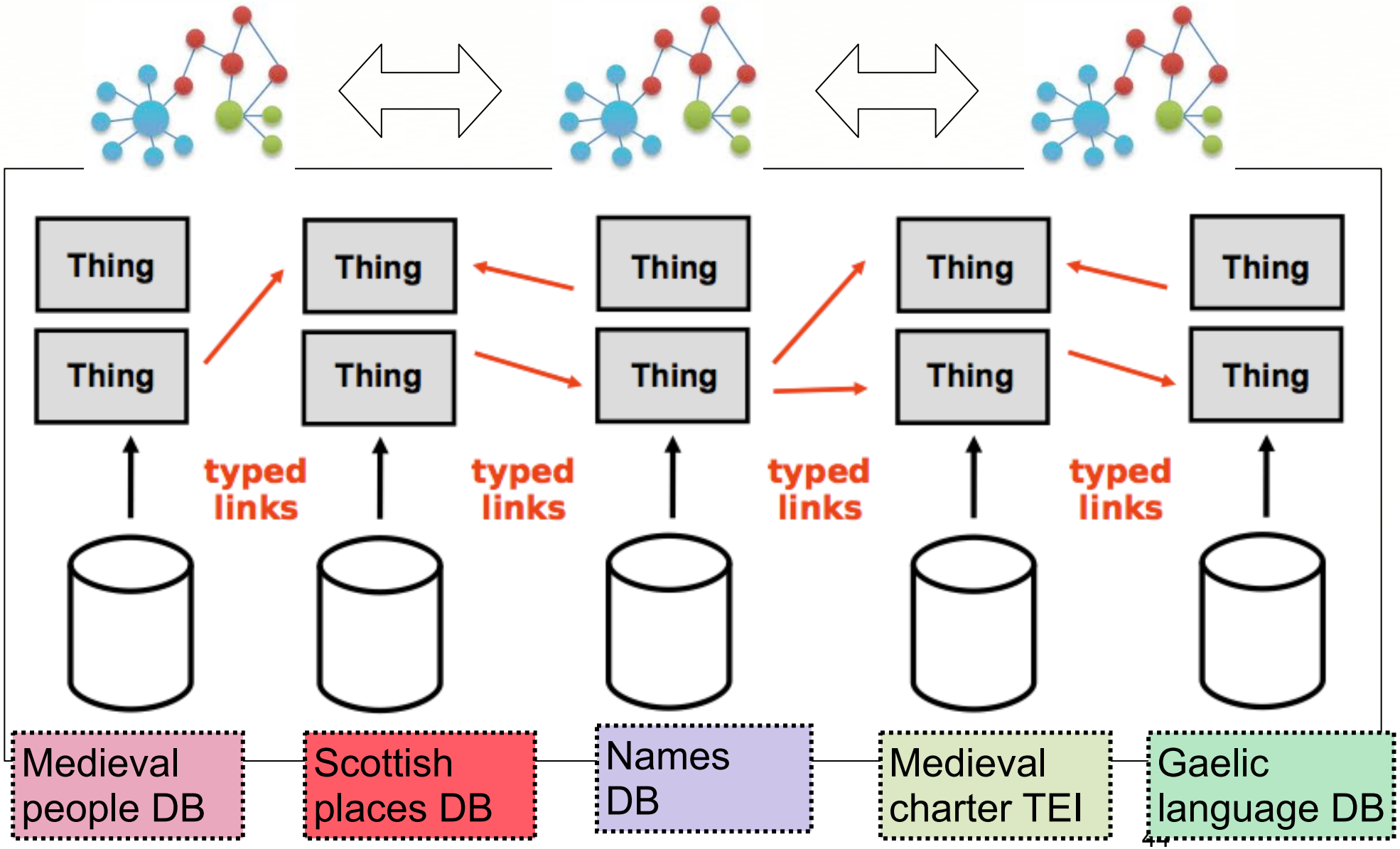
?

<**group**: **ScottishPeople**>
<relation: speak-language>
<language: gaelic>

Medieval people DB

Scottish places DB

Not one, but many ontologies (and inferences)!



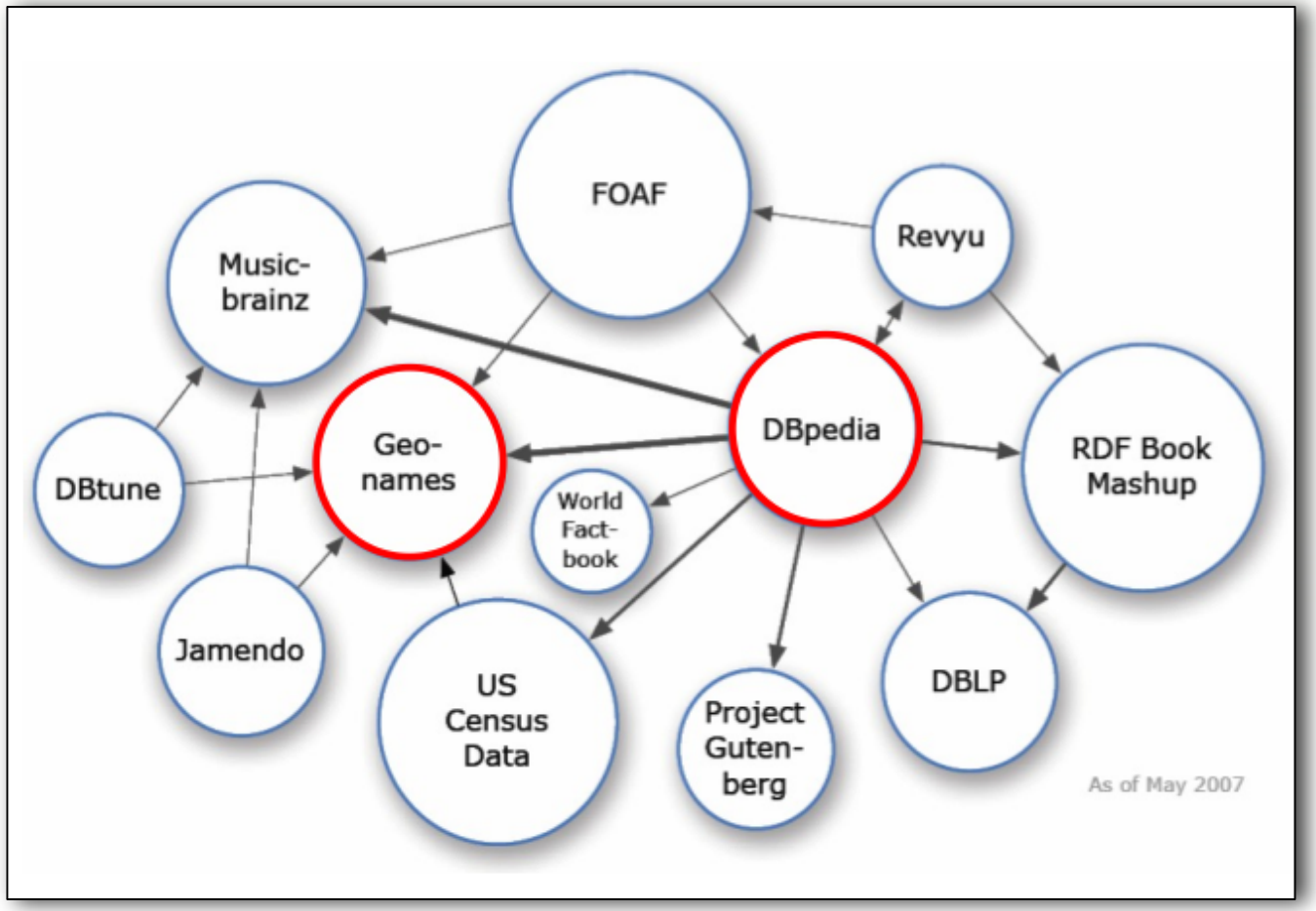
Recent developments: Linked Data (2007)

- Less ambitious version of the SW
  - less artificial intelligence: “a method of publishing structured data so that it can be interlinked and become more useful.”
  - more grassroots initiatives to build a ‘data web’

- 4 simple principles
  - Use URIs to identify things
  - Use HTTP URIs so that these things can be referred to and looked up ("dereferenced") by people and user agents.
  - Provide useful information about the thing when its URI is dereferenced, using standard formats such as RDF/XML
  - Include links to other, related URIs in the exposed data to improve discovery of other related information on the Web

The evolution of Linked Data, from 2007...

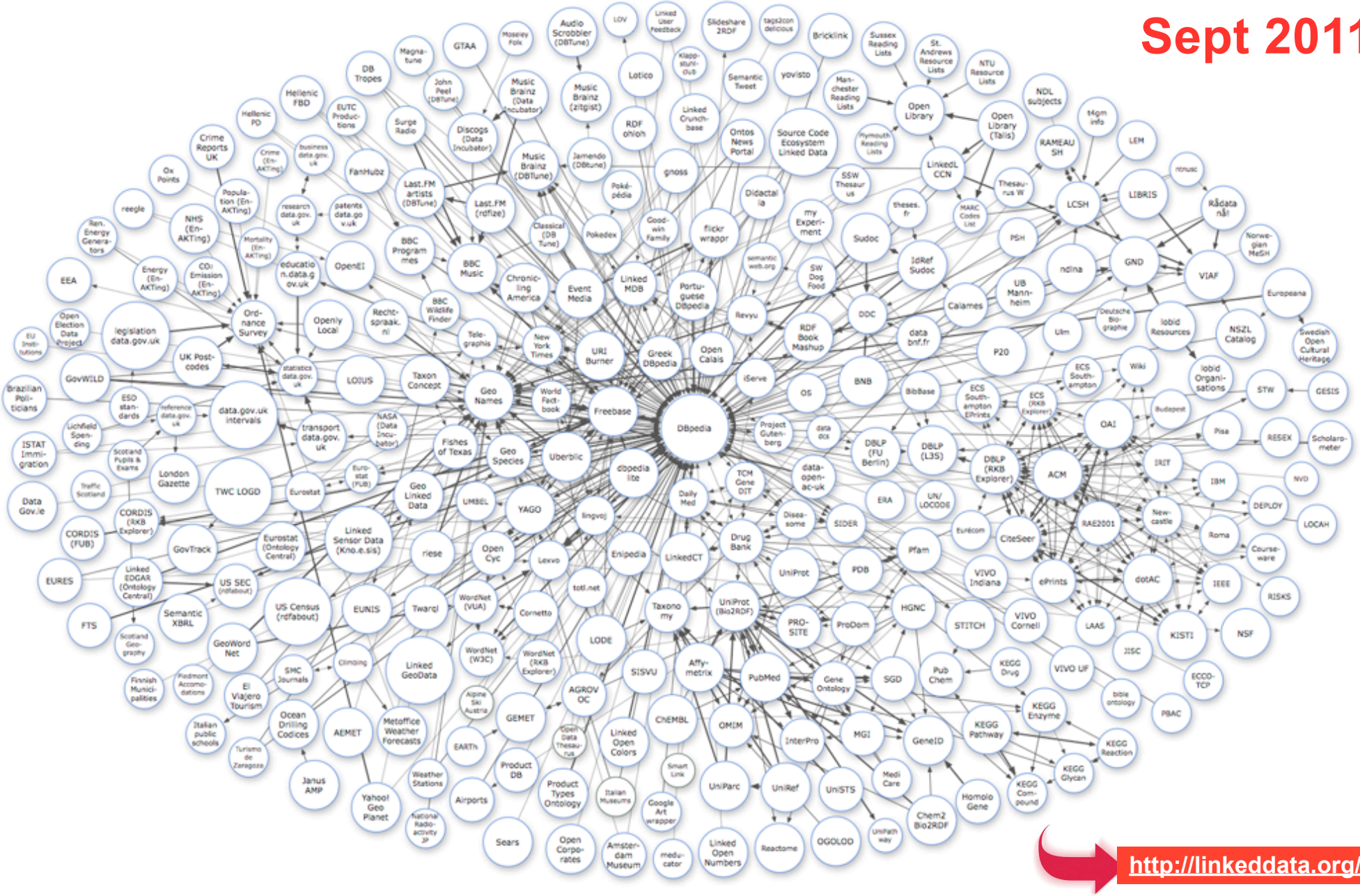
May 2007



http://linkeddata.org/

.. to 2011!

Sept 2011


DBpedia
Freebase
Geo Names
data.gov.uk intervals
legislation data.gov.uk
Ordnance Survey
Openly Local
UK Post-codes
GovWILD
statistics data.gov.uk
transport data.gov.uk
LOIUS
Taxon Concept
World Fact-book
New York Times
URI Burner
Greek DBpedia
Open Calais
iServe
OS
BNB
Project Guten-berg
data dcs
data-open-ac-uk
DBLP (FU Berlin)
DBLP (L3S)
DBLP (RKB Explorer)
ACM
OAI
ECS South-ampton
ECS South-ampton EPrints
ECS (RKB Explorer)
BibBase
Budapest
Pisa
RESEX
Scholaro-meter
STW
GESIS
lobid Organi-sations
lobid Resources
NSZL Catalog
Europeana
Swedish Open Cultural Heritage
VIAF
GND
ndlna
UB Mann-heim
Ulm
Deutsche Bio-graphie
P20
data bnf.fr
Calames
DDC
Sudoc
IdRef Sudoc
PSH
LCSH
LIBRIS
RÅdata nå!
Norwe-gian MeSH
RAMEAU SH
NDL subjects
t4gm info
LEM
Open Library
Open Library (Talis)
LinkedL CCN
Thesau-rus W
SSW Thesaur us
theses. fr
MARC Codes List
Source Code Ecosystem Linked Data
Ontos News Portal
Plymouth Reading Lists
Sussex Reading Lists
St. Andrews Resource Lists
NTU Resource Lists
Man-chester Reading Lists
Bricklink
Slideshare 2RDF
tags2con delicious
Semantic Tweet
yovisto
Linked Crunch-base
Klapp-stuhl-club
Lotico
LOV
Linked User Feedback
Audio Scrobbler (DBTune)
Music Brainz (zitgist)
Music Brainz (Data Incubator)
Music Brainz (DBTune)
Jamendo (DBtune)
RDF ohloh
gnoss
Pokédex
Poképédia
Good-win Family
flickr wrappr
Didactalia
my Experi-ment
SW Dog Food
semantic web.org
Revyu
RDF Book Mashup
Linked MDB
Event Media
Chronic-ling America
Classical (DB Tune)
BBC Music
BBC Programmes
BBC Wildlife Finder
Last.FM (rdfize)
Last.FM artists (DBTune)
Discogs (Data Incubator)
Surge Radio
FanHubz
Jamendo (DBtune)
Magna-tune
GTAA
Moseley Folk
John Peel (DBTune)
DB Tropes
EUTC Produc-tions
Hellenic FBD
Hellenic PD
Crime Reports UK
Ox Points
reegle
Rgn. Energy Genera-tors
NHS (En-AKTing)
Popula-tion (En-AKTing)
Crime (En-AKTing)
business data.gov.uk
research data.gov.uk
patents data.gov.uk
Mortality (En-AKTing)
Energy (En-AKTing)
CO2 Emission (En-AKTing)
EEA
Open Election Data Project
EU Institu-tions
Brazilian Poli-ticians
ESD stan-dards
Lichfield Spend-ing
ISTAT Immi-gration
Data Gov.ie
Traffic Scotland
Scotland Pupils & Exams
reference data.gov.uk
London Gazette
TWC LOGD
CORDIS (FUB)
CORDIS (RKB Explorer)
EURES
FTS
Linked EDGAR (Ontology Central)
Scotland Geo-graphy
Finnish Munici-palities
GovTrack
US SEC (rdfabout)
Semantic XBRL
US Census (rdfabout)
GeoWord Net
Piedmont Accomo-dations
El Viajero Tourism
Italian public schools
Ocean Drilling Codices
SMC Journals
Climbing
Linked GeoData
AEMET
Janus AMP
Yahoo! Geo Planet
Metoffice Weather Forecasts
Weather Stations
National Radio-activity JP
Airports
Product DB
Sears
Eurostat (Ontology Central)
Eurostat (FUB)
Eurostat
Linked Sensor Data (Kno.e.sis)
riese
EUNIS
Twarql
WordNet (VUA)
WordNet (W3C)
WordNet (RKB Explorer)
Alpine Ski Austria
GEMET
EARTh
Product Types Ontology
Open Data Thesau-rus
Linked Open Colors
Italian Museums
Google Art wrapper
Open Corpo-rates
Amster-dam Museum
medu-cator
Linked Open Numbers
Reactome
OGOLOD
UniPath way
Chem2 Bio2RDF
Homolo Gene
KEGG Com-pound
KEGG Glycan
KEGG Reaction
KEGG Pathway
KEGG Enzyme
KEGG Drug
Pub Chem
Medi Care
GeneID
MGI
InterPro
UniSTS
UniRef
UniParc
Smart Link
OMIM
ChEMBL
Linked Open Colors
SISVU
Affy-metrix
PubMed
Gene Ontology
SGD
AGROV OC
LODE
totl.net
Cornetto
Taxono my
UniProt (Bio2RDF)
PRO-SITE
ProDom
HGNC
STITCH
VIVO Cornell
VIVO UF
bible ontology
PBAC
ECCO-TCP
KISTI
NSF
JISC
LAAS
dotAC
ePrints
VIVO Indiana
RAE2001
Course-ware
IEEE
RISKS
IBM
DEPLOY
LOCAH
NVD
Roma
New-castle
IRIT
CiteSeer
Eurécom
UN/ LOCODE
ERA
Pisa
Budapest
Daily Med
TCM Gene DIT
Disea-some
SIDER
Drug Bank
UniProt
PDB
Pfam
LinkedCT
Enipedia
dbpedia lite
lingvoj
Lexvo
YAGO
UMBEL
Uberblic
Geo Species
Fishes of Texas
NASA (Data Incu-bator)
Taxon Concept
Geo Names
Freebase
DBpedia
iServe
OS
BNB
RDF Book Mashup
Revyu
Portu-guese DBpedia
Linked MDB
Event Media
Chronic-ling America
BBC Music
BBC Programmes
BBC Wildlife Finder
Tele-graphis
New York Times
URI Burner
Greek DBpedia
Open Calais
Rechtspraak.nl
OpenEI
education.data.gov.uk
Ordnance Survey
Openly Local
legislation data.gov.uk
UK Post-codes
data.gov.uk intervals
transport data.gov.uk
statistics data.gov.uk
LOIUS
Taxon Concept
Geo Names
World Fact-book
Freebase
DBpedia
Uberblic
dbpedia lite
Geo Species
Fishes of Texas
NASA (Data Incu-bator)
Eurostat (FUB)
Eurostat
Linked Sensor Data (Kno.e.sis)
riese
EUNIS
Twarql
WordNet (VUA)
WordNet (W3C)
WordNet (RKB Explorer)
Cornetto
totl.net
LODE
AGROV OC
SISVU
Affy-metrix
PubMed
Gene Ontology
SGD
Taxono my
UniProt (Bio2RDF)
PRO-SITE
ProDom
HGNC
STITCH
VIVO Cornell
VIVO UF
bible ontology
PBAC
ECCO-TCP
KISTI
NSF
JISC
LAAS
dotAC
ePrints
VIVO Indiana
RAE2001
Course-ware
IEEE
RISKS
IBM
DEPLOY
LOCAH
NVD
Roma
New-castle
IRIT
CiteSeer
Eurécom
UN/ LOCODE
ERA
Pisa
Budapest
Daily Med
TCM Gene DIT
Disea-some
SIDER
Drug Bank
UniProt
PDB
Pfam
LinkedCT
Enipedia
dbpedia lite
lingvoj
Lexvo
YAGO
UMBEL
Open Cyc
Linked Data
Eurostat (Ontology Central)
GovTrack
US SEC (rdfabout)
Semantic XBRL
US Census (rdfabout)
GeoWord Net
El Viajero Tourism
Ocean Drilling Codices
SMC Journals
Climbing
Linked GeoData
AEMET
Janus AMP
Yahoo! Geo Planet
Metoffice Weather Forecasts
Weather Stations
Airports
Product DB
Sears
GEMET
EARTh
Product Types Ontology
Open Data Thesau-rus
Linked Open Colors
Italian Museums
Google Art wrapper
Open Corpo-rates
Amster-dam Museum
medu-cator
Linked Open Numbers
Reactome
OGOLOD
UniPath way
Chem2 Bio2RDF
Homolo Gene
KEGG Com-pound
KEGG Glycan
KEGG Reaction
KEGG Pathway
KEGG Enzyme
KEGG Drug
Pub Chem
Medi Care
GeneID
MGI
InterPro
UniSTS
UniRef
UniParc
Smart Link
OMIM
ChEMBL
Linked Open Colors
SISVU
Affy-metrix
PubMed
Gene Ontology
SGD
AGROV OC
LODE
totl.net
Cornetto

Conclusions: the ‘web of data’ IS happening

- An increasing number of people and institutions are ‘opening’ their data using SW approaches
  - soon it may become a ‘requirement’ than any publicly funded cultural heritage resource publishes its data in raw format too

- The technological side of things is quite elaborated
  - complex architecture and technologies
  - still in evolution
  - requires collaboration with IT people

- Domain experts (eg historians) are badly needed:
  - they provide the expertise needed for formalising the ‘meanings’ of terms
  - IT people can’t make this vision reality by themselves
  - particularly relevant in humanities disciplines

3. SW approaches in digital history

SW approaches in history: summary

1) Work aimed at **creating ontologies** that characterise history at large, or some specific historical domain;

2) Digital systems that **use ontologies** as a knowledge representation that makes **inference tasks** more efficient and transparent

3) Digital system that **use ontologies** and other SW technologies in order to facilitate **data integration** and knowledge sharing

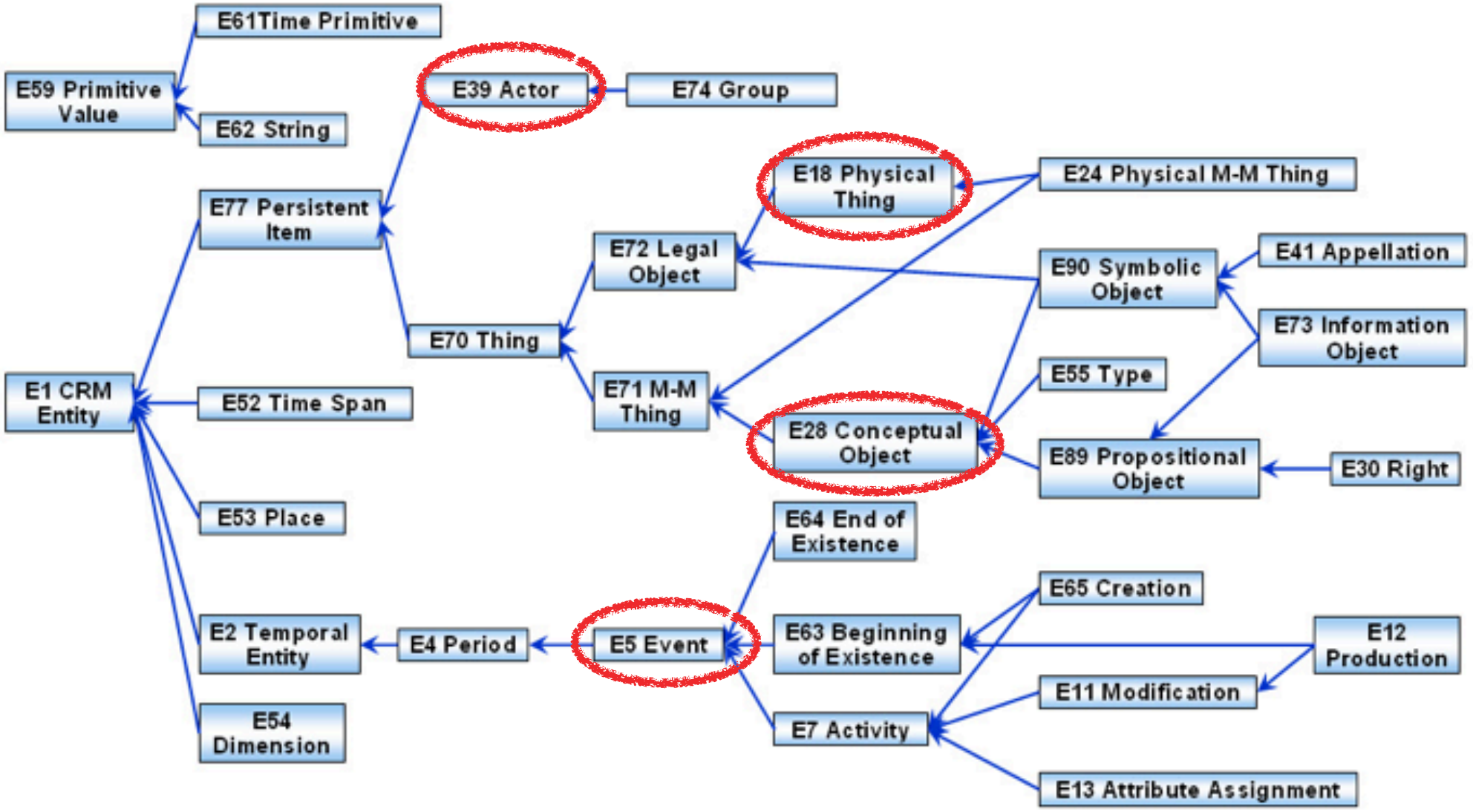
The CIDOC-CRM ontology

- A ‘semantic glue’ for cultural institutions
  - ontology aiming at bringing interoperability, provide the "semantic glue" needed to mediate between different sources of cultural heritage information
  - extensible, generic, focused on expressing the semantic contents of data such as that published by museums, libraries and archives.

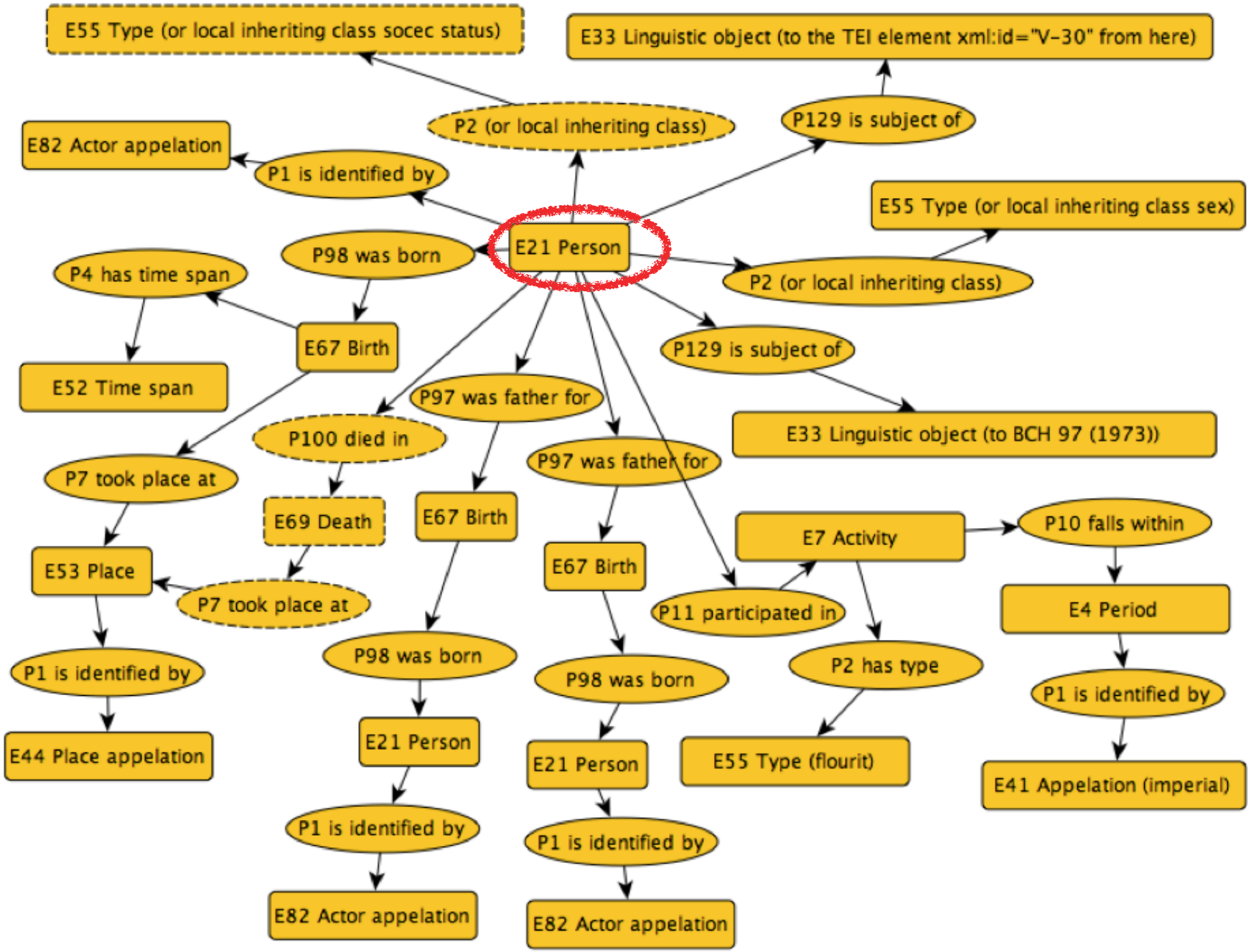
- A highly interdisciplinary work
  - originally emerged from the CIDOC Documentation Standards Group in the International Committee for Documentation of the International Council of Museums (1996)
  - has become the international standard (ISO 21127:2006) for the controlled exchange of cultural heritage information

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

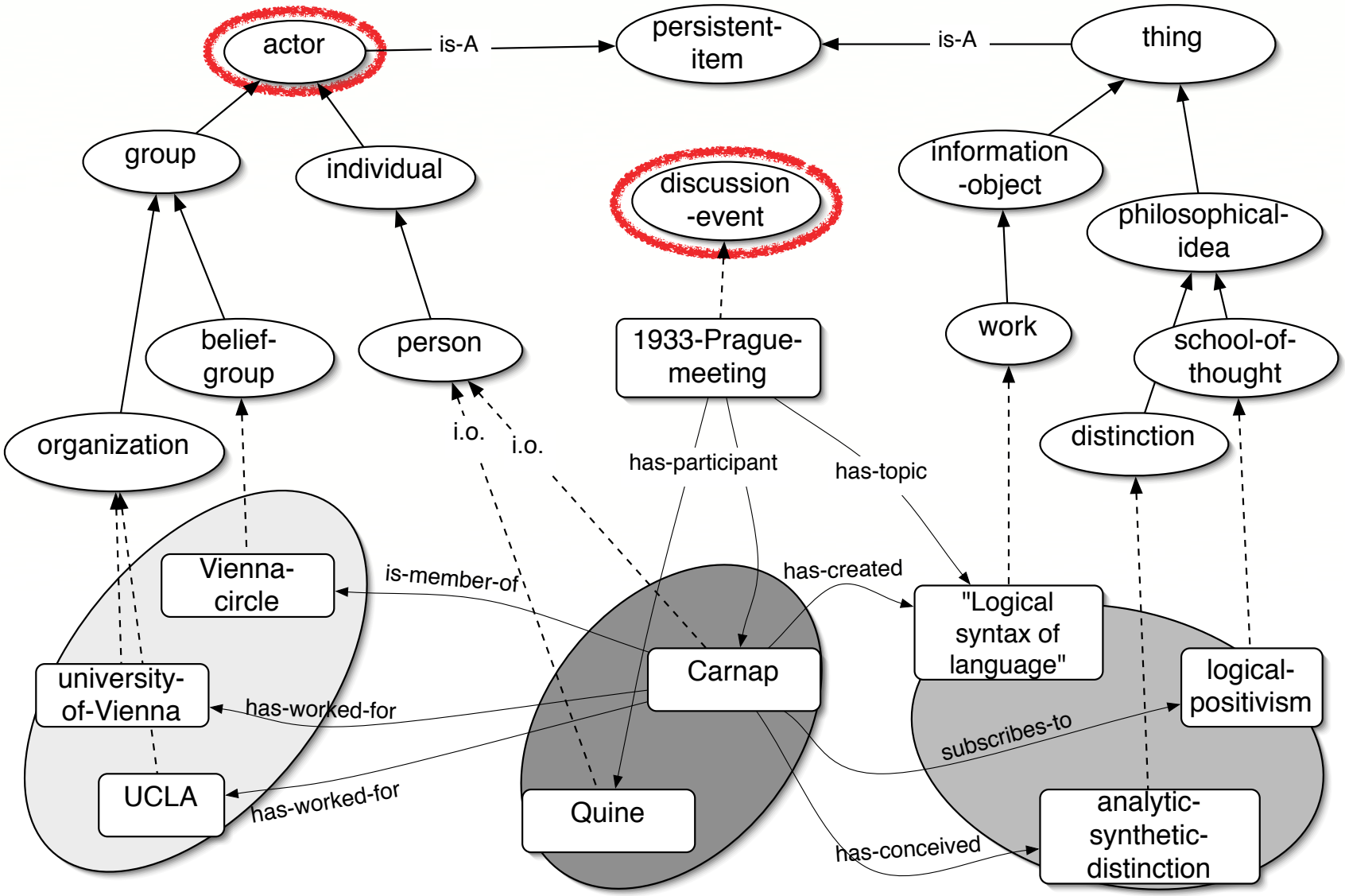
CIDOC-CRM: hierarchy of core classes



CIDOC-CRM: classes and relations



CIDOC-CRM: practical use via extension



http://philosurfical.open.ac.uk/

Henry III Fine Rolls project



- Home
- Project Information
- Fine Rolls
  - Translations
  - Images
  - Indexes
  - Search
- Style Book
- Commentary
- News
- Help
- Fine of the Month
- Book Publication
- Blog
- Conference
- Podcasts

Previous Next

3 HENRY III (28 October 1218–27 October 1219)

Fine Roll C 60/11 Contents

Membrane 12	Membrane 9	Membrane 5
Membrane 11	Membrane 8	Membrane 4
Membrane 10	Membrane 7	Membrane 4d.
Membrane 10d.	Membrane 6	Membrane 3

Fine Roll C 60/11, 3 HENRY III (1218–1219)

Images of roll C 60/11

Membrane 12

Image of membrane 12

Roll of fines from the third year of King Henry son of King John.

1 **30 Oct.** Westminster. *Norfolk.* Aubrey de Neville gives the king 2 m. for having a writ to attain summoned and taken at St. Edmund's before H. de Burgh, justiciar, between the same Aubrey security from Aubrey for those 2 m. to the king's use for this writ. [1] Witness the earl.

1. To judge from the compressed layout of the entries and the uniformity of the hand this rol

http://www.finerollshenry3.org.uk/home.html

Henry III Fine Rolls project: main info

- AHRC project (2009)

  - goal: publish in both print and digital edition the parchment rolls compiled between 1216 and 1248, which record mainly (but not only) offers of money made to King Henry III of England in exchange for a wide range of concessions and favours.

  - collaborative venture between King's College London and The National Archives of the United Kingdom

- Different types of 'metadata' for the rolls

  1) the physical structure of the roll—for instance, the fact that it is composed of a series of membranes stitched together;

  2) the structure of the English calendar, a concise translation of the Latin records, including county and date information concerning the record, body of each entry and witness lists;

  3) the semantic content of the roll—for instance, names of individuals, names of locations, and key themes mentioned in the text.

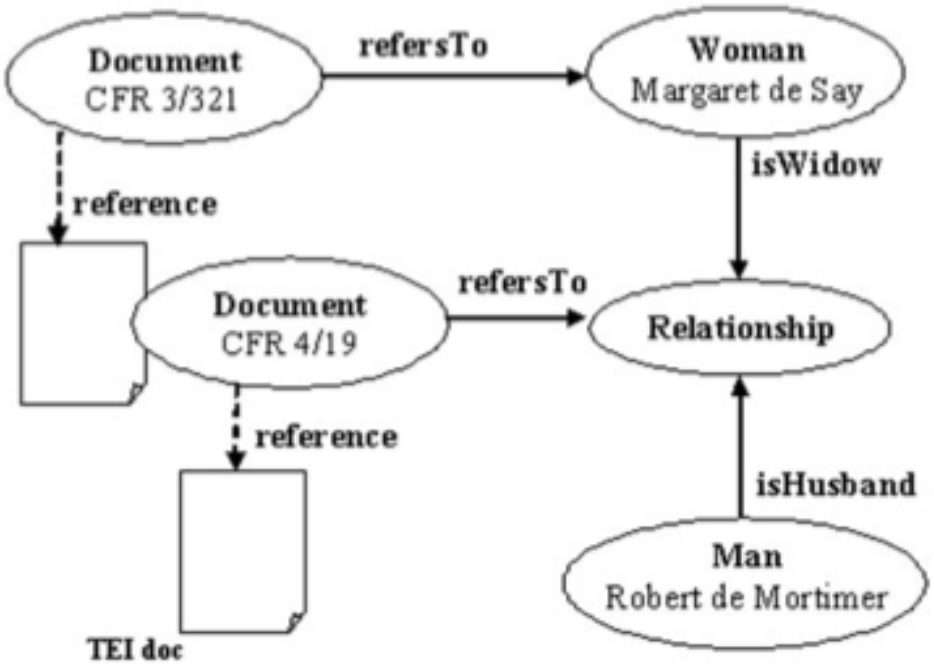
http://www.finerollshenry3.org.uk/home.html

Henry III Fine Rolls project: ontology

- Ontology as a ‘representation’ device

  - to express complex associations between entities in historical texts that have been marked up in XML, according to the Text Encoding Initiative guidelines.

  - for facilitating the interpretation of implicit and hidden associations in the sources of interest

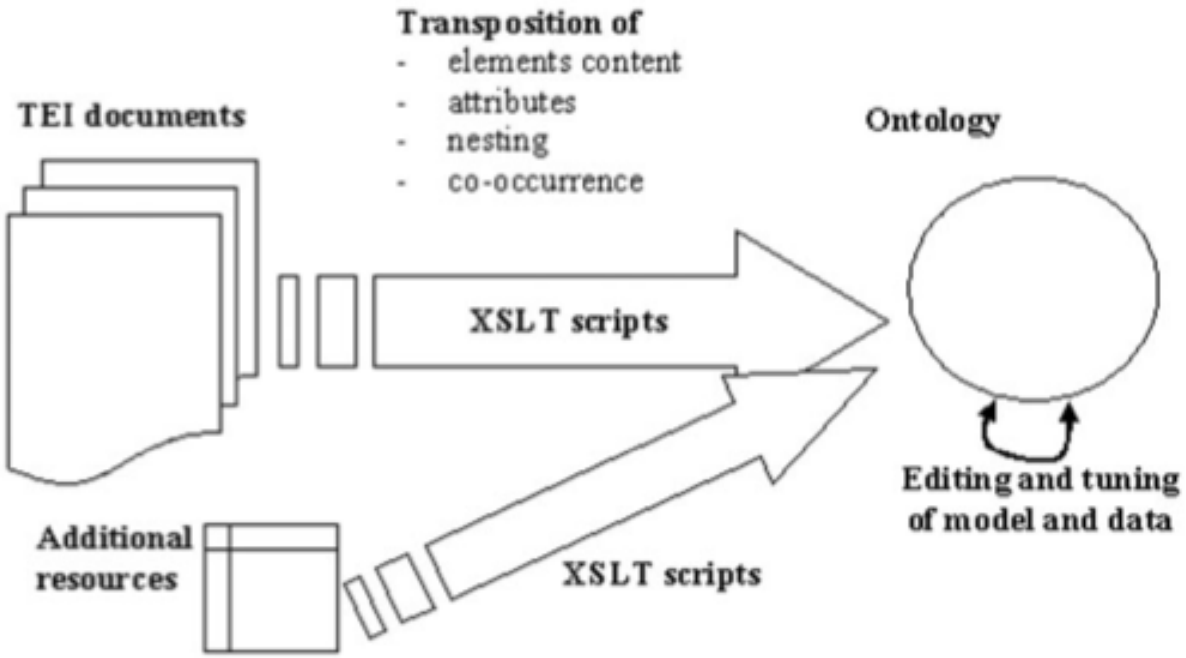


Henry III Fine Rolls project: ontology

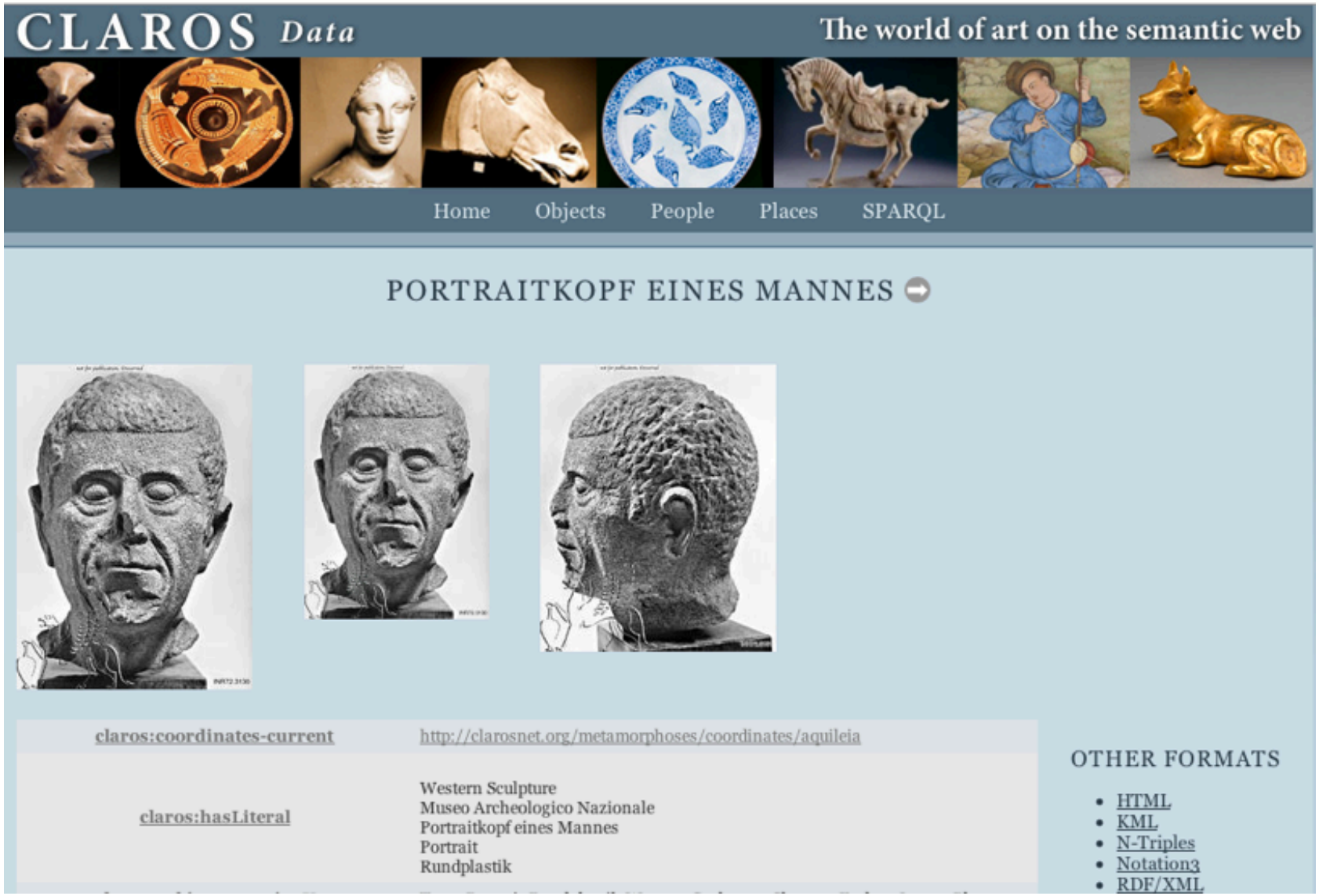
- Ontology as a ‘representation’ device

  - to express complex associations between entities in historical texts that have been marked up in XML, according to the Text Encoding Initiative guidelines.

  - for facilitating the interpretation of implicit and hidden associations in the sources of interest



Claros: SW for classical art



www.clarosnet.org/

Claros: SW for classical art

- Collaborative research initiative led by the University of Oxford

  - goal: use datasets in Classics and Classical Art to exploit the potential of ICT for public service

  - International data federation project: Faculty of Classics, Oxford, Beazley Archive, Lexicon of Greek Personal Names, University of Cologne, Arachne, Research Sculpture Archive, German Archaeological Institute, Berlin Archaeological Institute, Berlin Lexicon Iconograhicum Mythologiae Classicae, Paris.

- 2 million records and images in total

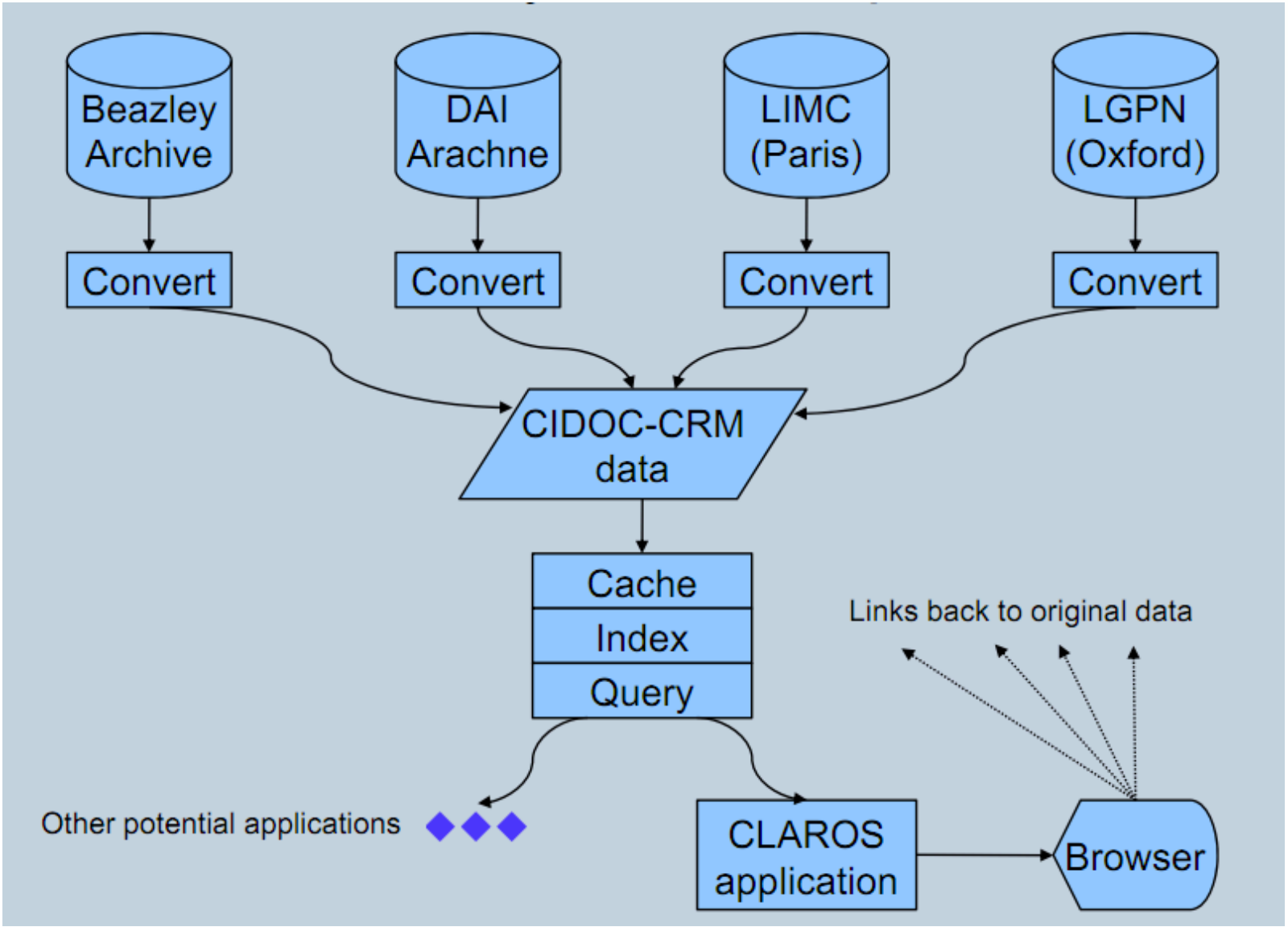
  Pottery records, Engraved gem and cameo records, Plaster casts records , Antiquarian photographs, information about individuals and names, Sculpture images, images of mythological and religious records, iconography etc..

- Was possible thanks to Semantic Technologies

No changes required to existing databases or programs. Interchange of of data is achieved by export of underlying data to CIDOC-CRM.

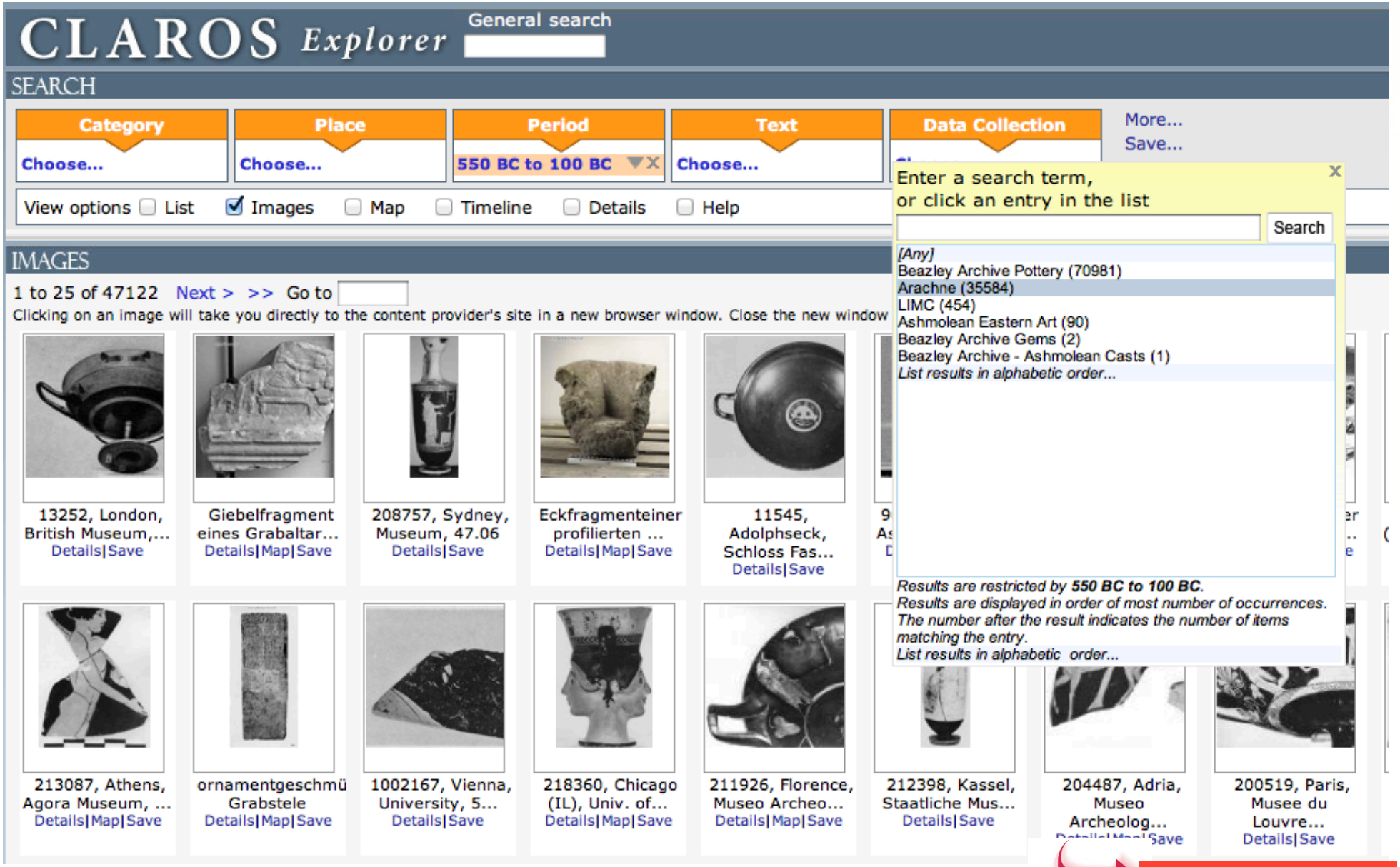
www.clarosnet.org/

Claros: SW for classical art



Adapted from “Digital imaging: objects. The Beazley Archive, CLAROS and the world of ancient art” presentation slides

Claros: example of an integrated search



www.clarosnet.org/

Europeana: SW on a large scale



Explore Europe's cultural collections

einstein

Search

Refine your search

Help

Matches for: einstein

Filter your search:

- By media type
  - IMAGE (334)
  - SOUND (76)
  - TEXT (70)
  - VIDEO (36)
- By language
- By date
- By country
- By copyright
- By provider

Results 1 - 12 of 516 Page: 1 2 3 4 5 6 →

Like



Discussion Baffie et Einstein



Colpi d'obiettivo: Princeton. Milano. To...



Elementarbuch der Hebräischen Sprache



Portrait of Erwin F Freundlich (188...



Albert Einstein, cosmos and religion



Colpi d'obiettivo. Princeton. Roma.



Professor Doctor Hendrik Antoon Lorentz



Blue and White Pa Documenting the

http://www.europeana.eu/

Europeana: SW on a large scale

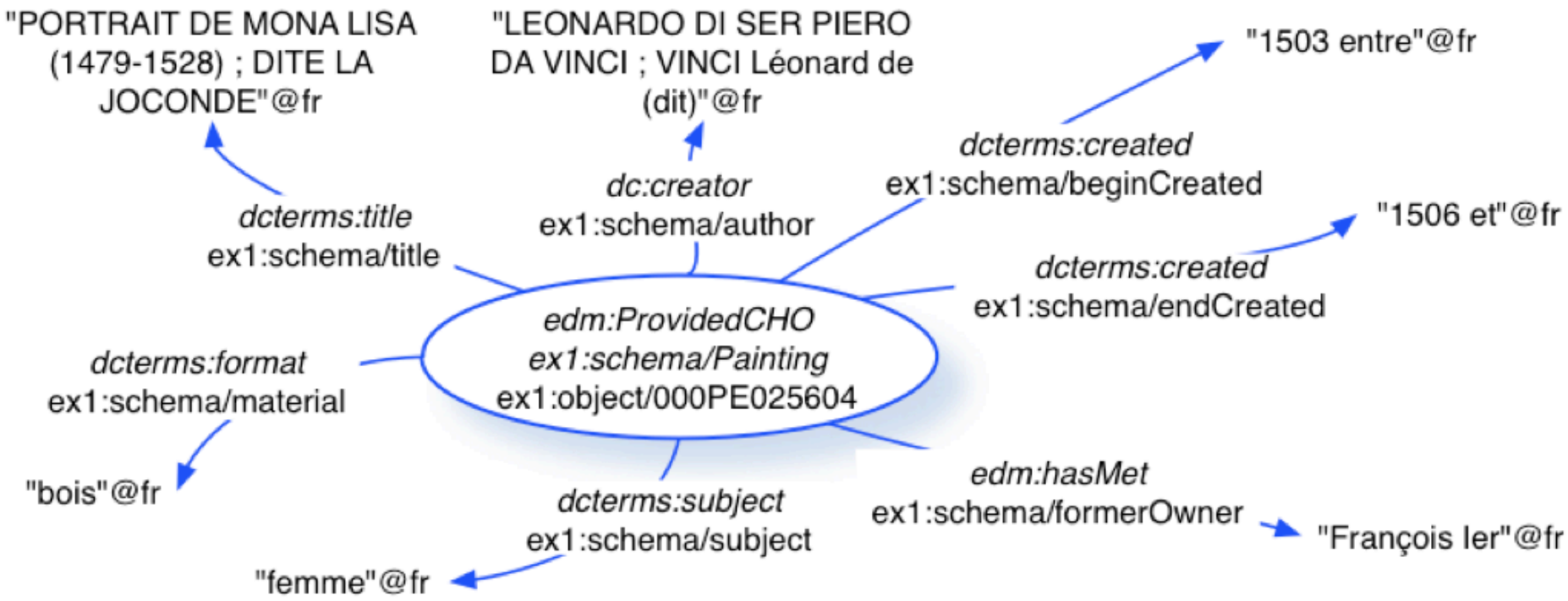
- Huge EU project (2008)
  - an interface to millions of books, paintings, films, museum objects and archival records that have been digitised throughout Europe.

- Approach similar to Claros, but on a larger scale
  - Around 1500 institutions across Europe have contributed to Europeana.
  - assembled collections let users explore Europe's cultural and scientific heritage from prehistory to the modern day.

- Several ontologies have been used/created

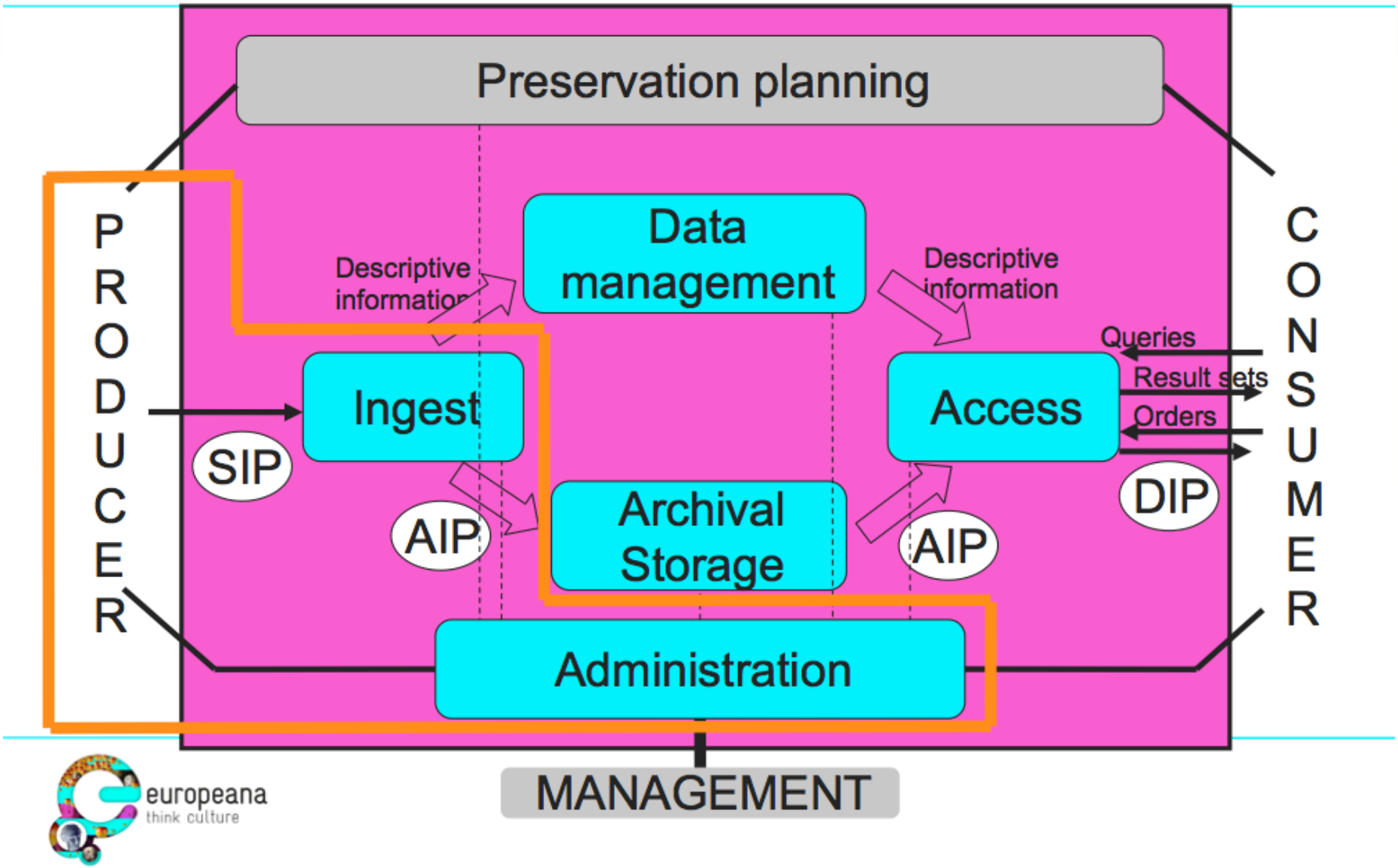
http://www.europeana.eu/

Europeana: ontologies for data integration



Adapted from Europeana Data Model Primer, 2011, http://www.europeana-libraries.eu/web/europeana-project/technicaldocuments/

Europeana: system design

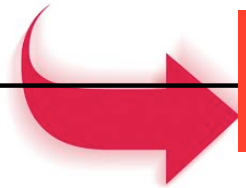


Adapted from Content ingestion, Master Class session, The Europeana Plenary Conference: Creation, Collaboration and Copyright: September 14/15 2009

4. Hands on session: find a use-case for your own ‘semantic’ mash-up!

Hands on session..

Source	Rationale	Mash-up
eg Claros	extract all pieces constructed in Egypt between 100 and 200 BC	we can.....
eg Europeana	extract all documents describing social life in Egypt between 100 and 200 BC	



http://goo.gl/Ebhzl

Thanks for the attention

Questions?