Francesco Beretta
(CNRS UMR5190 LARHRA – Université de Lyon)

Data modelling IV:

Historical information and foundational ontologies

ACDH Digital Prosopography Summer School

Wien, 7 July 2020

Francesco Beretta (CNRS/Université de Lyon), 7 July 2020 CC BY-NC-SA 4.0

Findable

Accessible

Interoperable

Re-usable

«There is an urgent need to improve the infrastructure supporting the ***reuse*** of scholarly data »

Wilkinson, Mark D., Michel Dumontier, Ijsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. "
The FAIR Guiding Principles for Scientific Data Management and Stewardship." Scientific Data 3 (March 15, 2016): 160018.

The FAIR Data Principles

To be **Interoperable**:

I1. (meta)data use a *formal, accessible, shared, and broadly applicable language for* ***knowledge representation***.

I2. (meta)data use *vocabularies that follow FAIR principles*.

I3. (meta)data include qualified references to other (meta)data.

To be **Re-usable**:

R1. meta(data) have a plurality of accurate and relevant attributes.

R1.1. (meta)data are released with a *clear and accessible data usage license*.

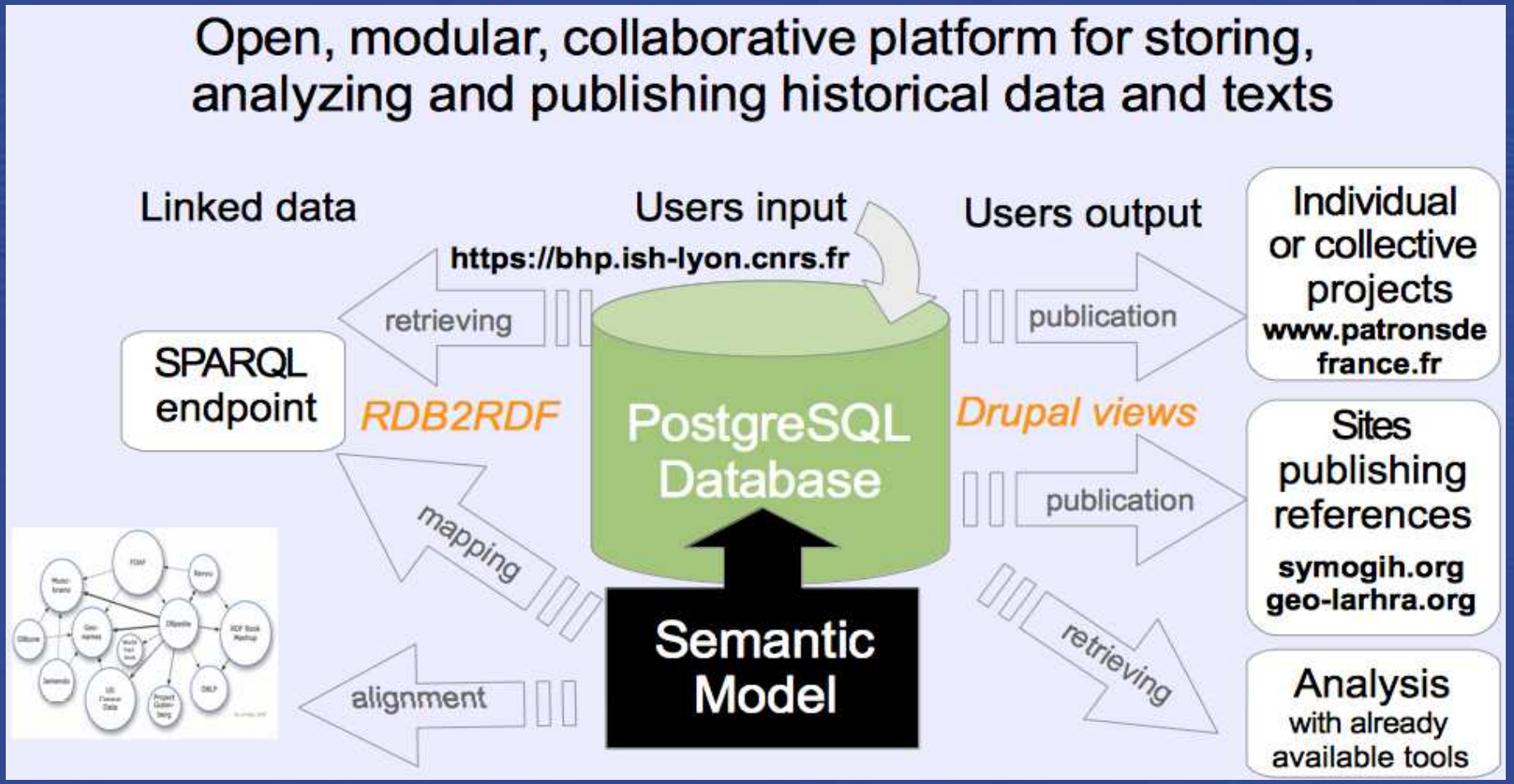
R1.2. (meta)data are associated with their *provenance*.

R1.3. (meta)data meet ***domain-relevant community standards***.

“An ontology is
a formal explicit specification
of a shared conceptualization
of a domain of interest”

- « Formality – ... a knowledge representation language that is based on the grounds of **formal semantics**. »
- « Consensus – … an agreement on a domain conceptualization among people in a community. »
- « Conceptuality – … in terms of conceptual symbols that can be intuitively grasped by humans, as they correspond to the elements in their **mental models**. »
- « Domain Specificity – ... limited to knowledge about a particular **domain of interest**. »

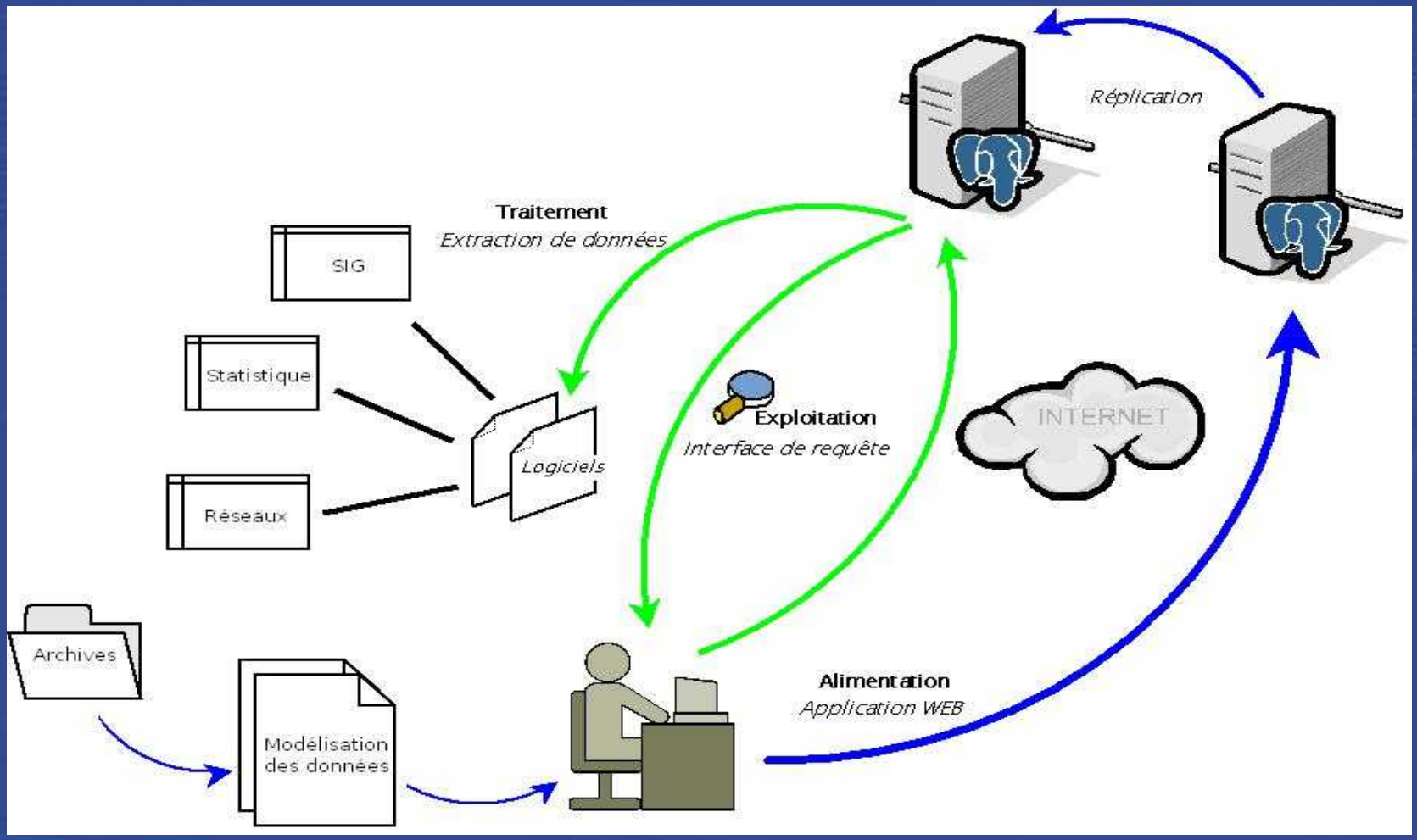
[Domingue et al. 2011, p. 510-511]



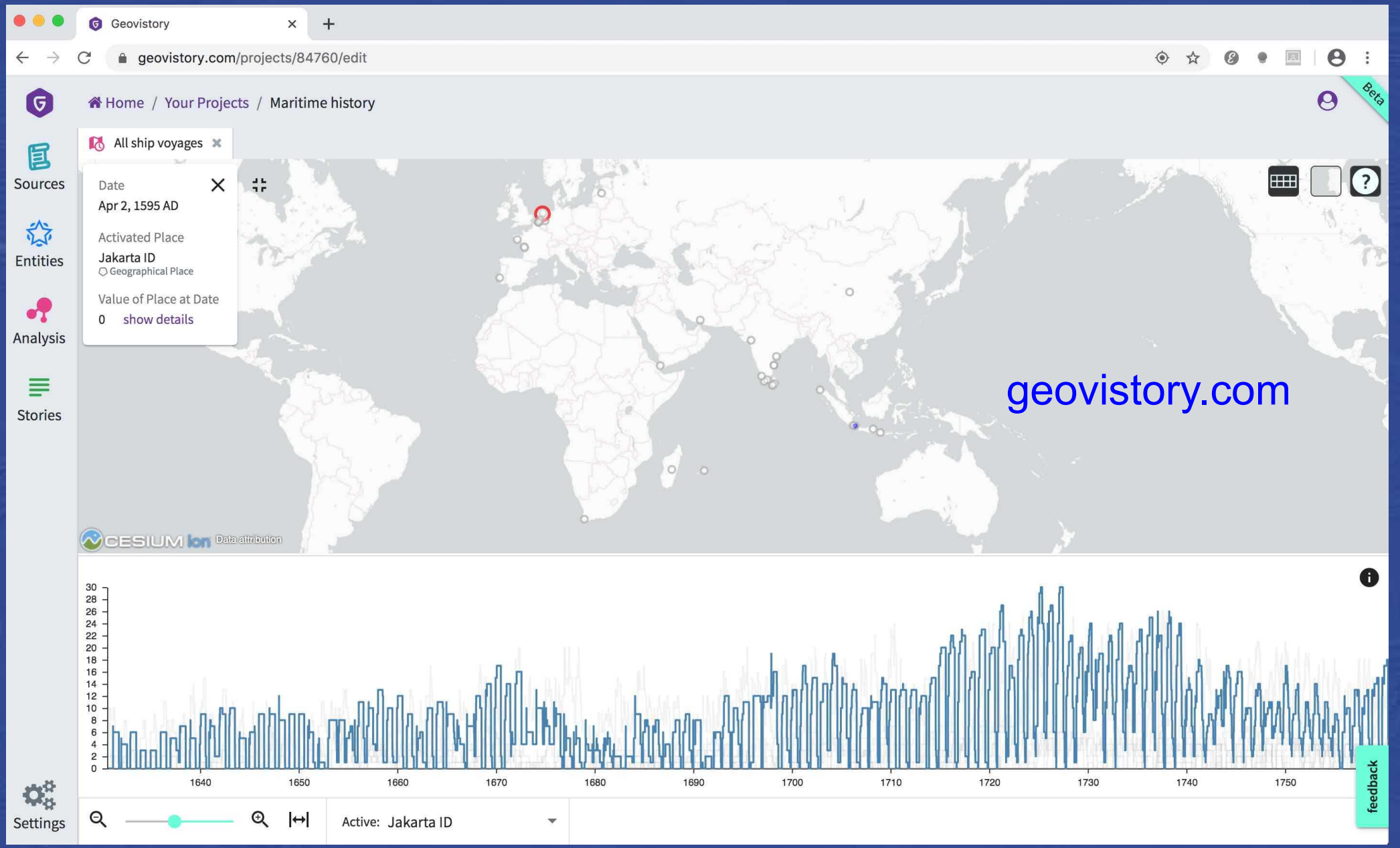
The *symogih.org* project was started in 2008.

About 50 scholars and students, and 15 research project, used or are currently using the collaborative database to store and share historical information

The *symogih.org* project : modular system for historical information management

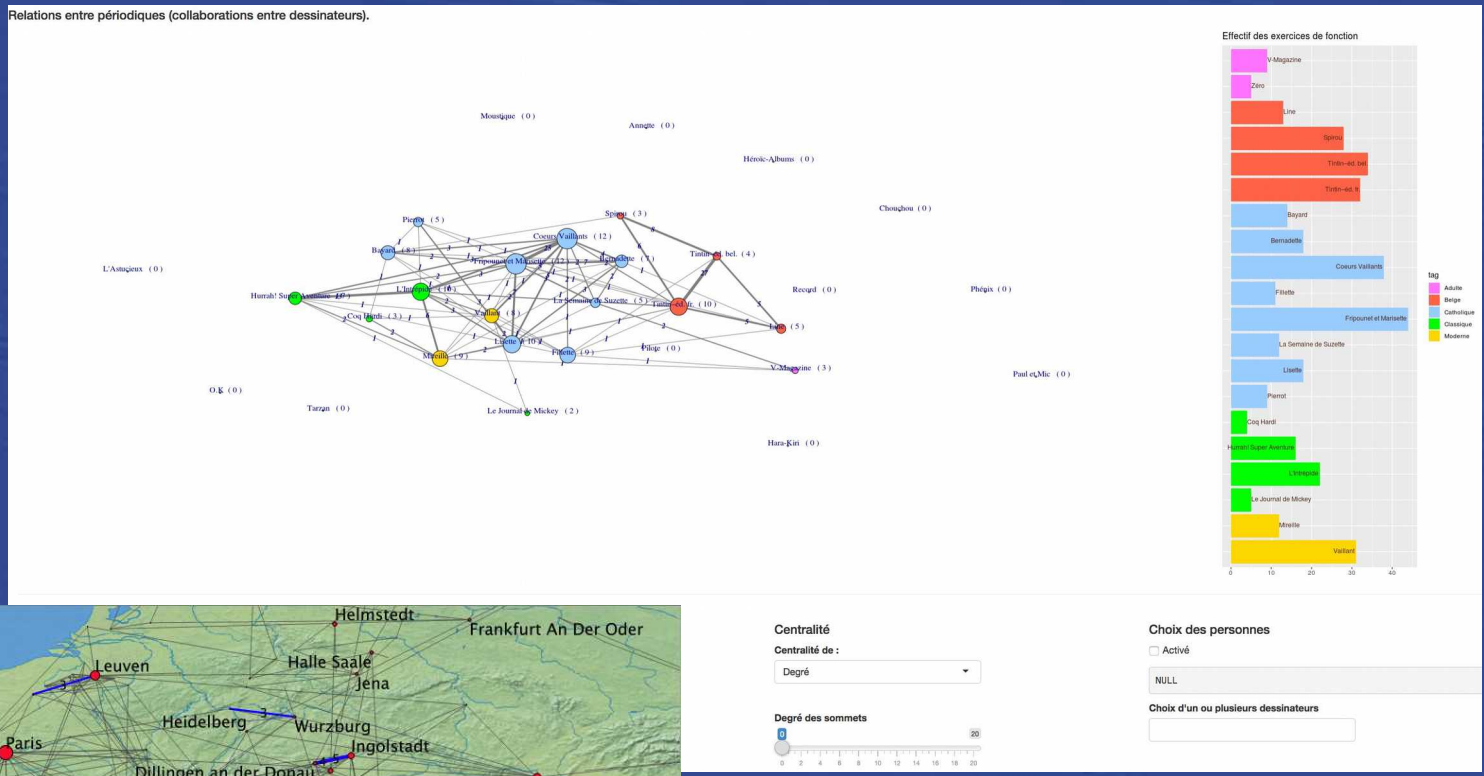


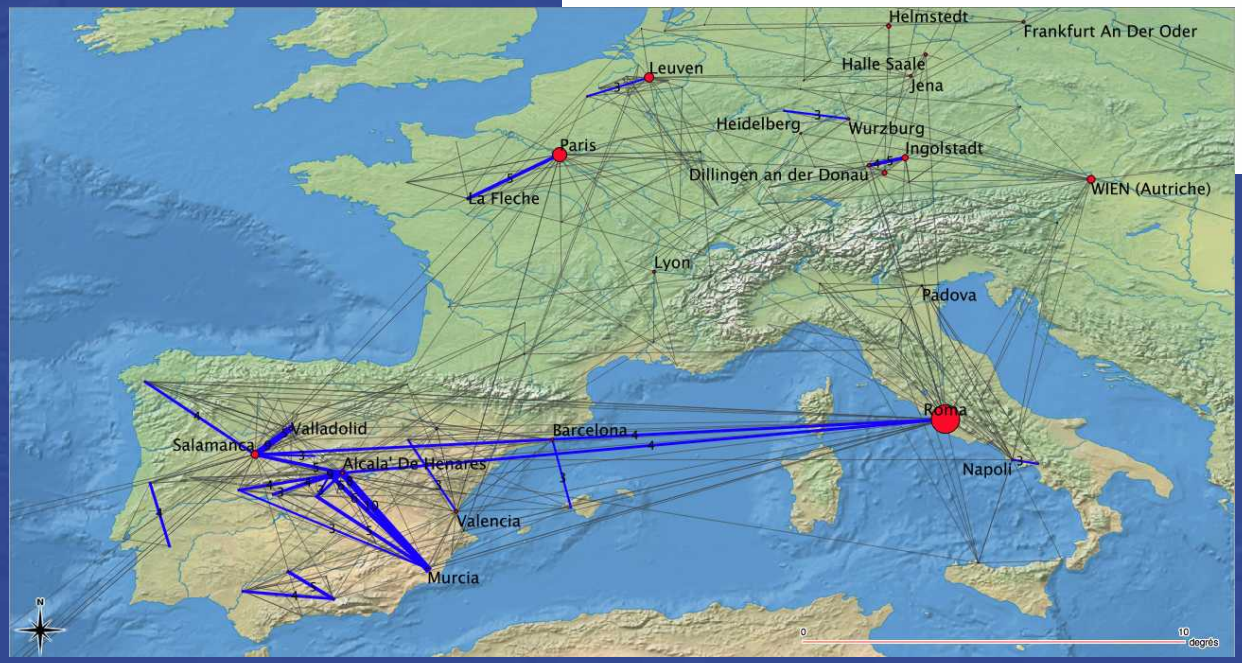
A collaborative and cumulative information system for storing historical data: produce, share, visualise and analyse stuctured historical data



Geovistory : a new VRE for the symogih.org project

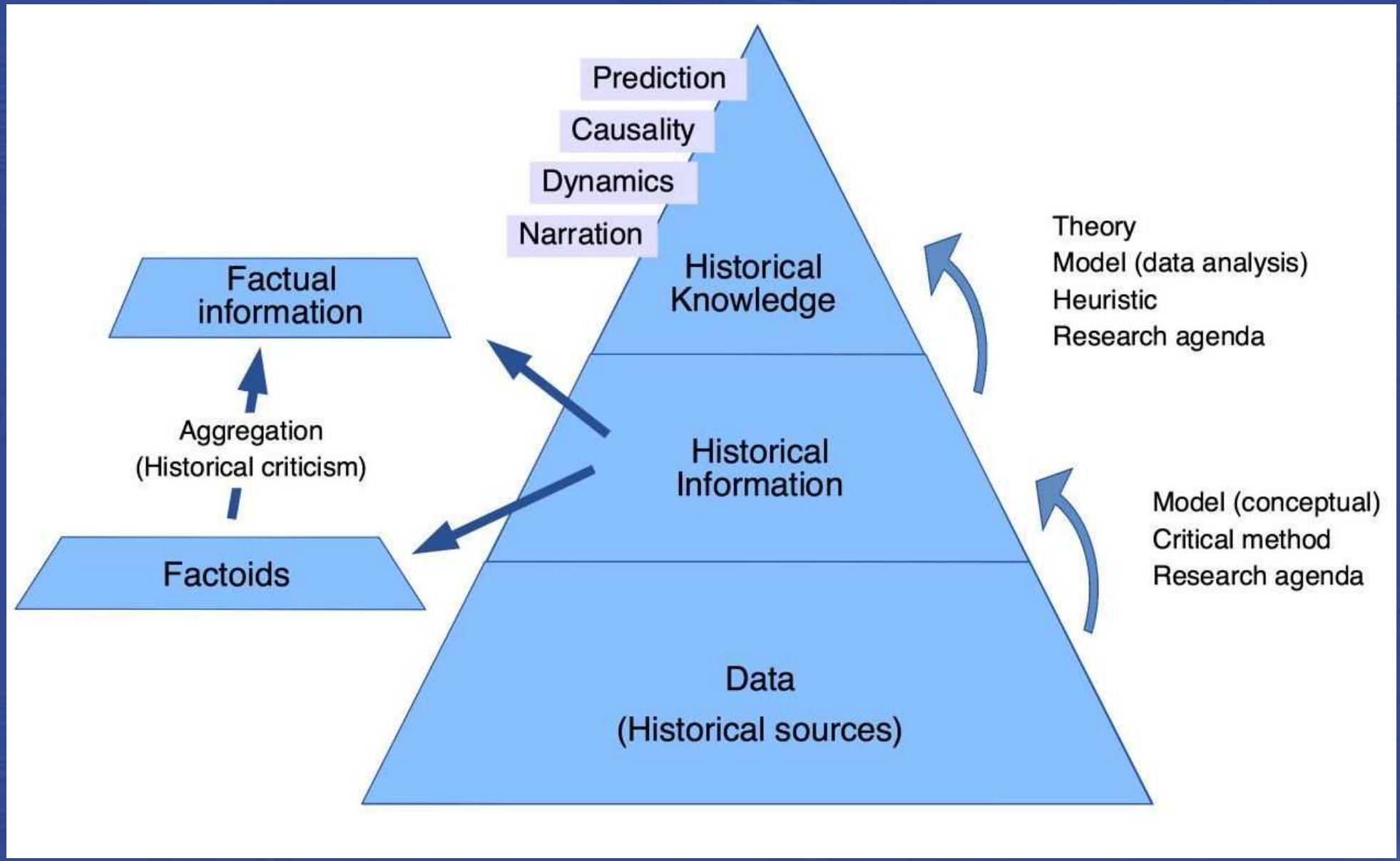
Interactive and collaborative data analysis
Huma-Num RStudio-Shiny :https://tinyurl.com/phn-shiny





Local, individual analysis
QGIS and network analysis

The relevant domain : historical knowledge production



“In 1592, Galileo Galilei was hired by the University of Padua, where he taught mathematics until 1610.”

Letter by A to B, 11 March 1616 / Dictionary of astronomers, 2020

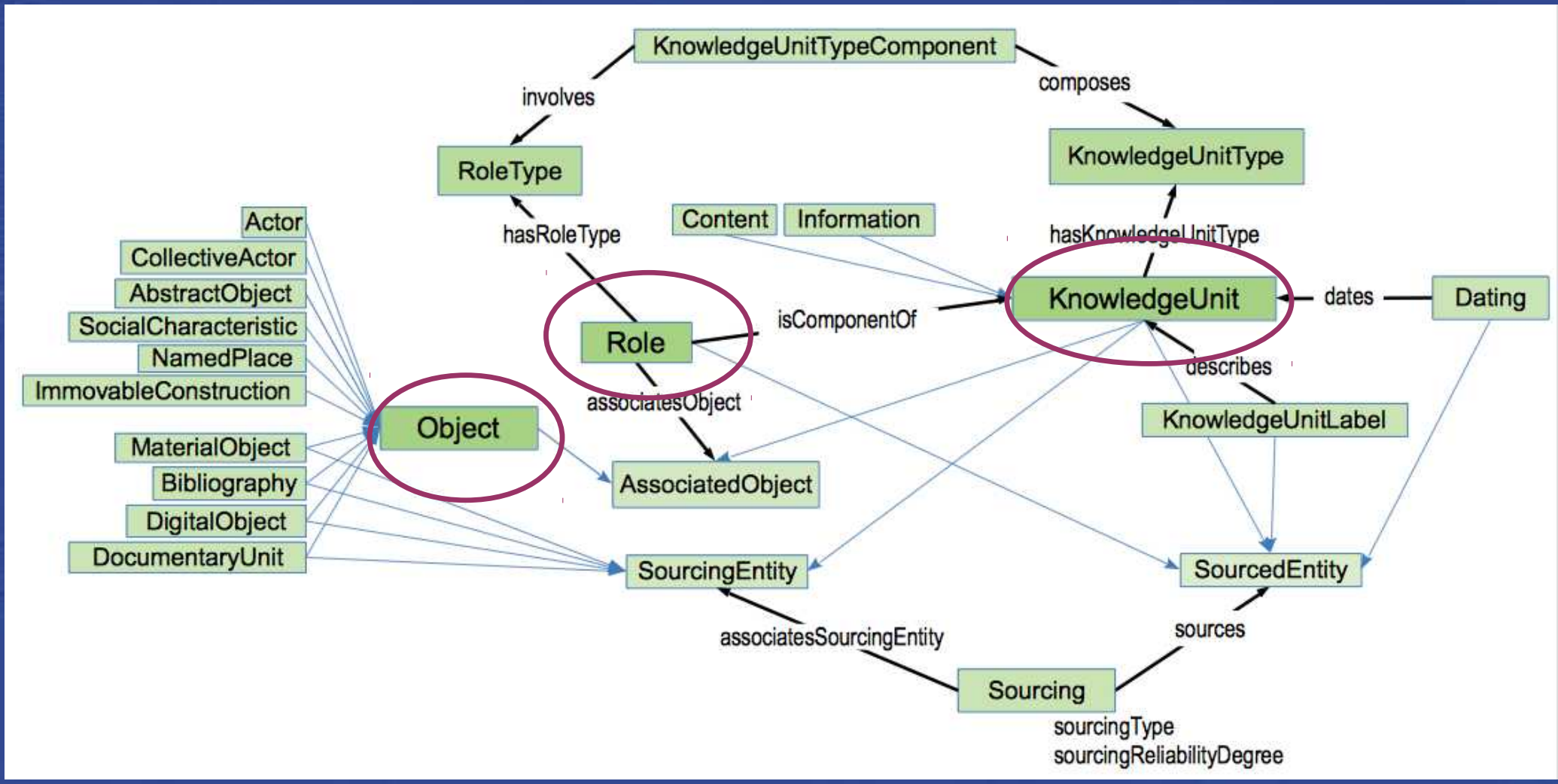
“Galileo Galilei was convicted in 1633 not because of heresy but for political reasons, in the context of the Thirty Years' War.”

Dictionary of astronomers, 2020

“In 1592, Galileo Galilei was hired by the University of Padua, where he taught mathematics until 1610.”

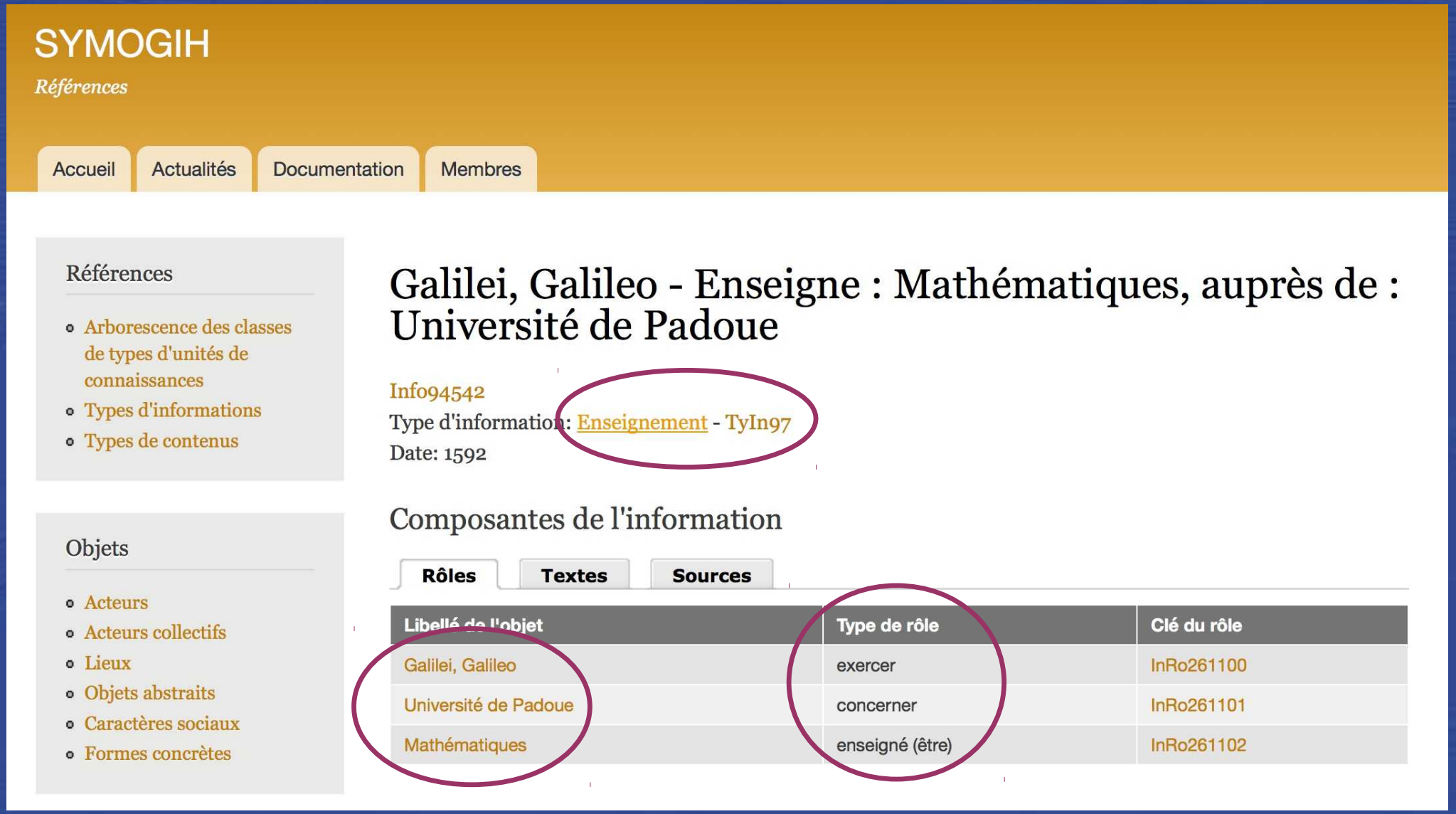
Possible factual information / modelling patterns:

- Galileo Galilei taught mathematics at the University of Padua from 1592 to 1610
- that Galileo from 1592 resided in the city of Padua
- that he was hired by and member of the University
- that he held the title of professor regardless of whether or not he was effectively teaching



The symogih.org ontology

SPARQL Endpoint



Galileo Galilei taught mathematics at the University of Padua from 1592 to 1610

The symogih.org project's modelling patterns (more then 150 in 10 years)

SYMOGIH

Références

Accueil | Documentation | Membres

Références

- Arborescence des classes de types d'unités de connaissances
- Types d'informations
- Types de contenus

Objets

- Acteurs
- Acteurs collectifs
- Objets abstraits
- Caractères sociaux

Classes de

Chercher une

- Biographi
  - Ense
  - Exer
  - Fin
  - Lien
  - Loca
  - Rite
  - Vie
  - Vie

Enseignement

TyIn97

Exercer la fonction d'enseigner, avec indication de l'institution auprès de laquelle s'exerce l'enseignement et des matières enseignées.
Il s'agit d'un cas particulier du TyIn 'Exercice d'une fonction' : cf. Classe TyIn 'Exercice d'une fonction'.
Attention : ne pas renseigner le lieu si on peut localiser l'institution elle-même.

Liste des types de rôles associés

Libellé du type de rôle	Clé du TyRo	Description
concerner	TyRo21	Institution auprès de laquelle s'exerce l'enseignement. Ce rôle a été gardé pour être l'équivalent du TyIn7 : Exercice d'une fonction.
enseigné (être)	TyRo131	La matière enseignée (un objet abstrait). On peut en associer plusieurs si on enseigne en même temps plusieurs matières. En revanche, il faut créer plusieurs informations si les enseignements des différentes matières se succèdent ou si le contexte institutionnel est différent
exercé (être)	TyRo47	Qualification de l'enseignement : professeur, chargé de cours, etc. Ce rôle a été gardé pour être l'équivalent du TyIn7 : Exercice d'une fonction.
exercer	TyRo12	Ce rôle a été gardé pour être l'équivalent du TyIn7 : Exercice d'une fonction.
localiser	TyRo8	Ne pas renseigner si l'institution auprès de laquelle s'effectue l'enseignement est déjà localisée.
occasionner la fin	TyRo176	Associe l'information ou le AbOb qui explique la fin de l'enseignement
origine (être l')	TyRo16	Associe l'information (nomination, élection, ...) ou l'objet abstrait qui indiquent la cause de l'enseignement
typer	TyRo98	A utiliser dans le contexte de ce TyIn pour spécifier la nature de l'enseignement grâce à un AbOb (cours magistral, séminaire, etc.).

MCD disponible(s)

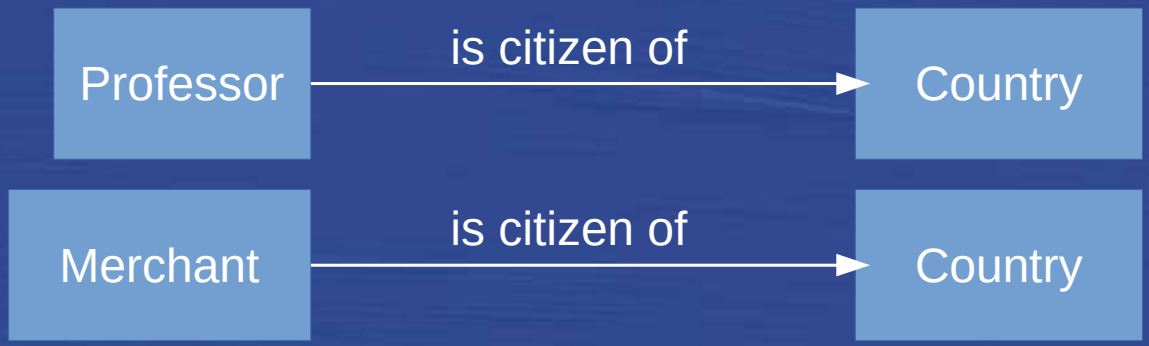
Télécharger ce MCD

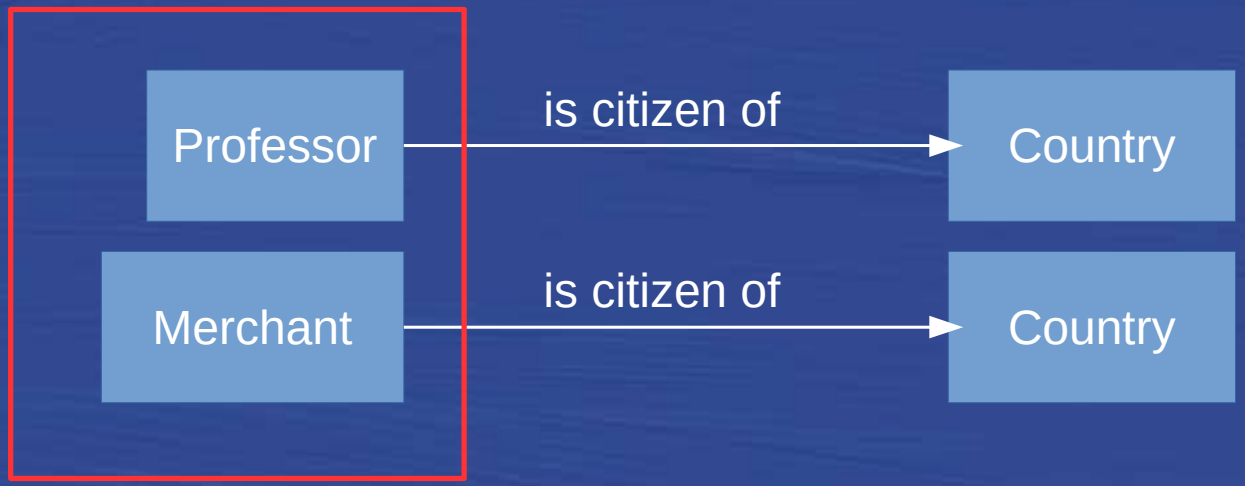
Patterns tree

“In 1592, Galileo Galilei was hired by the University of Padua, where he taught mathematics until 1610.”

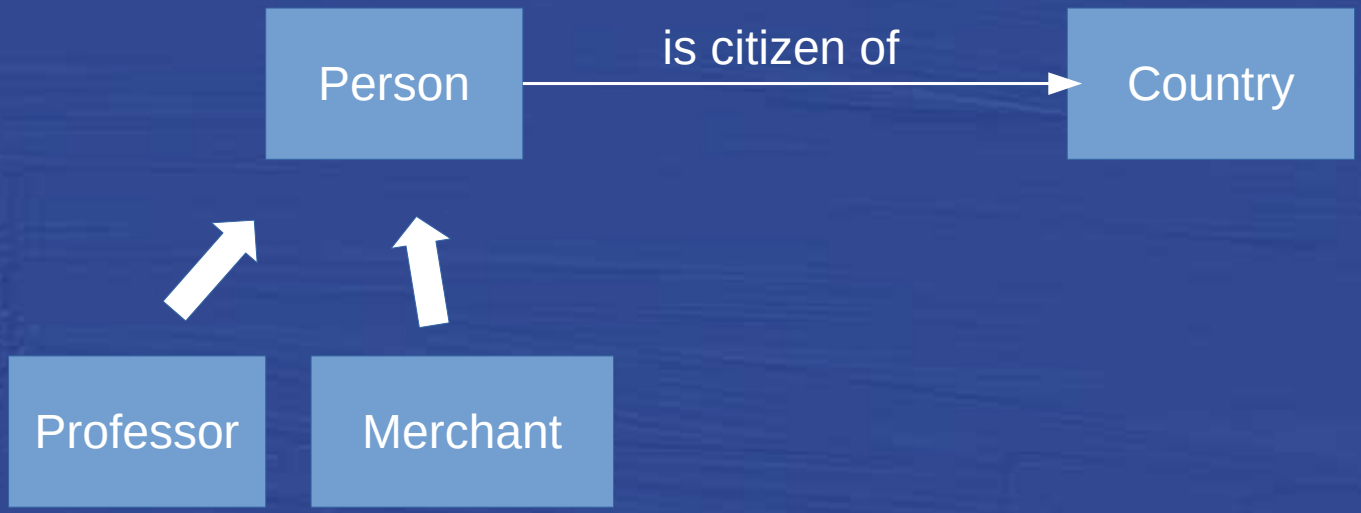
Possible factual information / modelling patterns:

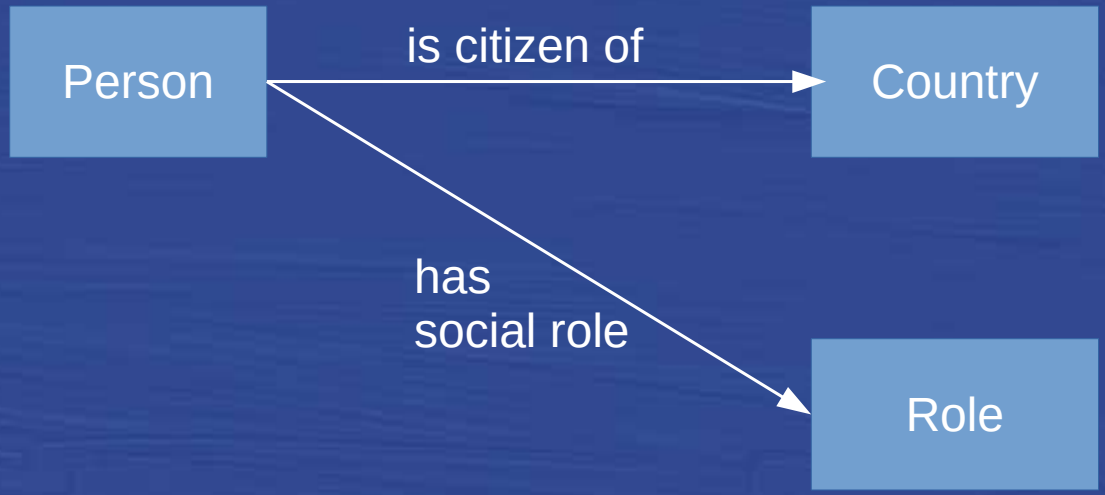
- Galileo Galilei taught mathematics at the University of Padua from 1592 to 1610
- that Galileo from 1592 resided in the city of Padua
- that he was hired by and member of the University
- that he held the title of professor regardless of whether or not he was effectively teaching

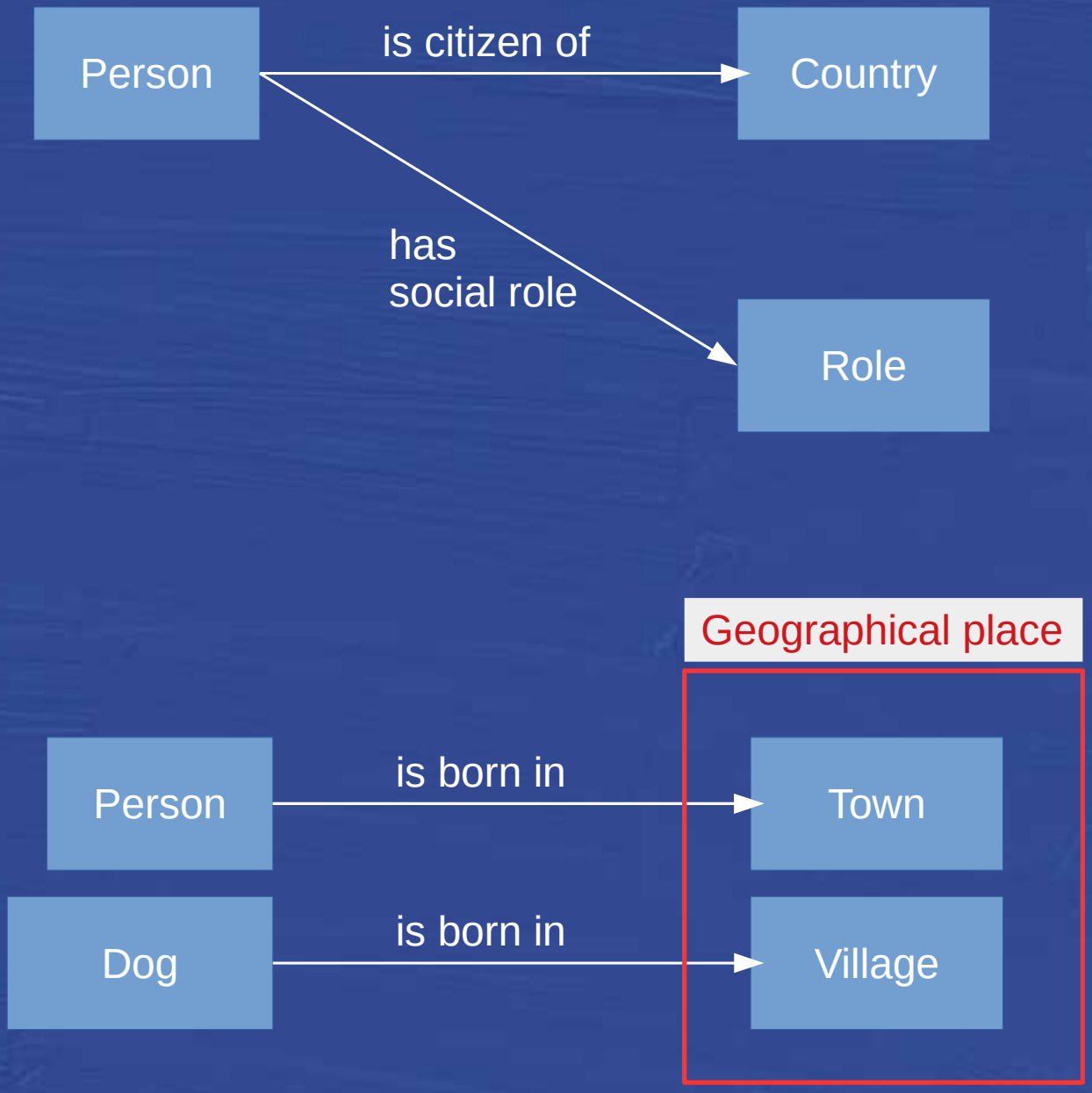

Professor
is citizen of
Country
Merchant
is citizen of
Country

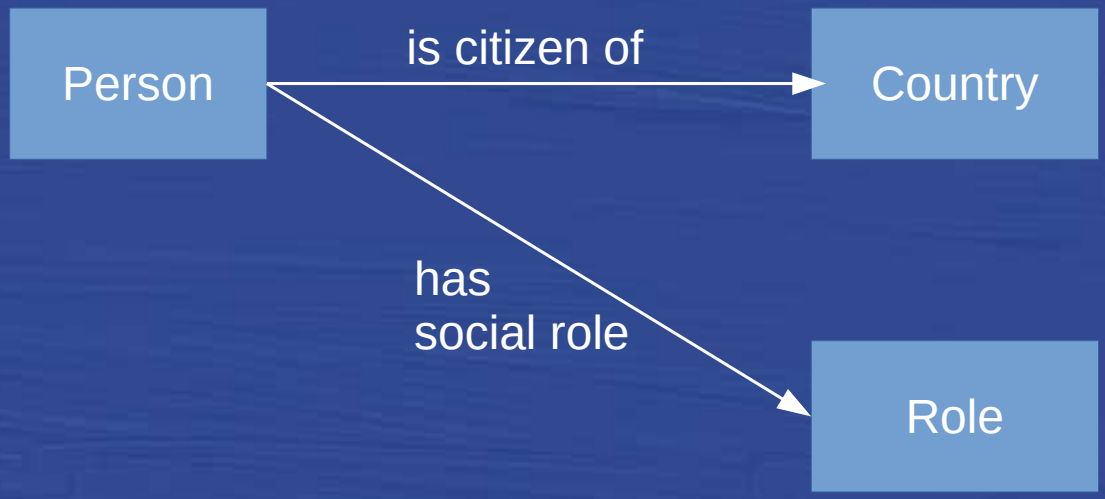


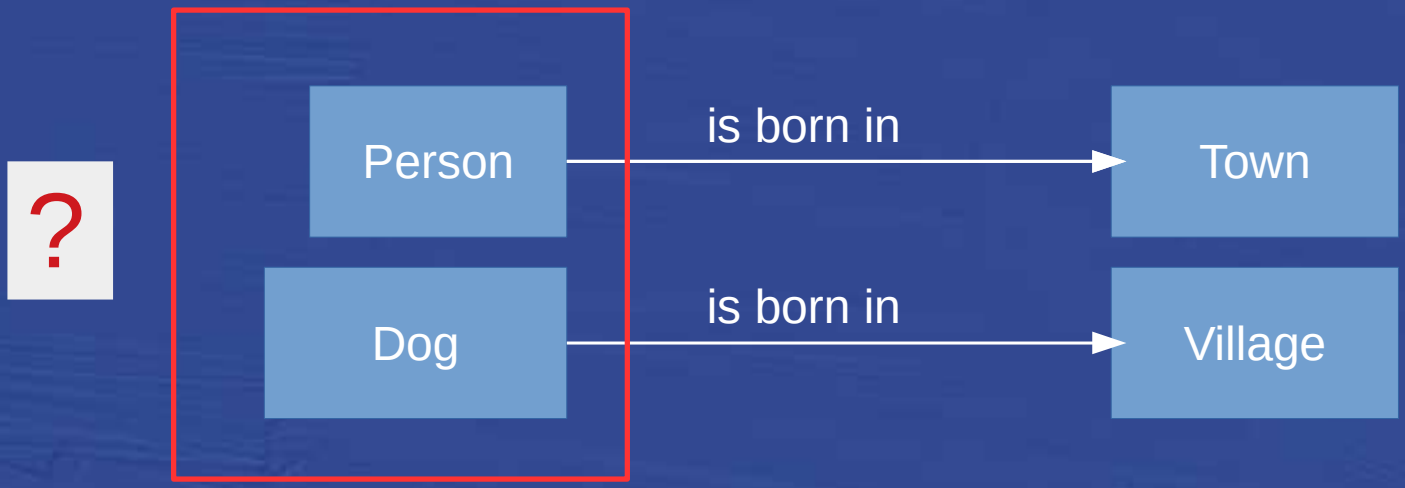
Every professor is also a merchant

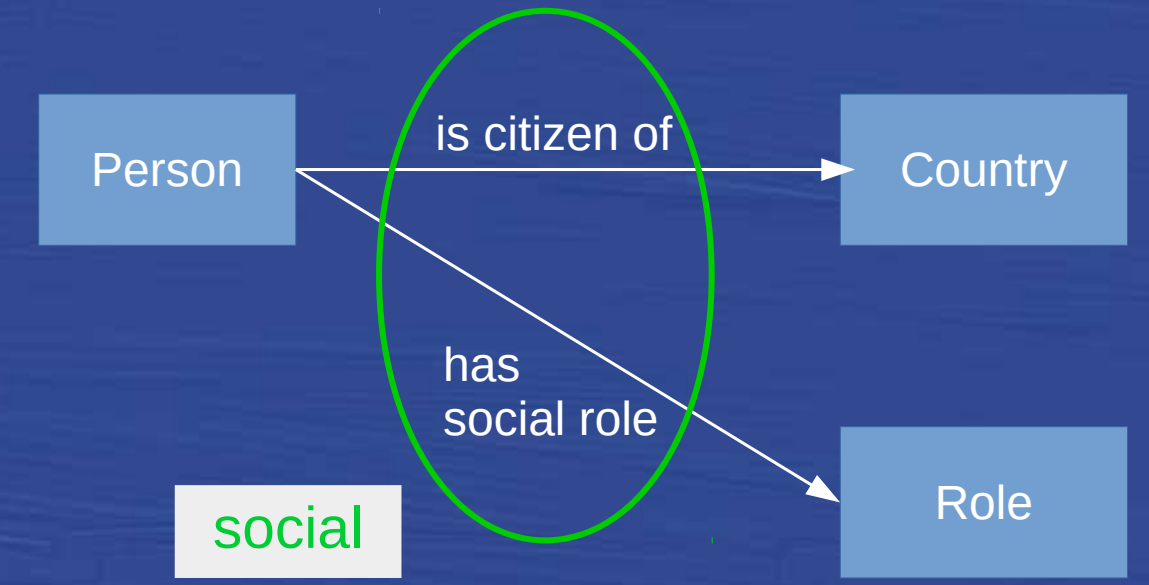

Person
is citizen of
Country
Professor
Merchant

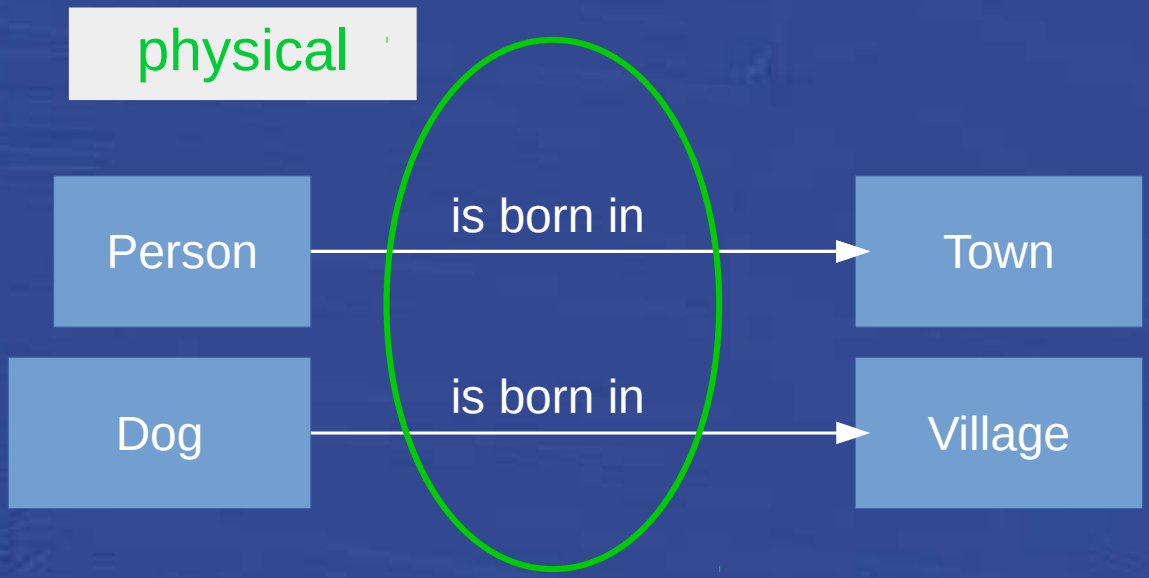

Person
is citizen of
Country
has
social role
Role


Person
is citizen of
Country
has
social role
Role
Geographical place
Person
is born in
Town
Dog
is born in
Village


Person
is citizen of
Country
has
social role
Role


?
Person
is born in
Town
Dog
is born in
Village


Person
is citizen of
Country
has
social role
Role
social


physical
Person
is born in
Town
Dog
is born in
Village

Formalization

Object-Oriented Languages	OWL and RDF
Domain models consist of classes, properties and instances (individuals). Classes can be arranged in a **subclass hierarchy with inheritance**. Properties can take objects or primitive values(literals) as values.	
Classes and Instances	
Classes are regarded as **types for instances**.	Classes are regarded as **sets of individuals** [defined by common properties].
Each instance has one class as its type. Classes cannot share instances.	Each individual can belong to multiple classes. [multiple inheritance]
Properties, Attributes and Values	
Properties are defined locally to a class (and its subclasses through inheritance).	Properties are stand-alone entities that can exist without specific classes.
…	...
Closed world: If there is not enough information to prove a statement true, then it is assumed to be false.	Open world: If there is not enough information to prove a statement true, then it may be true or false.

[Holger Knublauch et al. W3C 2006]

OWL Class Hierarchies and Disjunctiveness

```
:Book a owl:Class .
:Writer a owl:Class .
:Novel a owl:Class ;
       rdfs:subClassOf :Book .
:Poet a owl:Class ;
       rdfs:subClassOf :Writer .

:Book owl:disjointWith :Writer .
```

$$\text{Novel} \sqsubseteq \text{Book}$$
$$\text{Poet} \sqsubseteq \text{Writer}$$
$$\text{Book} \sqcap \text{Writer} \sqsubseteq \bot$$

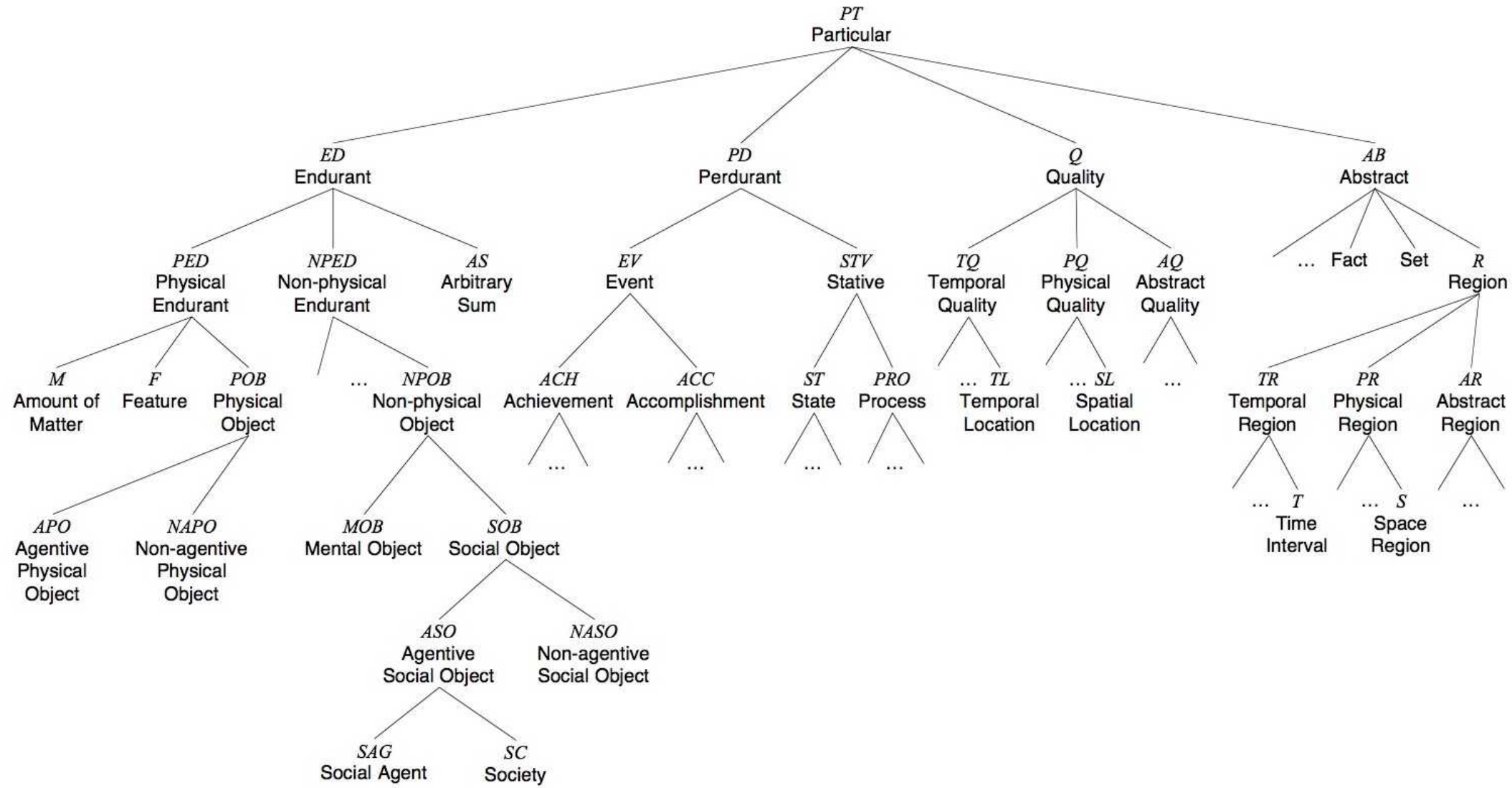
- via inference it can be entailed that „Novel“ and „Poet“ are also disjoint classes.

OWL DL and Description Logic

Harald Sack, "4.3 Classes, Instances and Properties in OWL". Knowledge Engineering with Semantic Web Technologies, OpenHPI Tutorials, 2019

Conceptualization

Foundational ontologies
were developed to support the
verification and improvement
of the **conceptualization** of a **domain of discourse**.



Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) – a foundational ontology designed in 2002 in the context of the WonderWeb EU project, developed by Nicola Guarino and his associates at the Laboratory for Applied Ontology (LOA) – WonderWeb Deliverable D18, p.14

« **Axiomatic ontologies** come in different forms and can have different levels of generality, but a special relevance is enjoyed by the so-called **foundational ontologies**, which **address very general domains**.

To capture (or at least approximate) such subtle distinctions we need an explicit representation of the so-called **ontological commitments** about the meaning of terms, in order to **remove terminological and conceptual ambiguities**. A rigorous logical axiomatisation seems to be unavoidable in this case, as it accounts [...] for the formal structure of the **domain to be represented**.

One of the goals of the WONDERWEB project is the development of a **library of such foundational ontologies**, systematically related to each other in a way that makes the rationales and alternatives underlying different ontological choices as explicit as possible.

Foundational ontologies are ultimately devoted to facilitate mutual understanding and inter-operability among people and machines.

The first module of our foundational ontologies library is a **Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE)**.

DOLCE has a clear cognitive bias, in the sense that it aims at capturing the **ontological categories underlying natural language** and human common-sense.

DOLCE is an **ontology of particulars**, in the sense that its **domain of discourse is restricted to them**. »

WonderWeb Deliverable D18, p. 2-3/13

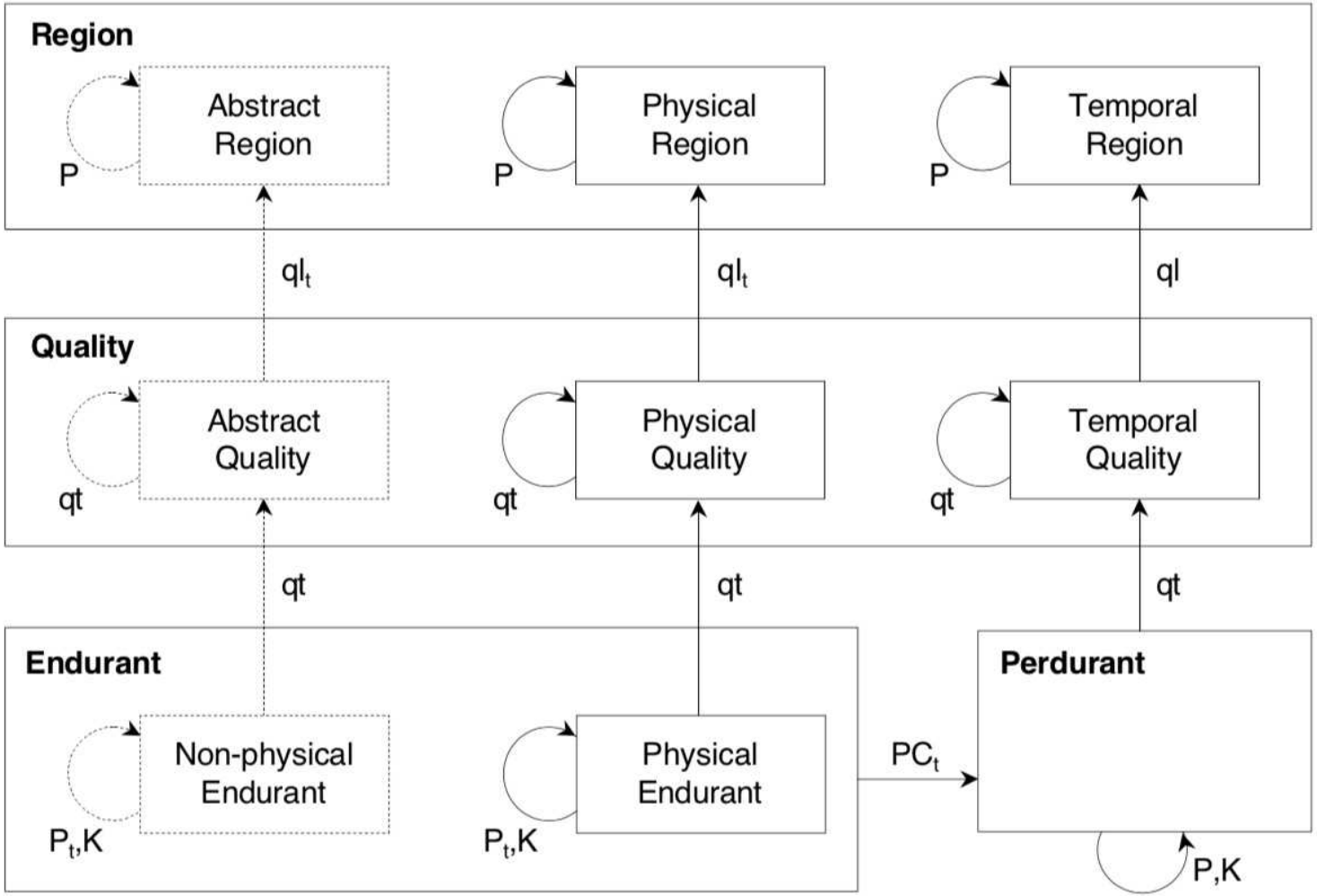


Figure 4: Primitive relations between basic categories (the dotted lines to the left indicate that we are less confident with what concerns non-physical endurants.

WonderWeb Deliverable D18, p.25

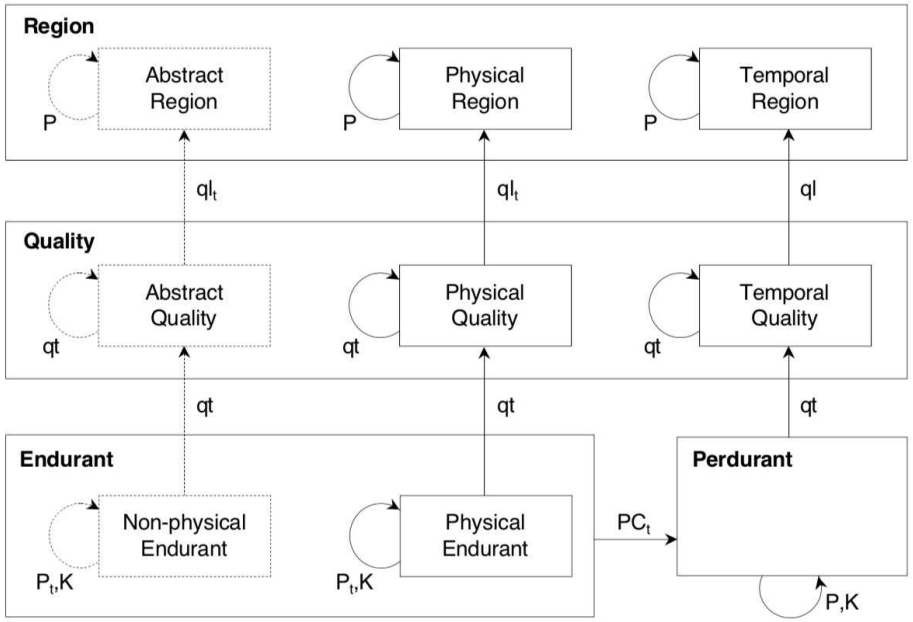


Figure 4: Primitive relations between basic categories (the dotted lines to the left indicate that we are less confident with what concerns non-physical endurants.

4.2.7 Participation

(Dd63) $\mathsf{PC}_C(x,y) \triangleq \exists t(\mathsf{PRE}(y,t)) \wedge \forall t(\mathsf{PRE}(y,t) \rightarrow \mathsf{PC}(x,y,t))$ *(Const. Participation)*

(Dd64) $\mathsf{PC}_T(x,y,t) \triangleq PD(y) \wedge \forall z((\mathsf{P}(z,y) \wedge \mathsf{PRE}(z,t)) \rightarrow \mathsf{PC}(x,z,t))$
(Temporary Total Participation)

(Dd65) $\mathsf{PC}_T(x,y) \triangleq \exists t(\mathsf{ql}_T(t,y) \wedge \mathsf{PC}_T(x,y,t))$ *(Total Participation)*

(Dd66) $\mathsf{mpc}(x,y) \triangleq x = \sigma_t z(\mathsf{PC}_T(z,y))$ *(Maximal Participant)*

(Dd67) $\mathsf{mppc}(x,y) \triangleq x = \sigma_t z(\mathsf{PC}_T(z,y) \wedge PED(z))$ *(Maximal Physical Participant)*

(Dd68) $\mathsf{lf}(x,y) \triangleq x = \sigma z(\mathsf{PC}_T(y,z))$ *(Life)*

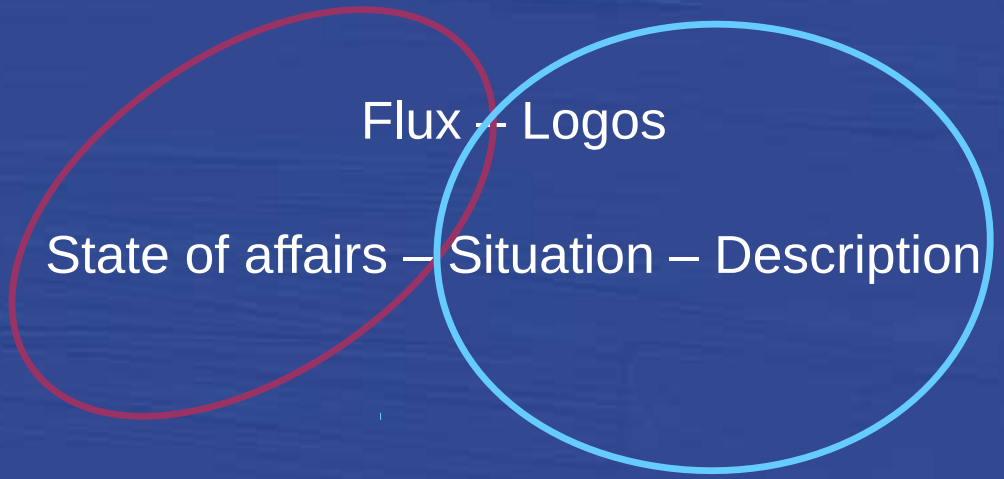
WonderWeb Deliverable D18, p.30

Descriptions and Situations (DnS)

Flux – Logos

State of affairs – Situation – Description

Descriptions and Situations (DnS)



Descriptions and Situations (DnS)

Flux – Logos

State of affairs – Situation – Description

Descriptions and Situations (DnS)

Flux – Logos

State of affairs – Situation – Description

Model social entities, roles, collectives, etc.

« If you want to express the legal constraints imposed by norms and regulations on the domain of your ground ontology, you have to extend the latter and add to it a DnS description of social reality under a legal perspective. Such DnS description makes it possible to describe the ideal (legal) view on the behaviour of your **social entities** (a situation), according to a given legal system. […]

Reified constraints and theories are classified as social objects, which hold various properties like, for instance, a(n indirect) location in space and time. »

Bottazzi E., Catenacci C., Gangemi A., Lehmann J.(2006), p. 194-195

Descriptions and Situations (DnS)

Flux – Logos

State of affairs – Situation – Description

Model information production in the context of a scientific discipline

« Constructive DnS (hereafter, c.DnS). Secondly, our definition of intentional col- lectives is exteded with normative elements. This provides us with the conceptual means to define typologies of normed **intentional collectives**.

[It] provides the formal means to specify the **epistemological perspective** from which the **entities of the domain** are considered.

We define knowledge collectives [as] **intentional normative collectives** whose unifying bundle is an epistemic influence that has a paradigm as part.»

Gangemi A., Lehmann J., Catenacci C. (2008), p. [3/34]

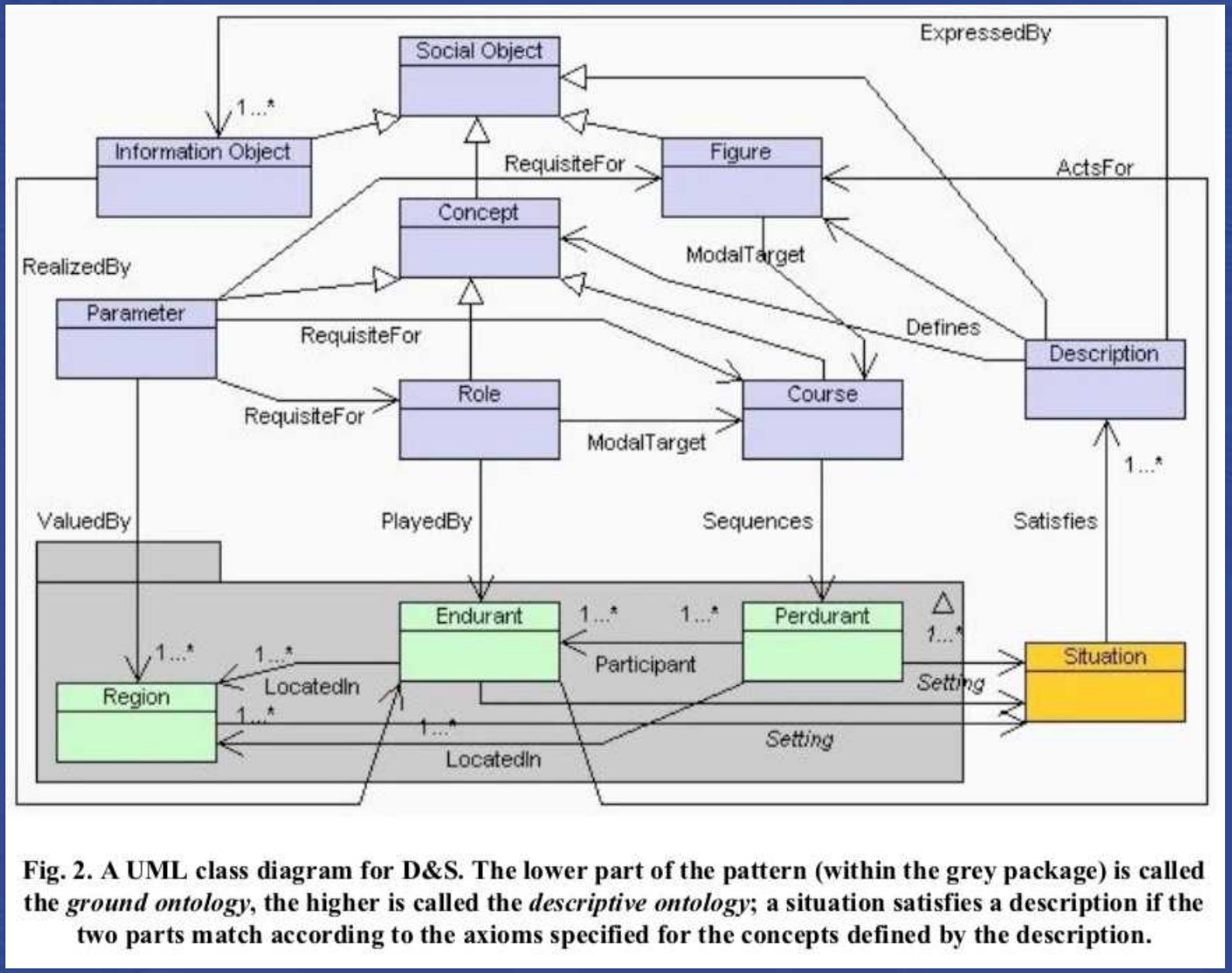
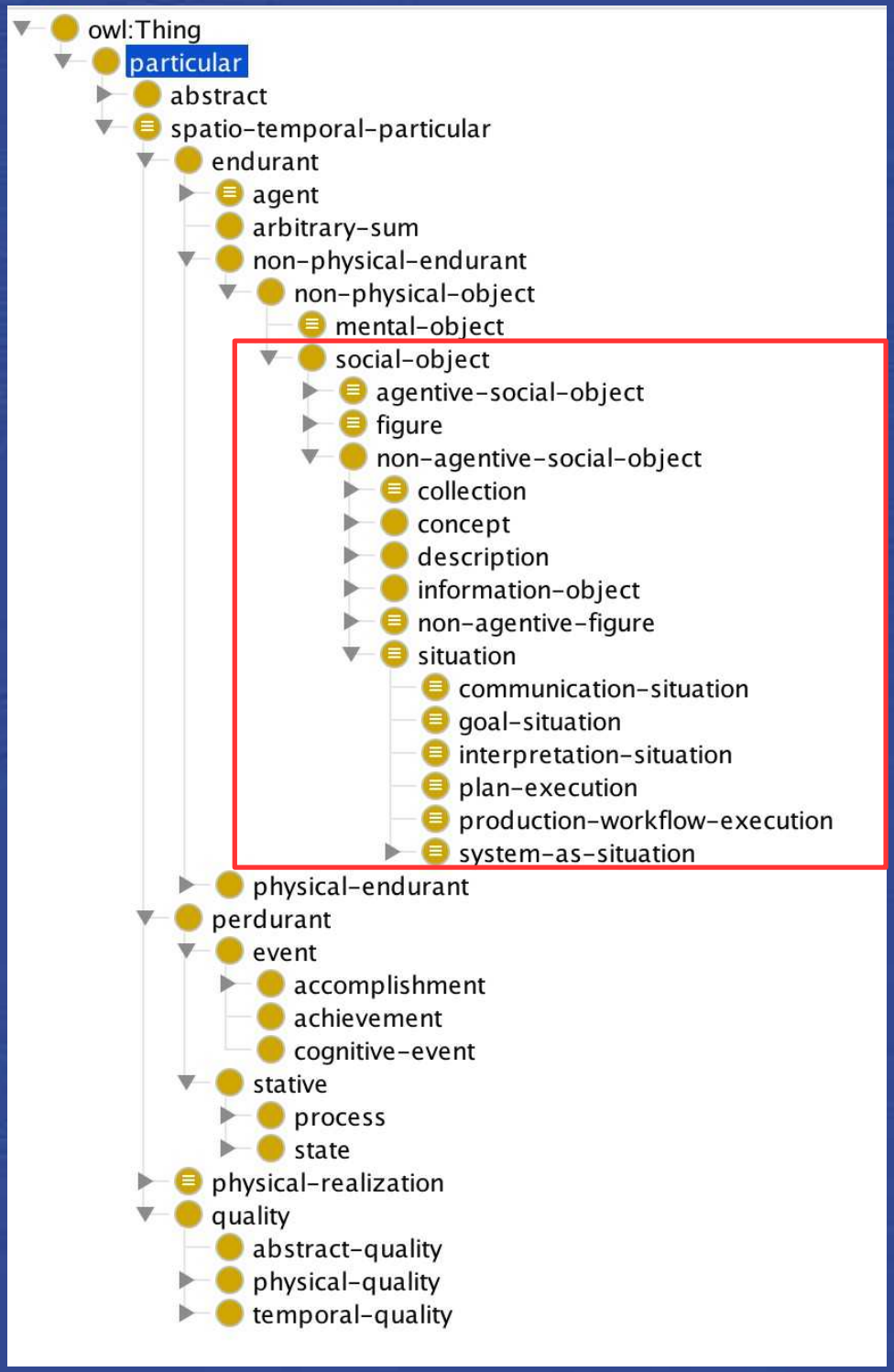


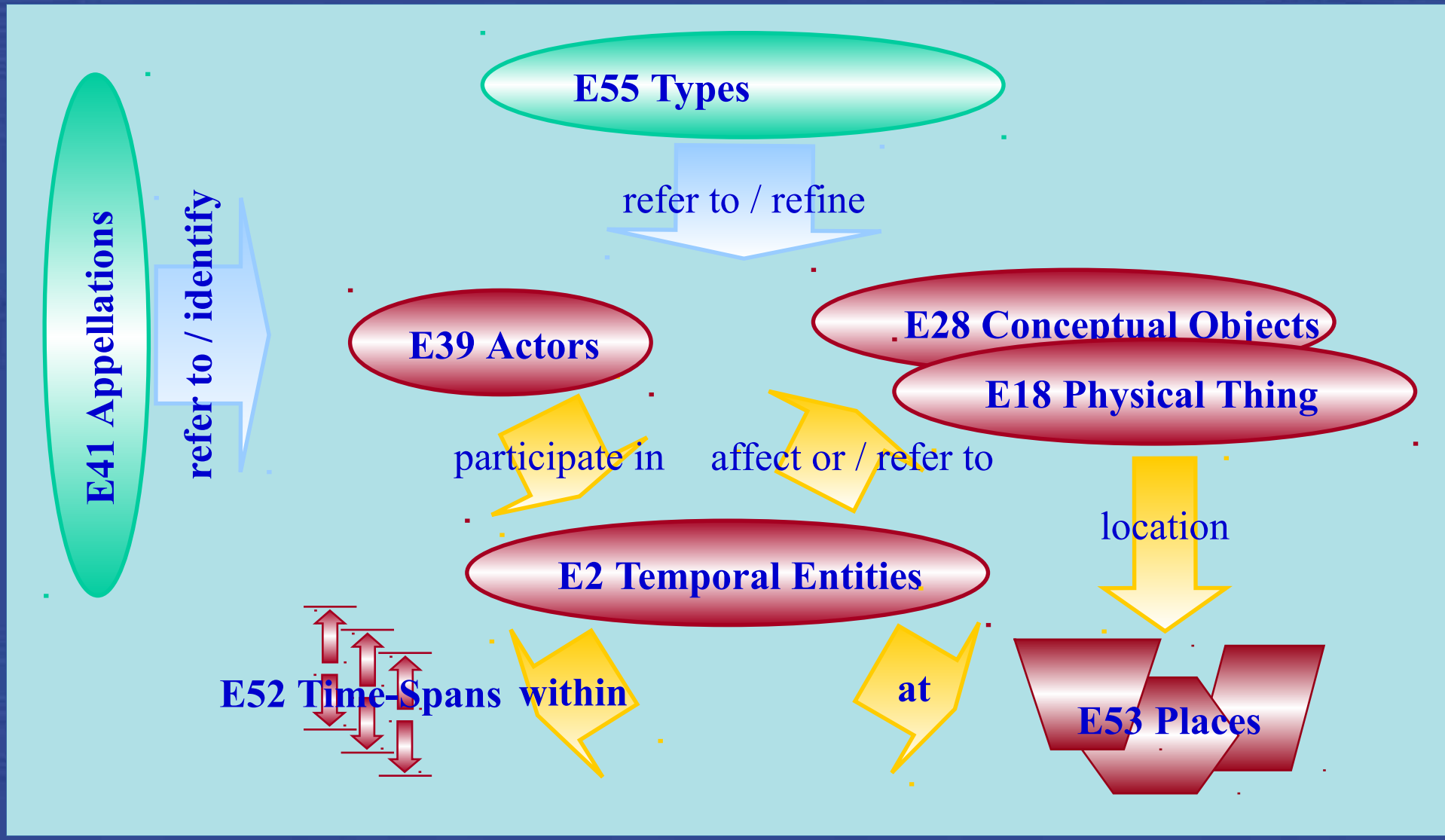
Fig. 2. A UML class diagram for D&S. The lower part of the pattern (within the grey package) is called the *ground ontology*, the higher is called the *descriptive ontology*; a situation satisfies a description if the two parts match according to the axioms specified for the concepts defined by the description.

Bottazzi E., Catenacci C., Gangemi A., Lehmann J.(2006) (from pre-print, not in published version)

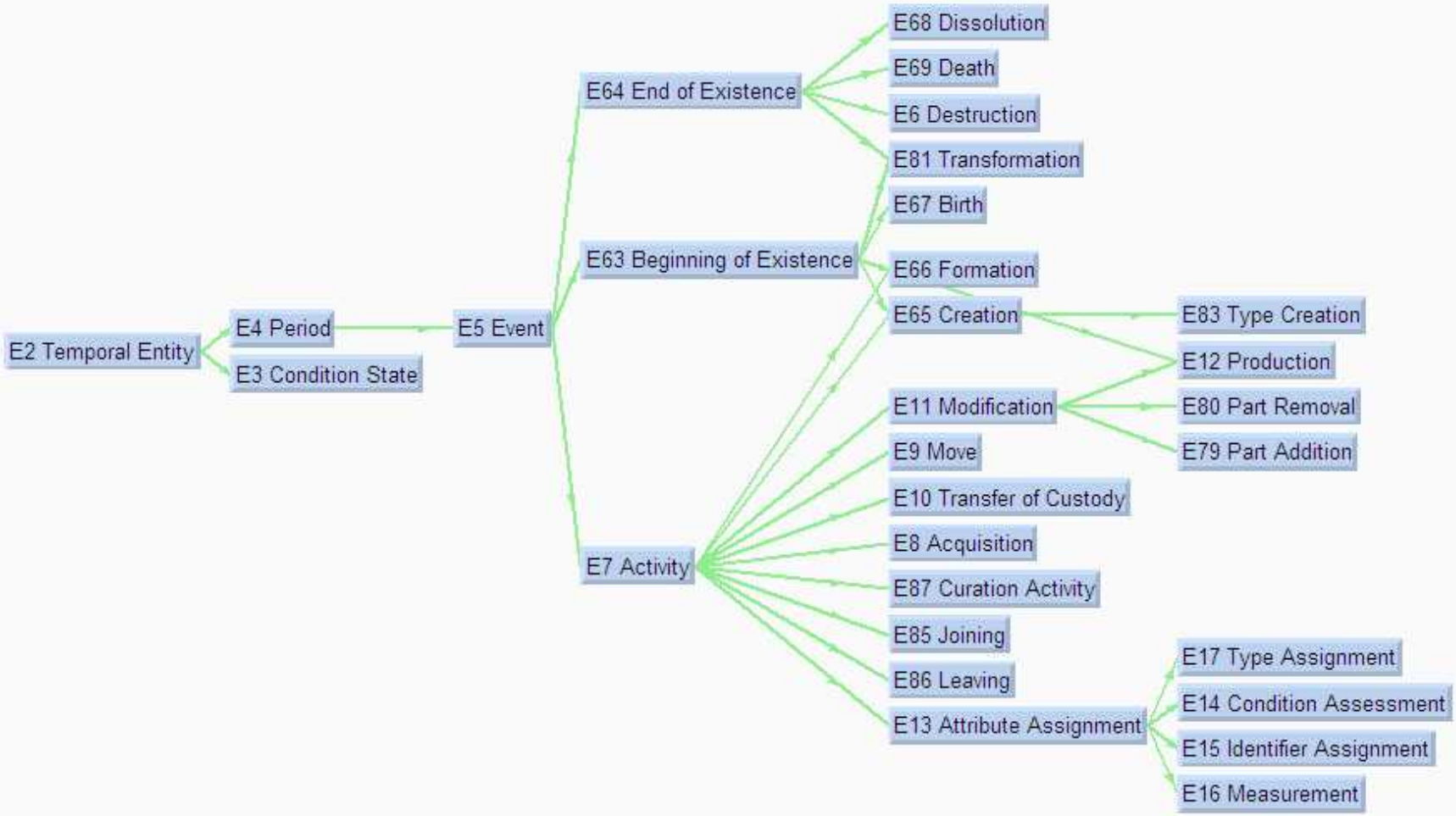


DOLCE Lite PLUS

The CIDOC CRM (ISO21127:2006)
A semantic framework that provides interoperability between different sources of cultural heritage information



Stephen Stead (2008)



Stephen Stead (2008)

Stephen Stead (2008)

« This document is the formal definition of the CIDOC **Conceptual Reference Model** (“CRM”), a **formal ontology** intended to facilitate the **integration**, mediation and interchange of heterogeneous cultural heritage information.

Since 2000, development of the CRM has been officially delegated by ICOM-CIDOC to the **CIDOC CRM Special Interest Group**, which collaborates with the ISO working group ISO/TC46/SC4/WG9 to bring the CRM to the form and status of an International Standard. [ISO21127:2006, renewed in 2014: ISO21127:2014]

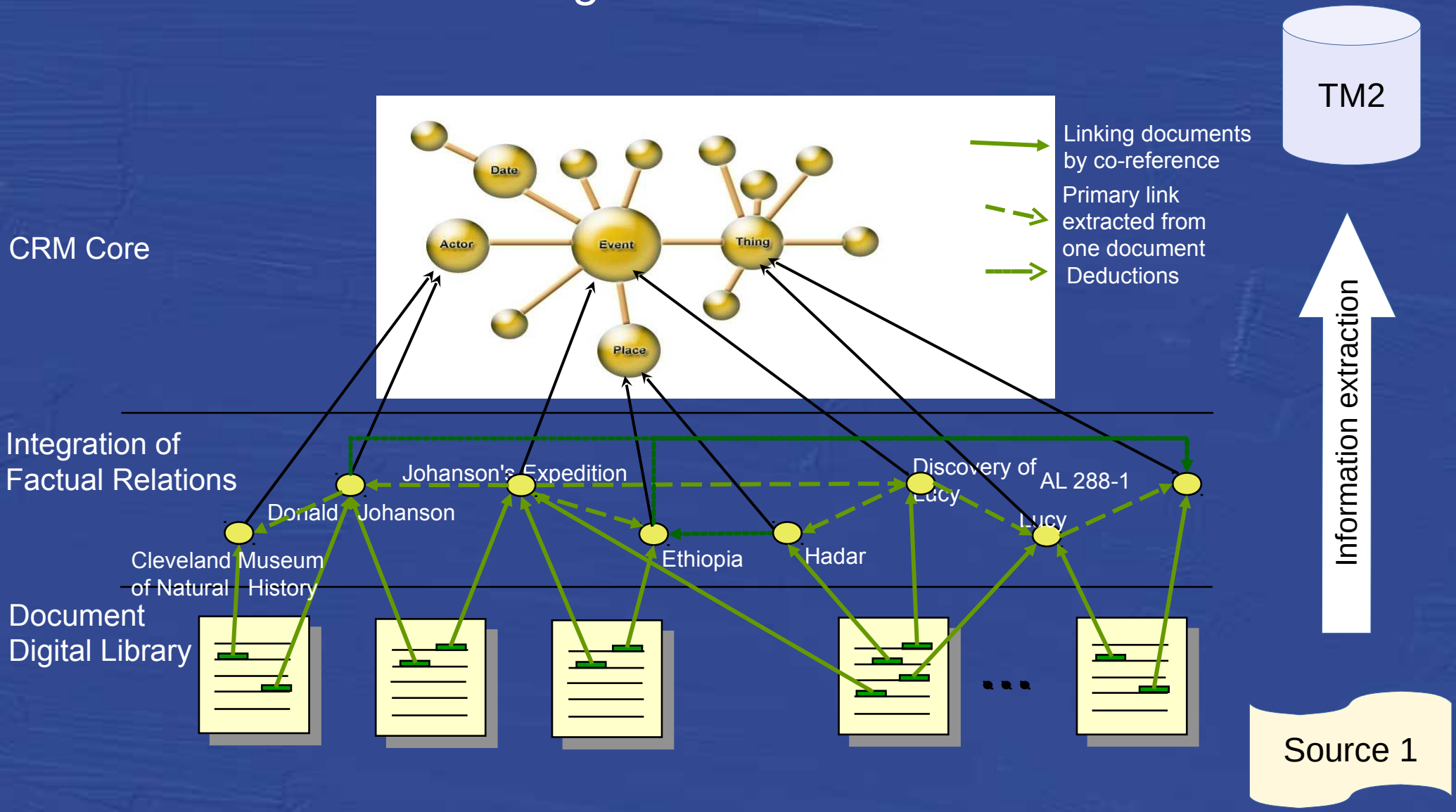
The primary role of the CRM is to enable **information exchange and integration** between heterogeneous sources of **cultural heritage information**. It aims at providing the semantic definitions and clarifications needed to transform disparate, localised information sources into a coherent global resource, be it within a larger institution, in intranets or on the Internet.

The CRM is an **ontology** in the sense used in computer science.

It has been expressed as an **object-oriented semantic model**, in the hope that this formulation will be comprehensible to both documentation experts and information scientists alike, while at the same time **being readily converted to machine-readable formats** such as RDF Schema, KIF, DAML+OIL, OWL, STEP, etc.»

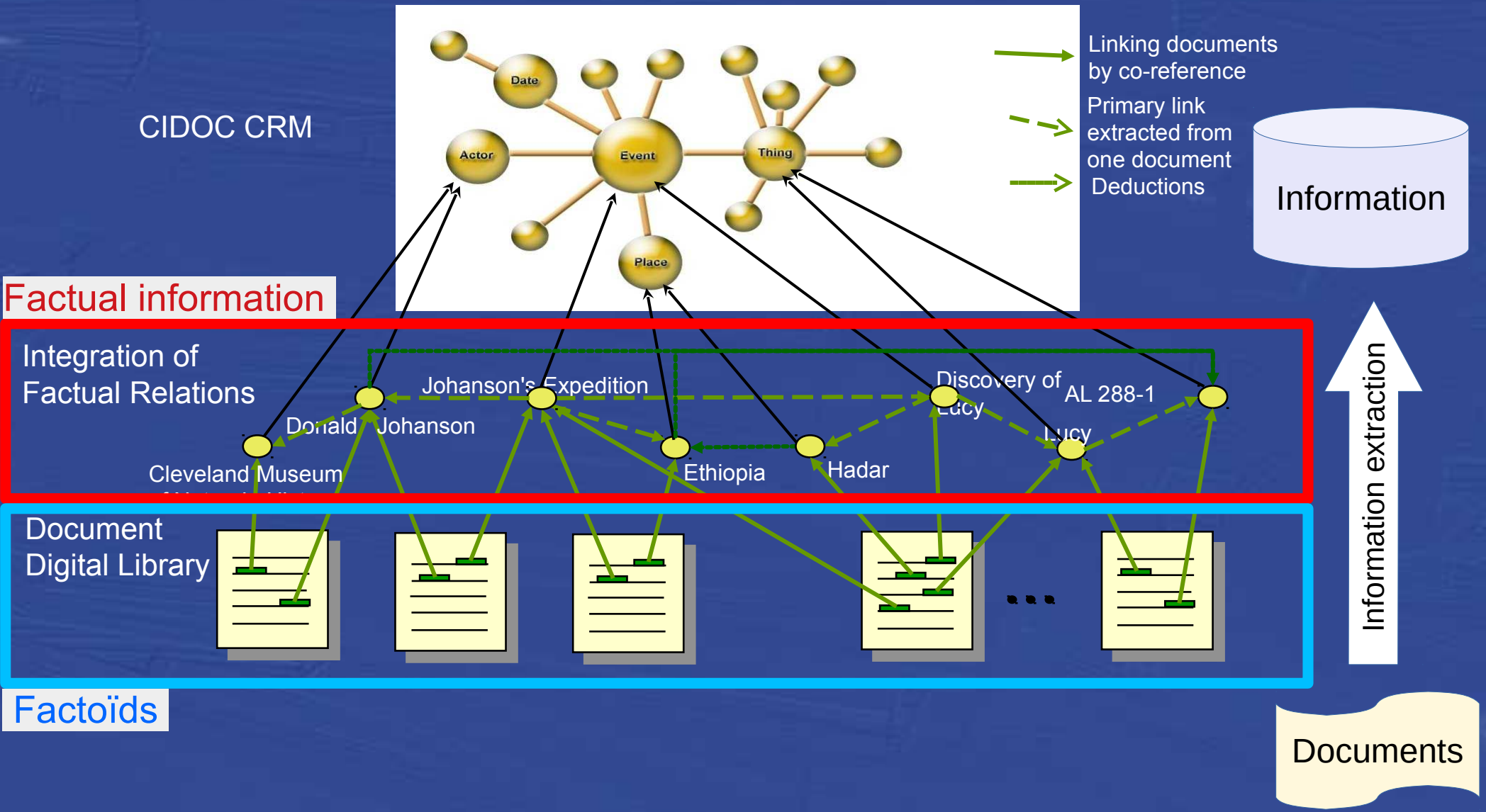
CIDOC CRM 6.2.1

Integration of information extracted from documents using the CIDOC CRM

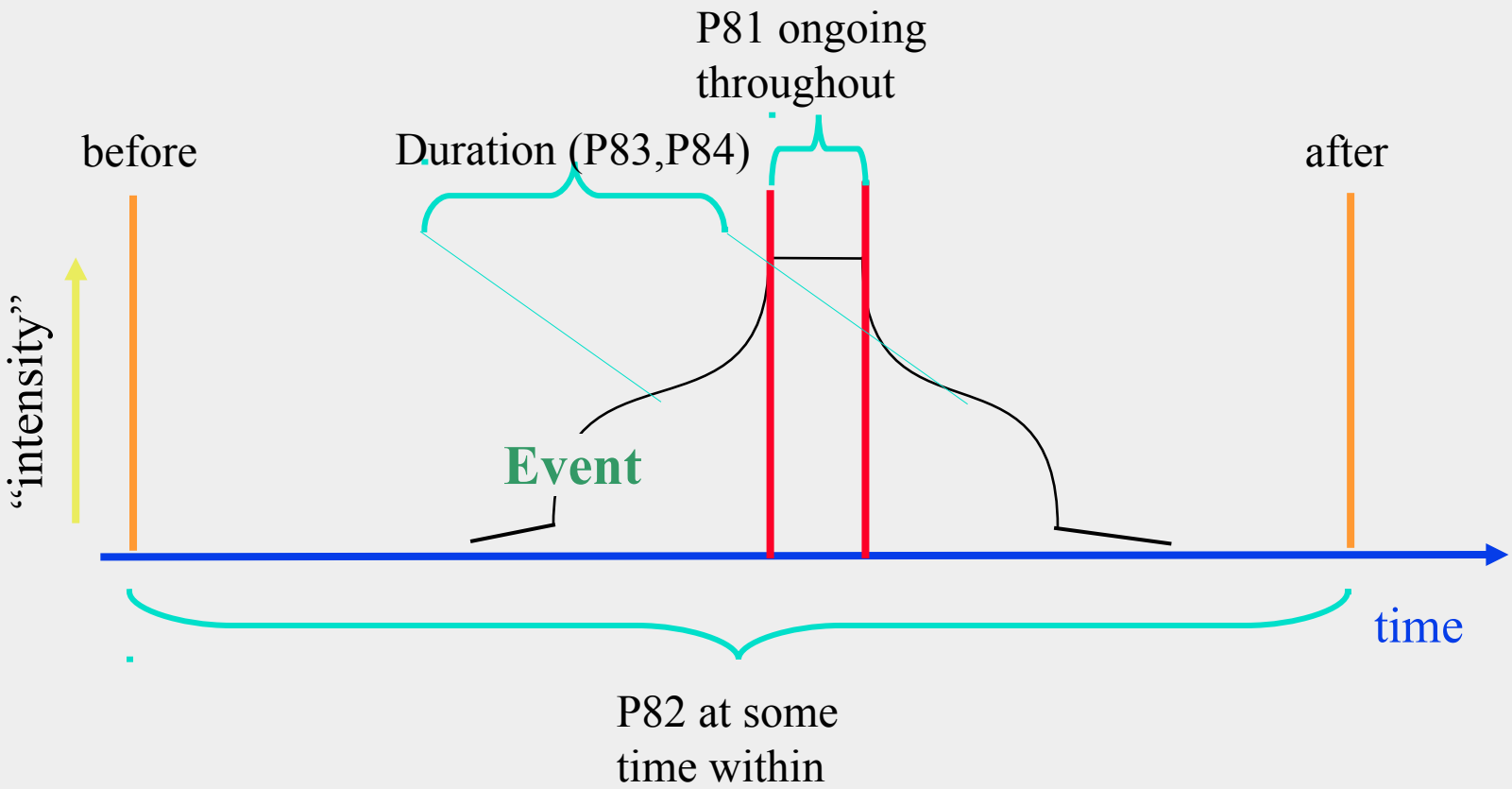


Stephen Stead (2008). Cf. Doerr M. (2003), fig. 4, p. 81

Integration of information extracted from documents using the CIDOC CRM



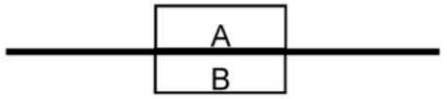
The CIDOC CRM
Time Uncertainty, Certainty and Duration



Stephen Stead (2008)

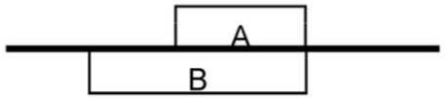
The CIDOC CRM P114-120 ‘Allen’ properties (James F. Allen)

P114 is equal in time to



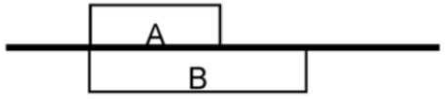
$A_s = B_s$ & $A_e = B_e$ ⇔ ($A_{ss} = B_{ss}$ & $A_{se} = B_{se}$) & ($A_{es} = B_{es}$ & $A_{ee} = B_{ee}$)

P115 finishes



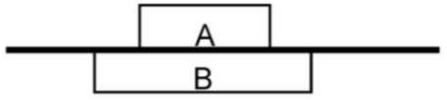
$A_s > B_s$ & $A_e = B_e$ ⇔ ($B_{ss} < A_{ss}$ & $B_{se} < A_{se}$) & ($B_{es} = A_{es}$ & $B_{ee} = A_{ee}$)

P116 starts



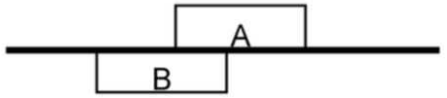
$A_e < B_e$ & $A_s = B_s$ ⇔ ($A_{ss} = B_{ss}$ & $A_{se} = B_{se}$) & ($A_{ee} < B_{ee}$ & $A_{es} < B_{es}$)

P117 occurs during



$B_s < A_s$ & $A_e < B_e$ ⇔ ($B_{ss} < A_{ss}$ & $B_{se} < A_{se}$) & ($A_{es} < B_{es}$ & $A_{ee} < B_{ee}$)

P118 overlaps

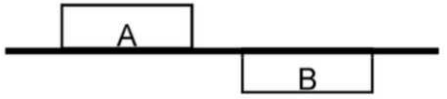


$B_s < A_s$ & $A_s < B_e$ & $B_e < A_e$ ⇔ ($B_{ss} < A_{ss}$ & $B_{se} < A_{se}$) & ($A_{ss} < B_{es}$ & $A_{se} < B_{ee}$) & ($B_{es} < A_{es}$ & $B_{ee} < A_{ee}$)

P119 meets in time with

$A_e = B_s$ ⇔ $A_{es} = B_{ss}$ & $A_{ee} = B_{se}$

P120 occurs before



$A_e < B_s$ ⇔ $A_{es} < B_{ss}$ & $A_{ee} < B_{se}$

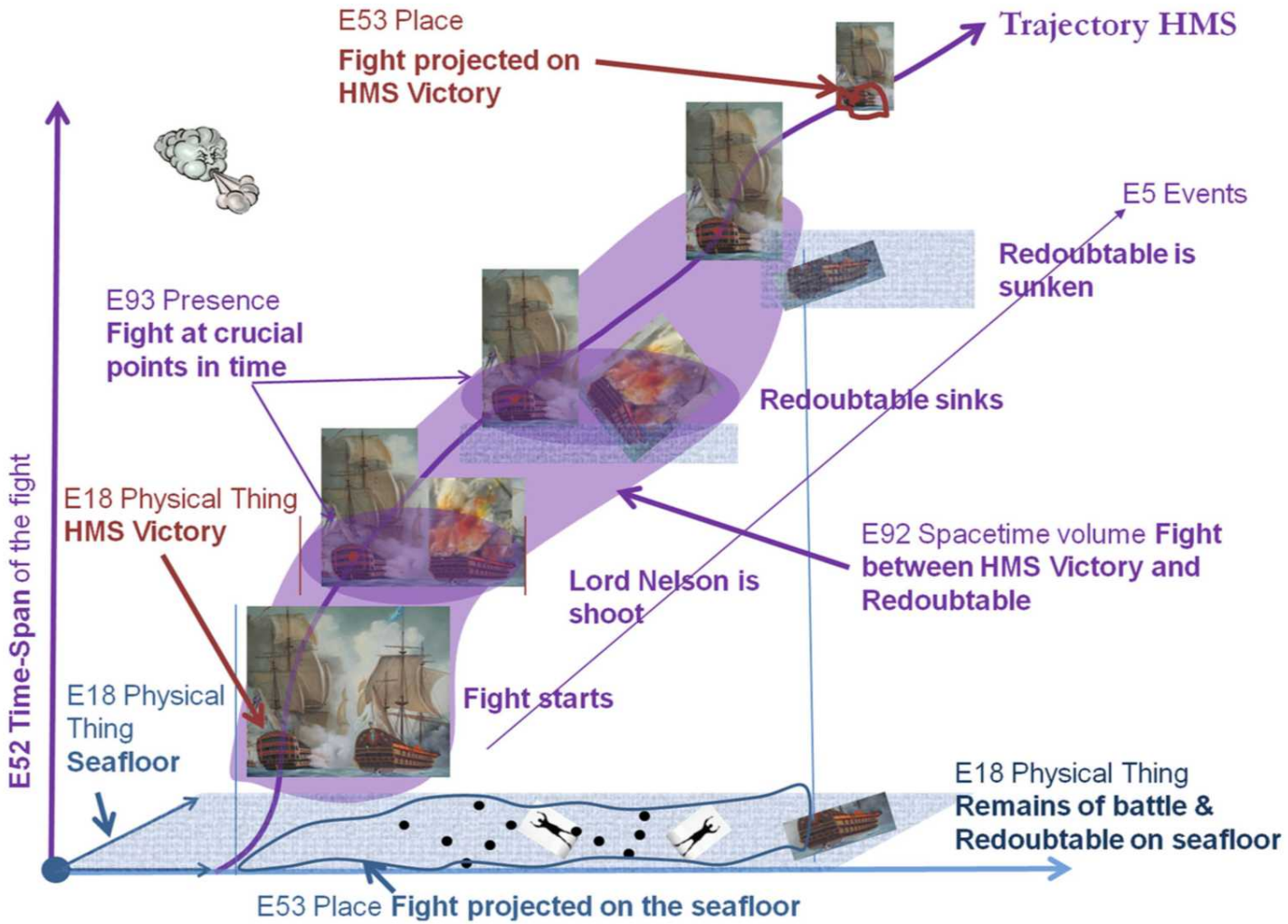


Fig. 2 The fight of the HMS Victory and the Redoubtable in the Battle of Trafalgar illustrating Spacetime volume (E92) and Presence (E93) and their projection to different Places (E53)

DOI 10.1007/s00799-016-0192-4

CRMgeo: A spatiotemporal extension of CIDOC-CRM Gerald Hiebel · Martin Doerr · Øyvind Eide

Francesco Beretta
(CNRS UMR5190 LARHRA – Université de Lyon)

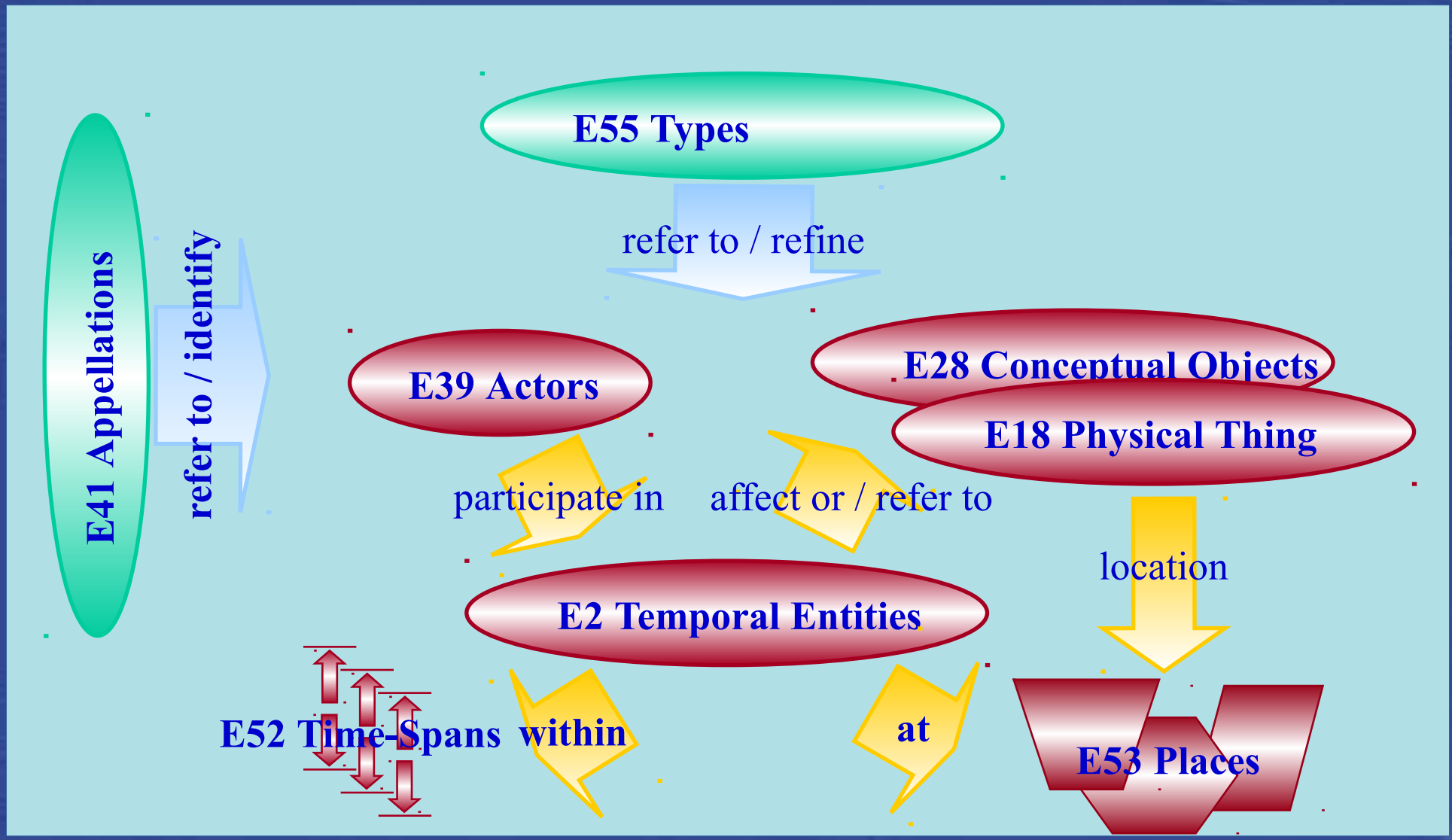
Data modelling V:

Exploring and extending the CIDOC CRM

ACDH Digital Prosopography Summer School

Wien, 7 July 2020

The CIDOC CRM (ISO21127:2006)
A semantic framework that provides interoperability between different sources of cultural heritage information



Stephen Stead (2008)

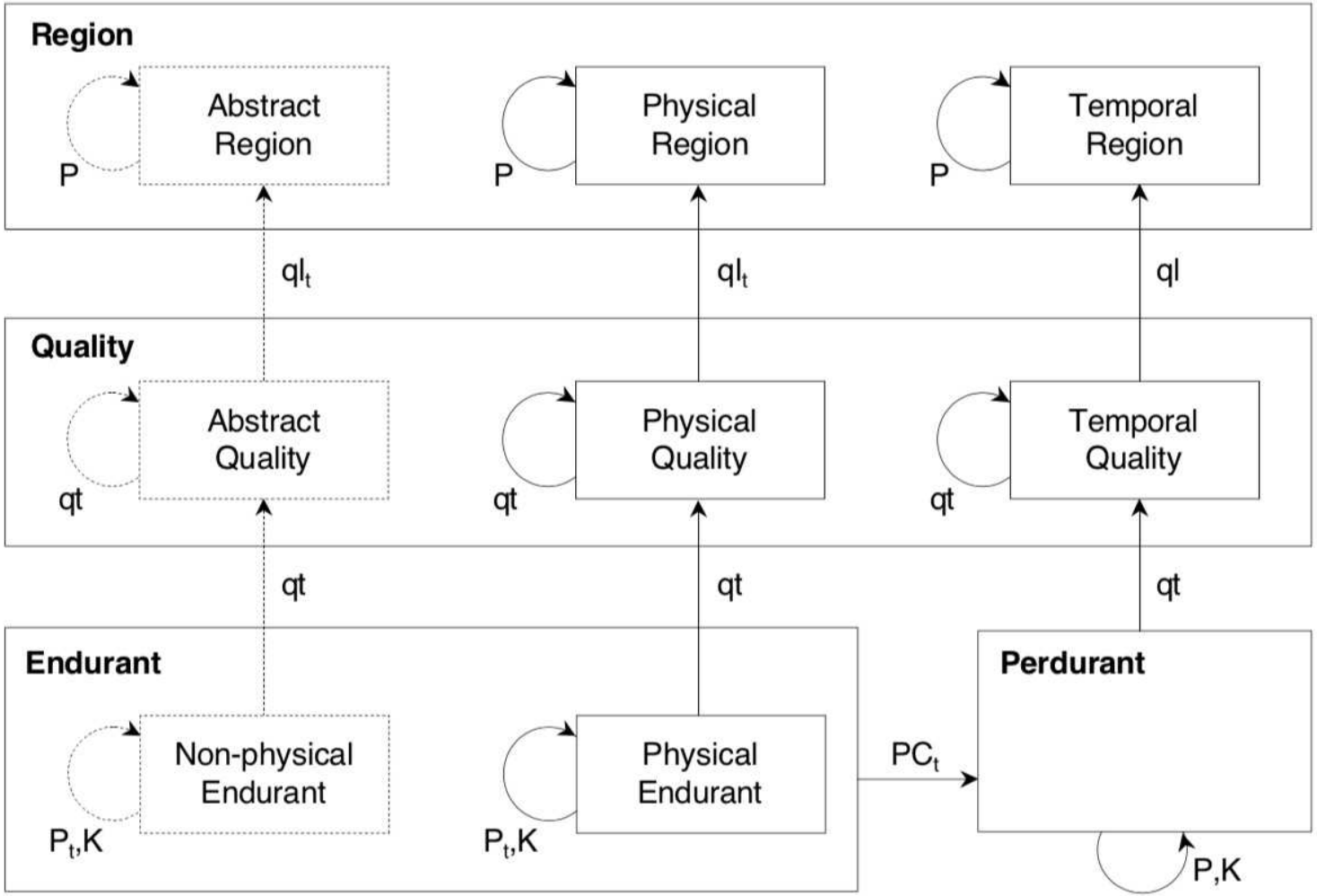
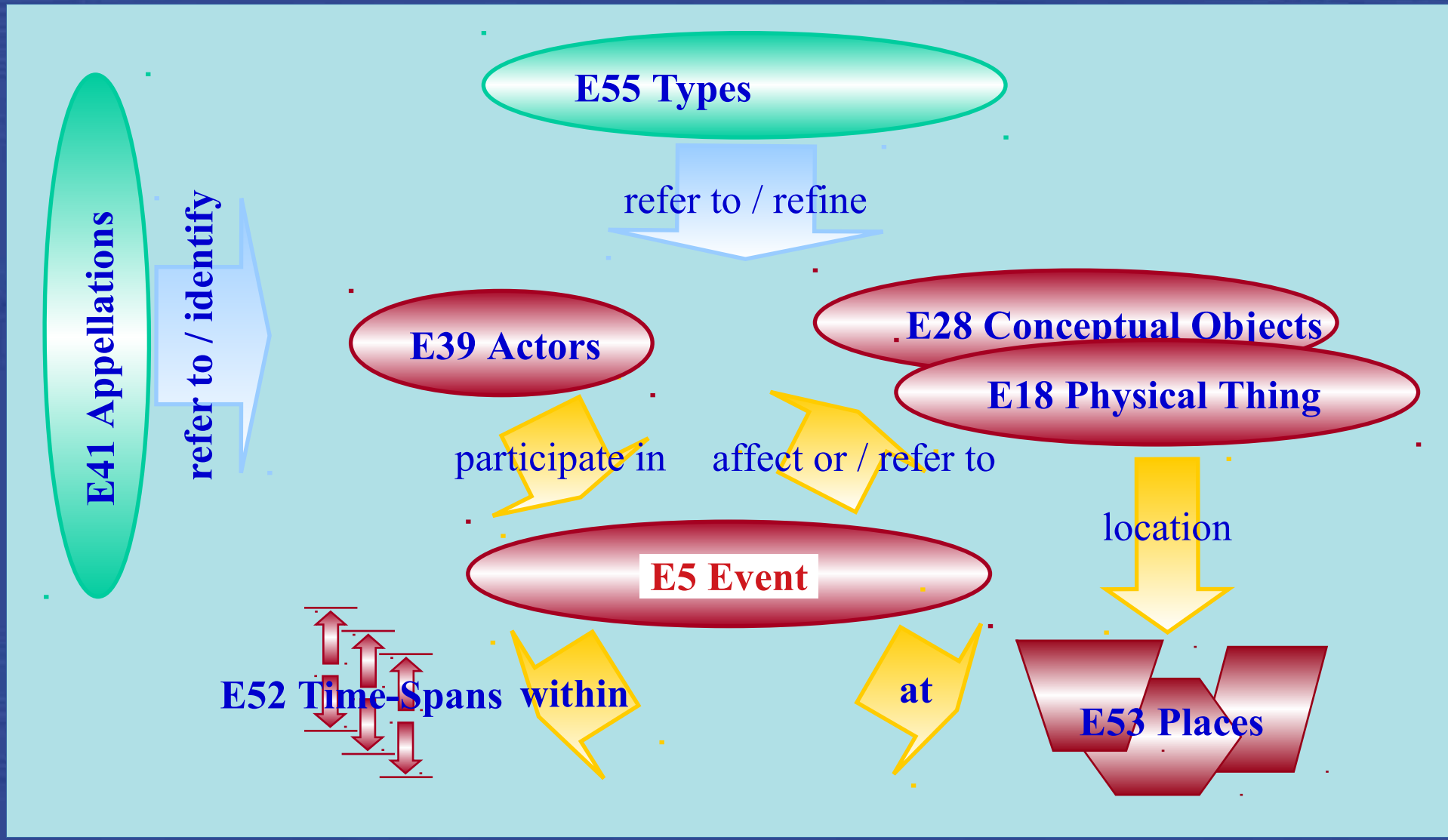
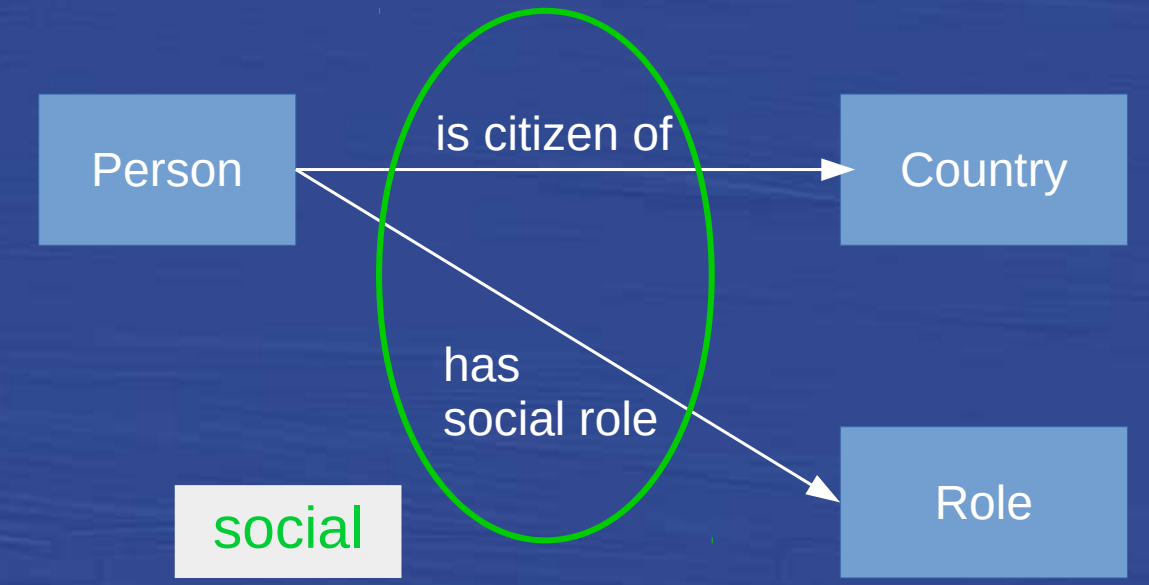


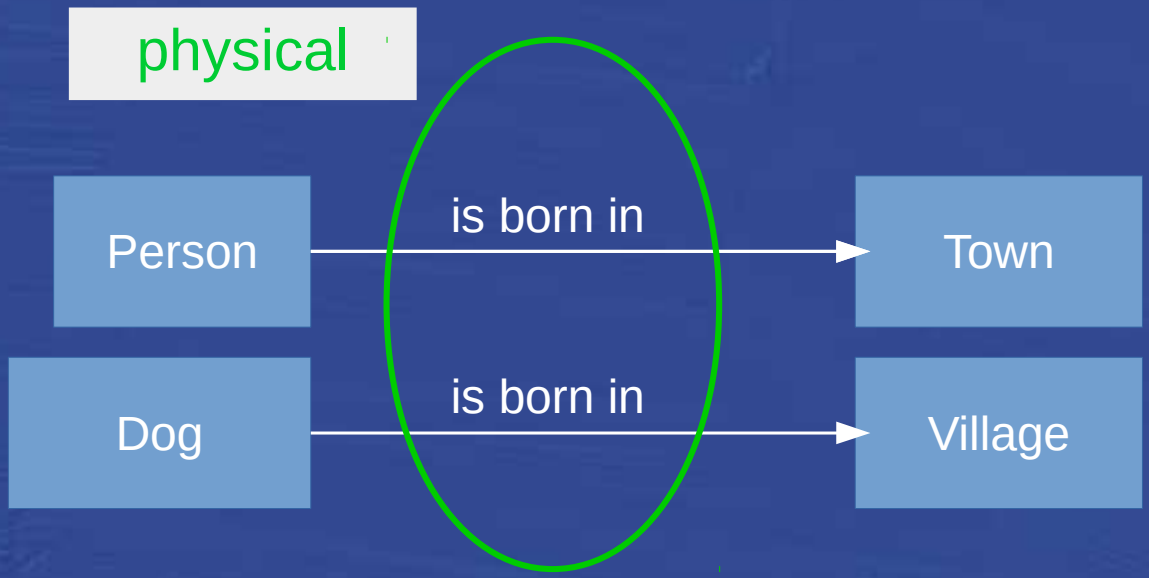
Figure 4: Primitive relations between basic categories (the dotted lines to the left indicate that we are less confident with what concerns non-physical endurants.

WonderWeb Deliverable D18, p.25

The CIDOC CRM (ISO21127:2006)
A semantic framework that provides interoperability between different sources of cultural heritage information




Person
is citizen of
Country
has
social role
Role
social


physical
Person
is born in
Town
Dog
is born in
Village

“Alexander Bach war Justizminister (Justizministerium) von 18.07.1848 bis 11.10.1848, im Rahmen des Ministeriums Wesseberg”

Resource 8008 : Alexander Bach – Resource 2007 : Justizministerium – Resource 8210 : Ministerium Wessenberg – Resource [331] : Justizminister

<rdf:Description rdf:about="https://mpr.acdh.oeaw.ac.at/events/**joined_pc**/8216">
<rdf:type rdf:resource="http://www.cidoc-crm.org/cidoc-crm/**PC144_joined_with**"/>
<cidoc:P01_has_domain rdf:resource="https://mpr.acdh.oeaw.ac.at/events/joined/8216"/>
<cidoc:P02_has_range rdf:resource="https://mpr.acdh.oeaw.ac.at/entity/2007"/>
<cidoc:**P144.1_kind_of_member**
rdf:resource="https://vocabs.acdh-dev.oeaw.ac.at/mprthesaurus/331"/>
</rdf:Description>
…
<rdf:Description rdf:about="https://mpr.acdh.oeaw.ac.at/events/joined_pc/8217">
<cidoc:P144.1_kind_of_member rdf:resource="https://vocabs.acdh-dev.oeaw.ac.at/mprthesaurus/59"/>
<cidoc:P02_has_range rdf:resource="https://mpr.acdh.oeaw.ac.at/entity/8210"/>
<cidoc:P01_has_domain rdf:resource="https://mpr.acdh.oeaw.ac.at/events/joined/8217"/>
<rdf:type rdf:resource="http://www.cidoc-crm.org/cidoc-crm/PC144_joined_with"/>
</rdf:Description>

<rdf:Description rdf:about="https://mpr.acdh.oeaw.ac.at/events/left/8217">
<rdf:type rdf:resource="http://www.cidoc-crm.org/cidoc-crm/E86_Leaving"/>
<cidoc:P4_has_time-span rdf:resource="https://mpr.acdh.oeaw.ac.at/appellation/left/date/8217"/>
<cidoc:P146_seperated_from rdf:resource="https://mpr.acdh.oeaw.ac.at/entity/8210"/>
<cidoc:P145_seperated rdf:resource="https://mpr.acdh.oeaw.ac.at/entity/8008/"/>
</rdf:Description>

« The CIDOC CRM on the other hand **does not include a general construct to model time periods** during which certain properties of an object are **static**.

Only the more specific notion of a Condition State is provided for describing phases during which an object's condition is assumed static e.g., a building was in ruins.

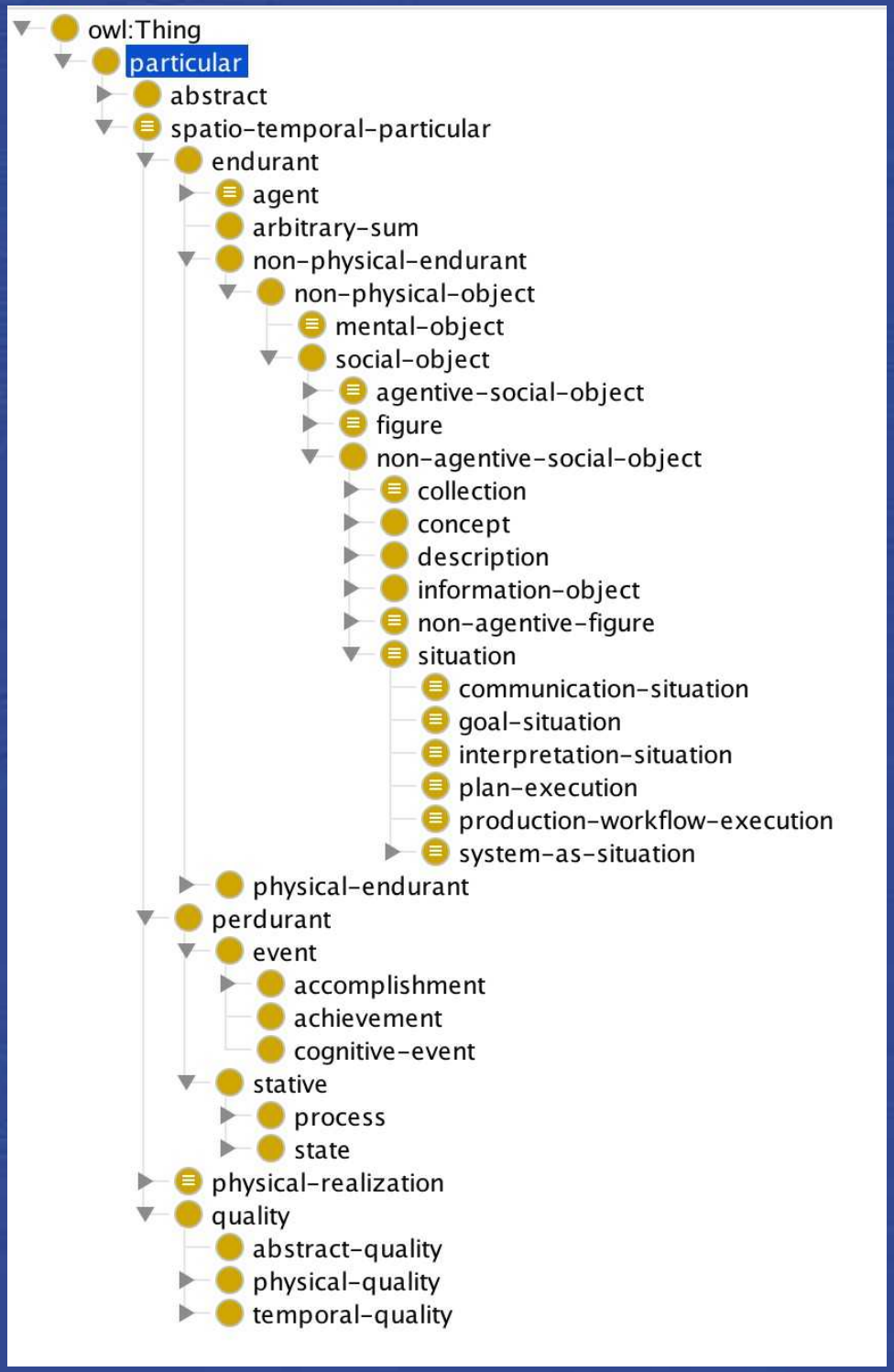
The CRM encourages the description of transitions rather than the states or situations in between them.

This approach was inspired by **considerations of modern physics**, which considers Events as complex interactions with non-negligible temporal extent, leading into phases of stability with fuzzy (if any) temporal boundaries.

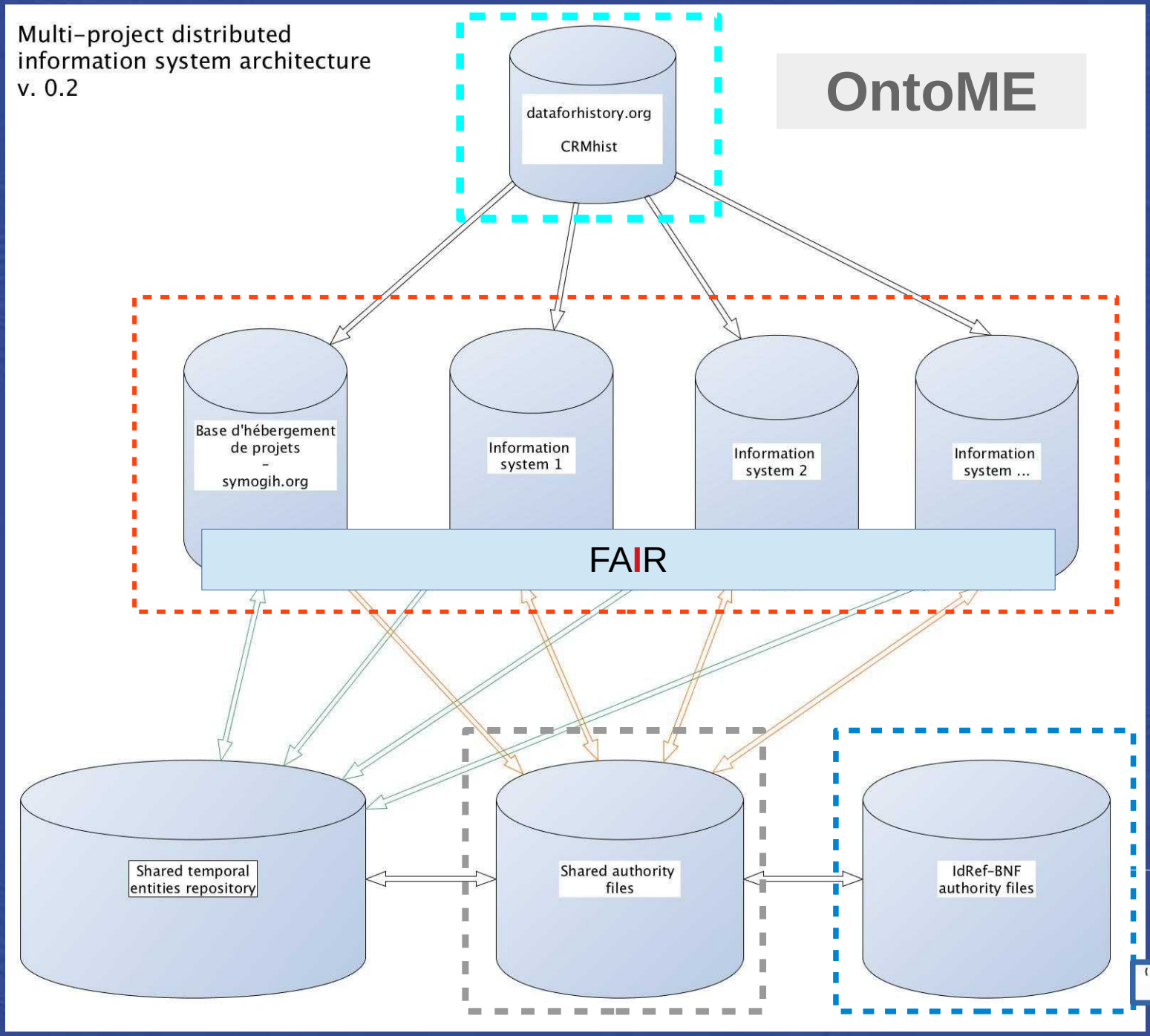
The observations necessary to acquire knowledge can themselves be regarded as Events that are part of the observed system.

It seems that an analogous argument can be **applied to historical research**. »

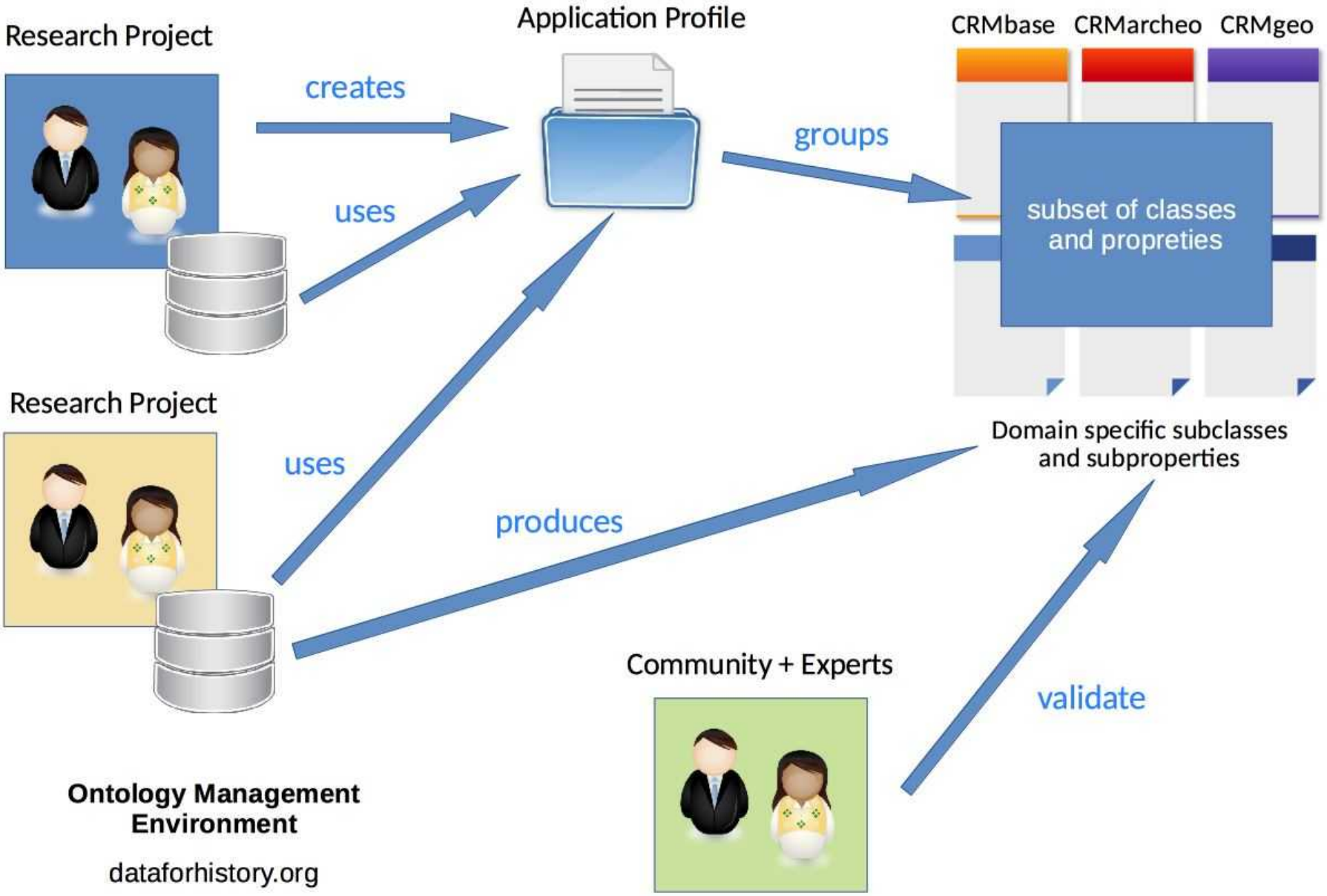
Doerr M., Hunter J., Lagoze C. (2003)



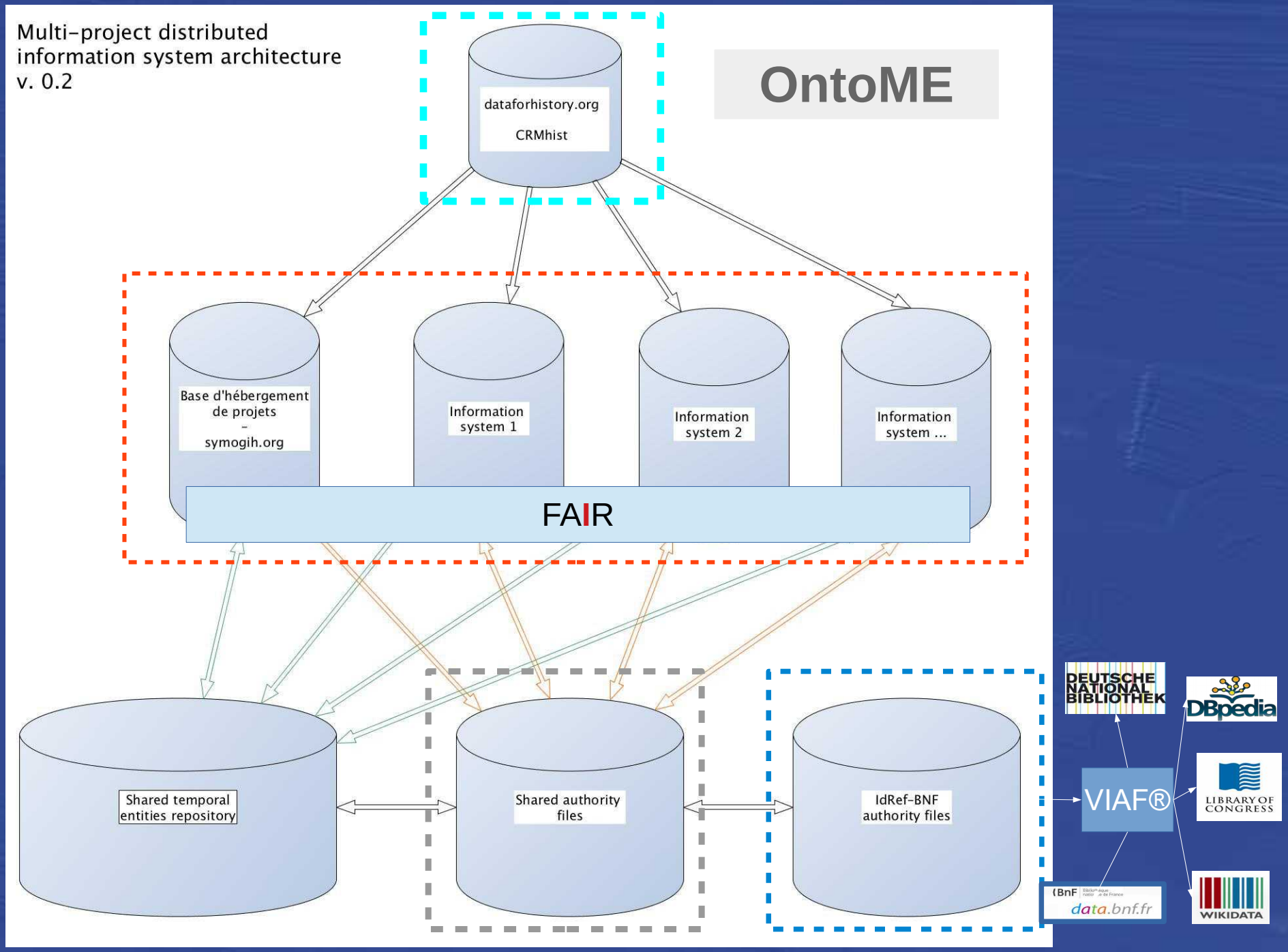
DOLCE Lite PLUS



dataforhistory.org



ontome.dataforhistory.org



dataforhistory.org