

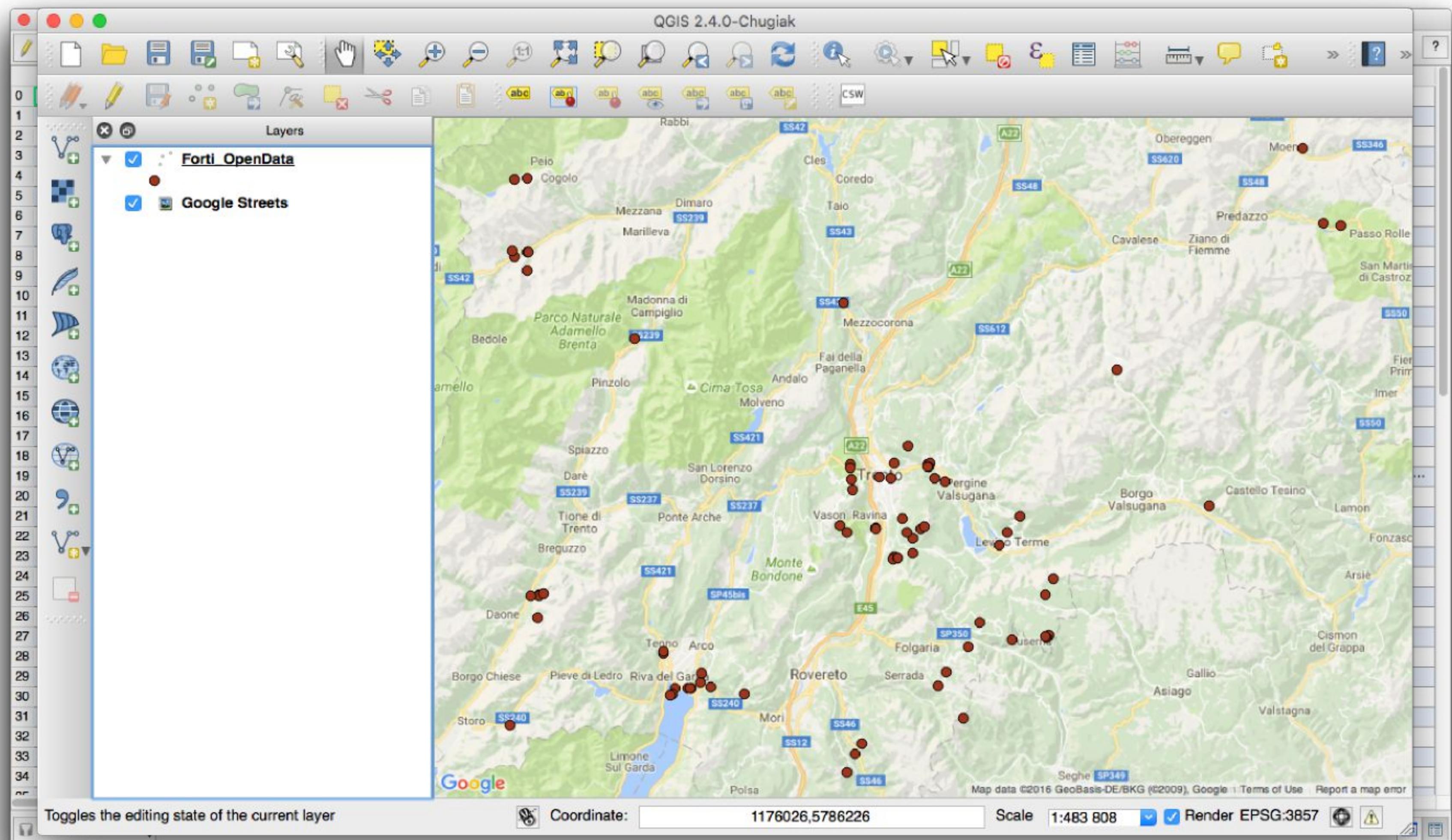
Information and its representation



Stratigraphy units









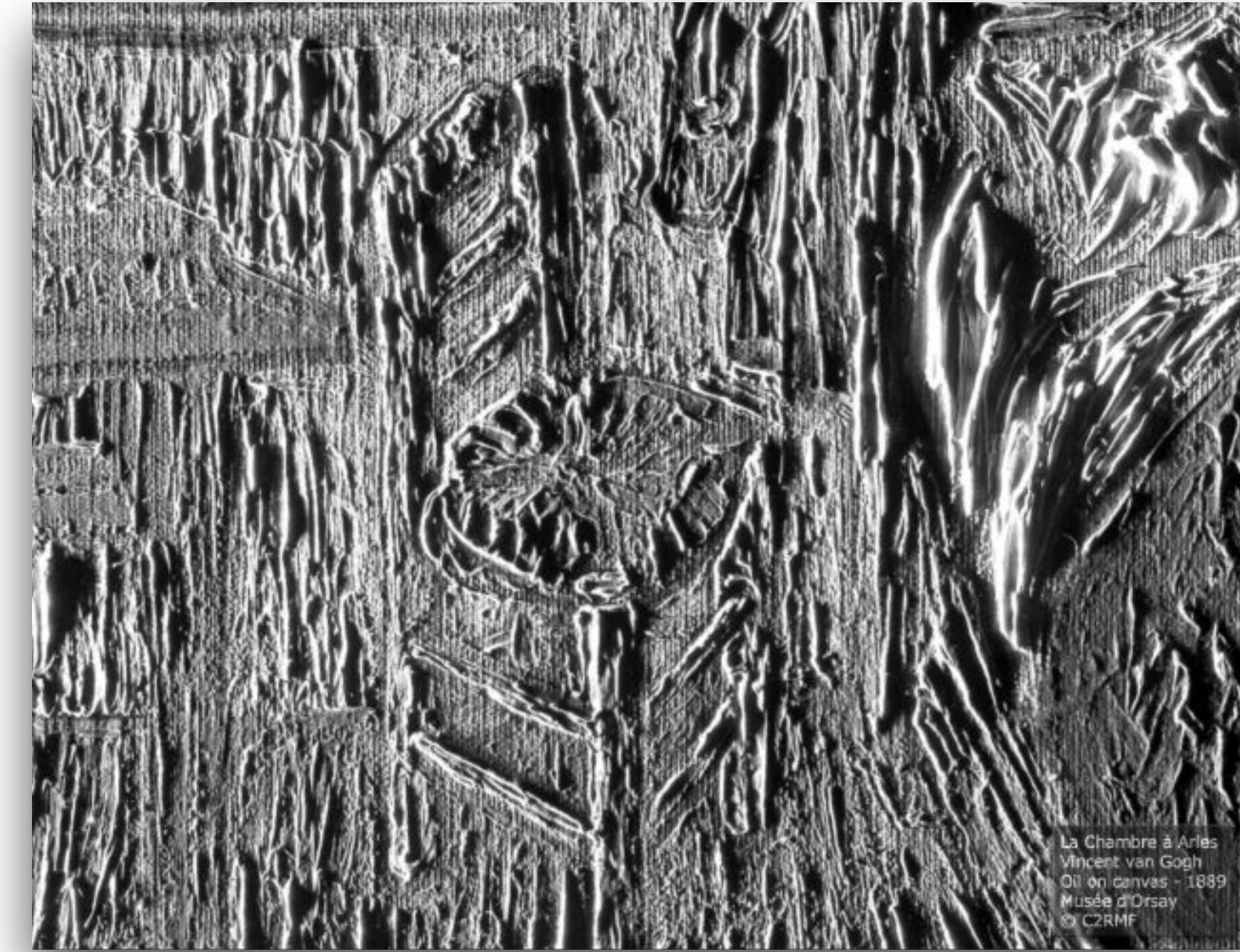
X-Ray



Original

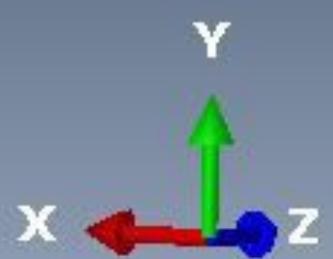
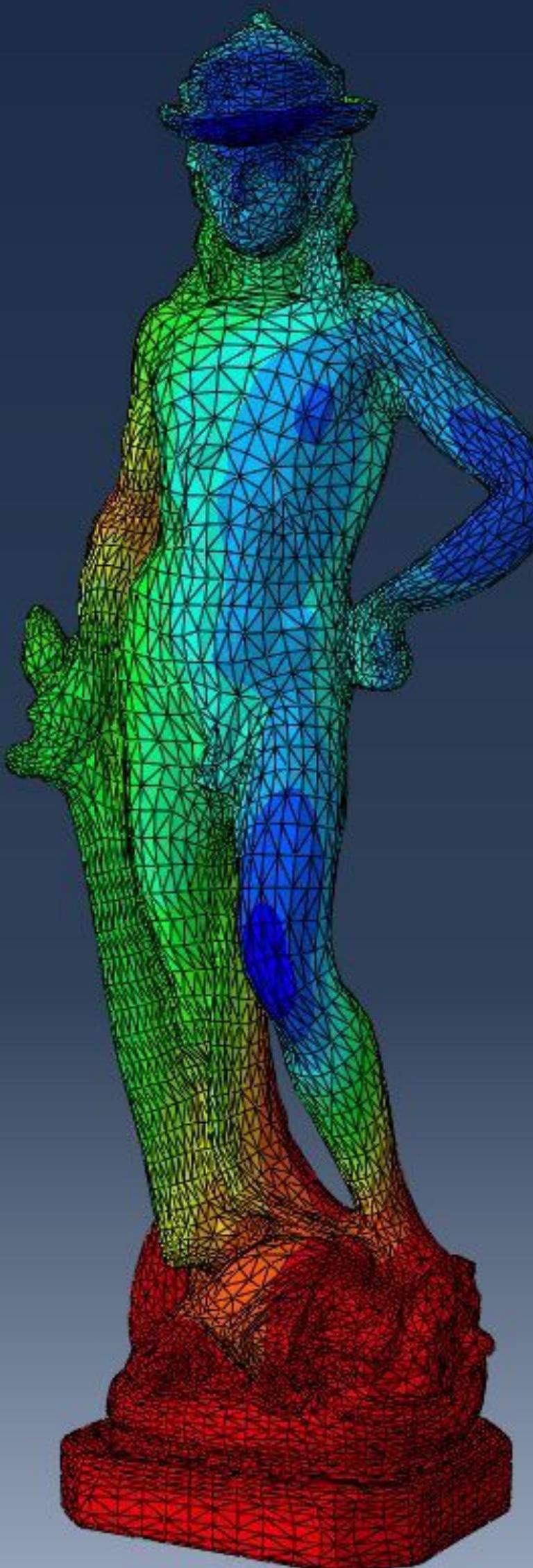
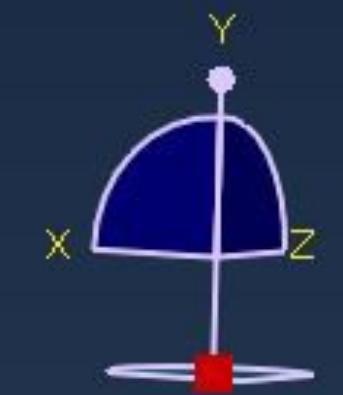


Raking Light



Infrared



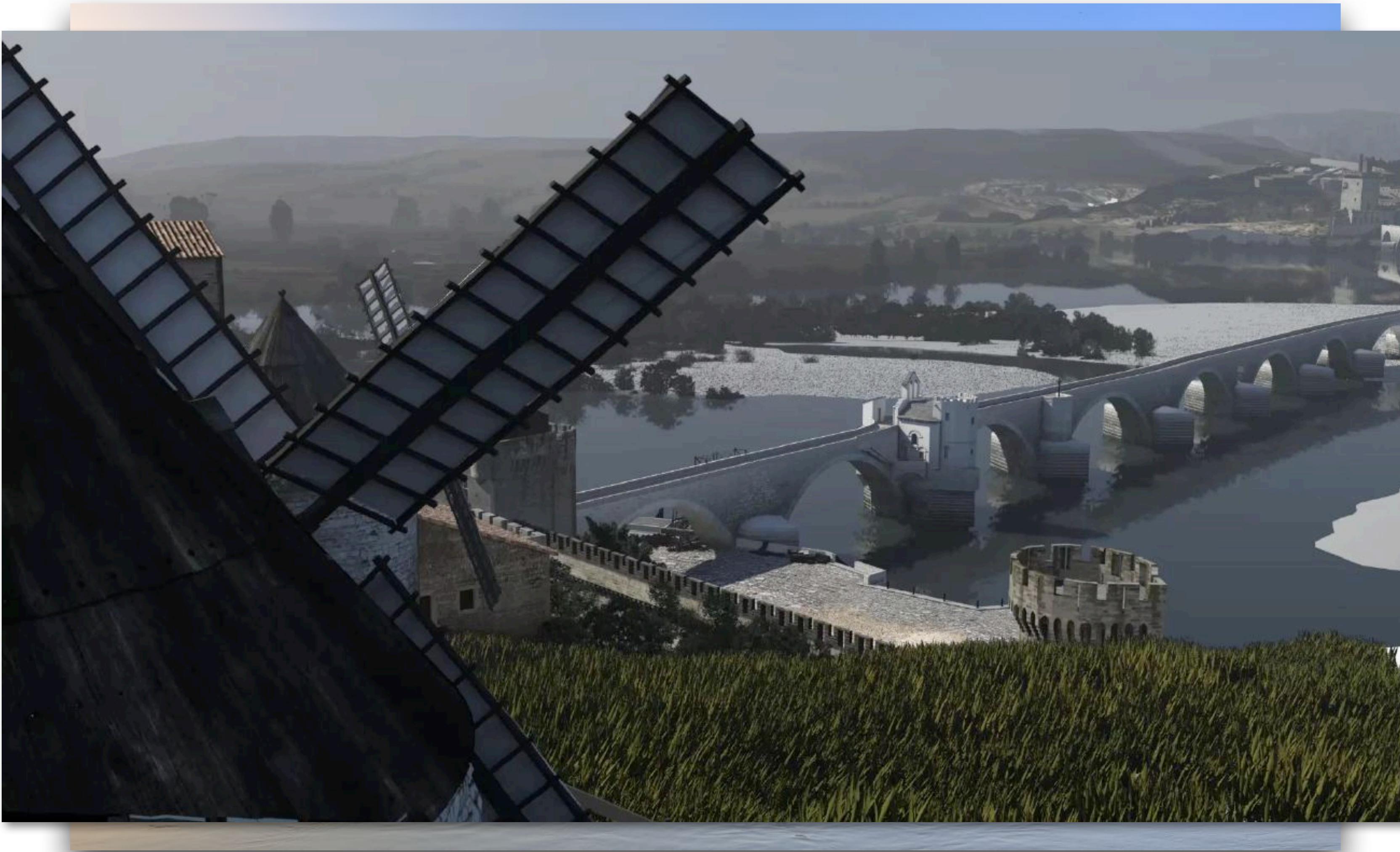


ODB: david.odb Abaqus/Standard 6.14-5 Mon Jun 27 17:10:59 ora legale Europa occidentale 2016

Step: Step-1
Increment 100: Step Time = 1.000
Primary Var: U, U2







Description :

Date de création : 1878

Nombre total de personnes actuellement occupées dans l'entreprise

160 pour DAUM France

20 pour DAUM Etats-Unis et Asie

CRISTALLERIE			
Etapes de la production	Description	Mode d'apprentissage (diplômes, formation en entreprise...)	N. de personnes dédiées à cette phase
Développement du produit et modelage	Les nouveaux produits sont développés au sein de l'atelier de création, trois alternatives de production de nouveaux objets en pâte de verre sont possibles : à partir d'une sculpture existante, qu'il faut adapter pour la reproduction en pâte de verre ; en collaboration avec un designer, le projet du designer est mis en forme en créant un premier exemplaire en plâtre ; création interne, sur la base d'un cahier des charges du directeur artistique on crée un premier exemplaire en plâtre. Il s'agit généralement d'objets à vocation utilitaire.	CAP, BEP, Bac Pro métiers de la céramique BTS modelage + formation interne	5 à Nancy 2 à Nancy 2 à VLC 1 à Nancy 6 à VLC 3 à VLC
Moulage	A partir du modèle, on l'étudie et on ajoute les éléments techniques nécessaires	Idem	2 à VLC



Joseph Bovet et ses élèves, Ecole Normale d'Hauterive – vers 1935 (@ Glasson/Musée gruérien, Bulle)

Les Fribourgeois sont à ce point épris de vocalises que l'on compte près d'un chanteur « organisé » pour 35 habitants dans ce canton. Chœurs d'église et chœurs profanes, mixtes ou non ; chœurs de jeunes et de moins jeunes rossignols, aux visées professionnelles ou ludiques : la Fédération fribourgeoise des Chorales rassemble près de 7'200 chanteurs, actifs dans 234 ensembles distincts... sans compter les formations éphémères, qui voient le jour autour de projets ponctuels, et les ensembles informels qui pratiquent le chant hors des structures associatives !

Cette densité exceptionnelle s'explique par une tradition séculaire solidement ancrée dans l'histoire régionale. Si le mouvement des Céciliennes s'est développé dans tous les cantons catholiques, c'est en effet à Fribourg – dans une société rurale fermement encadrée par le clergé – qu'il a trouvé son meilleur terrain. Un mouvement choral s'est cependant aussi développé indépendamment du contexte religieux – et parfois en réaction à ce dernier. La figure de l'abbé Joseph Bovet (1879–1951), a permis de fédérer les différents sons de cloche du canton, et son charisme régnait longtemps sur la vie chorale de toute la région.

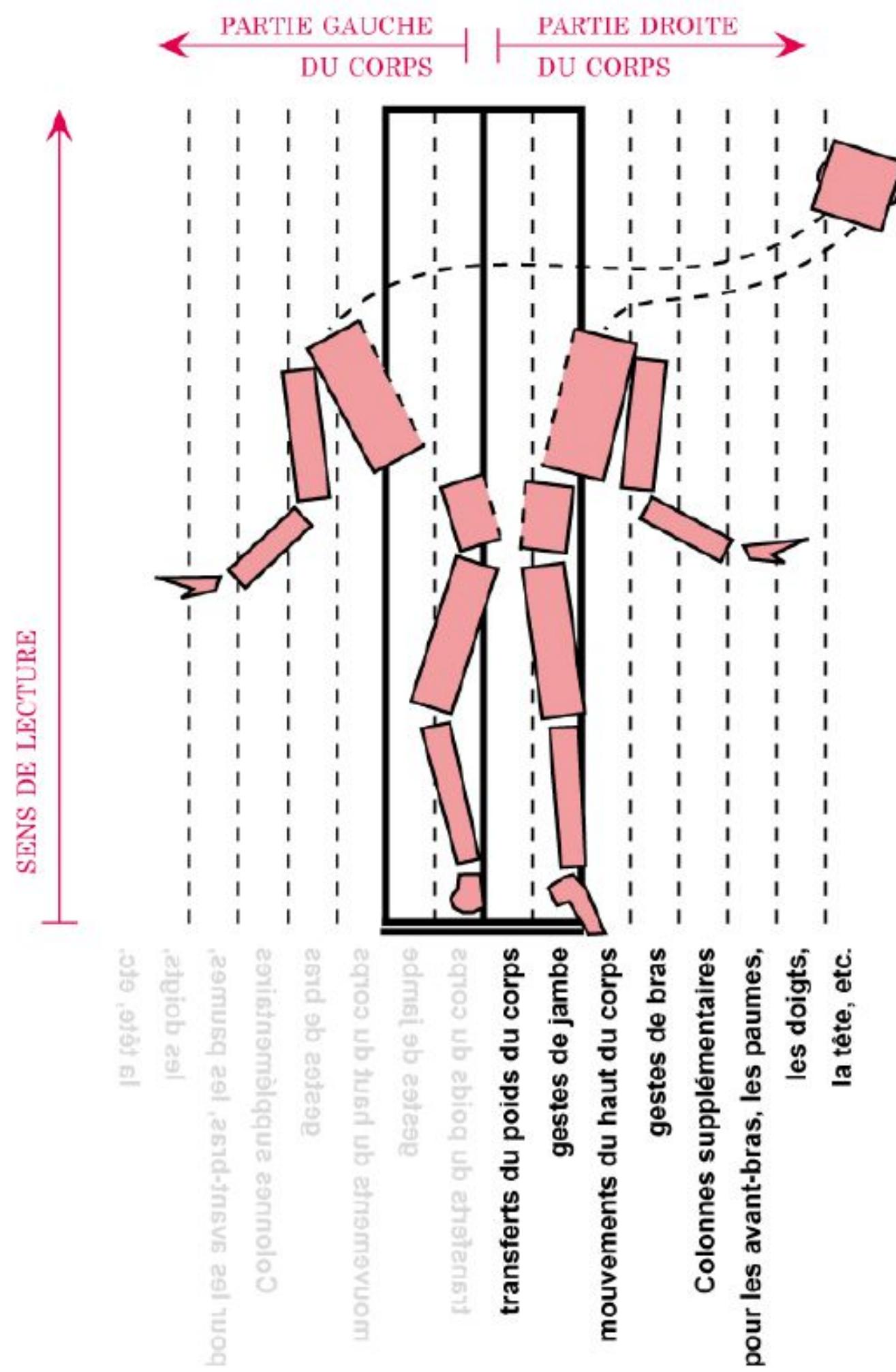
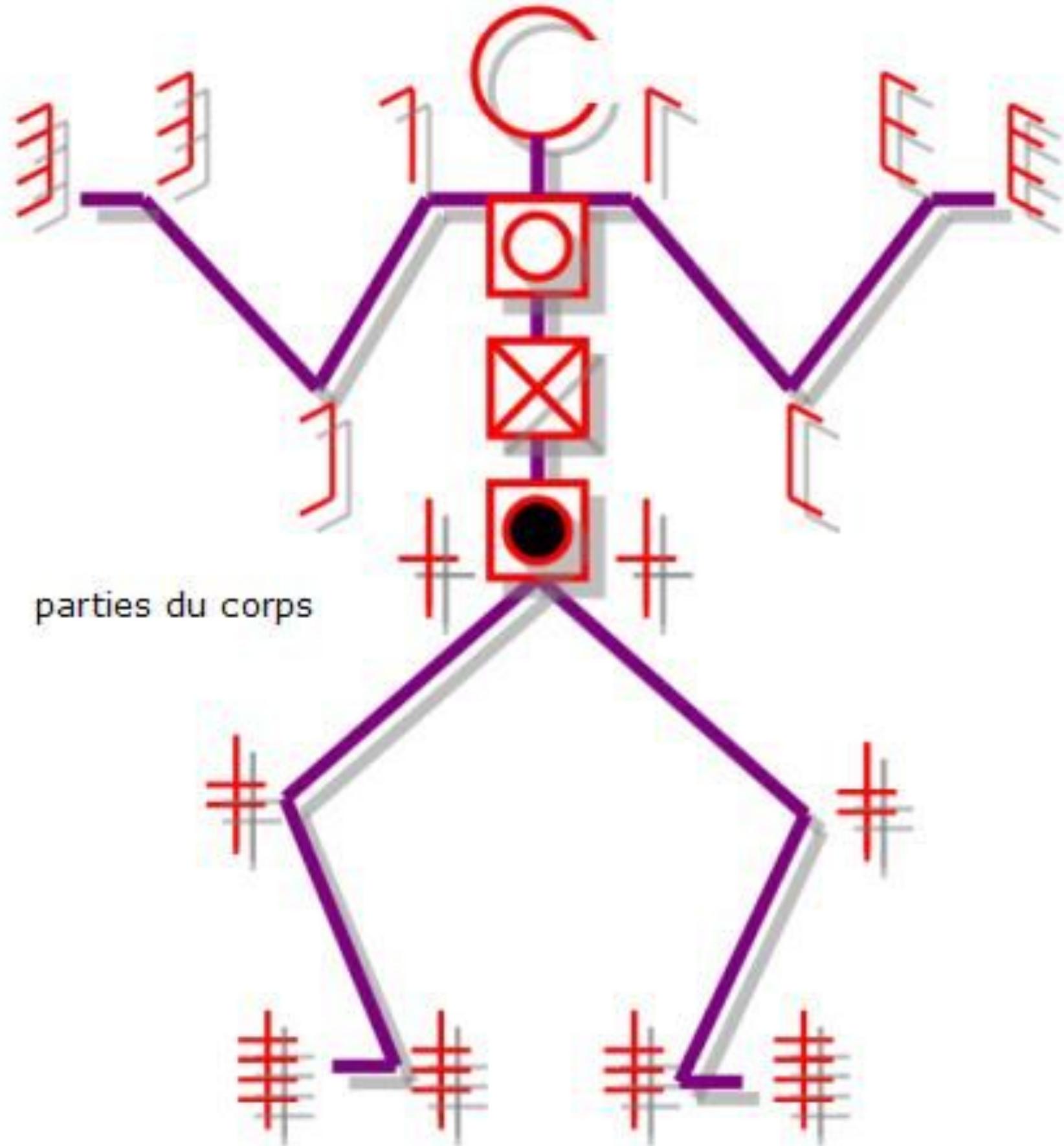
AT	ATTORE INDIVIDUALE	
ATT	ATTORE	
ATTI	Ruolo	#capocandeliere#
ATTZ	Nazionalità	italiana
ATTN	Nome	Pinna Antonio #Cico#
ATTS	Sesso	M
ATTE	Età	39
ATA	Annotazioni	Dirige i #portatori# con la voce, i movimenti del corpo e con l'ausilio de #li bacchetti# (un mazzo di spighe di grano inframmezzate da fiorellini bianchi). Prende parte alla #vestizione# e guida il #ballo# del candeliere addobbato.
AT	ATTORE INDIVIDUALE	
ATT	ATTORE	
ATTI	Ruolo	tamburino #tamburinaggiu#
ATTZ	Nazionalità	italiana
ATTN	Nome	Pinna Andrea
ATTS	Sesso	M

1. Història de la festa

La Festa Major i el Ball dels Cavallets, Gegants i Mulassa de Sant Feliu de Pallerols se celebra pels voltants de la diada de Pasqua Granada. Aquesta festa mòbil, que s'escau entre mitjan maig i mitjan juny, s'allarga cinc dies. S'inicia el divendres anterior a la diada de la Pasqua Granada i s'acaba el dimarts següent; enmig, - el dilluns- se celebra la Festa del Roser, de gran rellevància a la vila. A Sant Feliu, com en d'altres indrets de Catalunya, la Confraria de la Mare de Déu del Roser va arrelar amb força, sobretot a partir del segle XVI -després de la batalla de Lepant- i la devoció a la Mare de Déu del Roser va créixer considerablement. De fet, la Festa Major se celebra per la Festa del Roser, mantenint-se l'anomenat Ofici del Roser, nom amb que és coneguda la Missa de Festa Major.

Tot i seguir l'estructura generalitzada de les festes majors catalanes, la de Sant Feliu de Pallerols presenta alguns trets diferenciadors, el més emblemàtic dels quals és el *Ball dels Cavallets, Gegants i Mulassa*. Aquest ball, únic a Catalunya, és indissociable a la Festa Major del poble i representa el conjunt més complet d'entre tots els balls que, amb peces d'imatgeria, han arribat fins els nostres dies.

Els orígens de el *Ball dels Cavallets, Gegants i Mulassa* són incerts i desconeguts. La successió de guerres napoleòniques, carlines i civil que van afectar la zona de St. Feliu de Pallerols va comportar la desaparició de documentació i amb ella la impossibilitat de datar esdeveniments especials i fets històrics transcendentals. No obstant, diferents teories i hipòtesis -tant de caràcter llegendarí com teatral- sustenen la interpretació i motivació del Ball i algunes dades recollides en diferents indrets, han permès aproximar l'origen de la dansa.

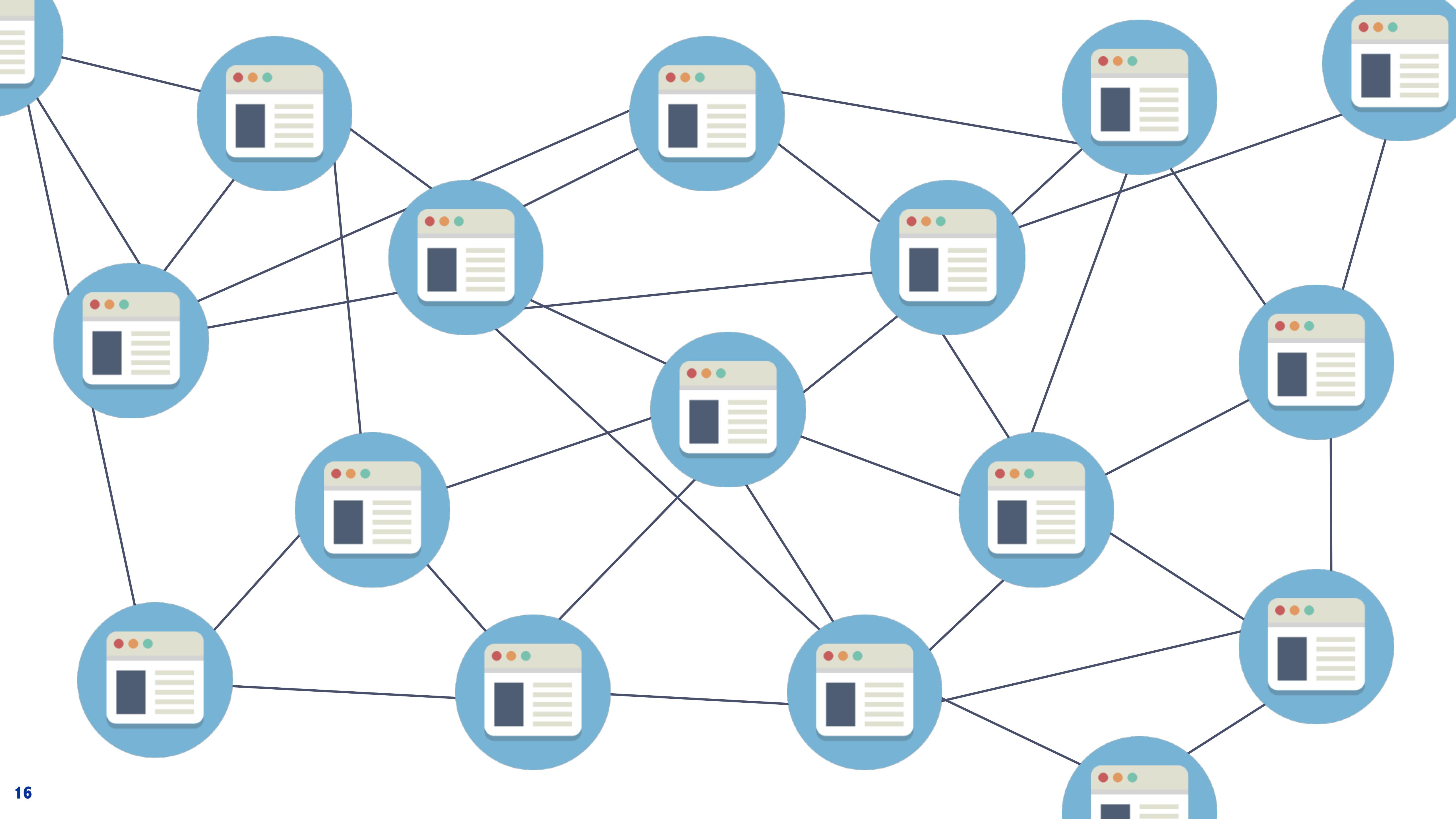


Colonnes pour les indocations complémentaires (orientations, parcours, etc.)

Structure not meaning

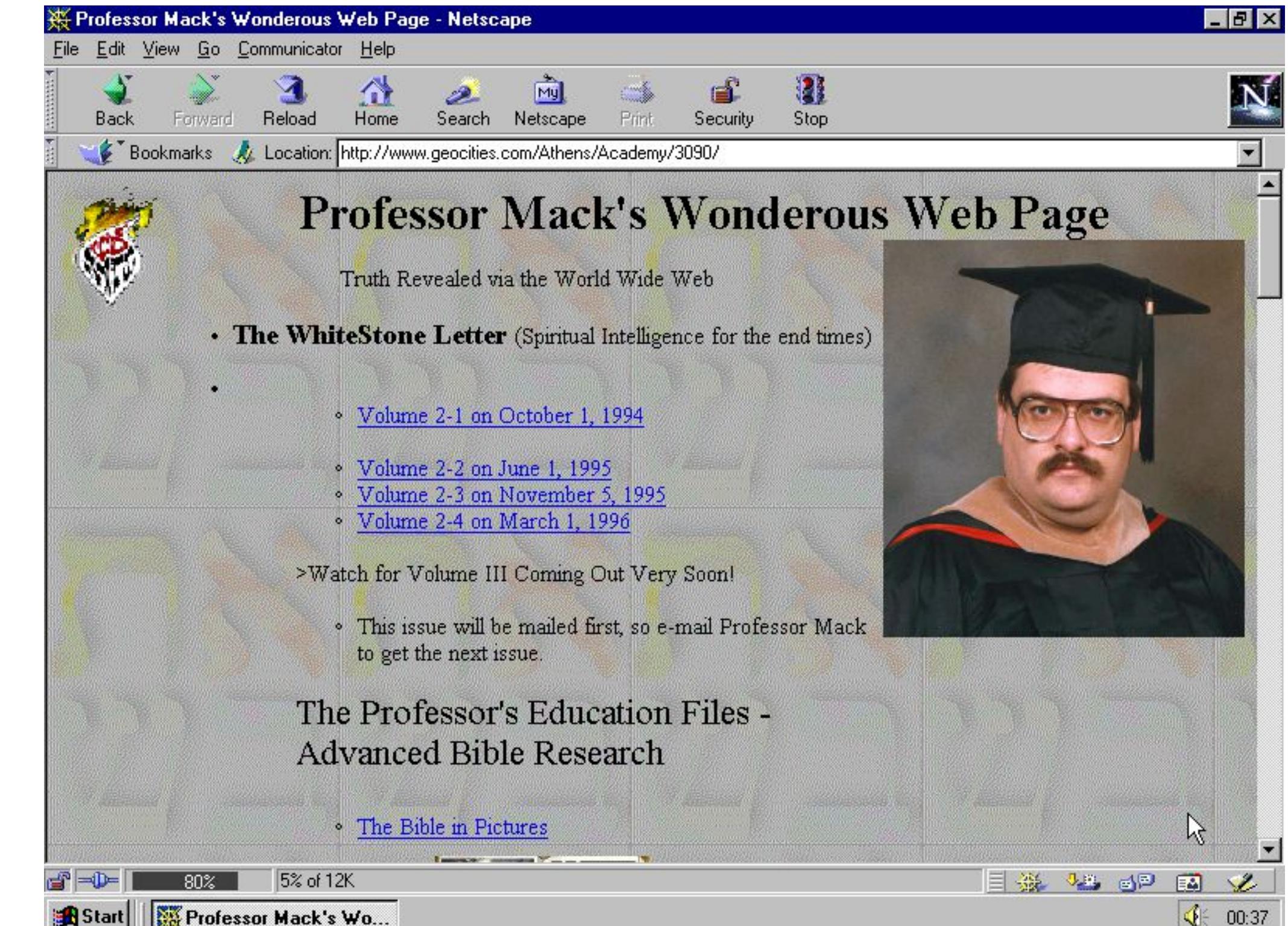
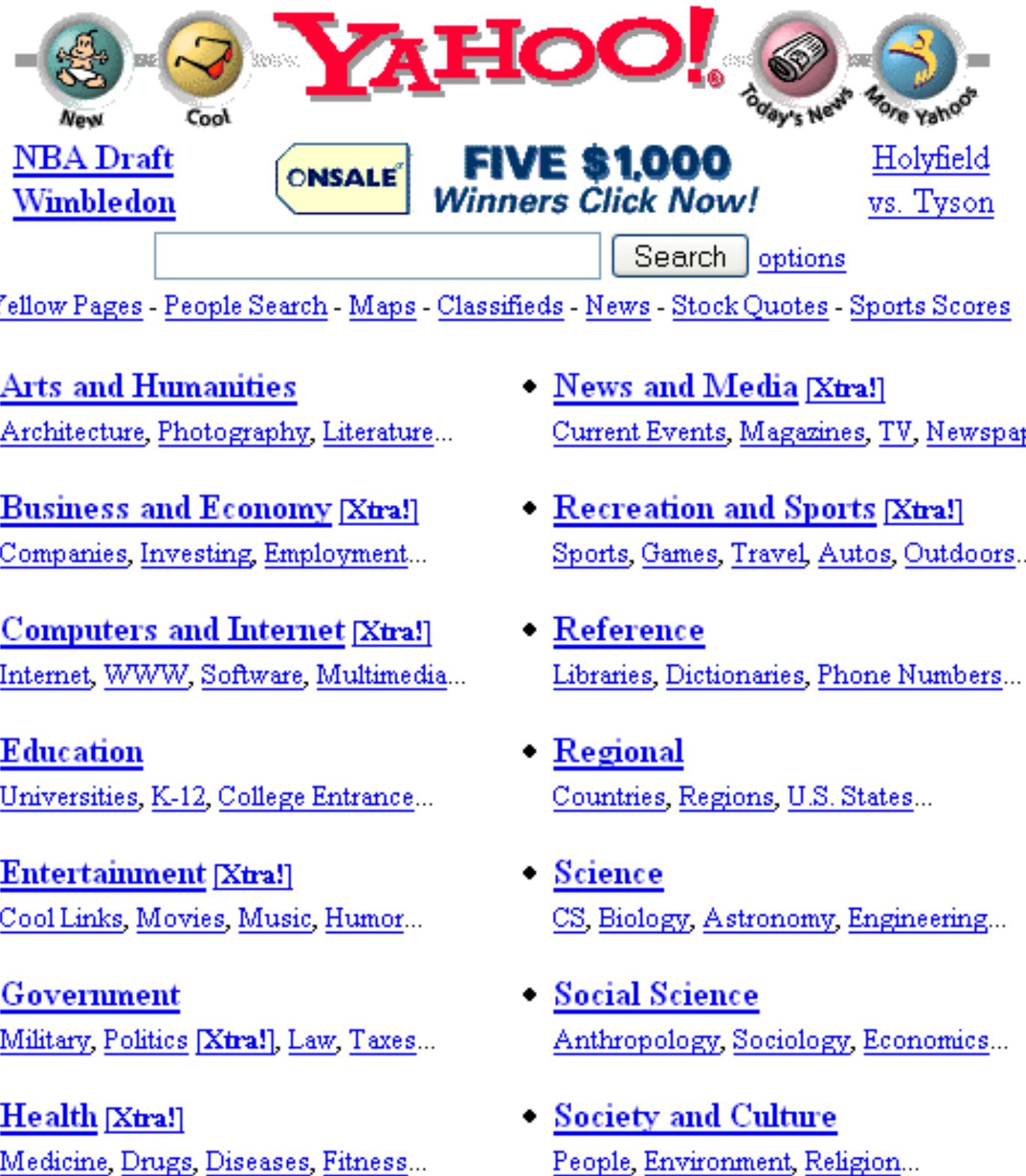






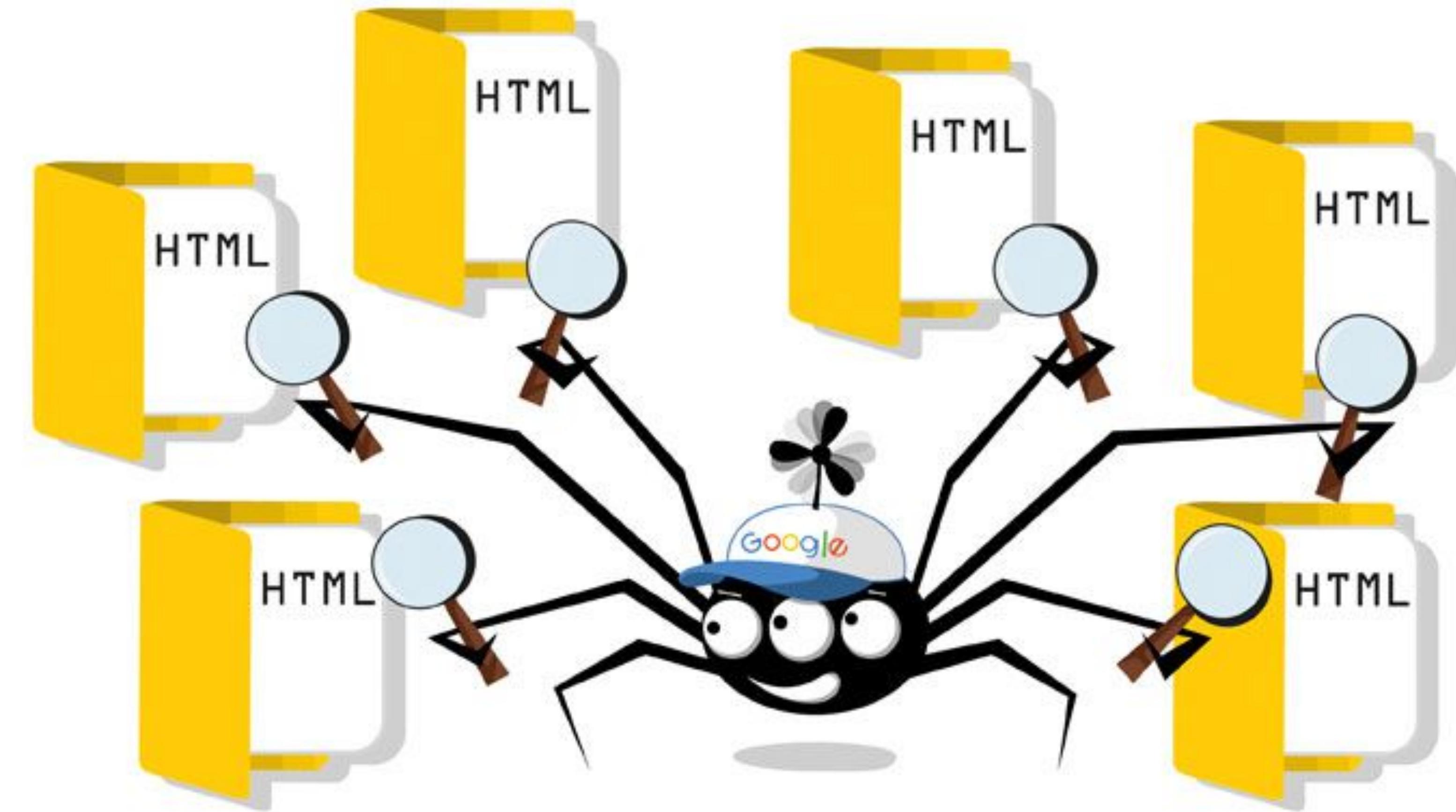


Type of information



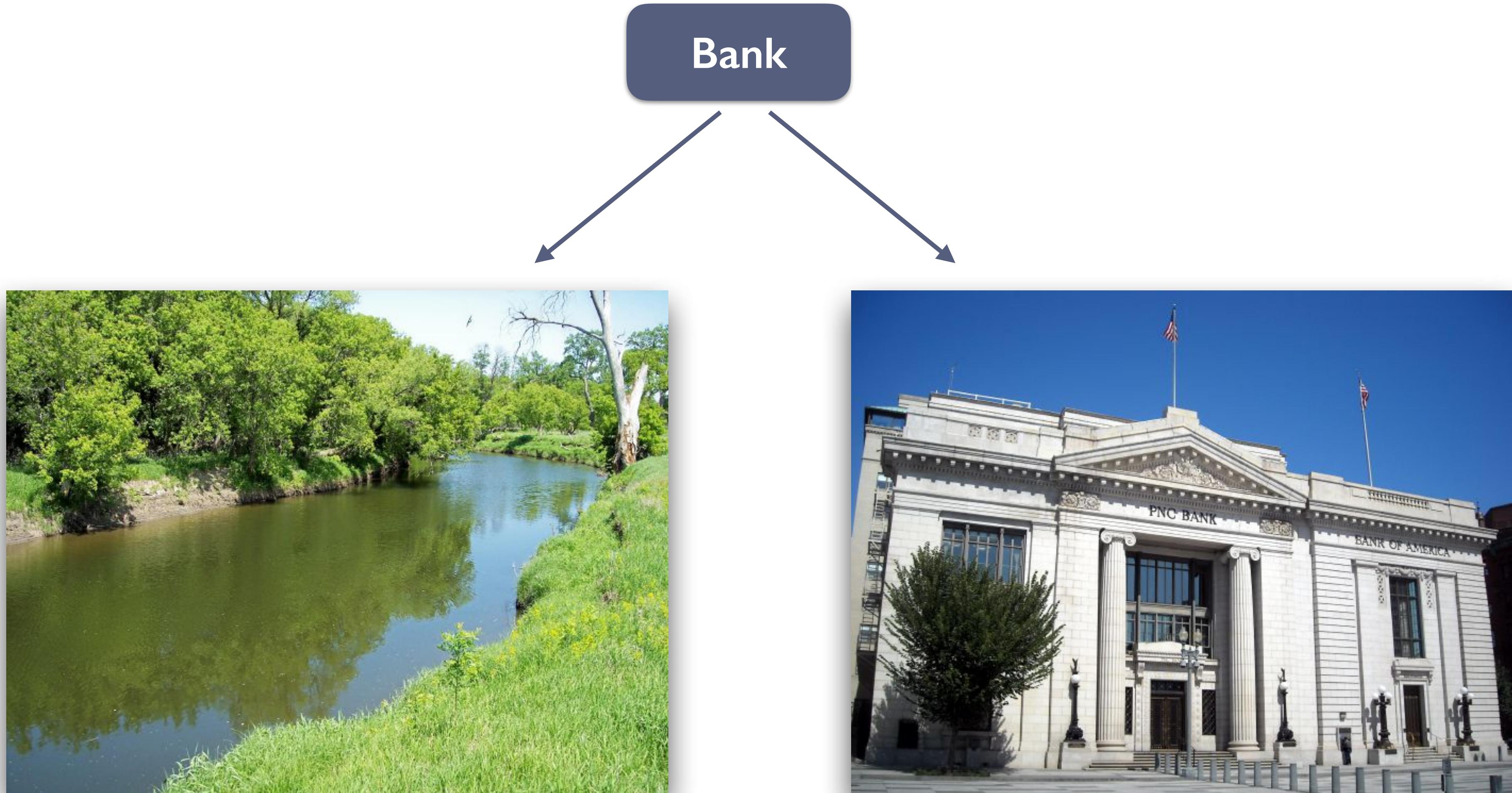
Directory

Information



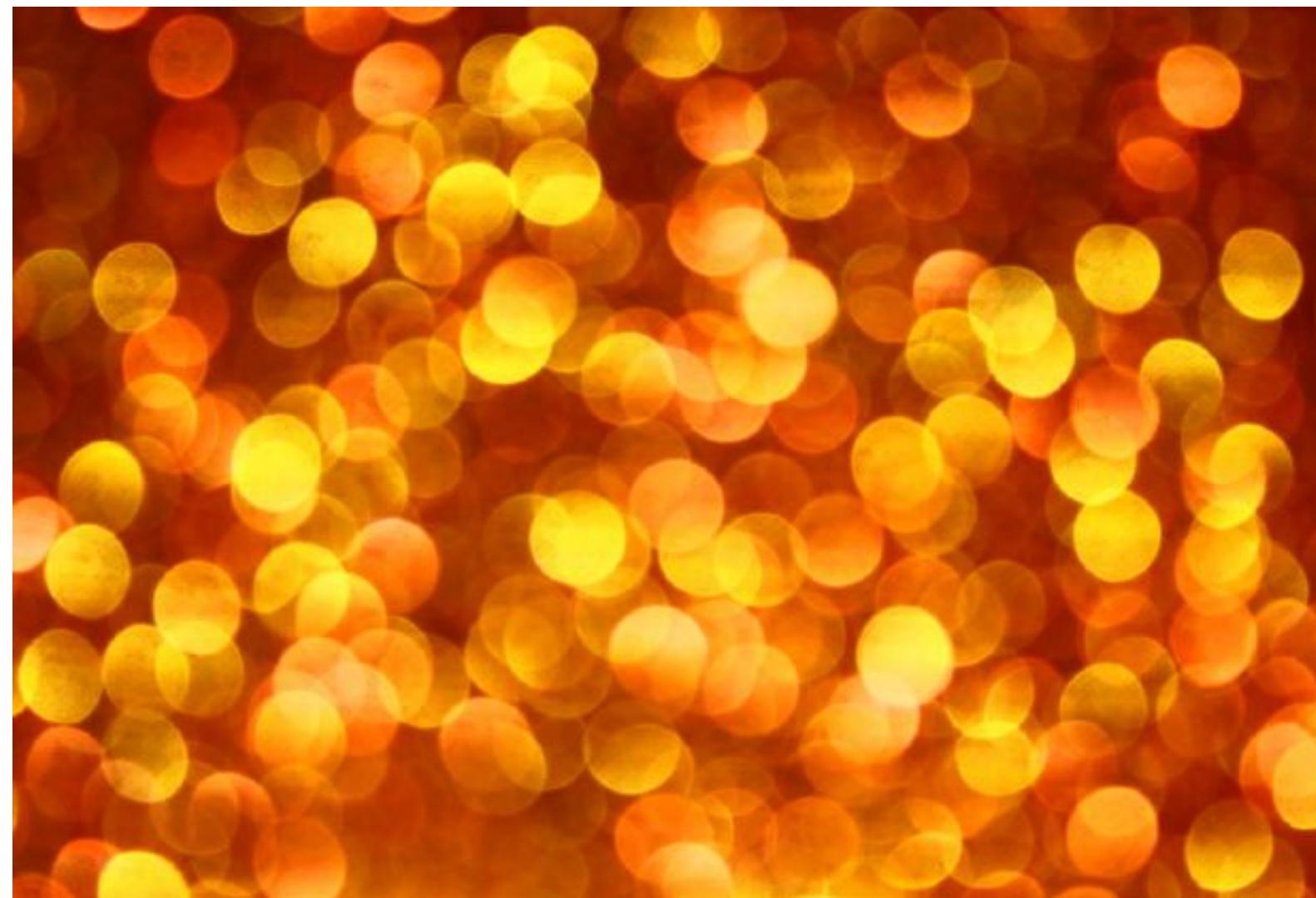
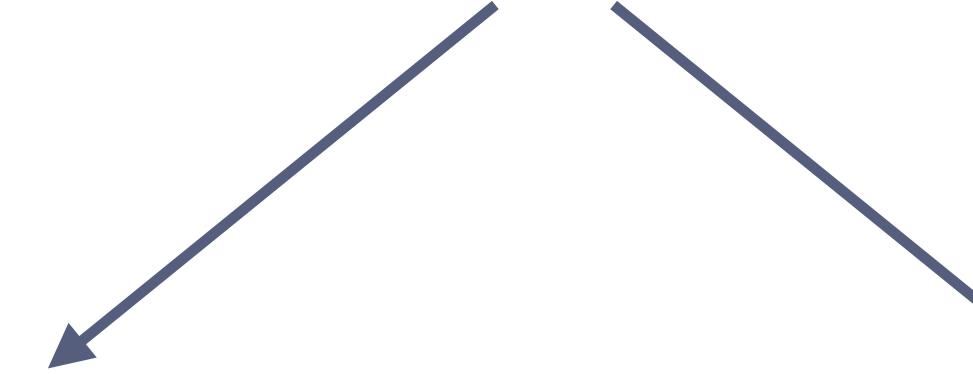
blue
gift
it
pole
leave
moment
sing
hine
it
chim
y tell
size
whisper
eat
through
inter
size
time
open
languid
ke
pu and
beauty
worship
the
wound
visit
largeous
delirious
f
w

Problem: homonym



Problem: polysemy

Warm



Linguistic



Tree	Albero	Arbre	Baum	Trae
Timber	Legno		Holz	
Wood	Bosco	Bois		
Forest	Foresta	Forêt	Wald	Skov

Categorisation: what is a chair



Categorisation: what is a chair



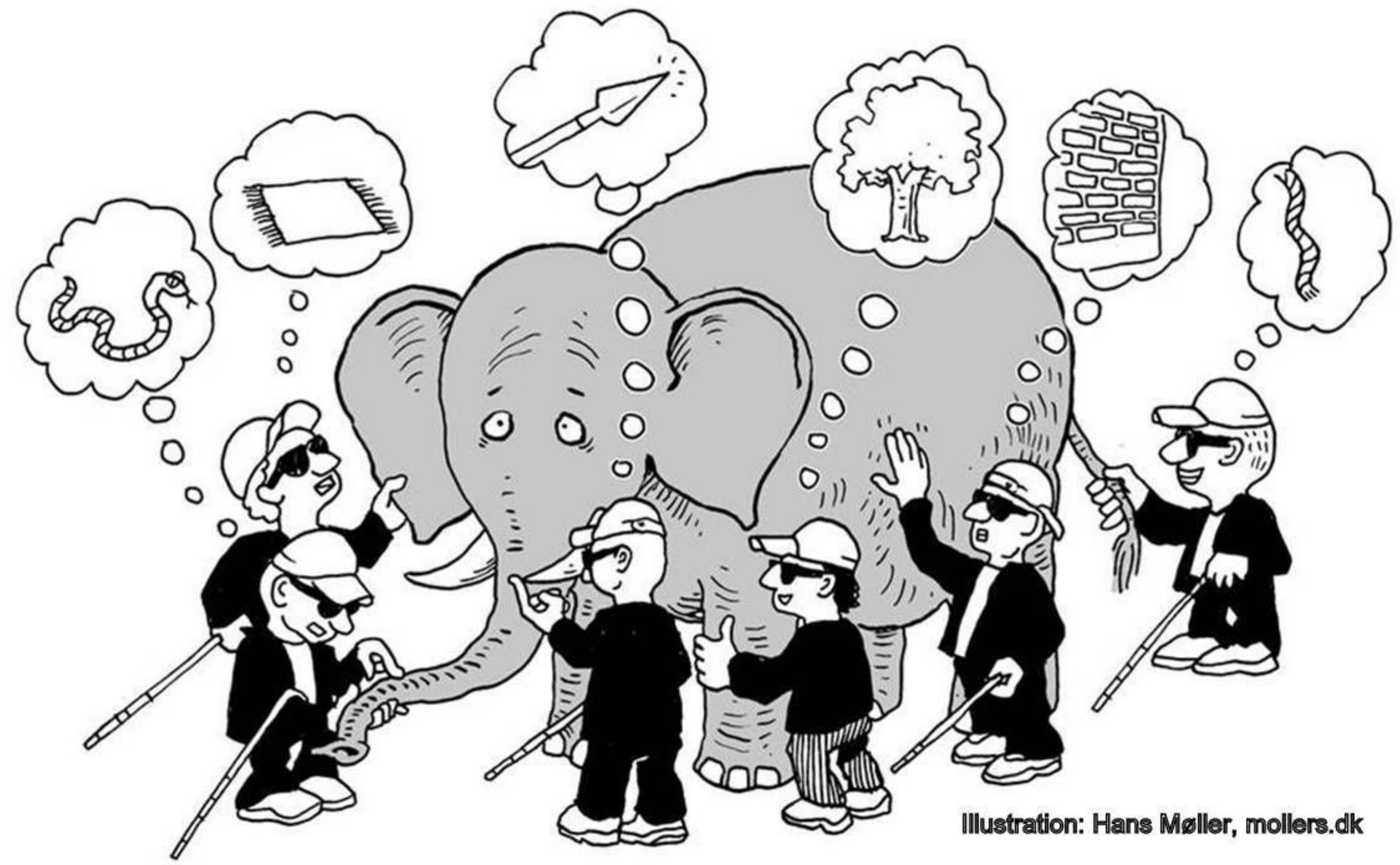
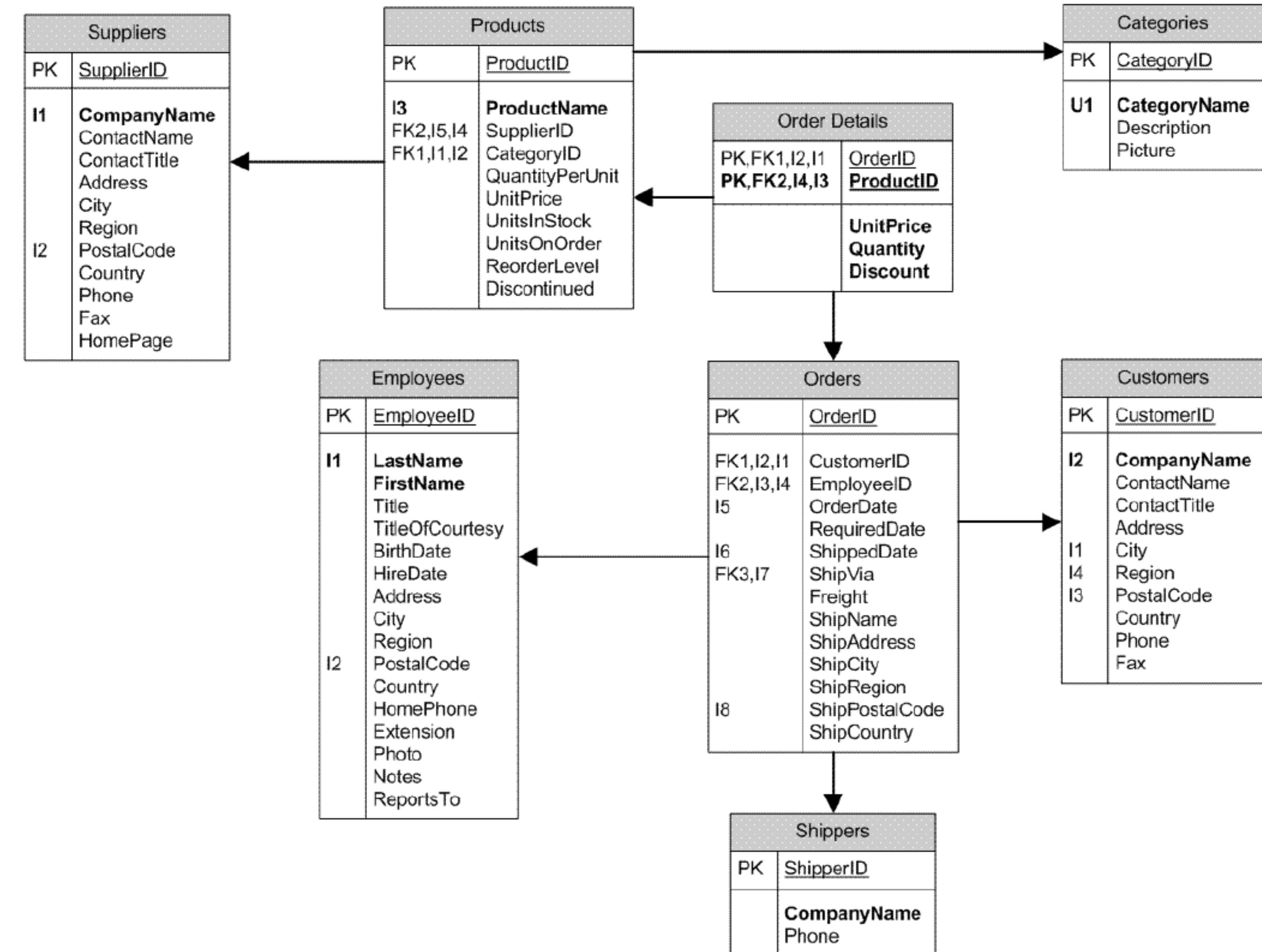


Illustration: Hans Møller, mollers.dk

About the object: Database schema



About the object: Database schema

Peritus Login

Records 15 51 / 1006 Found (Unsorted) Show All New Record Delete Record Find Sort Layout: RESUMES_tr...mpTransID View As: Preview Aa Edit Layout

Translators Clients Contacts Projects Forms Staff/Home **Applicants** Future Future Future Future Future Future Future Preferences

List Find Show All New Applicant PROFILE Résumé References RECEIVED 2 out of 3 RECEIVED out of EDITED out of Create New Translator Delete This Record/Translator

REFERENCES: Average 4.25 TESTS: Average

Main Info

Data From Web Submission Staff Referral Referred by 10/20/2007

First Name First Last Name Last Company Name Street City/State/Zip San Francisco Country USA E-Mail goodluck@lucas.org Other E-Mail Web Site Phone 514-2541 Fax Cell Phone Pager

Lang(s) translated from English - Australia English - British English - Canada English - US French - Canada Lang(s) translated into English - US French - France Addl near-native lang(s) German Native Country European Community Native Language English - US

References

Area(s) of Specialization (Prim) Arts and Humanities (General) -- Architecture -- Art -- Journalism Business (General)

Area(s) of Specialization (Sec) -- Gastronomy -- Linguistics -- Labor relations -- Film -- Music

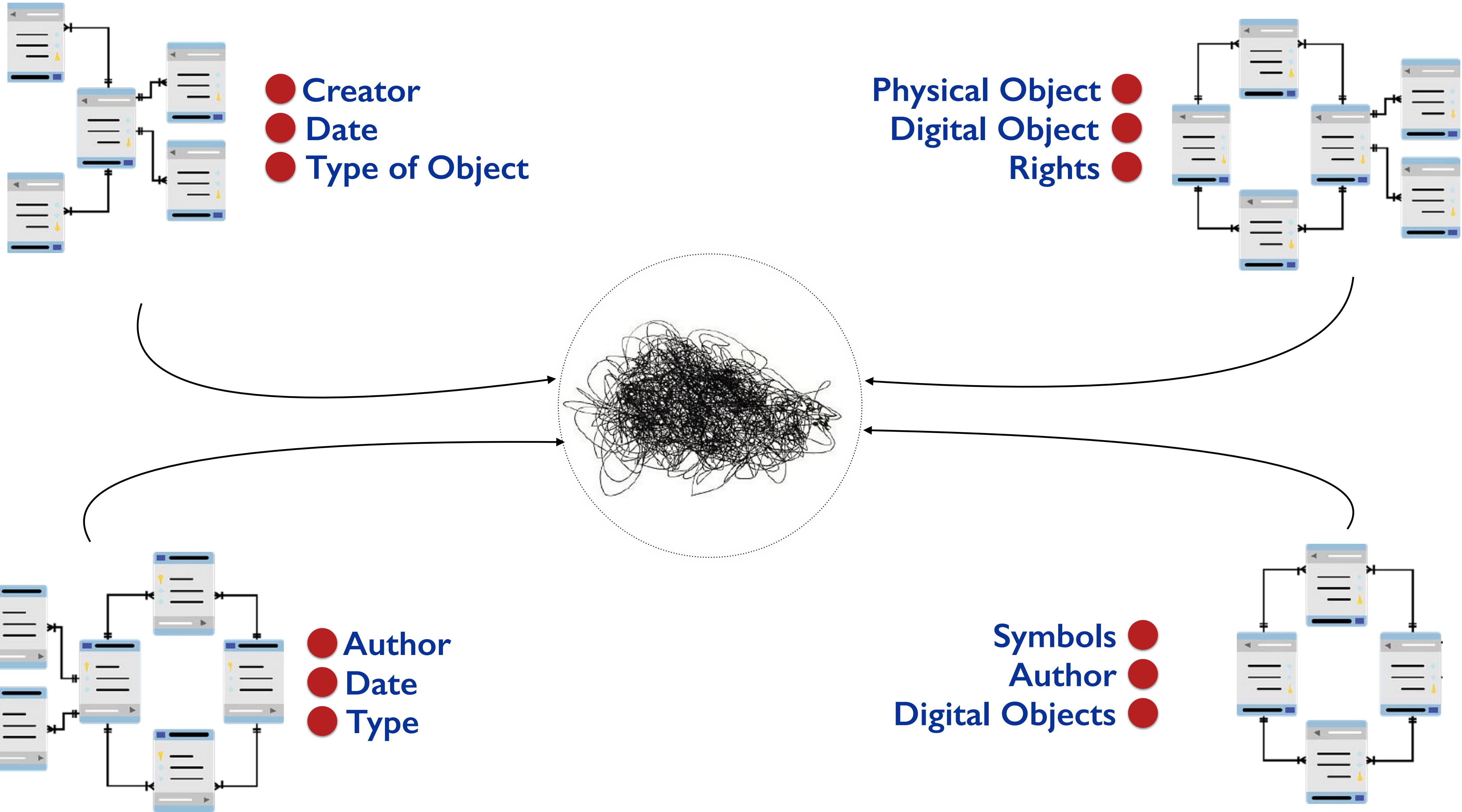
Work_Experience 1501 Characters 1990 - Now Independent translator and interpreter. Linguistic project management for major corporate clients (e.g., IBM; 5-10 Yrs Experience as Translator Translation Yes Court Interpreting Yes Editing Yes Transcribing Yes Proofreading Yes Voice-Over Yes Desktop Publishing No Localization Yes Copywriting Yes HTML Conference Interpreting Yes Simultaneous Interpreting Yes TM Tool Name & Version Trados 7 Hardware Acer (aspire 3610); Toshiba 4340 Software XP Pro; Office; Trados \$/Word \$0.13 \$Hour Edit/Proof \$35.00 \$Hour Interpreting \$100.00 Minimum Charge 50.00 Rush Charge Trados Rate 100% Match Trados Rate Fuzzy Match

Comments & Actions

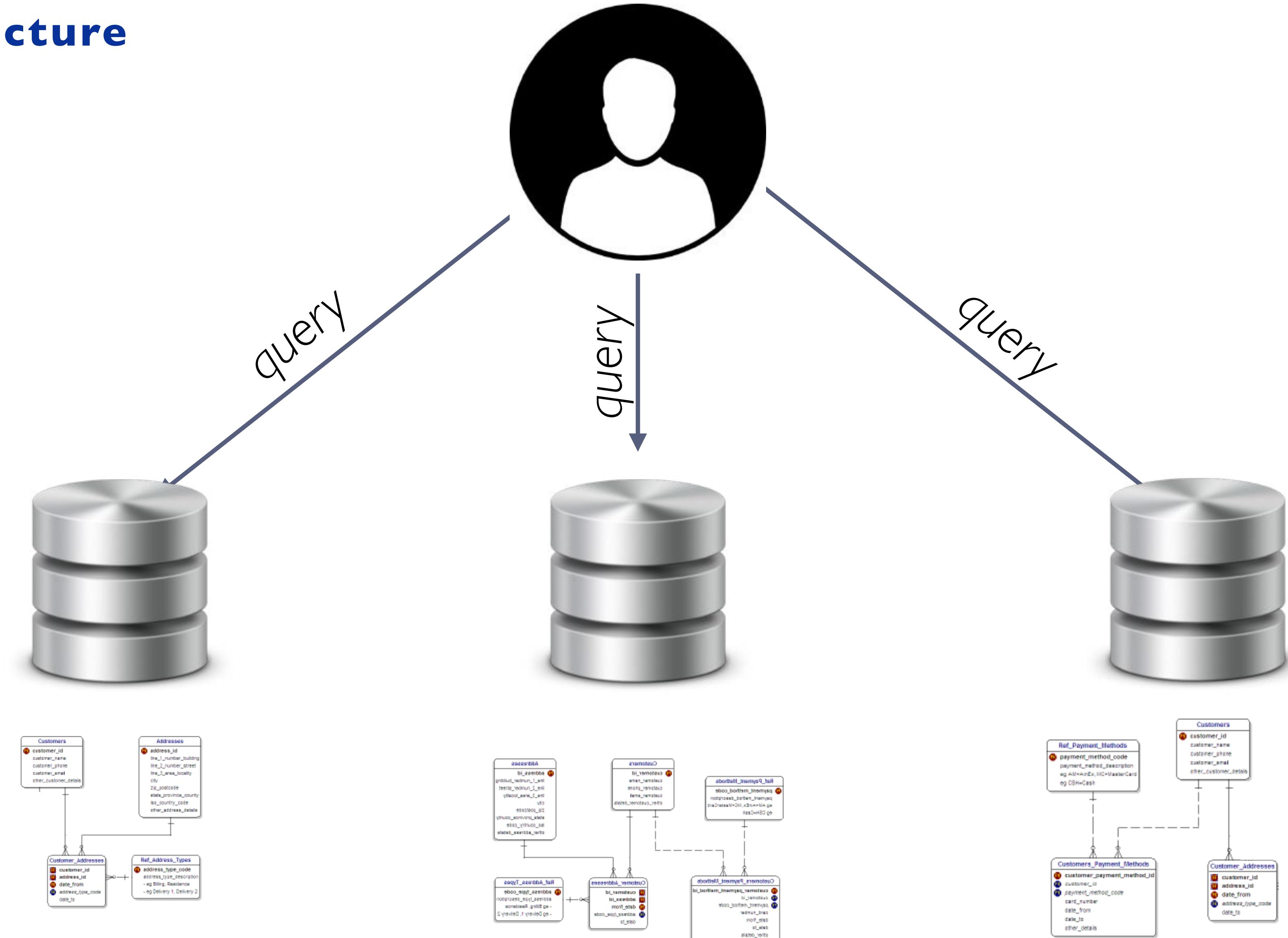
Education ATA Certified No ATA Certified Language BA; MA; Cambridge U.; RSA Celta Degrees & Certificates Comments Add comment Security Clearance No US Citizen Yes US Resident Yes Length of US Residency many years Other Residencies France Reference 1 Ref1 Reference 2 ref2 Reference 3 ref3 Built Build References RecordID 2117

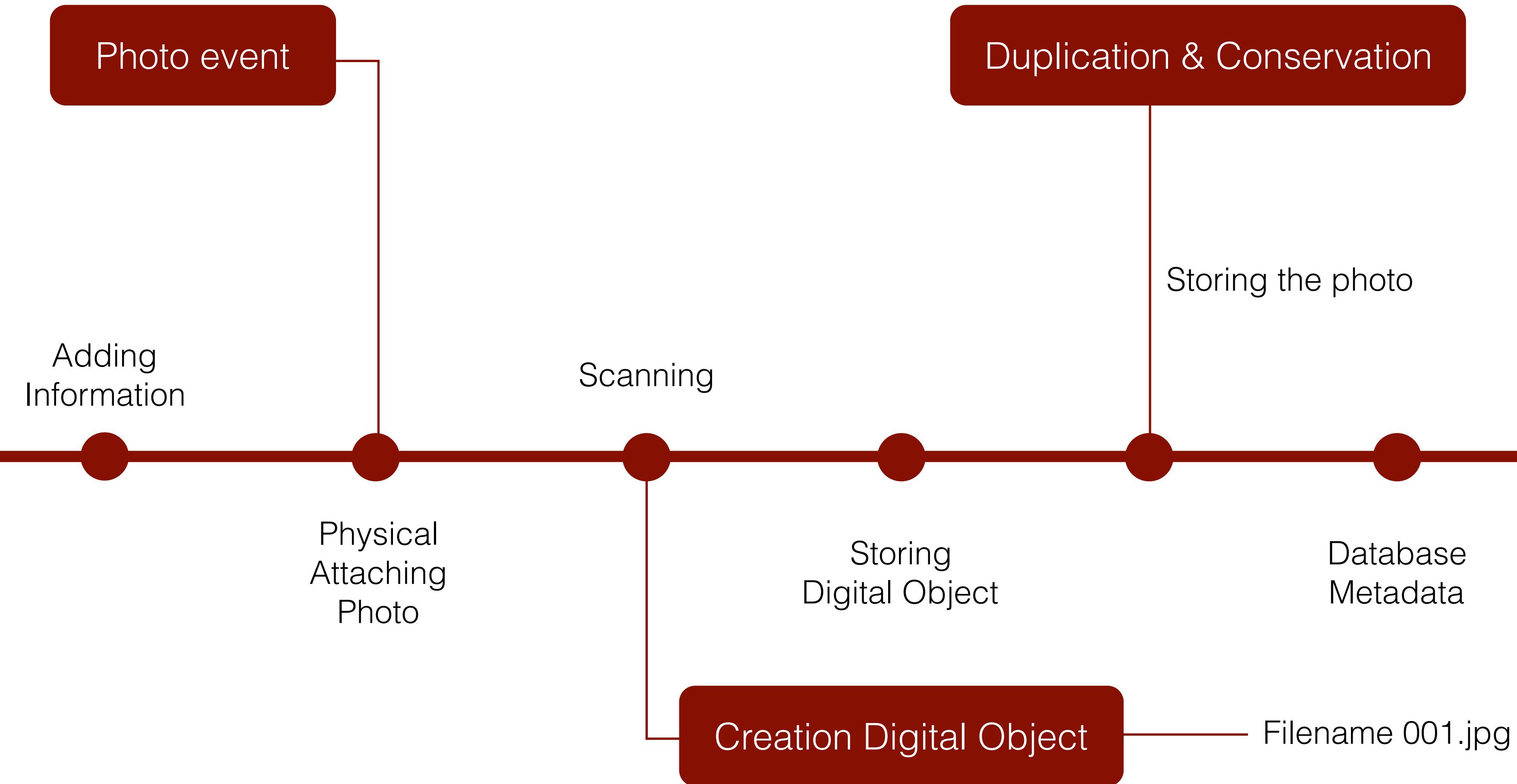
100 Browse



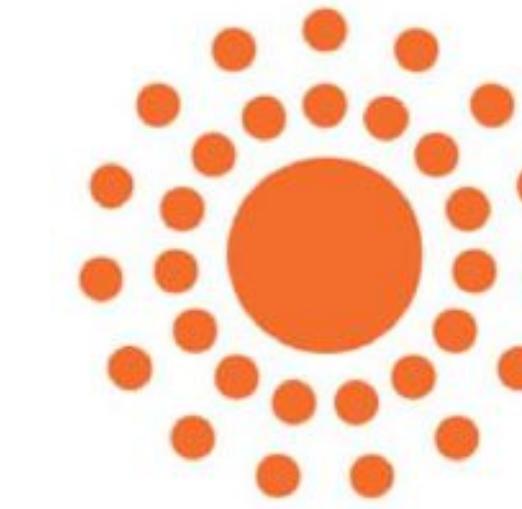


Problem: structure





About the object: metadata



```
<xs:element name="title" type="elementType"/>  
<meta name = "DC.Creator" content = "nicola">
```

Contributor	Coverage	Publisher	Subject	Identifier
-------------	----------	-----------	---------	------------

Creator	Date	Relation	Title	Language
---------	------	----------	-------	----------

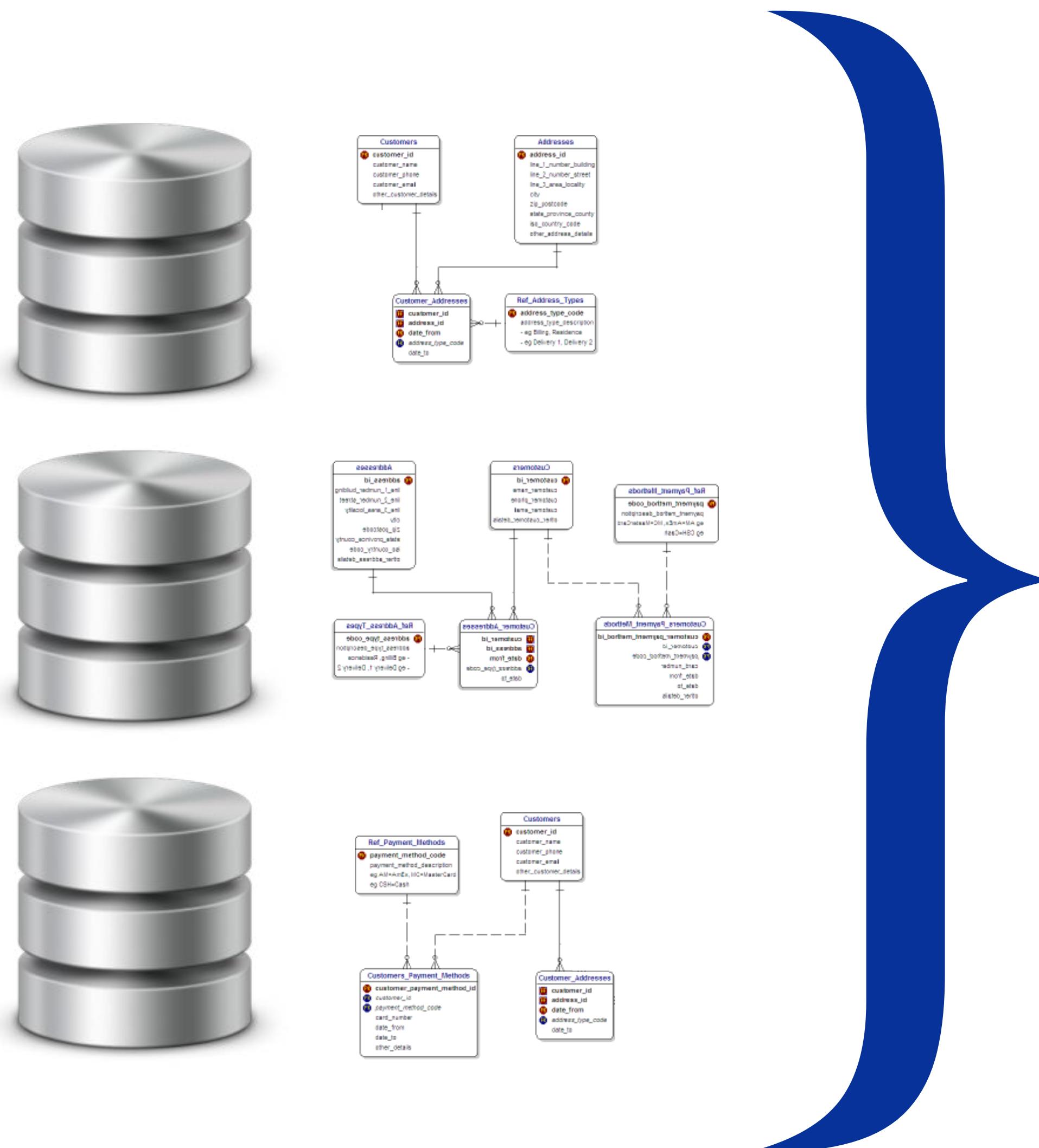
Description	Format	Rights	Type
-------------	--------	--------	------



- Defining standard vocabularies is difficult and time-consuming
- Once defined, standards don't adapt well
- Heterogeneous domains need a broad-coverage vocabulary
- People don't implement standards correctly anyway
- Vocabulary definitions are often ambiguous or circular



Problem: metadata and context



Limited set of Elements
+
Limited set of semantics



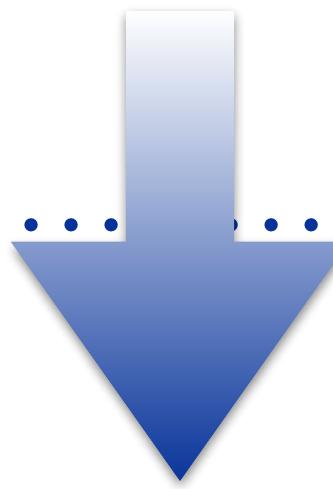
why it is difficult?

- Different formats, models, system
- Different names (polysemy, homonyms)
- Contradicting information
- Different conceptualisation of the domain

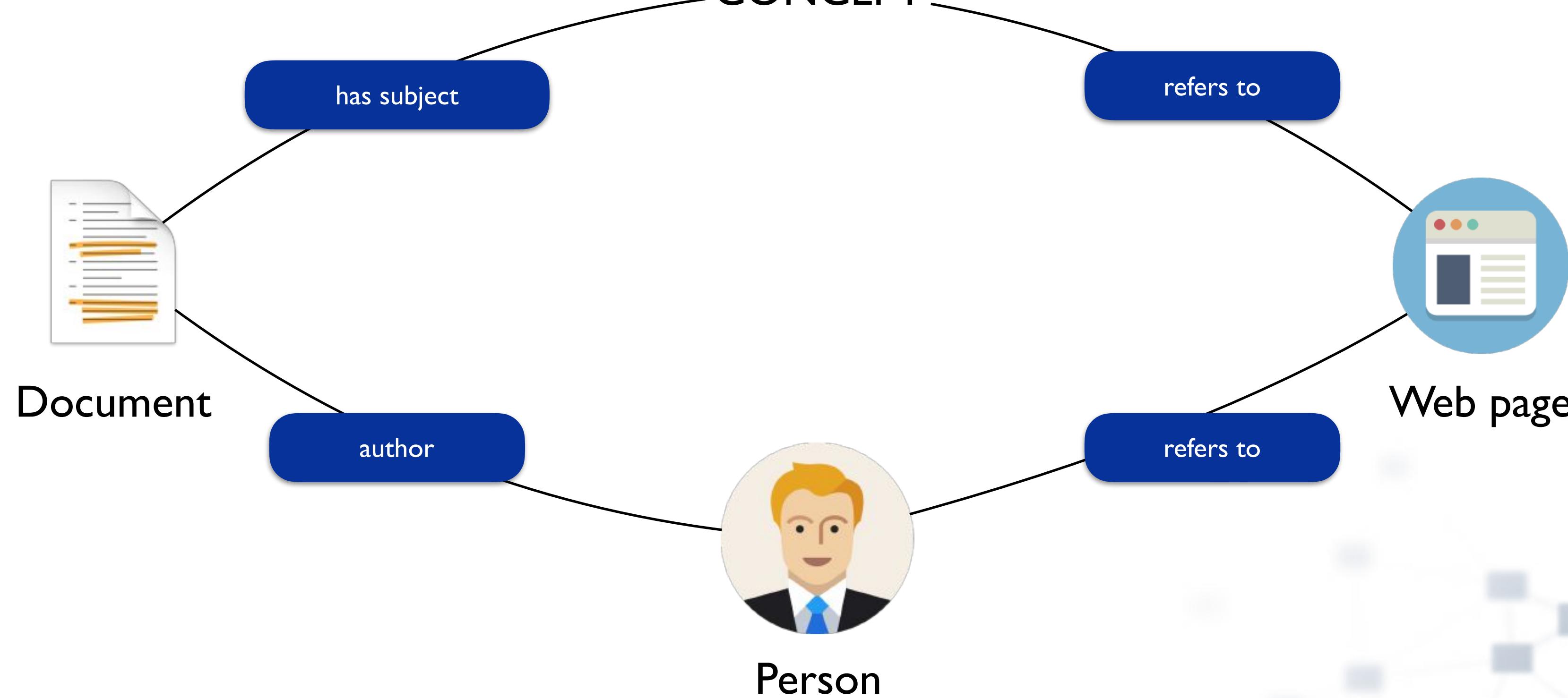




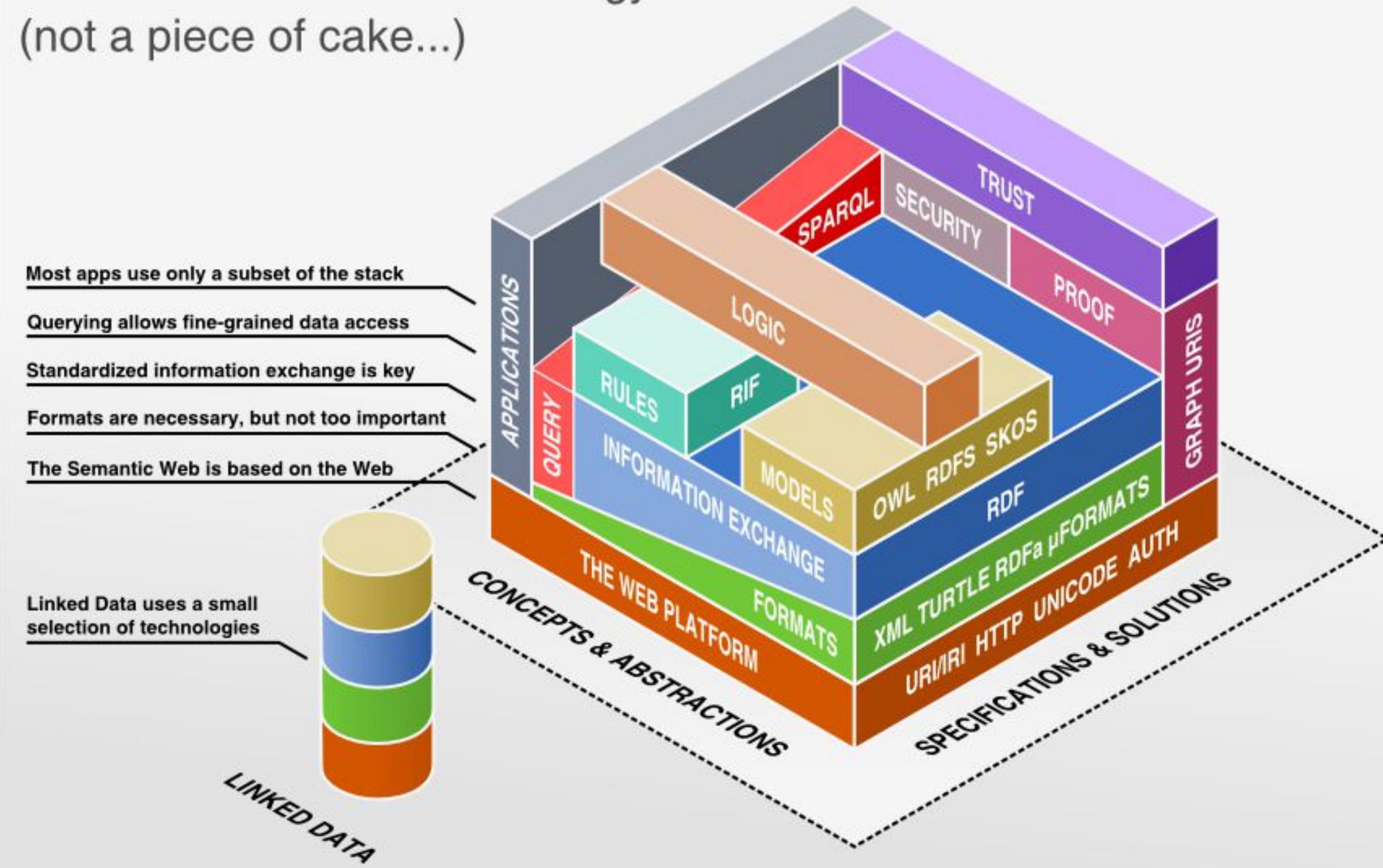
Wikipedia



CONCEPT



The Semantic Web Technology Stack (not a piece of cake...)



RDF: Resource Description Framework

- Data model: RDF
- Semantics: RDFS (Schema)
- Syntax: Turtle / RDFa/ RDF-XML



RDF: Resource Description Framework

- RDF is based on the idea of identifying things using Web identifiers (called Uniform Resource Identifiers, or URLs), and describing resources.
- A resource can be identified as a “thing” we want to talk about: a place, a person, a name, a webpage etc.
- Properties describe relationships between resource
- A statement declare to be composed by $\langle s, p, o \rangle$



{Subject} + {Predicate} + {Object}



Nicola



isA



Person

Engine

isPartOf

Car

Things

hasProperty

Value

IRI

IRI

IRI



```
<rdf:Description rdf:about="http://vocab.getty.edu/  
ulan/500115892">  
  <rdf:type rdf:resource="http://www.cidoc-crm.org/  
cidoc-crm/E21_Person"/>  
  <rdfs:label>Jacopo Torni</rdfs:label>  
  <crm:P129i_is_subject_of rdf:resource="https://  
collection.itatti.harvard.edu/resource/person/  
A00001629/hollis"/>  
  <crm:P1_is_identified_by>A00001629</  
crm:P1_is_identified_by>  
  <owl:sameas rdf:datatype="http://www.w3.org/2001/  
XMLSchema#anyURI">http://www.wikidata.org/entity/  
Q2632655</owl:sameas>  
  <skos:prefLabel>Torni, Jacopo</skos:prefLabel>  
</rdf:Description>
```

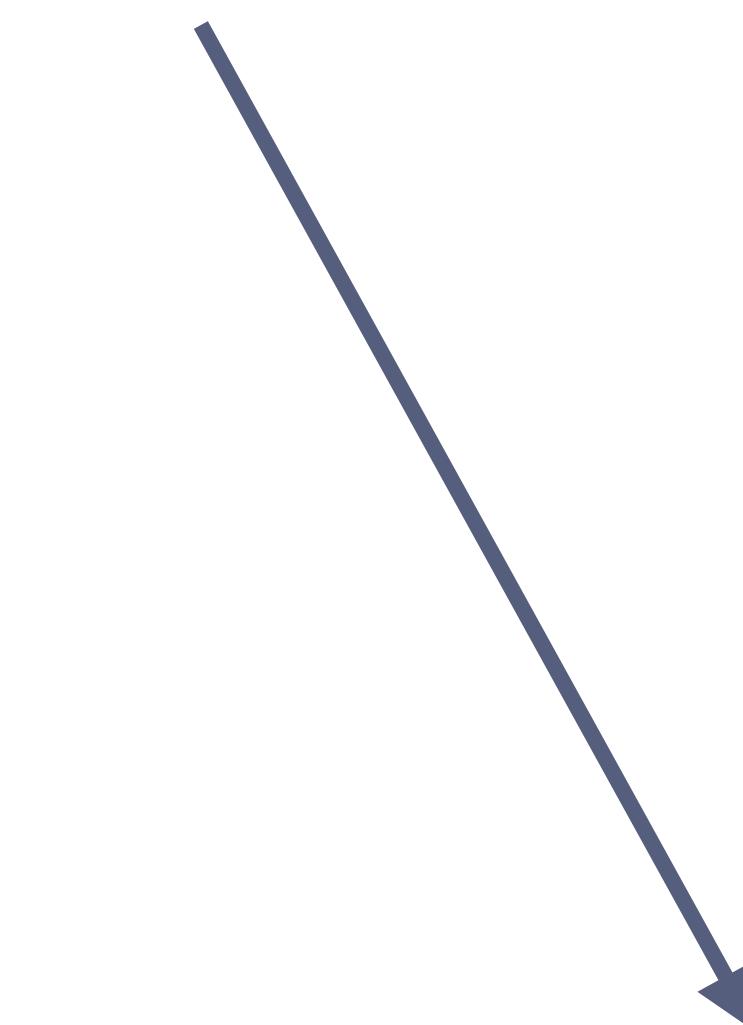
<http://vocab.getty.edu/ulan/500115892>

<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>

<http://www.cidoc-crm.org/cidoc-crm/E21_Person>

<http://www.w3.org/2000/01/rdf-schema#label>

“Jacopo Torni”



<http://www.w3.org/2002/07/owl#sameAs>

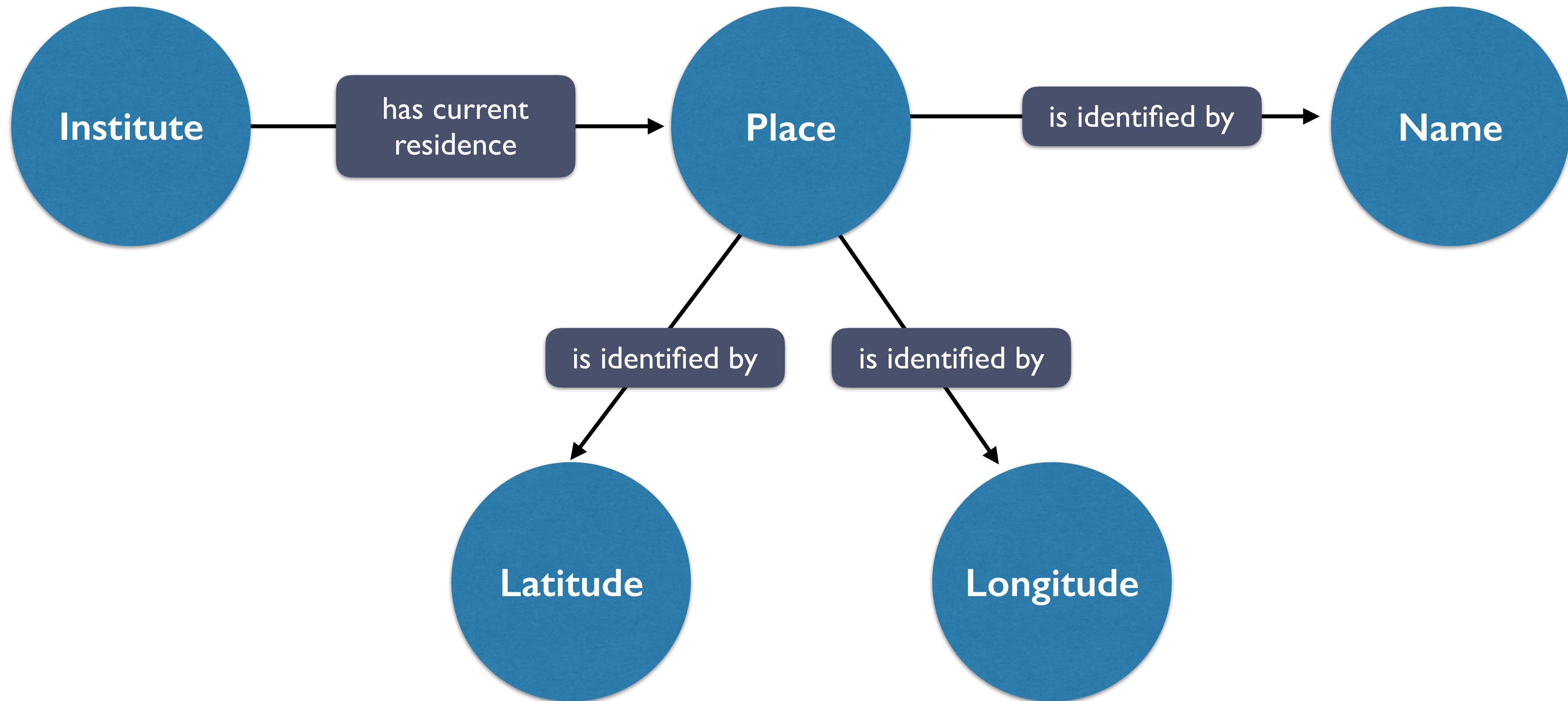
“<http://www.wikidata.org/entity/Q2632655>”^^<http://www.w3.org/2001/XMLSchema#anyURI>”

RDF: Resource Description Framework

- Predicates can be URI or Literals
- Literals are atomic values (strings or dates)
- A data type tells us whether we should interpret a value as string, a date, integer or some other type. It is recommended practice to use the data types defined by XML Schema
 - decimals - “1.23” <<http://www.w3.org/2001/XMLSchema#decimal>>
 - dates - “1982-08-30”^^<<http://www.w3.org/2001/XMLSchema#date>>
 - String - “Jacopo Torni”



```
<https://collection.itatti.harvard.edu/resource/institution/P00006239> <http://www.cidoc-crm.org/cidoc-crm/P74_has_current_or_former_residence> <https://collection.itatti.harvard.edu/resource/institution/Ravello/P00006239> .  
<https://collection.itatti.harvard.edu/resource/institution/Ravello/P00006239> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.cidoc-crm.org/cidoc-crm/E53_Place>  
<https://collection.itatti.harvard.edu/resource/institution/Ravello/P00006239> <http://www.cidoc-crm.org/cidoc-crm/P87_is_identified_by> <https://collection.itatti.harvard.edu/resource/institution/P00006239/Ravello{lng}> .  
<https://collection.itatti.harvard.edu/resource/institution/Ravello/P00006239> <http://www.cidoc-crm.org/cidoc-crm/P87_is_identified_by> <https://collection.itatti.harvard.edu/resource/institution/P00006239/Ravello/lat> .  
<https://collection.itatti.harvard.edu/resource/institution/P00006239/Ravello{lng}> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.cidoc-crm.org/cidoc-crm/E47_Spatial_Coordinates> .  
<https://collection.itatti.harvard.edu/resource/institution/P00006239/Ravello{lng}> <http://www.w3.org/2000/01/rdf-schema#label> "14.6125143" .  
<https://collection.itatti.harvard.edu/resource/institution/P00006239/Ravello{lng}> <http://www.cidoc-crm.org/cidoc-crm/P2_has_type> "Longitude" .  
<https://collection.itatti.harvard.edu/resource/institution/P00006239/Ravello/lat> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.cidoc-crm.org/cidoc-crm/E47_Spatial_Coordinates> .  
<https://collection.itatti.harvard.edu/resource/institution/P00006239/Ravello/lat> <http://www.w3.org/2000/01/rdf-schema#label> "40.6480381" .  
<https://collection.itatti.harvard.edu/resource/institution/P00006239/Ravello/lat> <http://www.cidoc-crm.org/cidoc-crm/P2_has_type> "Latitude" .
```



Information access

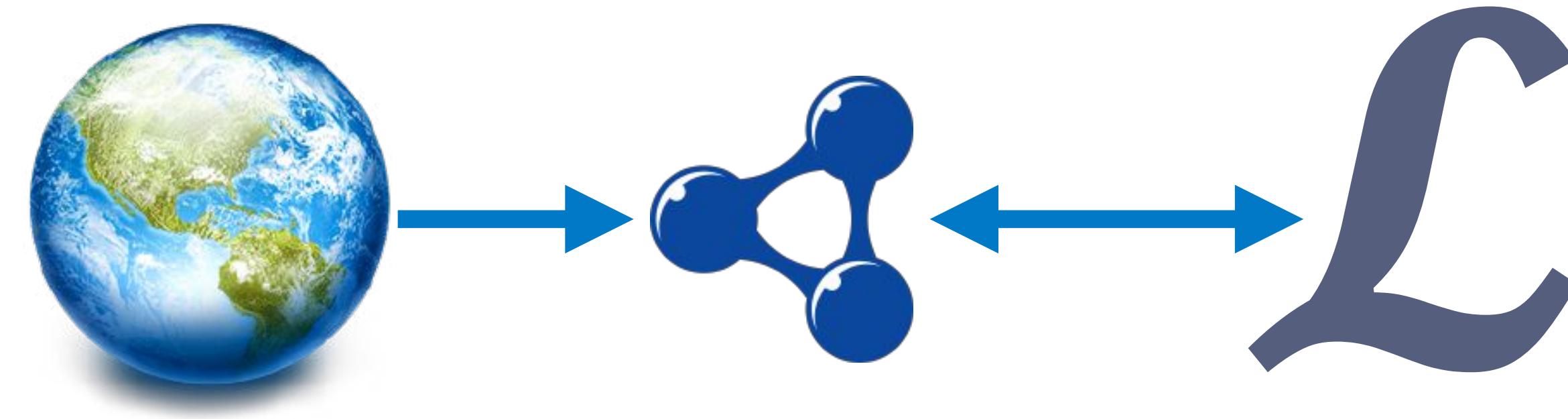
Information integration



Semantic Matching

Semantic Integration

About the object

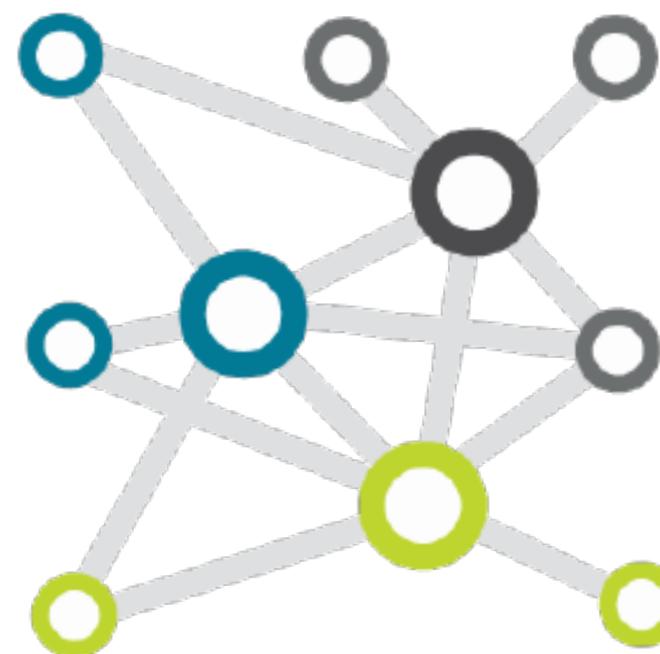


“An ontology is a formal, explicit specification of a shared conceptualization.”

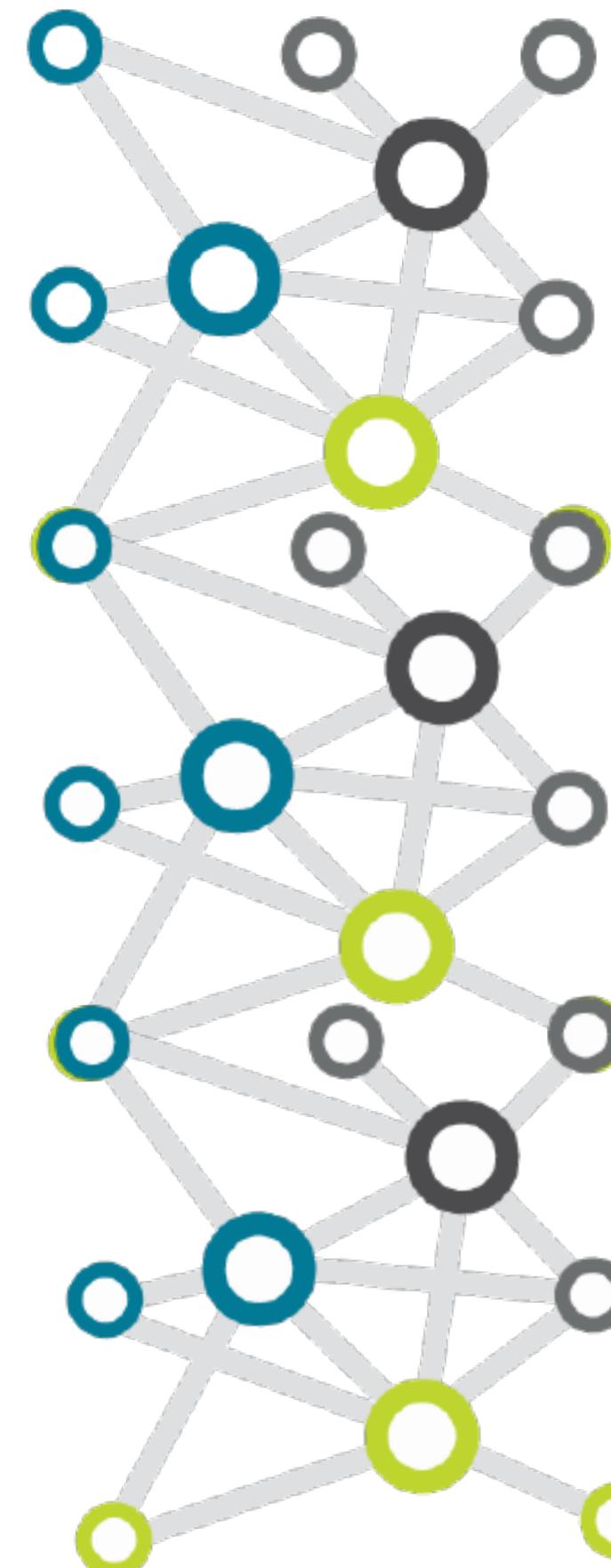
“logical theory accounting for the intended meaning of a formal vocabulary, i.e. its ontological commitment to a particular conceptualization of the world. The intended models of a logical language using such a vocabulary are constrained by its ontological commitment. An ontology indirectly reflects this commitment (and the underlying conceptualization) by approximating these intended models”

About the object

Language



Ontology



Knowledge



World



Ontological commitment

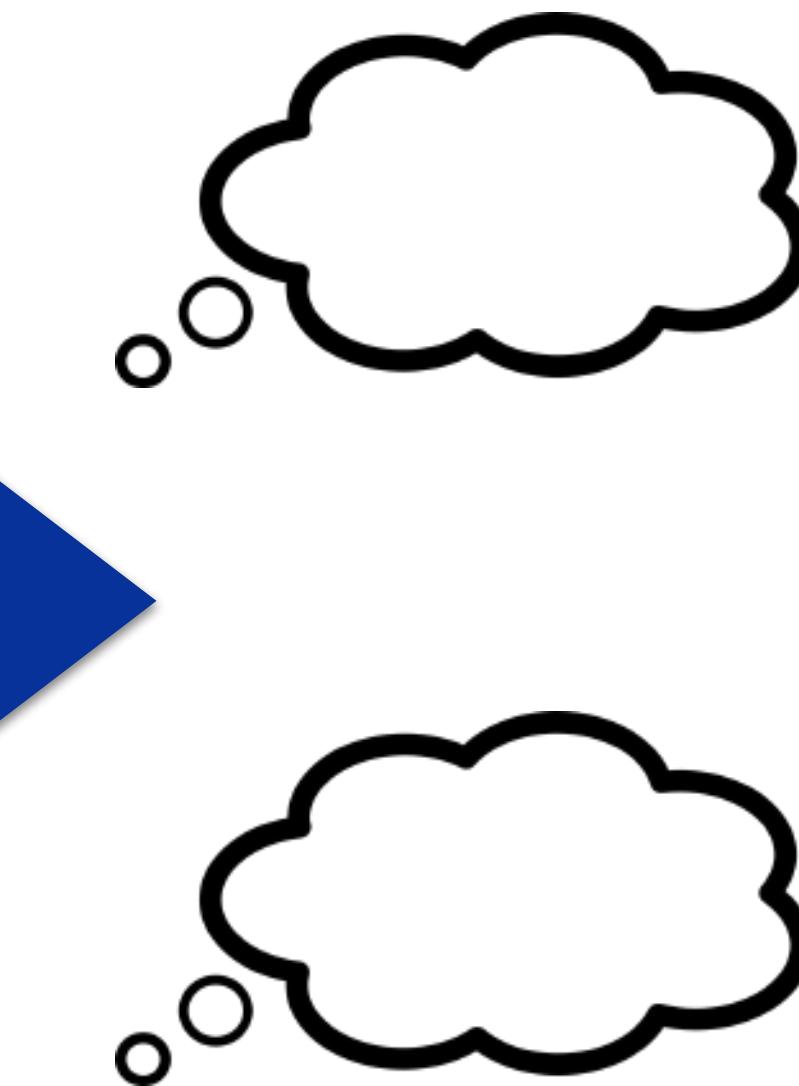
Language



Ontology



Knowledge



World



- What is a **contract**?
- What is a **document**?
- What is a **employee**?
- What is a **customer**?
- Can **people** make action?
- Can **organisation** make action?
- What is a **missing part**?
- What is a **hole**?
- What is a **passenger**?
- What is a **person**?



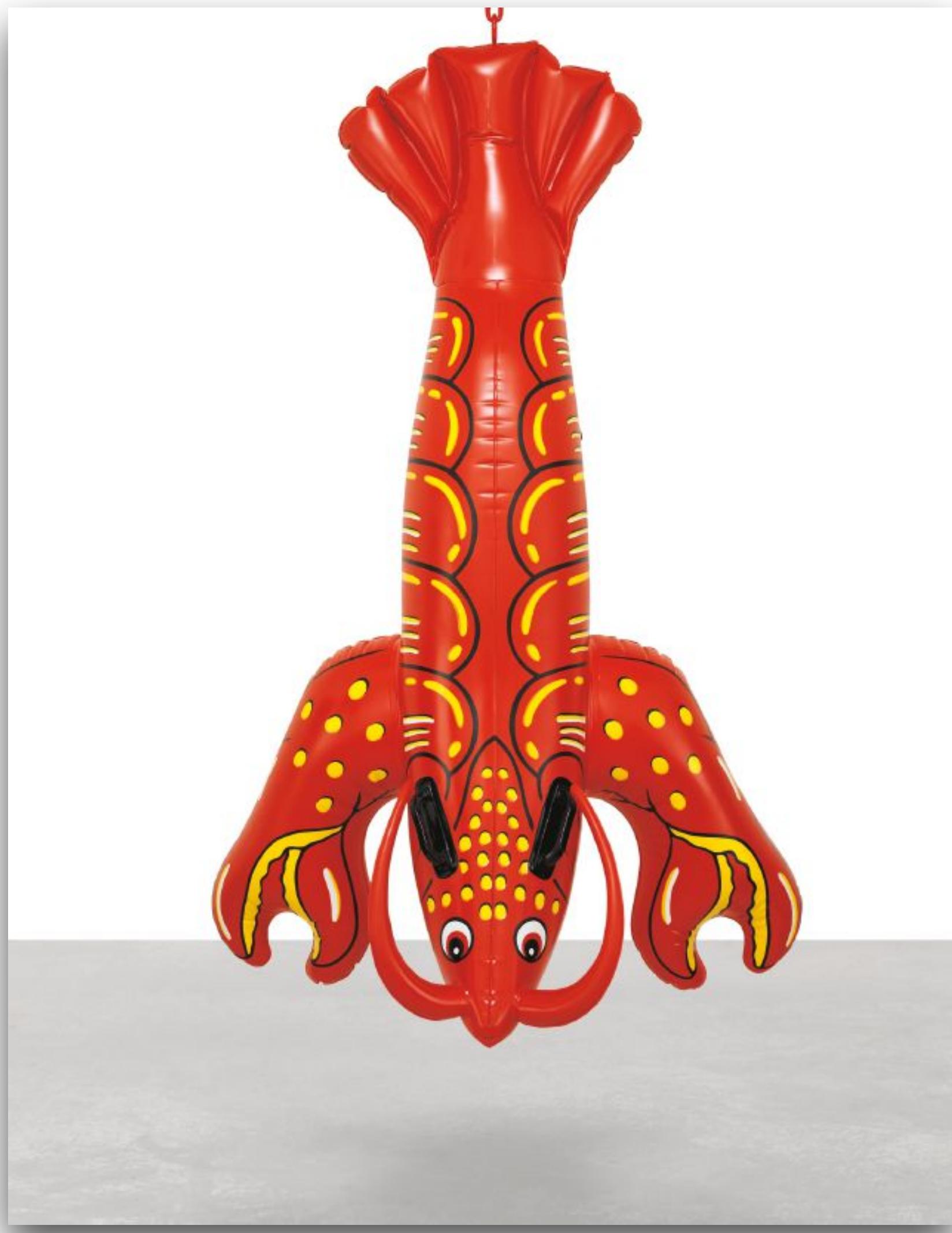
- **Part Of relationships**

Identity condition is defined by
the assembly of its parts or not?



- **Part Of relationships**
Is the collar influence the identity of my dog?





Classes and properties

Class

“A category of items that share one or more common traits serving as criteria to identify the items belonging to the class.”

Property

“A property serves to define a relationship of a specific kind between two classes [...] A property plays a role analogous to a grammatical verb, in that it must be defined with reference to both its domain and range, which are analogous to the subject and object in grammar.”

Instance

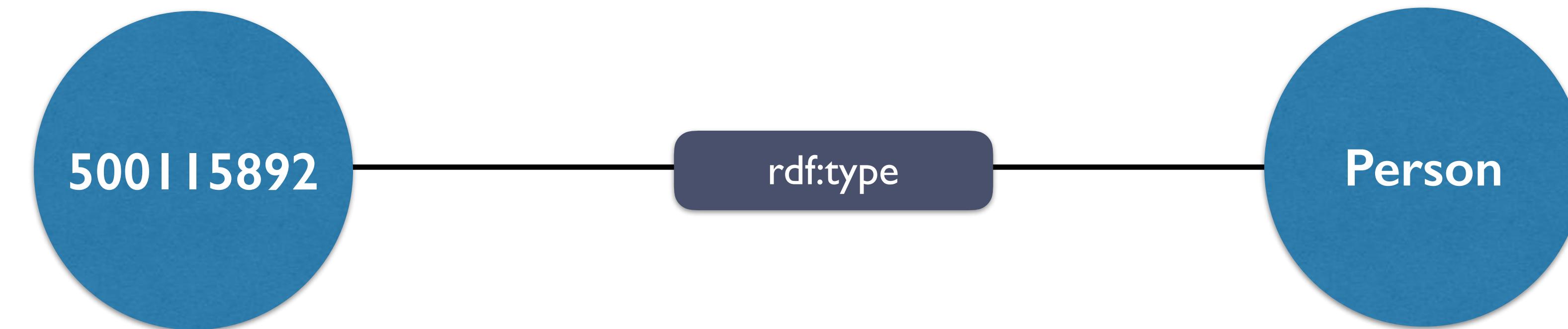
An instance of a class is a real world item that fulfils the criteria of the intension of the class



```
<rdf:Description rdf:about="http://vocab.getty.edu/ulan/500115892">
  <rdf:type rdf:resource="http://www.cidoc-crm.org/cidoc-crm/E21_Person"/>
  <rdfs:label>Jacopo Torni</rdfs:label>
  <crm:P129i_is_subject_of rdf:resource="https://collection.itatti.harvard.edu/resource/person/A00001629/hollis"/>
```

<http://vocab.getty.edu/ulan/500115892>

http://www.cidoc-crm.org/cidoc-crm/E21_Person



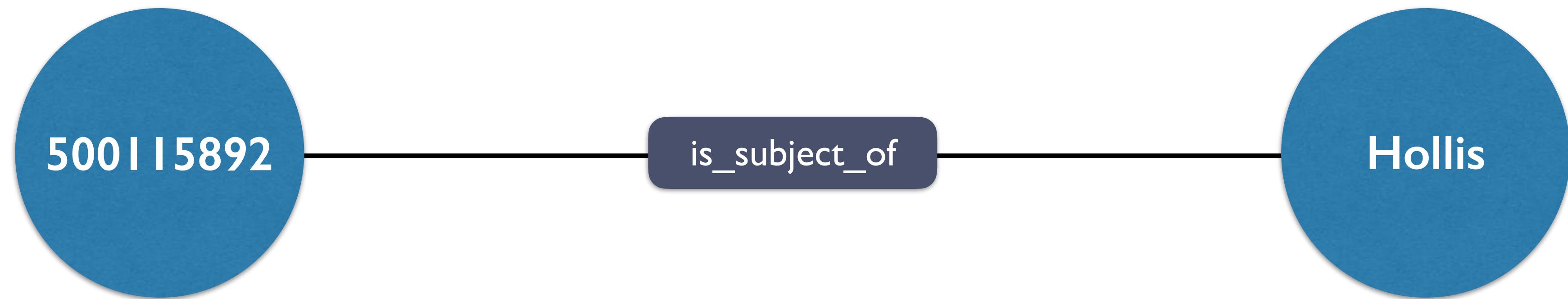
RDF: Resource Description Framework

- RDF provides us a way to define the relationship between instances and classes using a special property `rdf:type`.
- `rdf:type` impose restrictions on what can be stated in an RDF document
 - The range of the property it is used to declare that the values of a particular property are instances of a designated class
 - The domain of the property indicate that a particular property applies to a designated class



Property domain

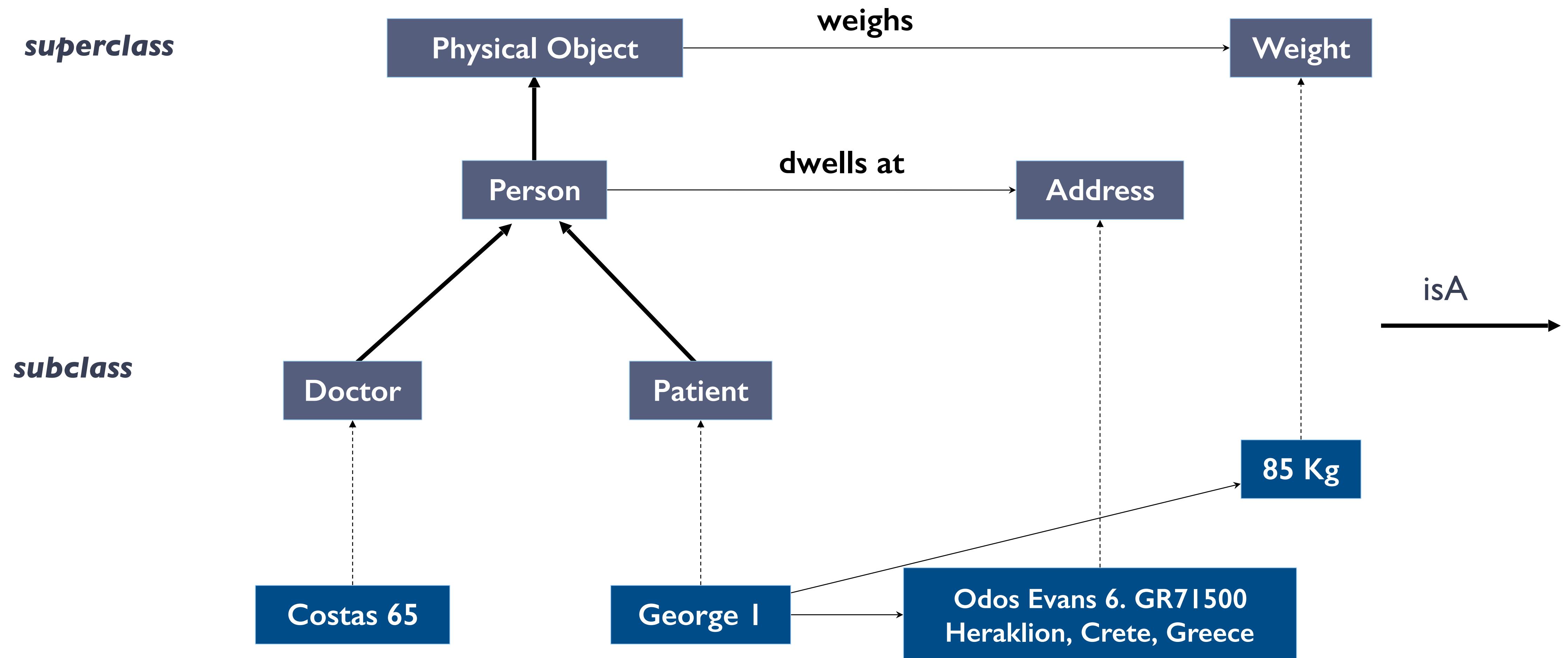
[<http://vocab.getty.edu/ulan/500115892>](http://vocab.getty.edu/ulan/500115892) [<http://www.cidoc-crm.org/cidoc-crm/E89_Propositional_Object>](http://www.cidoc-crm.org/cidoc-crm/E89_Propositional_Object)



Property range



- The Range is clearly wrong. In fact the property in CIDOC-CRM says:
 - Domain: E39 Actor
 - Range: E53 Place



- An instance of a class is an instance of all its superclasses.
- A subclass inherits the properties of all superclasses. (properties “move up”)

RDF: Resource Description Framework

- The subclass relationship defines a hierarchy of classes. In general, A is a subclass of B if every instance of A is also an instance of B.
 - Also properties can have sub-properties
 - Properties are inherited. Every property of A is also property of its subclass.
 - Possible to define new properties of a class without changing the class.
 - Classes help restrictions: Only people can have an address



RDF: Resource Description Framework

RDF provide the framework for the creation of statements
and RDF Schema provides modeling primitives to define
relationships between classes, properties and instances



The core classes are

- **rdfs:Resource**, the class of all resources
- **rdfs:Class**, the class of all classes
- **rdfs:Literal**, the class of all literals (strings)
- **rdf:Property**, the class of all properties

The core properties for defining relationships are

- **rdf:type**, which relates a resource to its class
- **rdfs:subClassOf**, which relates a class to one of its superclasses
- **rdfs:subPropertyOf**, which relates a property to one of its superproperties

RDF: Resource Description Framework

RDF is a formal language for expressing knowledge and, therefore, automated reasoners can deduce (infer) conclusions from the given knowledge, thus making implicit knowledge explicit.



Use URI to identify things

Use namespace prefix to shorten the uri

Statements always as <subject, predicate, object>

Ontology define what these URI are about:

- <http://www.cidoc-crm.org/cidoc-crm/E21_Person>



Literals

Literals are raw text unit that can be used instead of triple:

- identified by presence of “”
- possible to add language tag: “Nicola”@it .



Not unique name assumption

Same resource has several names

We need to explicitly tell the computer those means the same thing



Open World assumption

Any information can come at any point in future

We never know anything about the world



RDF: Resource Description Framework

However, the problem with RDFS is that it is limited roughly to a subclass hierarchy and a property hierarchy, with domain and range definitions.

Welcome OWL

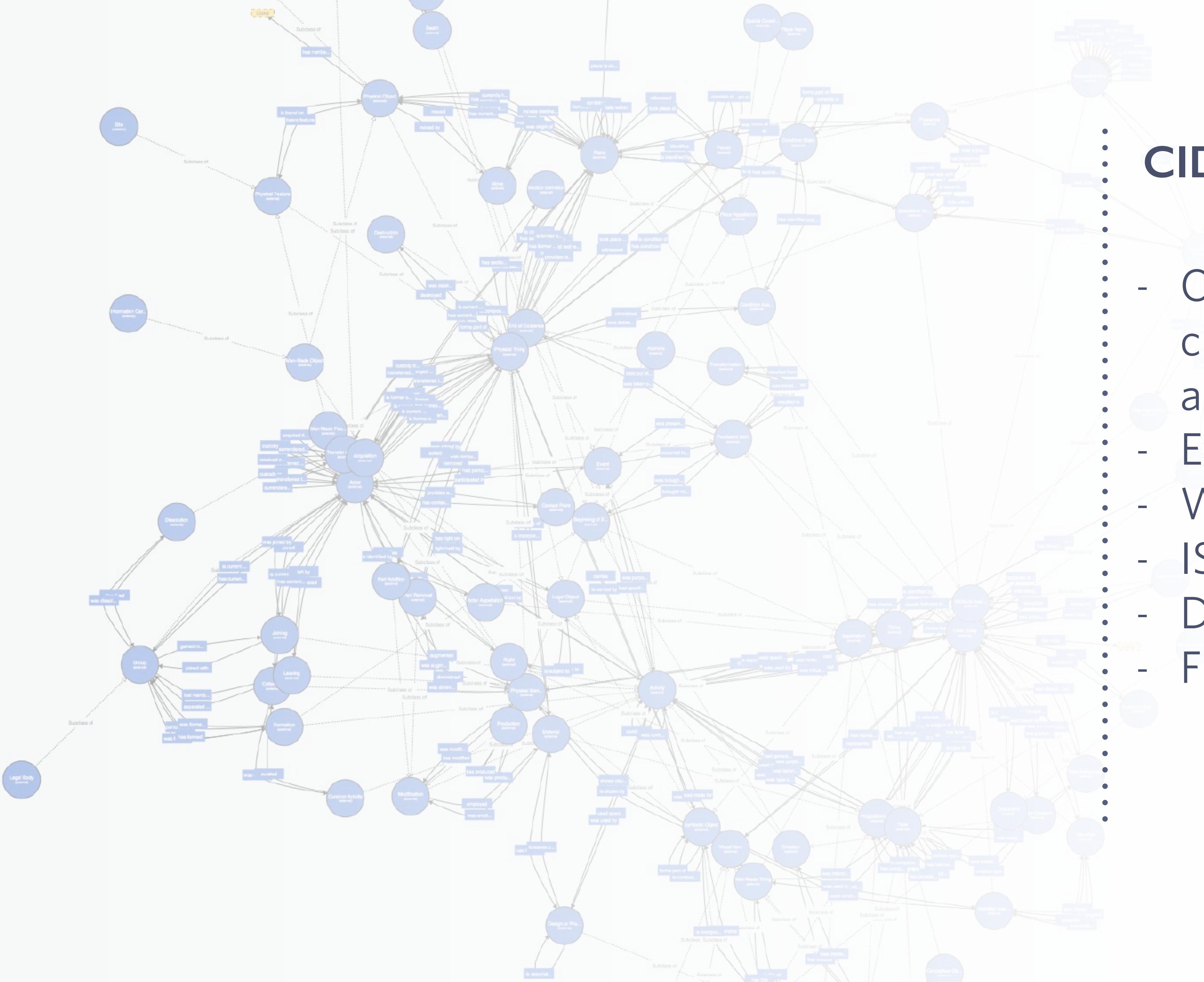


RDF: Resource Description Framework

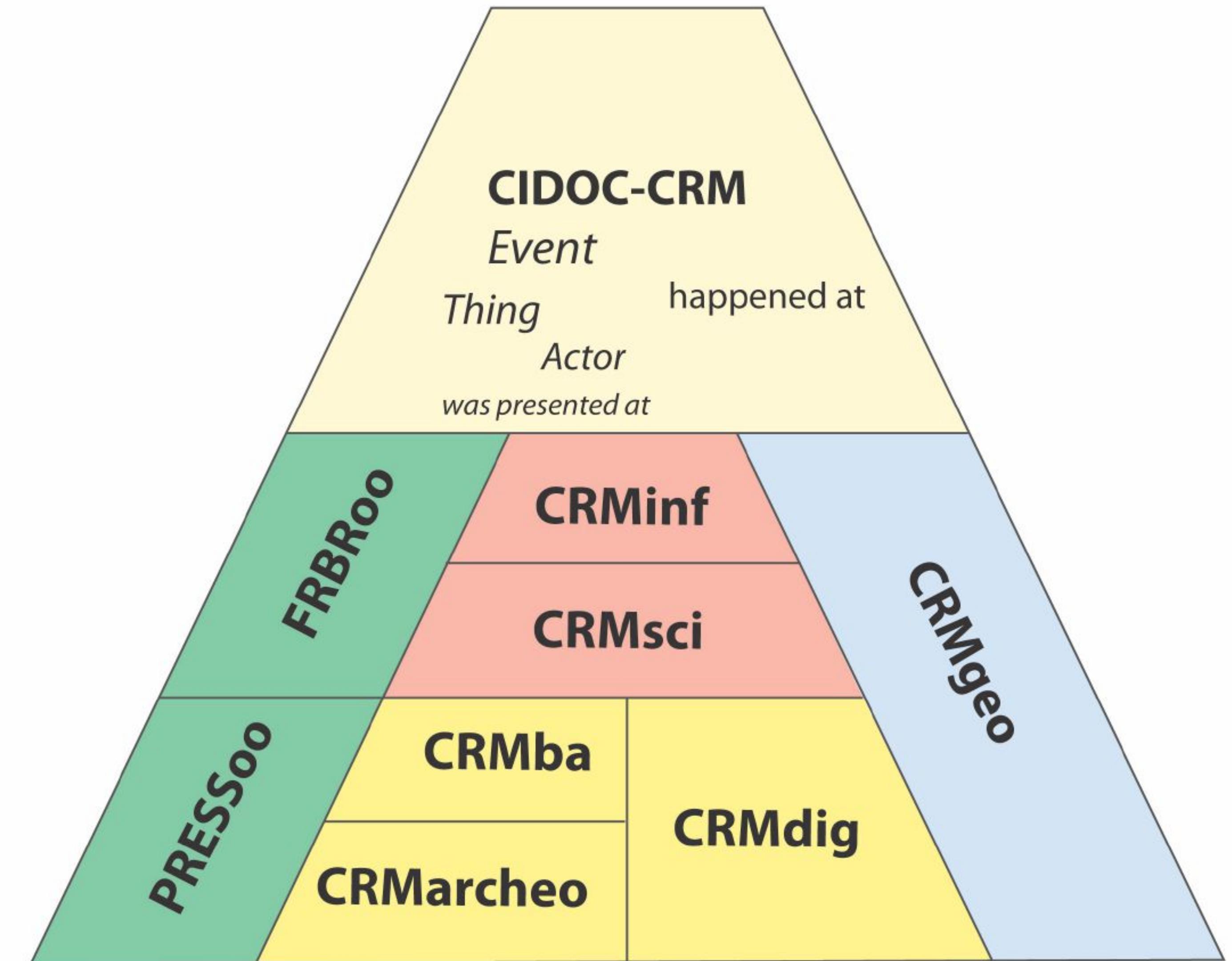
OWL (Web Ontology Language) builds on RDF and RDFS and uses an extension of their syntax. OWL aims to be more expressive than RDF using

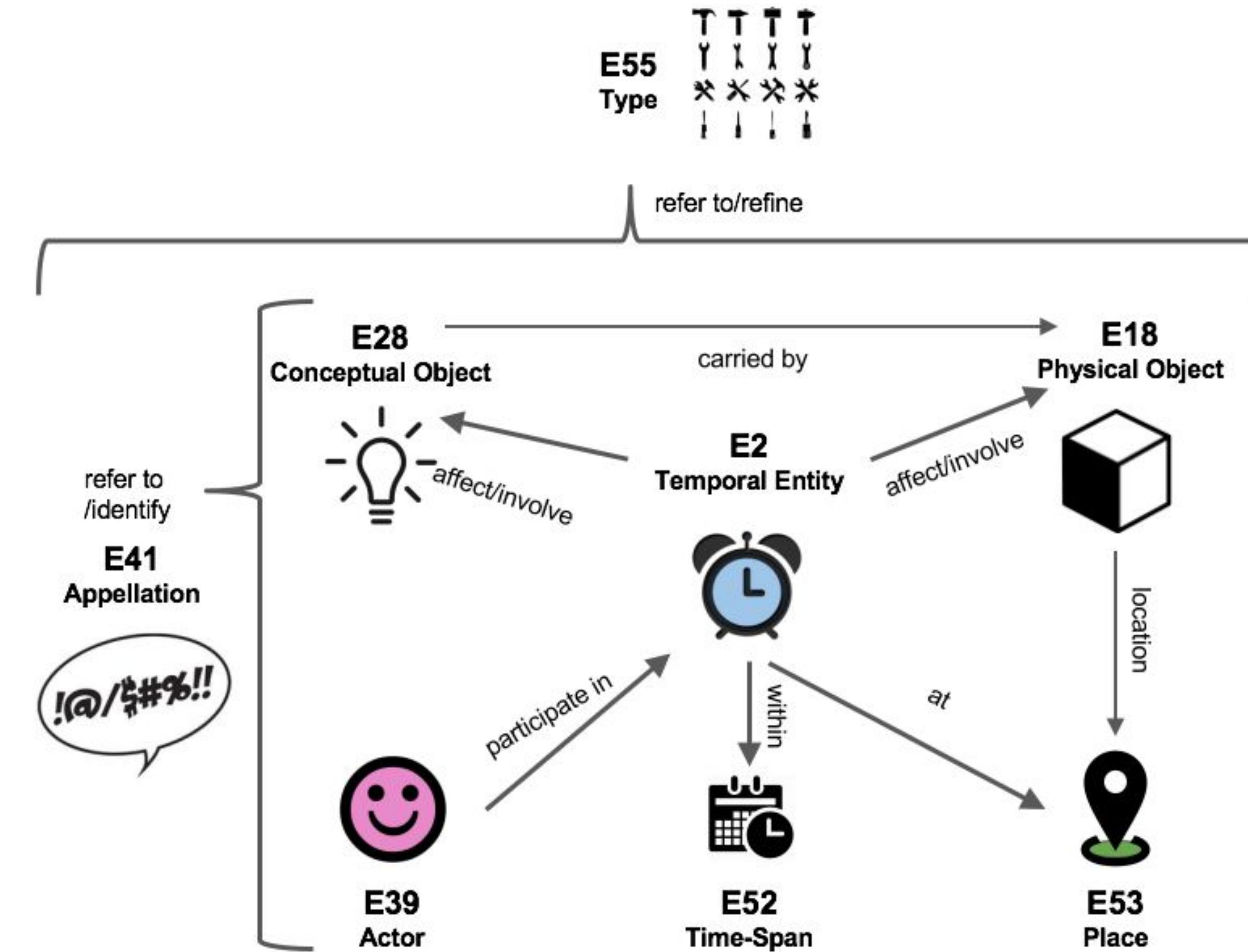


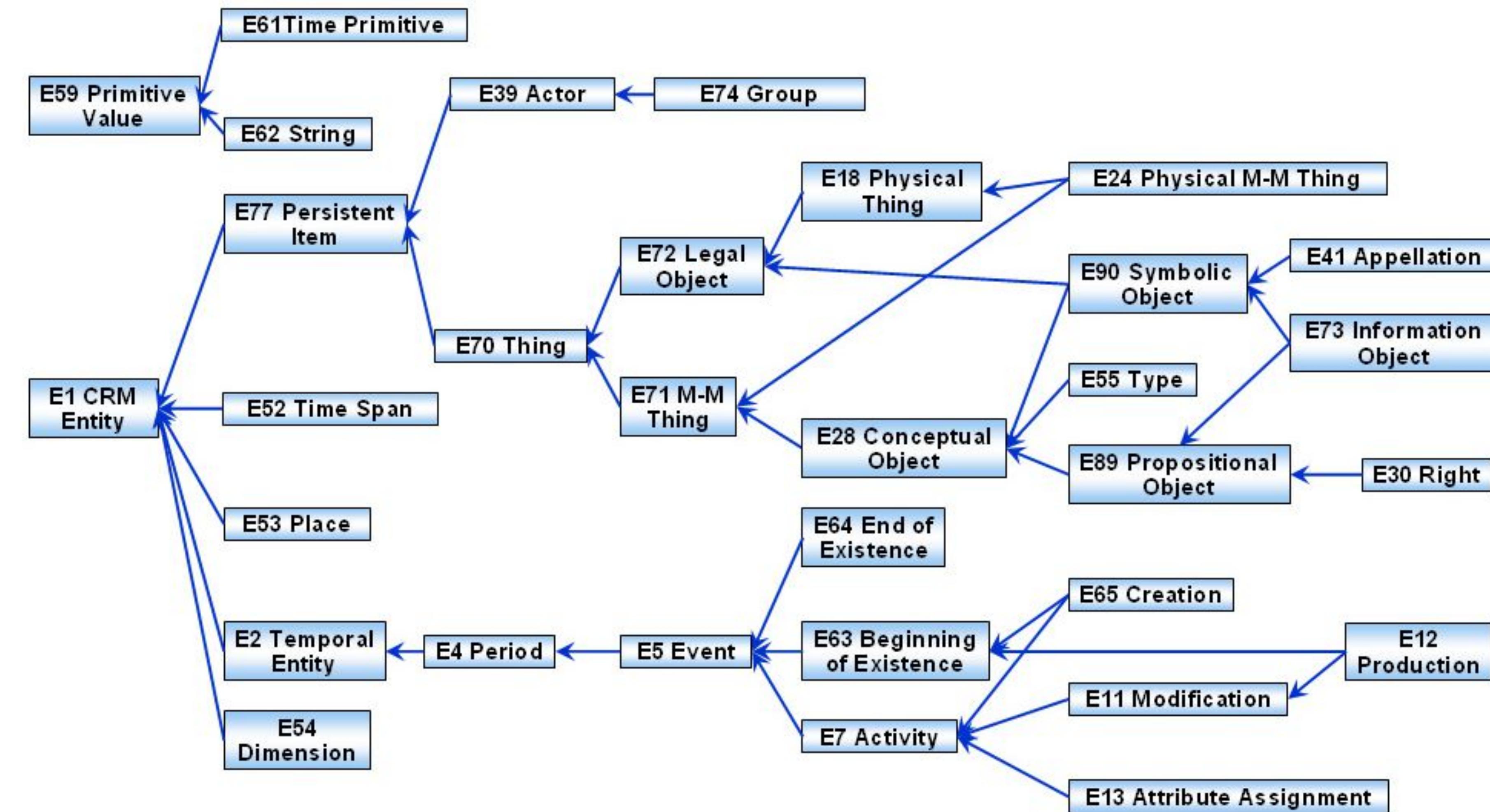
CIDOC-CRM



- Ontology for description of cultural processes and artefacts
- Event-based
- Widely adopted
- ISO standard
- Documentation and tools
- Flexible and extensible







SKOS—Simple Knowledge Organization System
is an ontology that can be used for thesauri, classification
schemes, subject heading lists, taxonomies, folksonomies
etc. (KOS).

The fundamental element of the SKOS vocabulary is the
concept. Concepts are the units of thought which underlie
many KOS



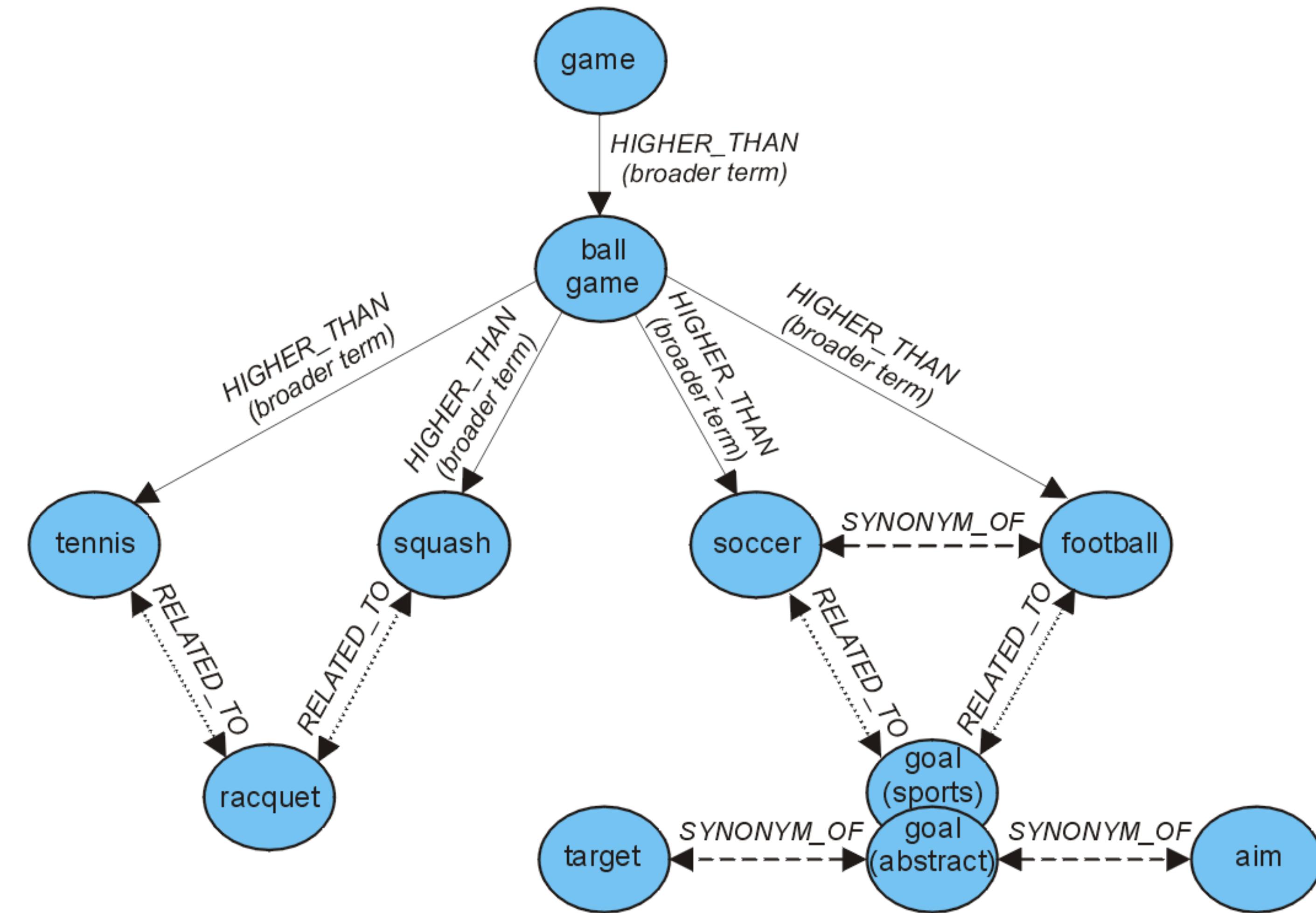
The first characterizations of concepts are the expressions that are used to refer to them in natural language: their labels. SKOS provides three properties to attach labels to conceptual resources: `skos:prefLabel`, `skos:altLabel` and `skos:hiddenLabel`.

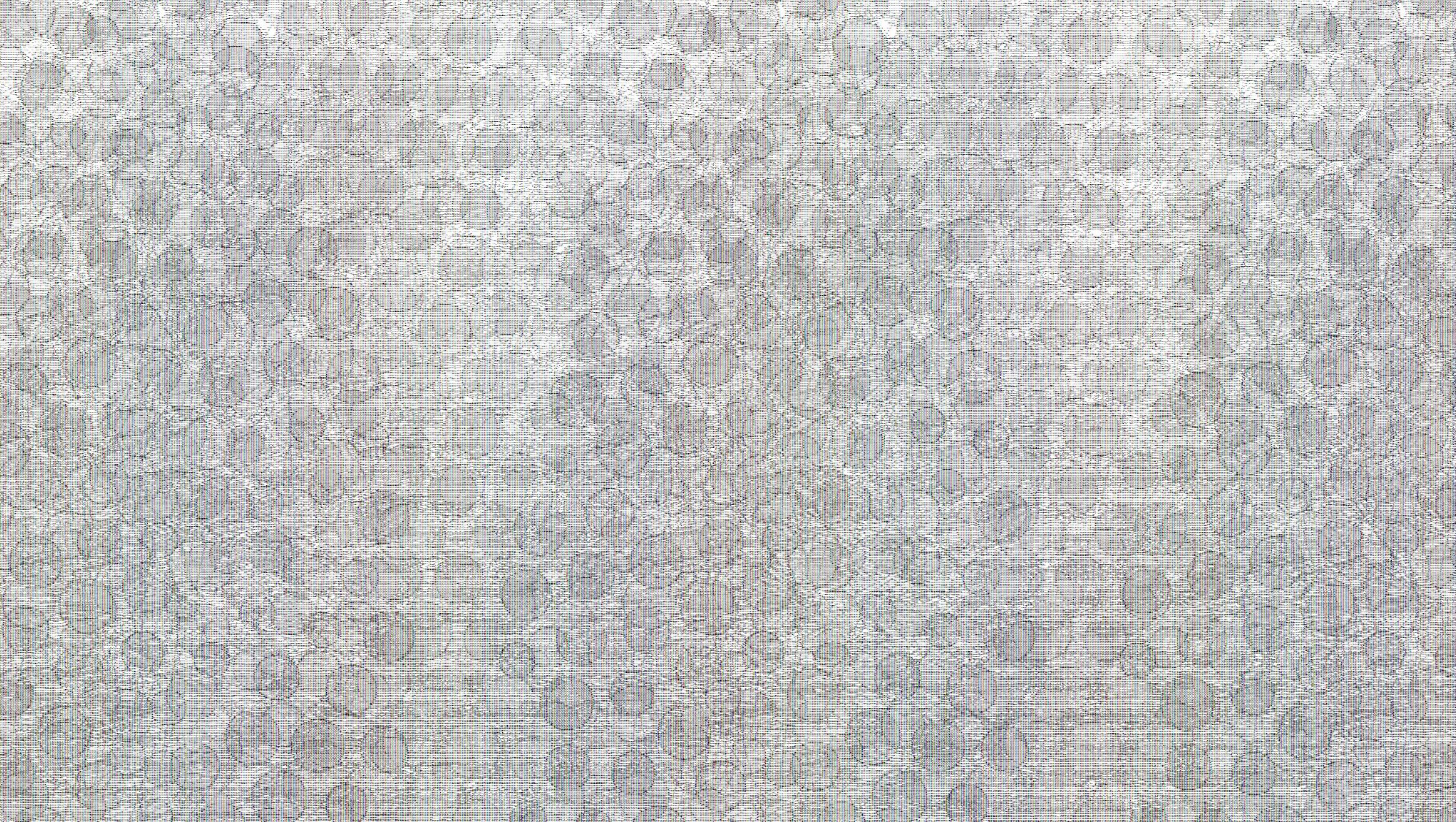


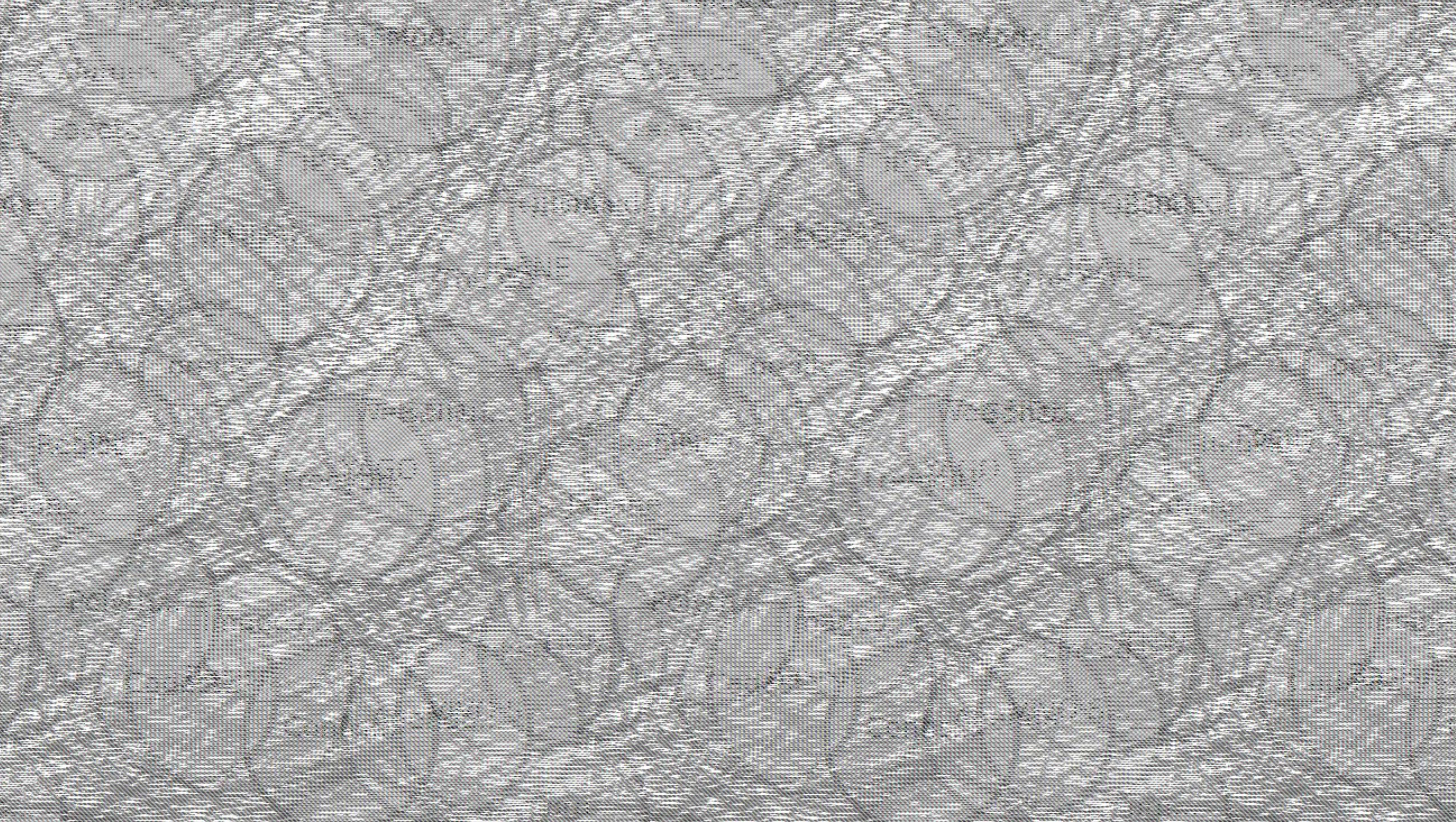
In KOSs semantic relations play a crucial role for defining concepts:

- **skos:broader** and **skos:narrower** enable the representation of hierarchical links, such as the relationship between one genre and its more specific species, or, depending on interpretations, the relationship between one whole and its parts;
- **skos:related** enables the representation of associative (non-hierarchical) links









<<https://municipality.org/person/giacomo>>

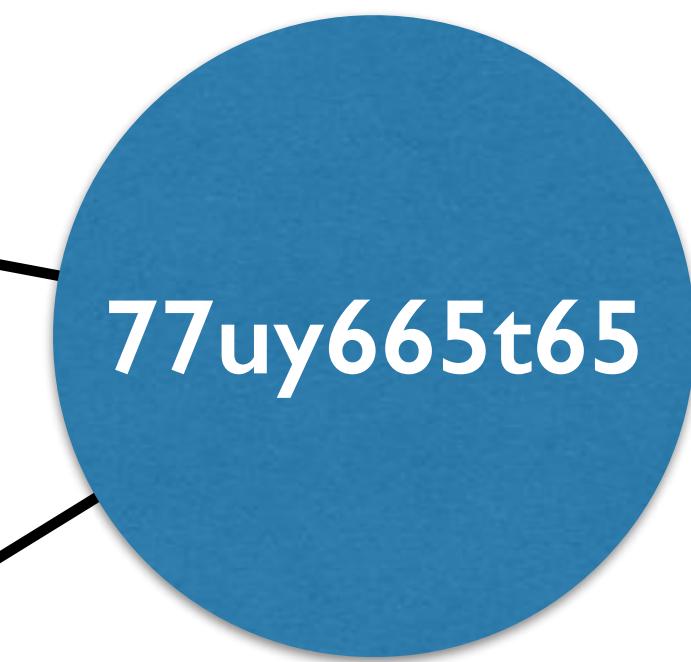


owl:sameAs

<<https://museum.org/8a1279e6>>



<<https://archive.org/77uy665t65>>



owl:sameAs



owl:sameAs

<<https://example.com/person/A00000822>>



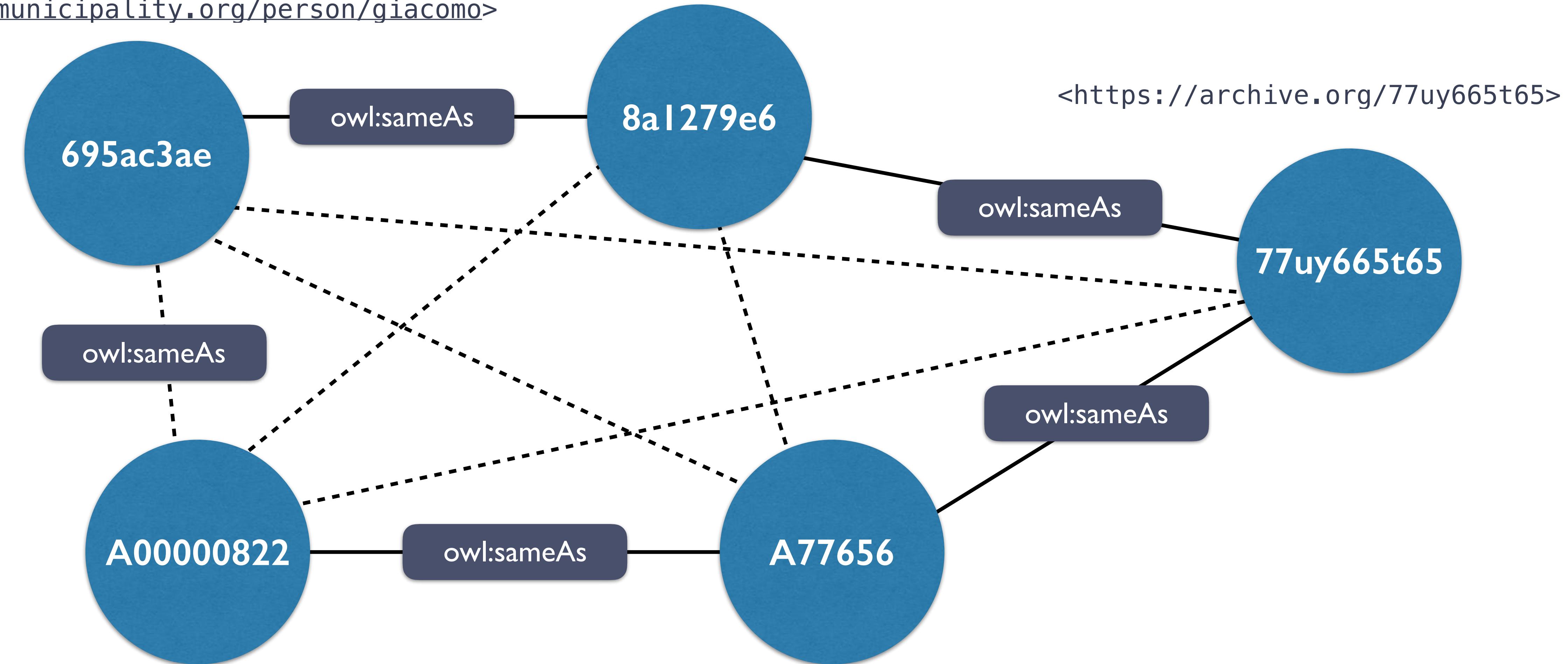
owl:sameAs

<<https://institution.org/A77656>>



<<https://municipality.org/person/giacomo>>

<<https://museum.org/8a1279e6>>



<<https://example.com/person/A00000822>>

<<https://institution.org/A77656>>



About 2,160,000 results (0.68 seconds)

Videos



[La Paranza dei Bambini \(2019\) - Trailer Ufficiale 90](#)

Vision Distribution
YouTube - Jan 18, 2019



[La Paranza dei Bambini, Clip del Film - HD - Film \(2019\)](#)

Coming Soon - Feb 12, 2019



[La Paranza dei Bambini \(2019\) - Trailer Ufficiale HD](#)

ScreenWeek TV | Trailer e...
YouTube - Jan 18, 2019



[More images](#)

Piranhas

2019 · Drama/Crime · 1 h 51m

[Play trailer on YouTube](#)

3.7/5
MyMovies

6.6/10
IMDb

4.2/5
Coming Soon

86% liked this film

Google users



Armed teenagers stalk the streets of Naples, eager to do their mob bosses' bidding.

Release date: February 13, 2019 (Italy)

Director: Claudio Giovannesi

Based on: [The Piranhas: The Boy Bosses of Naples](#); by Roberto Saviano

Language: Italian; Neapolitan

Production company: Palomar

Cast: Valentina Vannino, Carmine Pizzo, Ar Tem, Luca Nacarlo, MORE

Available on

YouTube

From €4.99

[La paranza dei bambini \(2019\) - MYmovies.it](#)

<https://www.mymovies.it/film/2019/la-paranza-dei-bambini/> ▾ [Translate this page](#)

Rating: 3.7 - 494 votes

Jul 4, 2018 - **La paranza dei bambini** - Un film di Claudio Giovannesi. Il racconto di un' innocenza compromessa, osservata con una compassione priva di ...

Regia di: Claudio Giovannesi

Uscita: mercoledì 13 febbraio 2019

Al cinema: 3 sale cinematografiche

Durata: 111 minuti

[Trailer La paranza dei bambini](#) · Il valore dell'attore non ... · [La paranza dei bambini](#)

[La paranza dei bambini \(2019\) - IMDb](#)

<https://www.imdb.com/title/tt8740778/> ▾

Rating: 6.6/10 - 1,165 votes

A gang of teenage boys stalk the streets of Naples armed with hand guns and AK-47s to do their mob bosses' bidding. ... Roberto Saviano (novel), Claudio Giovannesi (screenplay) ... Francesco Di Napoli, Viviana Aprea, Mattia Piano Del Balzo.

[La paranza dei bambini - Wikipedia](#)

https://it.wikipedia.org/wiki/La_paranza_dei_bambini ▾ [Translate this page](#)

La paranza dei bambini è il terzo romanzo di Roberto Saviano, pubblicato nel 2016 dalla casa editrice



ULAN + AAT



Iconclass



Geonames



VIAF



GND



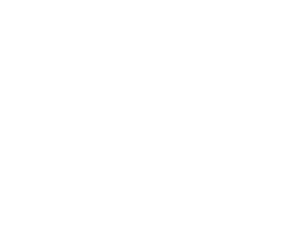
LOC



Inventaire



English Heritage



```

SELECT ?form_plu ?word_plu WHERE

{
  {
    ?l a ontolex:LexicalEntry ;
        dct:language ?language ;
        wikibase:lexicalCategory wd:Q1084 ;
        ontolex:lexicalForm ?form .
    ?language wdt:P218 'it'.

    ?s ?p ?form .
    ?form wikibase:grammaticalFeature wd:Q110786 ;
          ontolex:representation ?word .
    FILTER( regex(?word, "cane" ) ) . }

    ?o ontolex:lexicalForm ?form .
    ?o ontolex:lexicalForm ?form_plu .
    ?form_plu wikibase:grammaticalFeature wd:Q146786 ;
              ontolex:representation ?word_plu .

  }
}
  
```





OpenRefine

- Created by MetaWeb Technology as Gridworks in 2009
- Acquired by Google and rebranded as Google Refine
- Google continue its development: version 2.0 and 2.5
- Stop supporting in 2012
- OpenRefine is born: 2013



What is data Refining, Wrangling, Engineering...

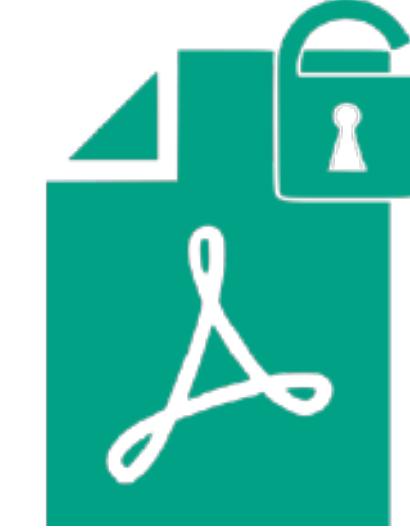
WHAT

- Removing duplicates
- Typos
- Joining field
- Data in wrong field
- Change of format (date)
- Encoding errors
- Join data sets
- Transpose row/column
- Enrichment

AIMS

- Integrating data
- Use & Re-use
- Sharing
- Analytics





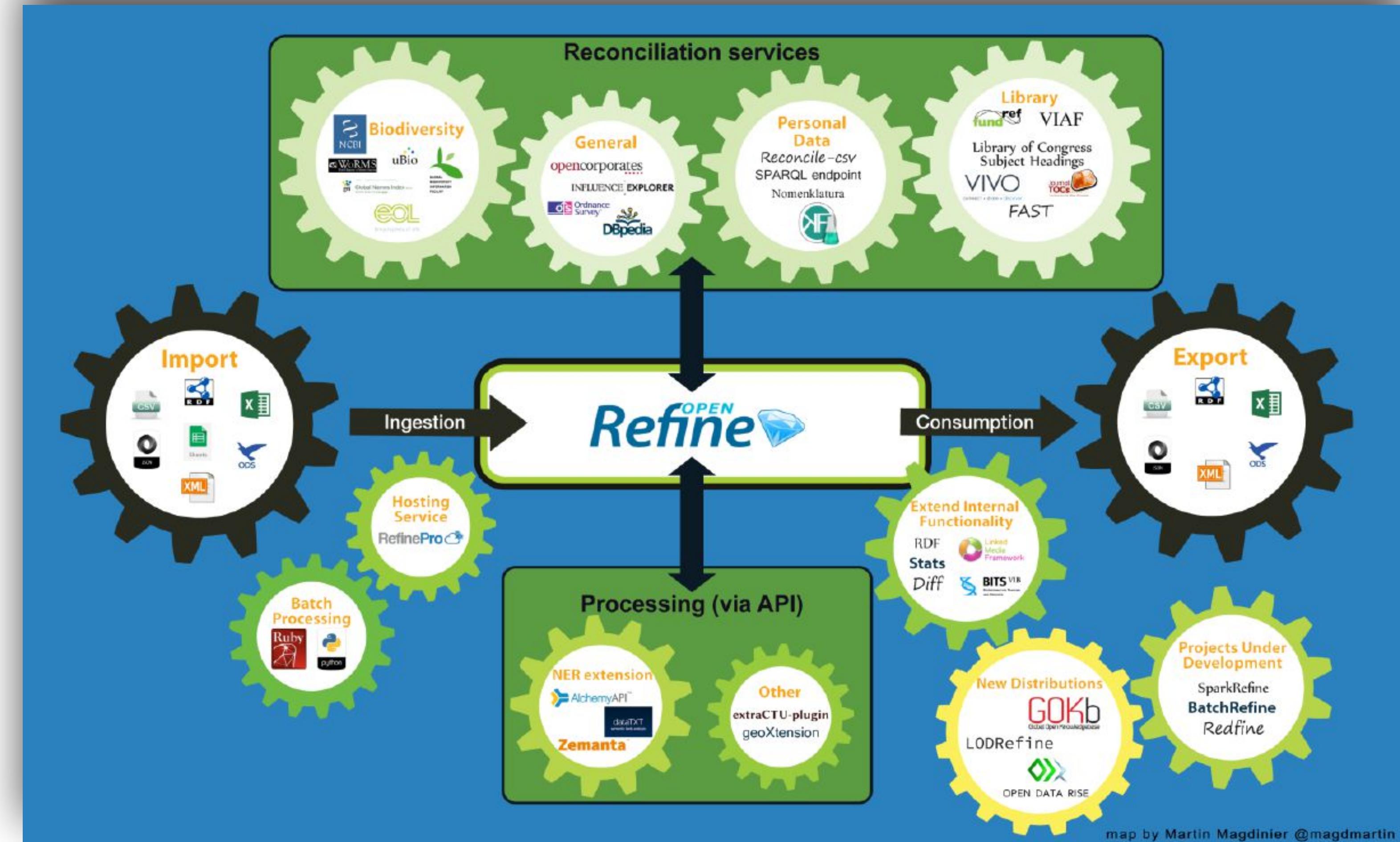
pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

CSVKit

Silk

tripleGEO

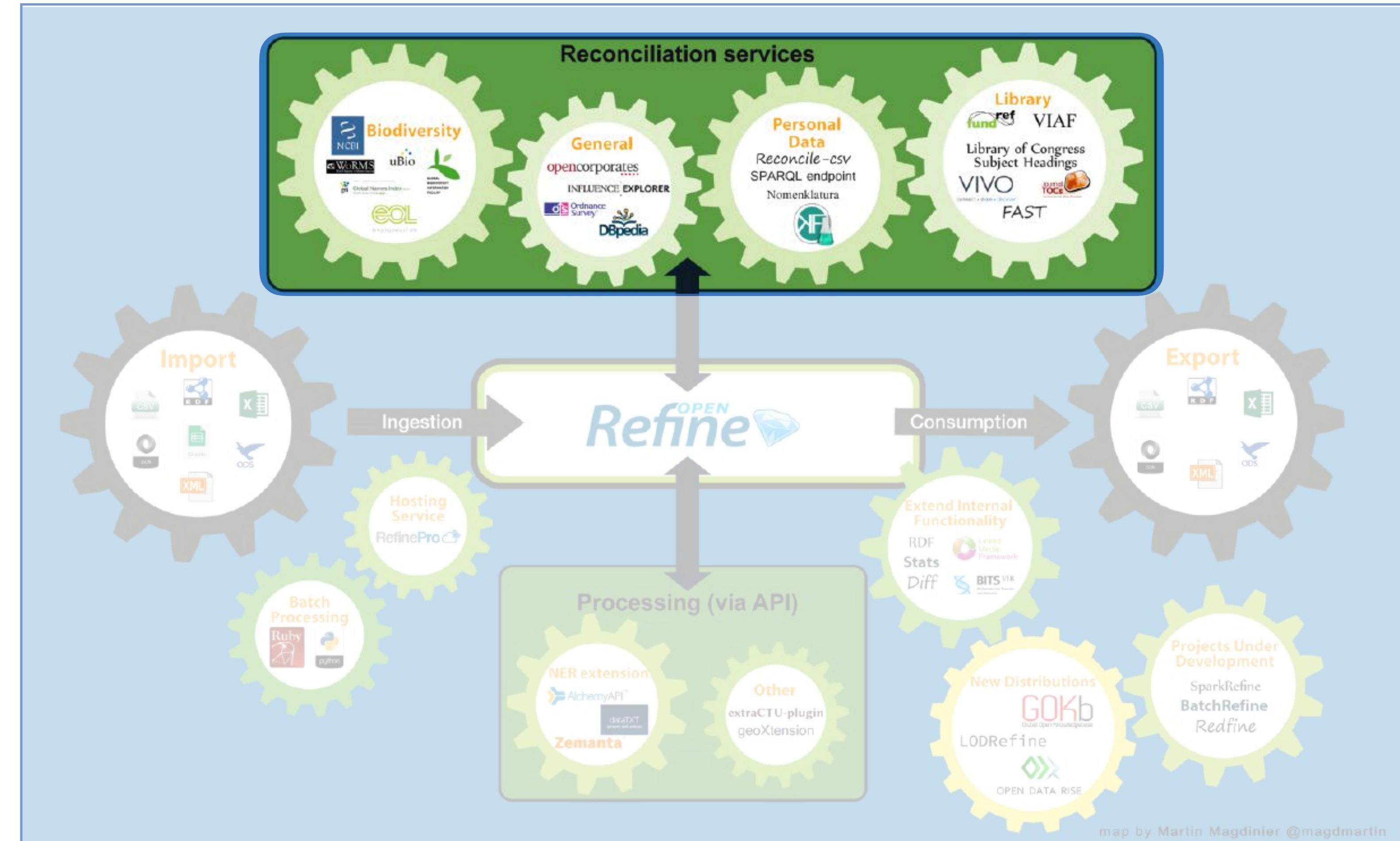


Nicola Carboni

Trasogno Roma. Leonardo tra antico e moderno

12 September 2019, Roma



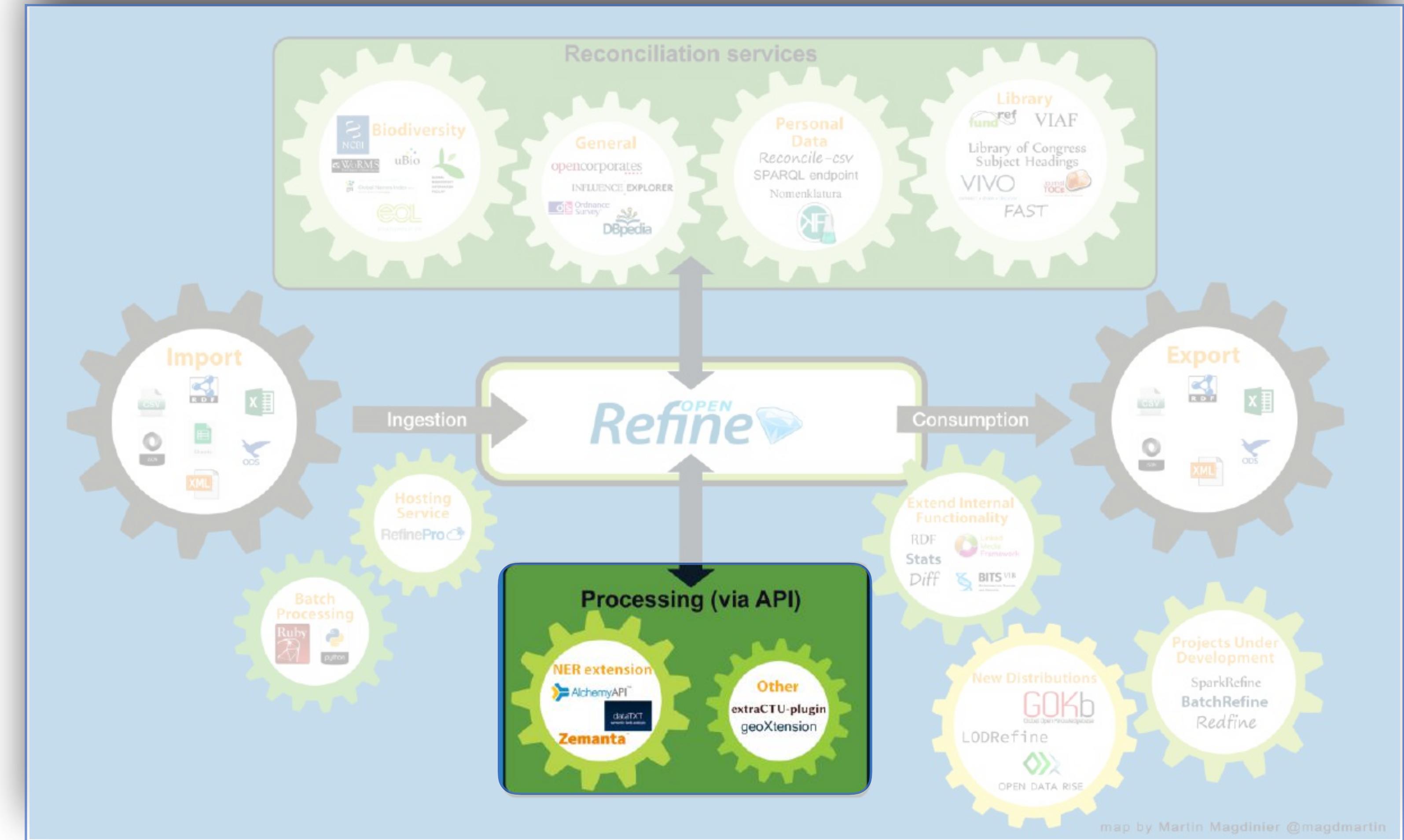


Nicola Carboni

Trasogno Roma. Leonardo tra antico e moderno

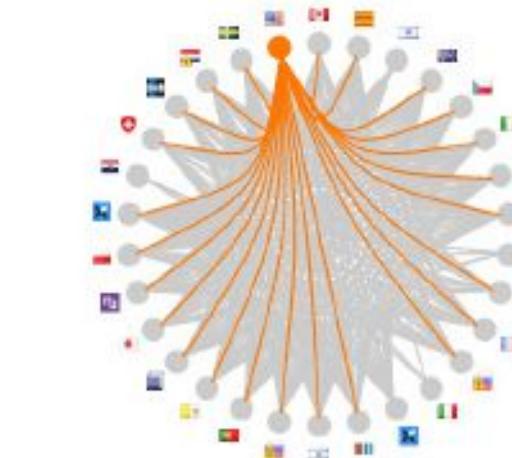
12 September 2019, Roma





Reconciliation

Compare value in my dataset with value from an external source. If they match link them and extract information



Ge^oNames



Wikidata



- Wikimedia knowledge base
- Single source of structured information for wiki*
- Organised with unique ID and attribute-value pair <attribute name, value>

http://bit.ly/workshop_leonardo

Nicola Carboni

Trasogno Roma. Leonardo tra antico e moderno

12 September 2019, Roma



Facet: Pattern in the data - birth - death date

(b. 1728, Paris, d. 1799, Paris)

(b. 1728, Wien, d. after 1778, Firenze)

(third quarter of 15th century in Valsesia)

(documented 1525-1551 in Milan)



Date - Add column based on this column→

```
import re
pattern = re.compile(r"\d{3,4}")
return pattern.findall(value)[0]
```

Date - Add column based on this column→

```
import re
pattern = re.compile(r"\d{3,4}")
return pattern.findall(value)[1]
```



Facet by blank for
object_creation_date_start
+
object_creation_date_end



numbers from centuries: Jython/Python

```
import re
pattern = re.compile(r"\d+(?=.)")
return pattern.findall(value)[0]
```



delete values using Jython

```
import re  
return re.sub(r"\d{3,4}", "", value)
```



replace for places

```
value.replace("'", "'").replace("'", "'").replace('ca', ''')  
    .replace('c.', "'").replace('?.', "'").replace('.','')  
.replace('before', "'").replace('/', "'").trim()
```



Reconcile

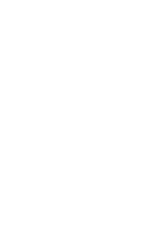
cell.recon.match.id

<https://tools.wmflabs.org/sqid/>



API commons

```
"https://commons.wikimedia.org/w/api.php?  
action=query&format=json&prop=imageinfo&iiprop=url&t  
itles=" + escape("File:" + value, 'url')
```



**value.parseJson().query
value.parseJson().pages + split by columns**



cell.recon.match.id

