Francesco Beretta
(CNRS UMR5190 LARHRA – Université de Lyon)

Interoperability of historical data and FAIR principles: an ontology management environment (OntoME) for sharing and aligning data models

ontome.dataforhistory.org

University of Amsterdam,
6 November 2018

Findable

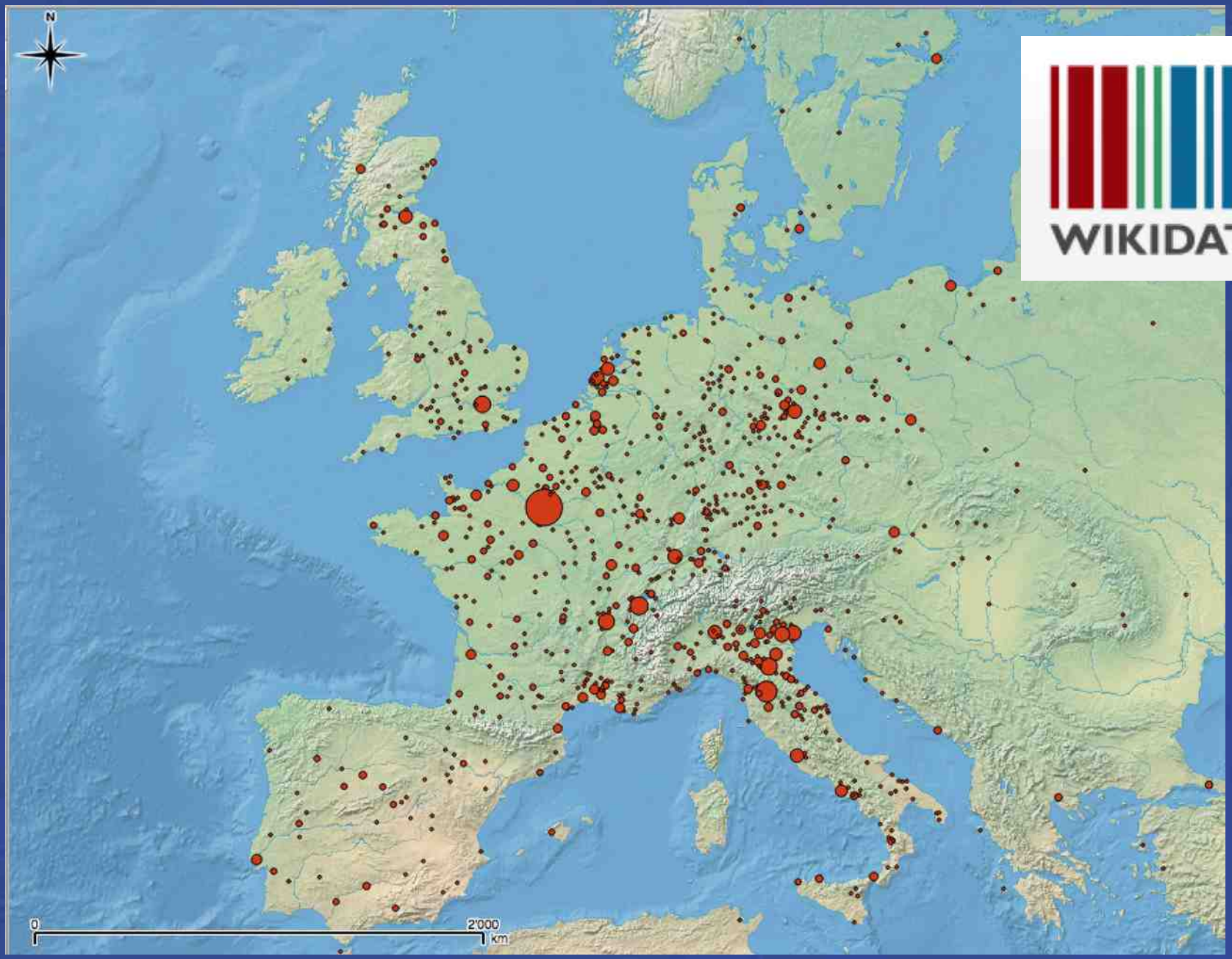
Accessible

Interoperable

Re-usable

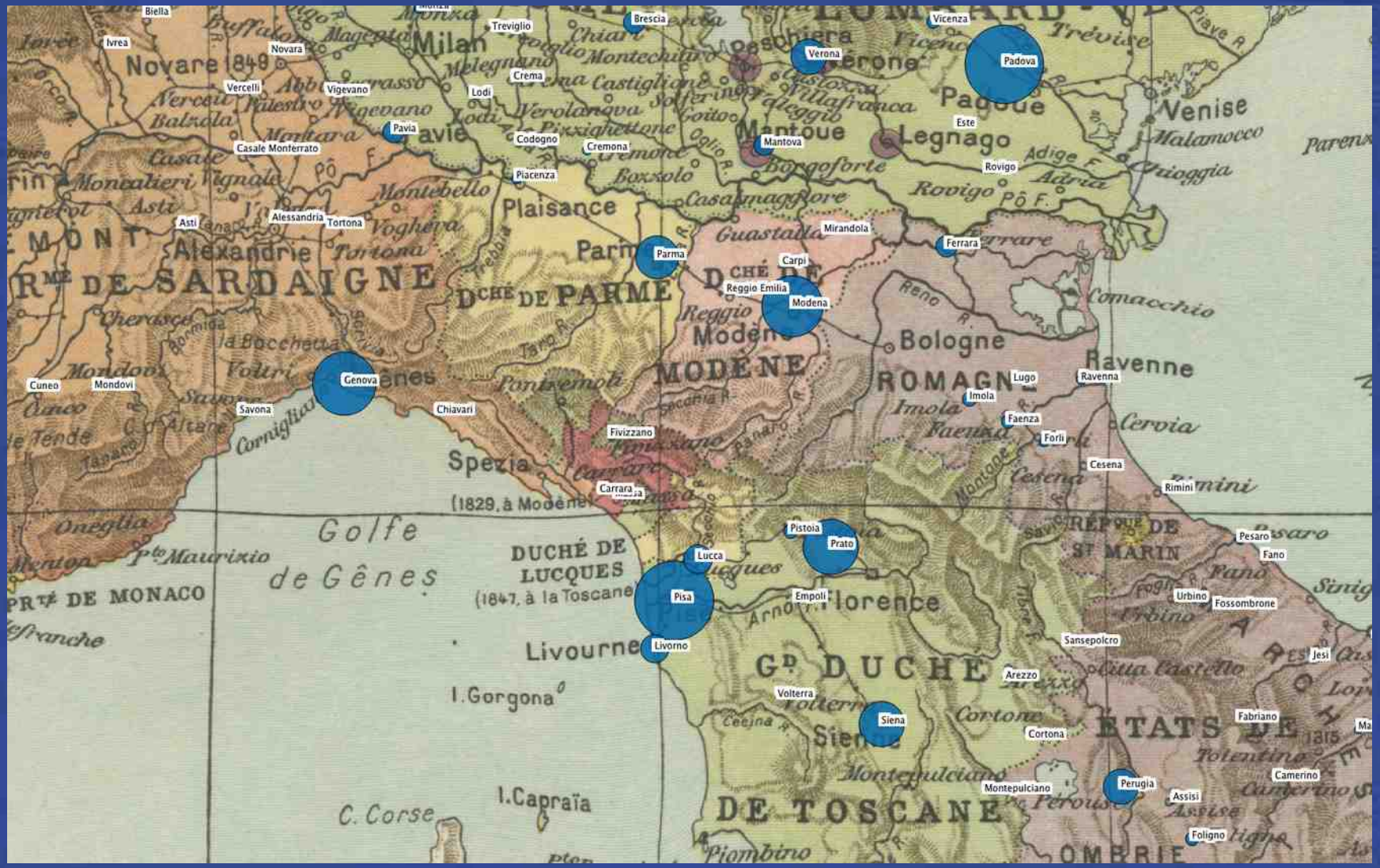
«There is an urgent need to improve the infrastructure supporting the reuse of scholarly data »

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.



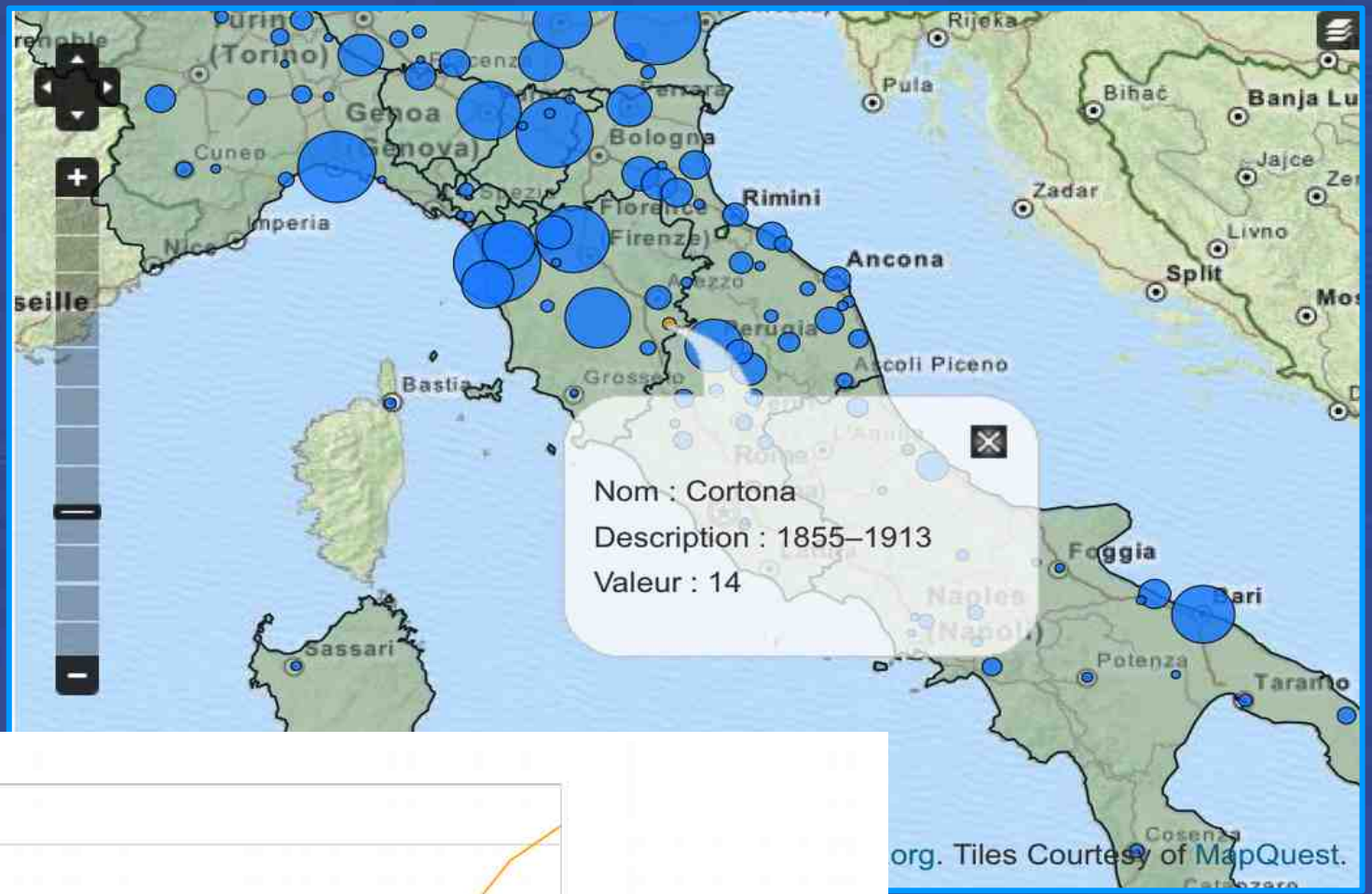
WIKIDATA

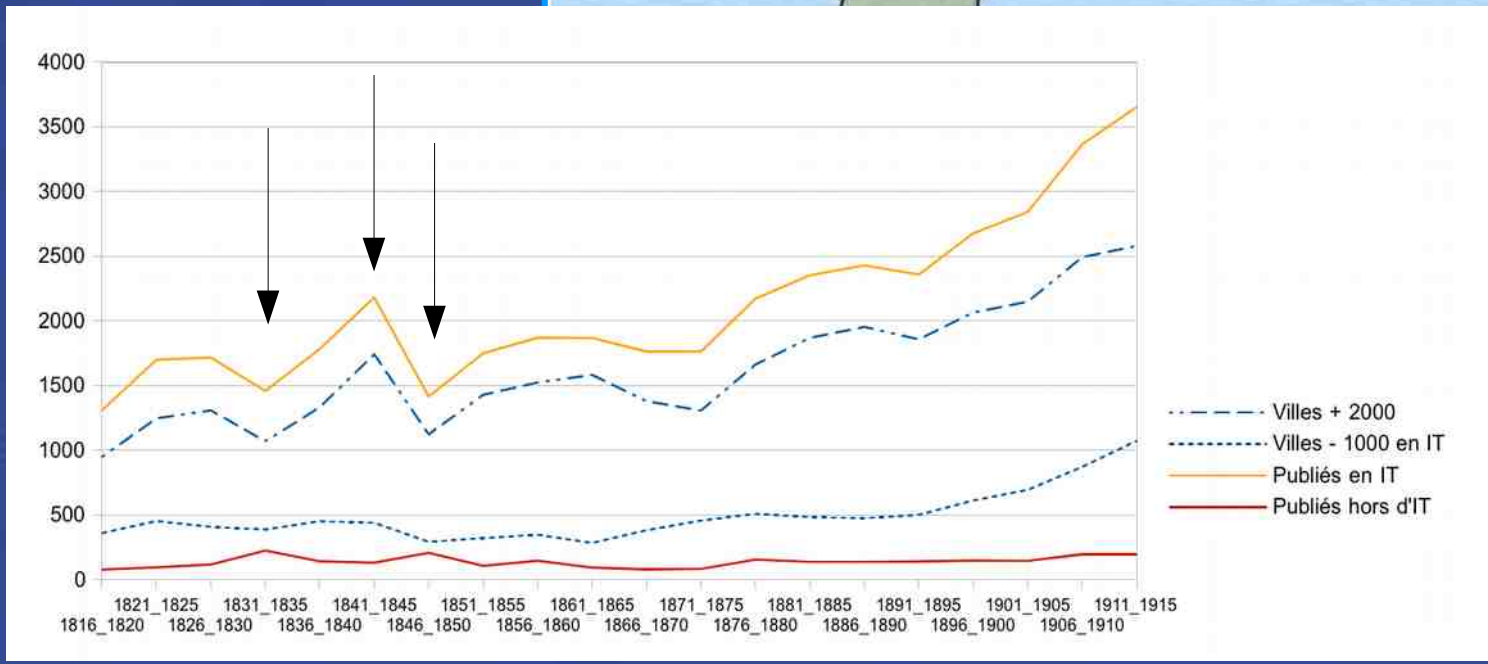
Data from Wikidata available for 1560 persons



SPARQL – endpoint B3Kat
http://lod.b3kat.de/sparql

Bayerische Staatsbibliothek, Bibliotheksverbund Bayern, Kooperative Bibliotheksverbund Berlin-Brandenburg





The FAIR data principles

To be **Findable**:

F1. (meta)data are assigned a globally unique and eternally *persistent identifier*.
F2. data are described with rich *metadata*.
F3. (meta)data are registered or indexed in a *searchable resource*.
F4. metadata specify the data identifier.

To be **Accessible**:

A1 (meta)data are retrievable by their identifier using a *standardized communications protocol*.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2 *metadata are accessible*, even when the data are no longer available.

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.

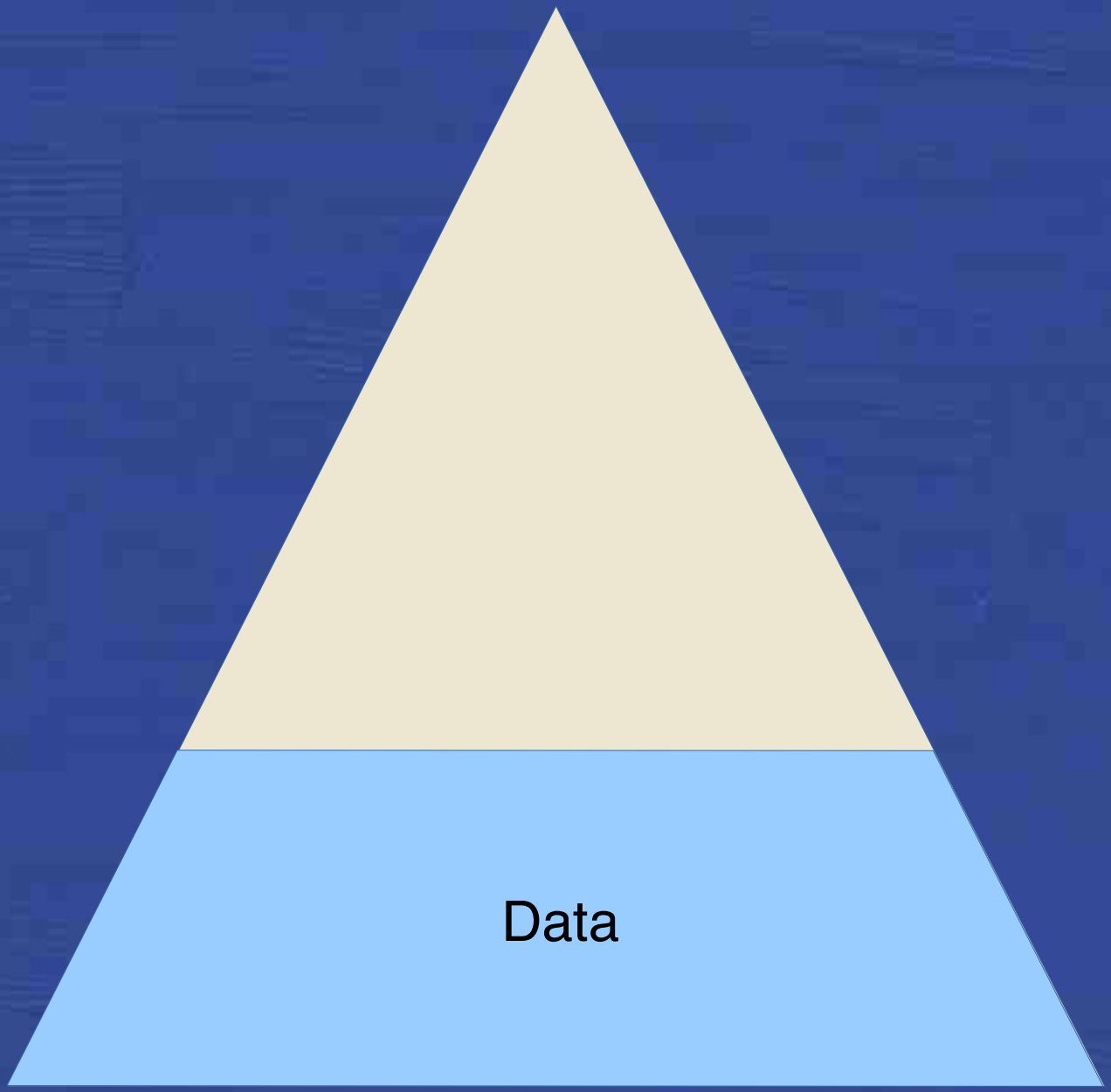
To be **Re-usable**:

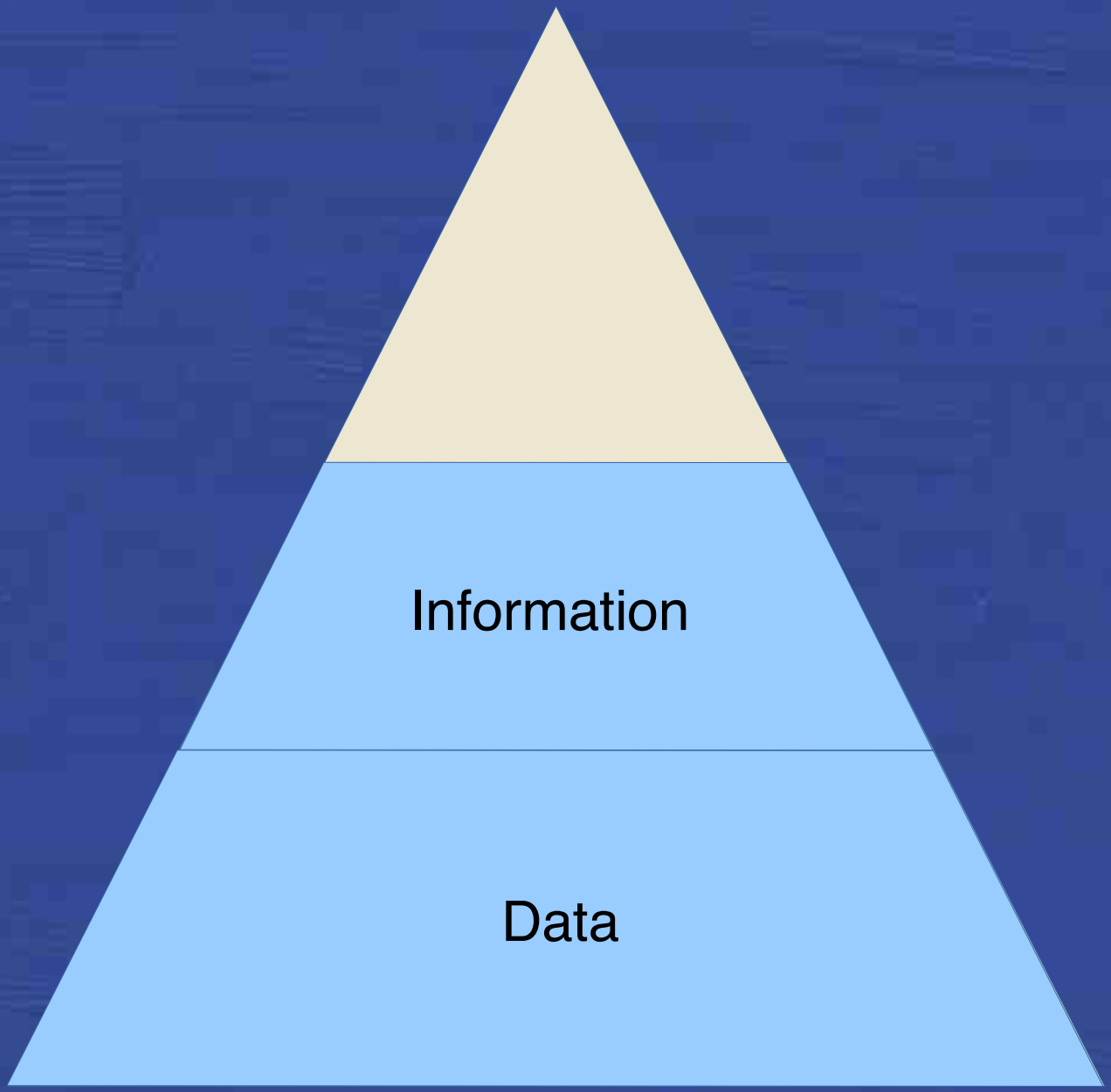
R1. meta(data) have a plurality of accurate and relevant attributes.

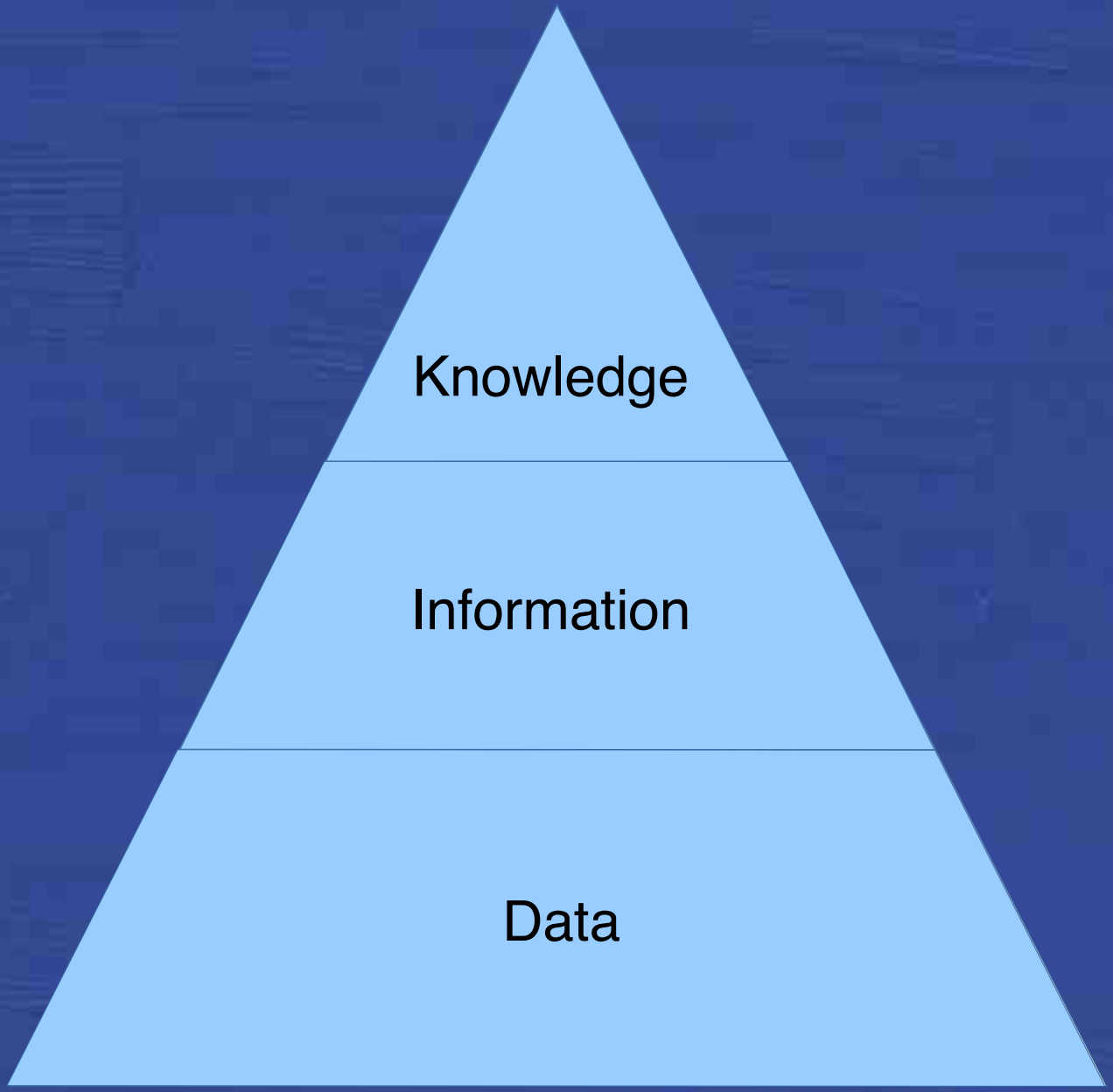
R1.1. (meta)data are released with a *clear and accessible data usage license*.

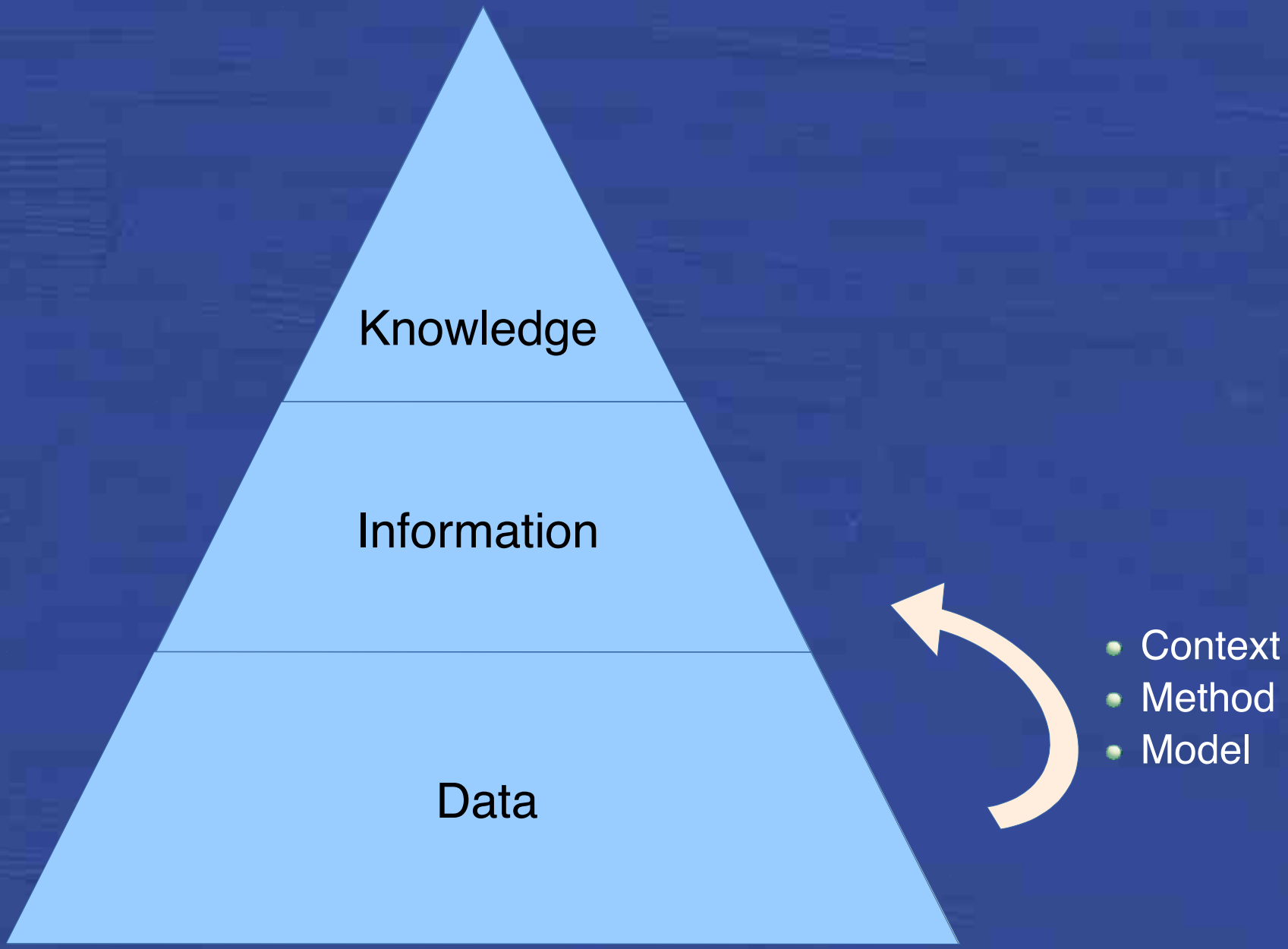
R1.2. (meta)data are associated with their *provenance*.

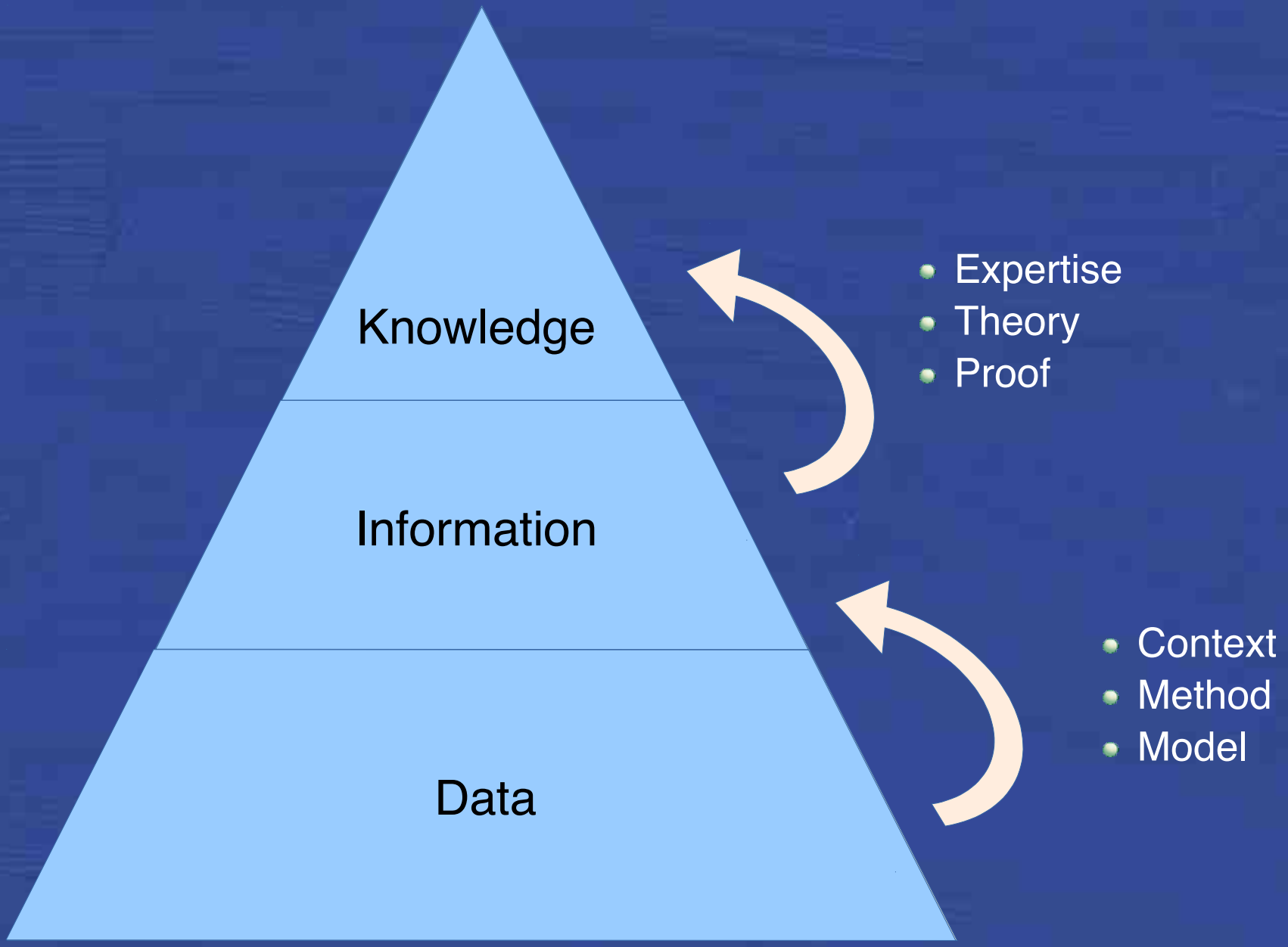
R1.3. (meta)data meet *domain-relevant community standards*.

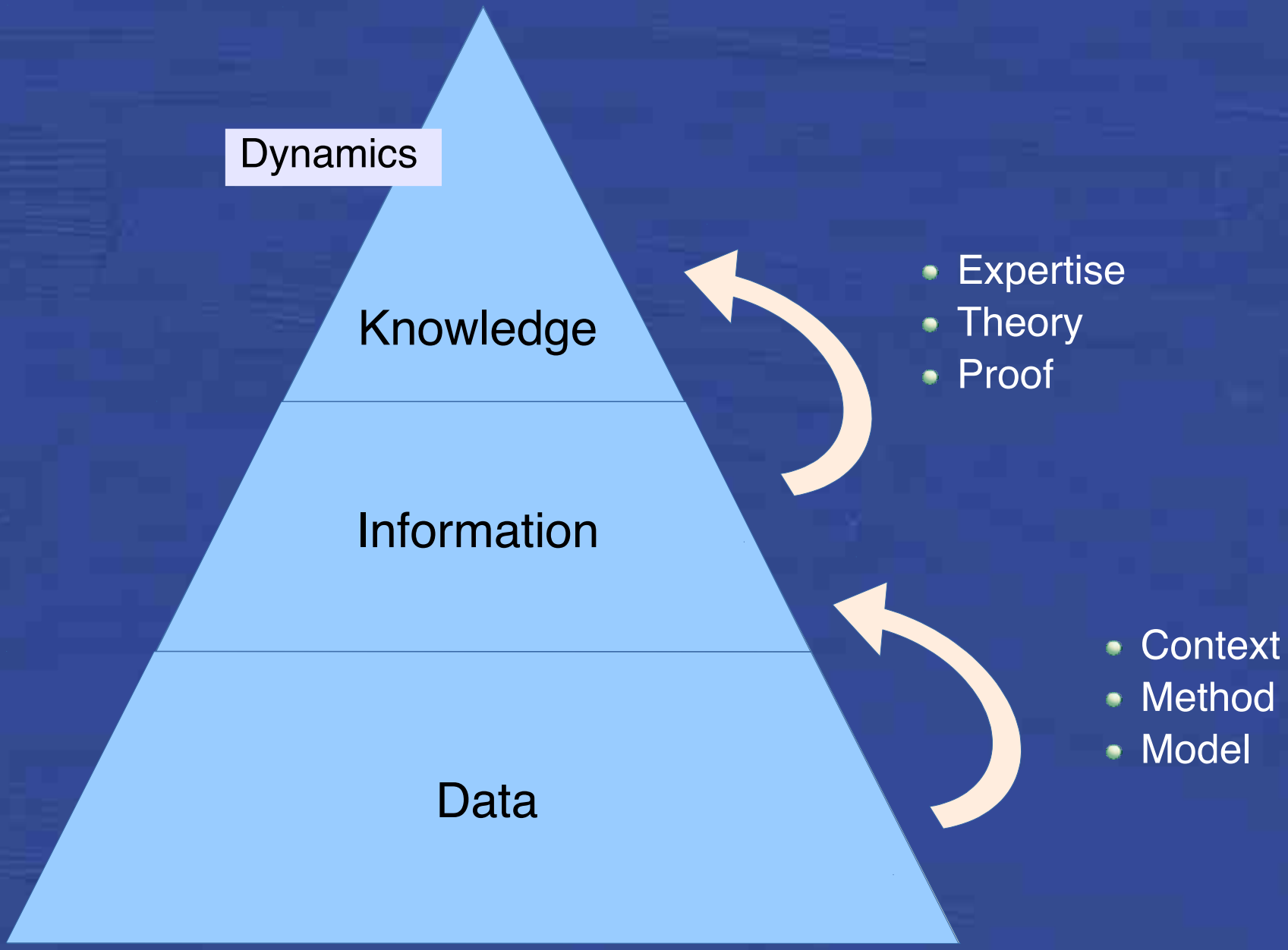

Data

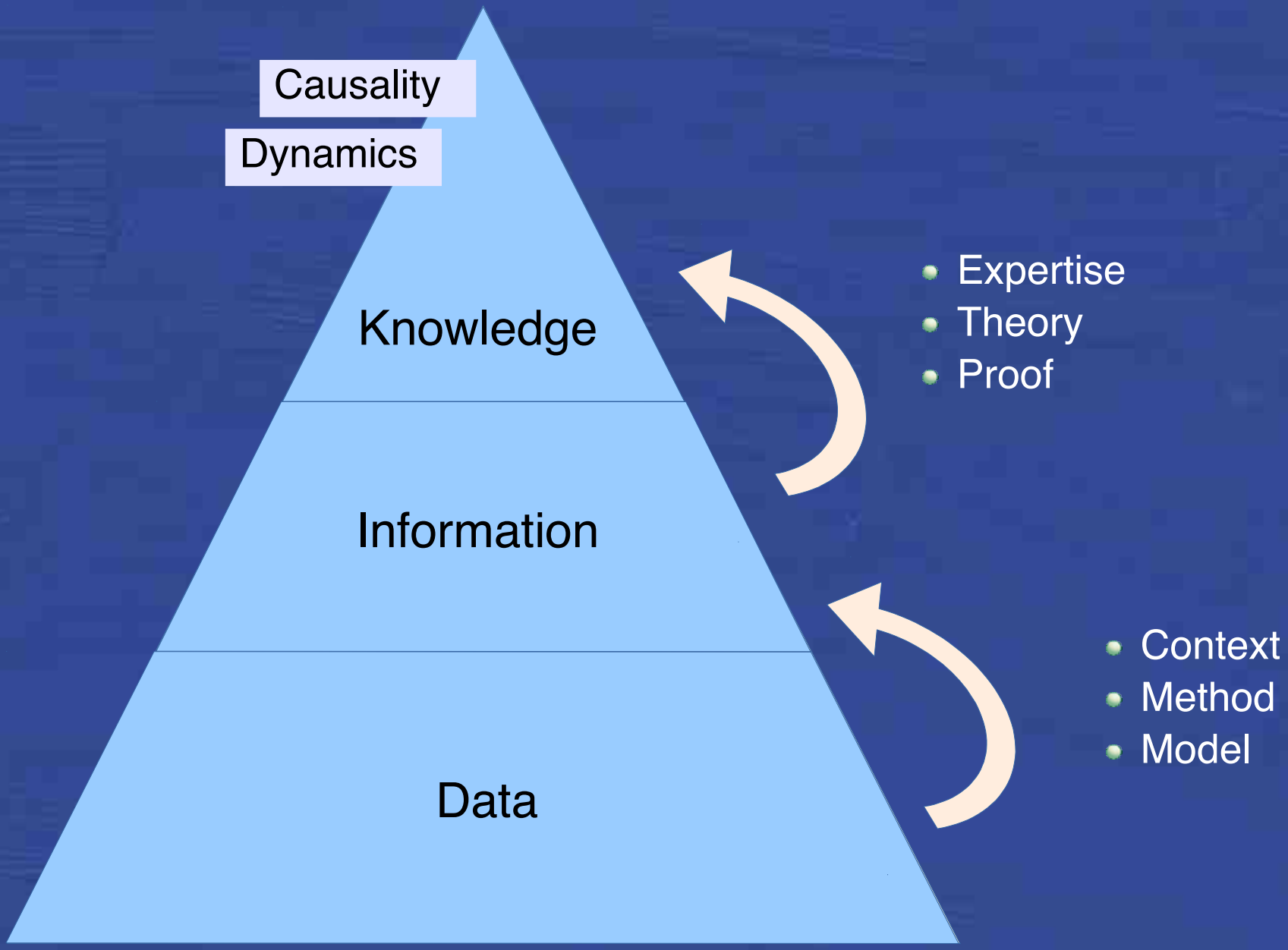

Information
Data

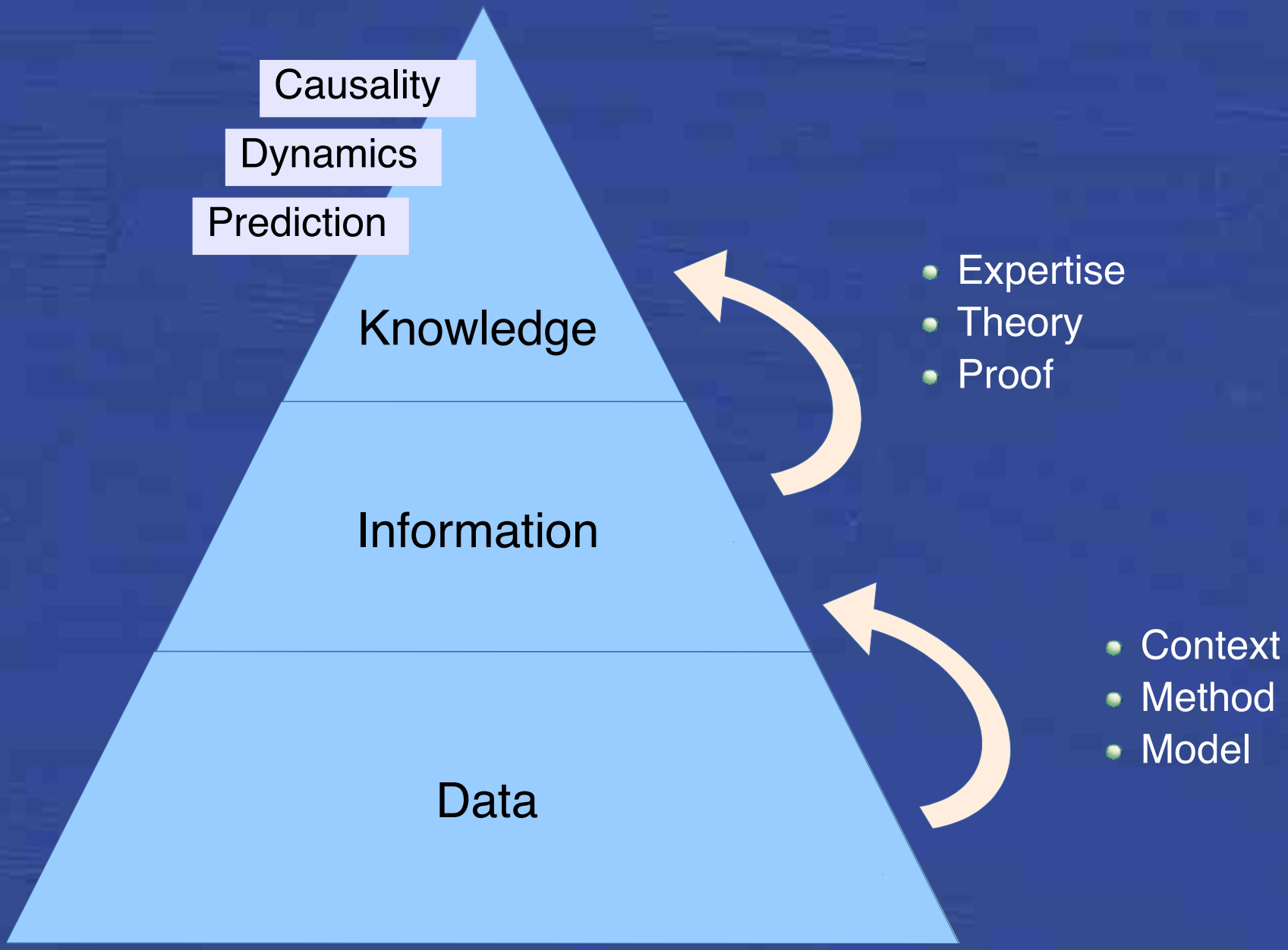

Knowledge
Information
Data

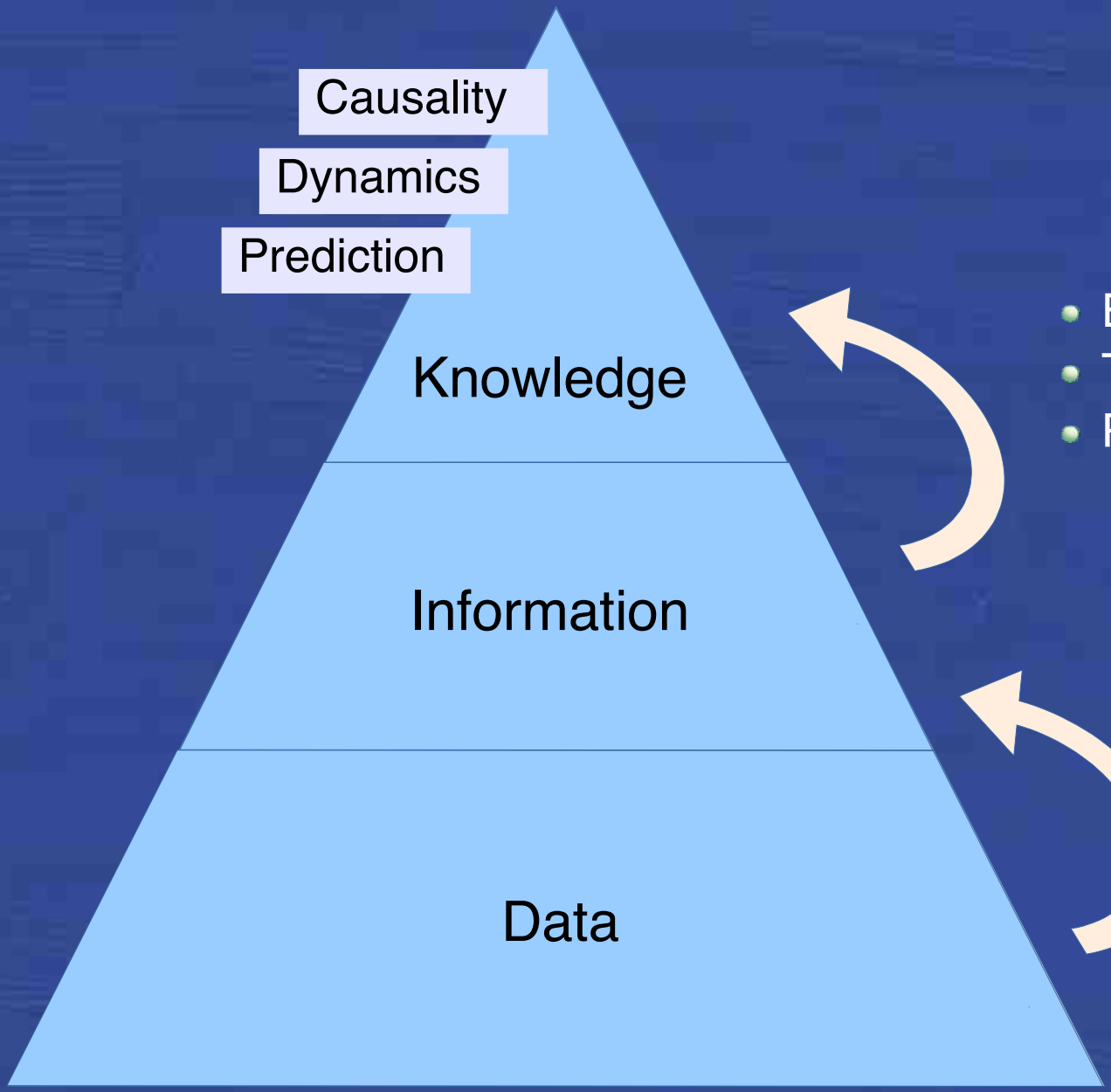

Knowledge
Information
Data
Context
Method
Model


Knowledge
Information
Data
Expertise
Theory
Proof
Context
Method
Model


Dynamics
Knowledge
Information
Data
Expertise
Theory
Proof
Context
Method
Model


Causality
Dynamics
Knowledge
Information
Data
Expertise
Theory
Proof
Context
Method
Model

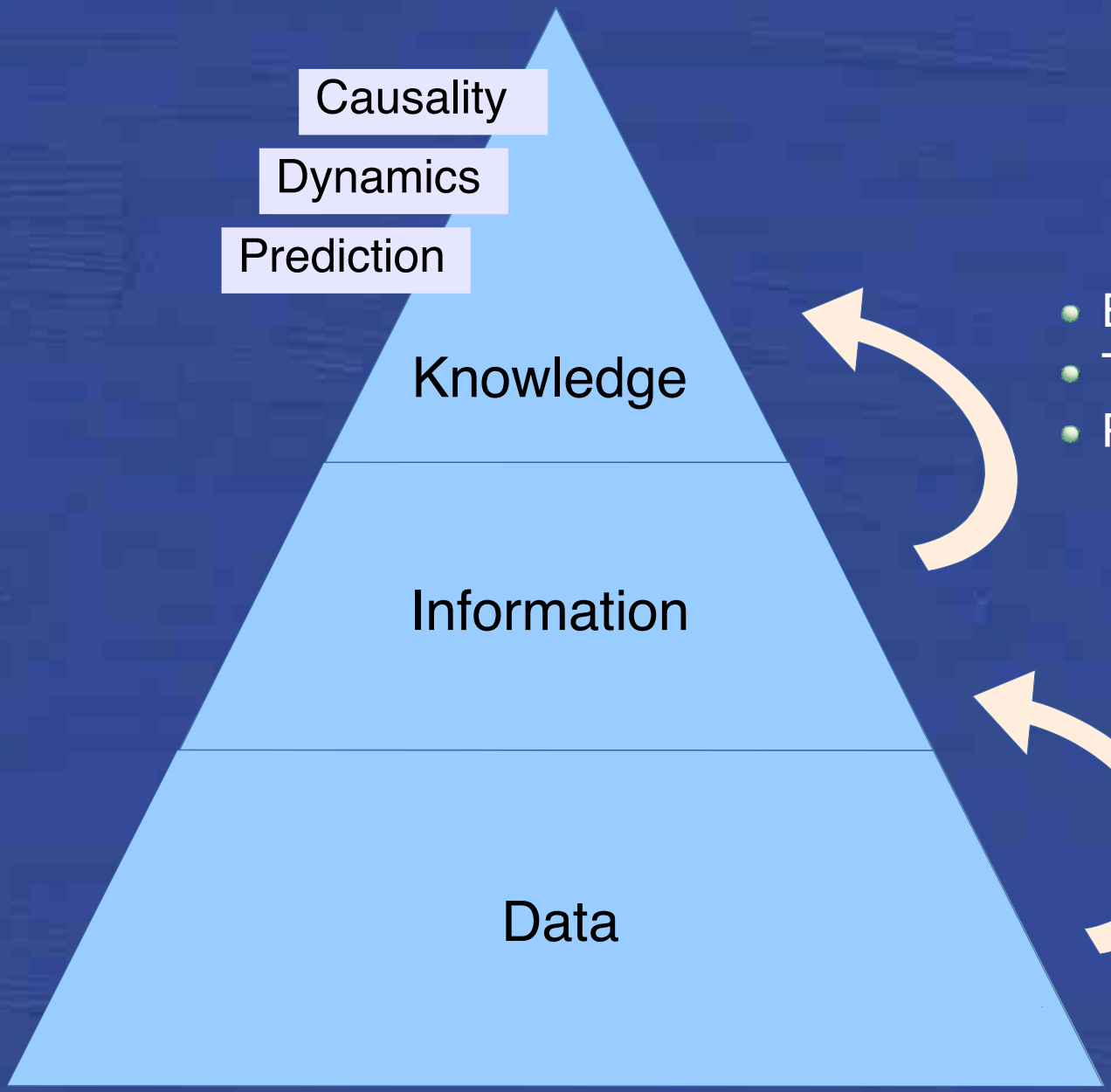

Causality
Dynamics
Prediction
Knowledge
Information
Data
Expertise
Theory
Proof
Context
Method
Model



- Expertise
- Theory
- Proof

Research agenda : gather relevant information

- Context
- Method
- Model

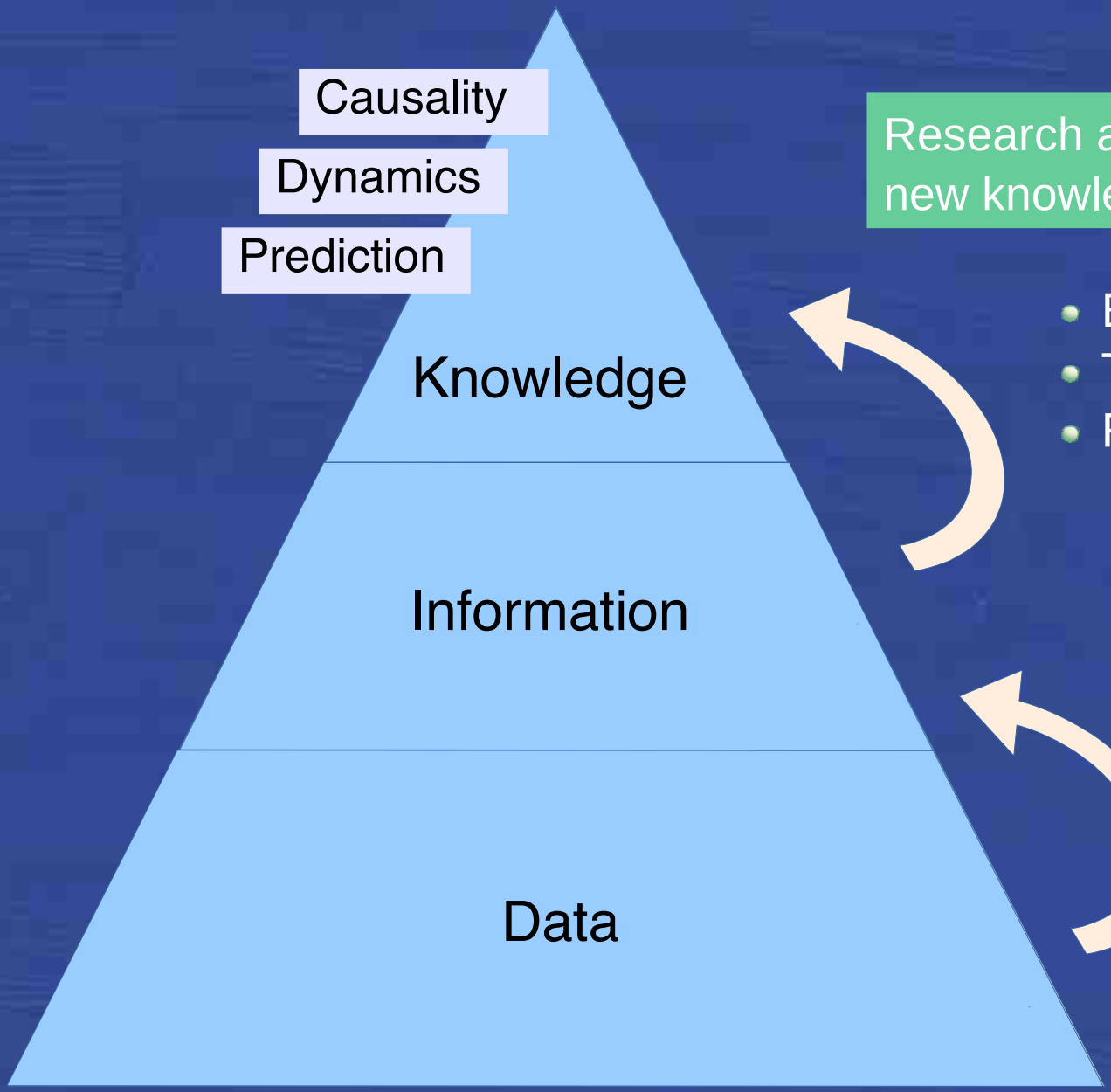


- Expertise
- Theory
- Proof

Research agenda : gather relevant information

- Context
- Method
- Model

Data production, cleansing and alignment



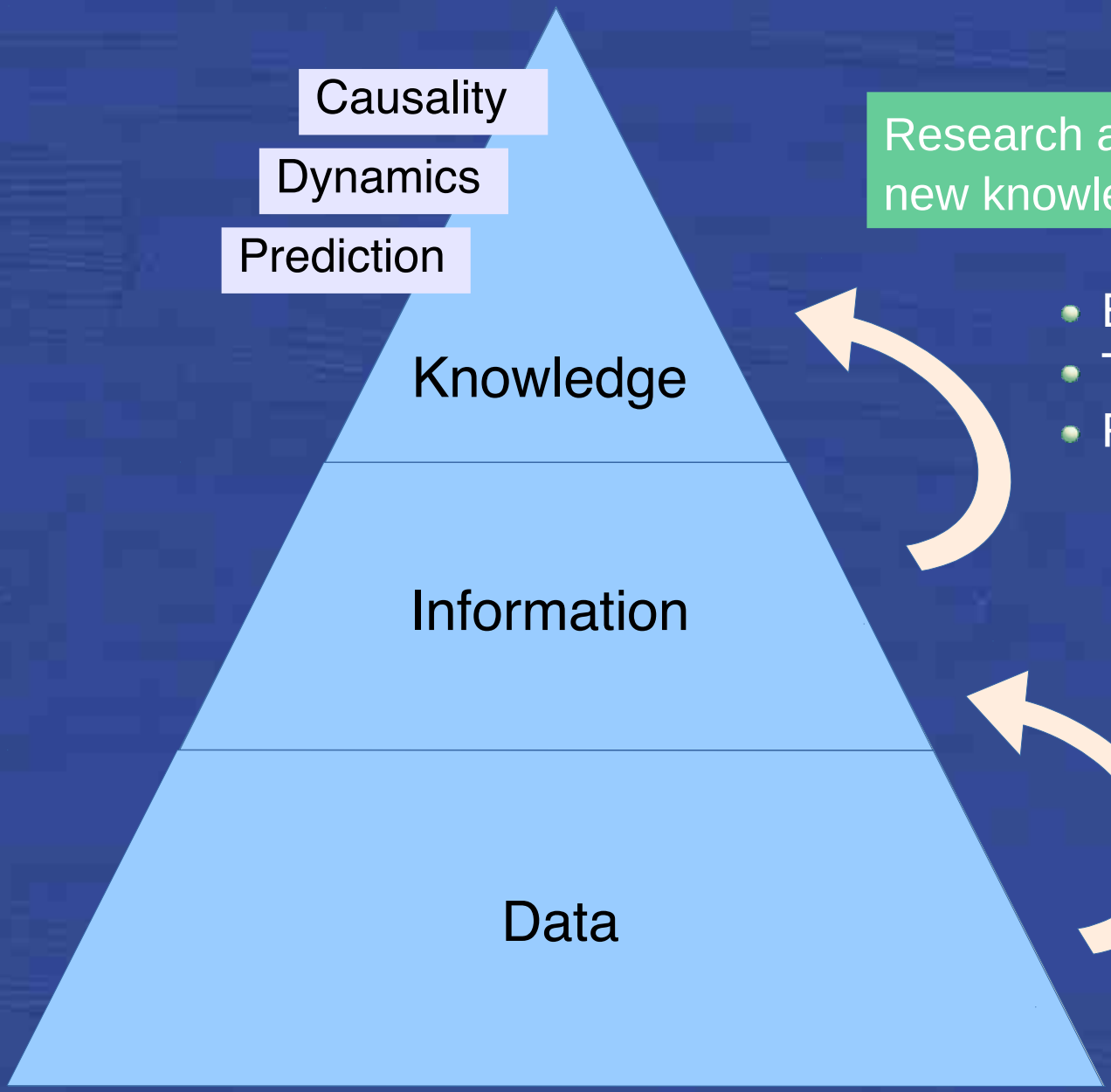
Research agenda : produce new knowledge about the past

- Expertise
- Theory
- Proof

Research agenda : gather relevant information

- Context
- Method
- Model

Data production, cleansing and alignment



Research agenda : produce new knowledge about the past

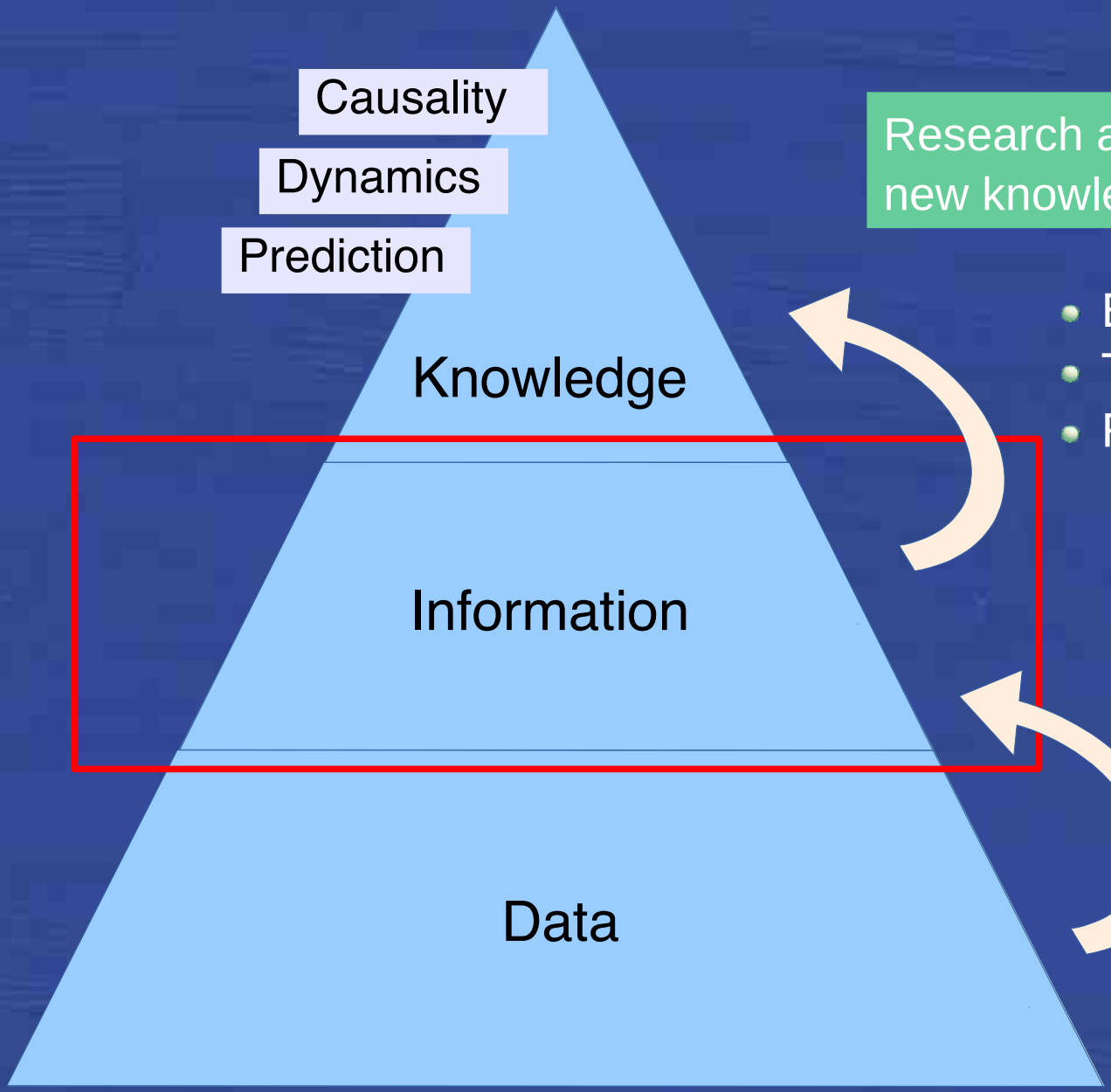
- Expertise
- Theory
- Proof

Analysis
Visualisation
Storytelling

Research agenda : gather relevant information

- Context
- Method
- Model

Data production, cleansing and alignment



Research agenda : produce new knowledge about the past

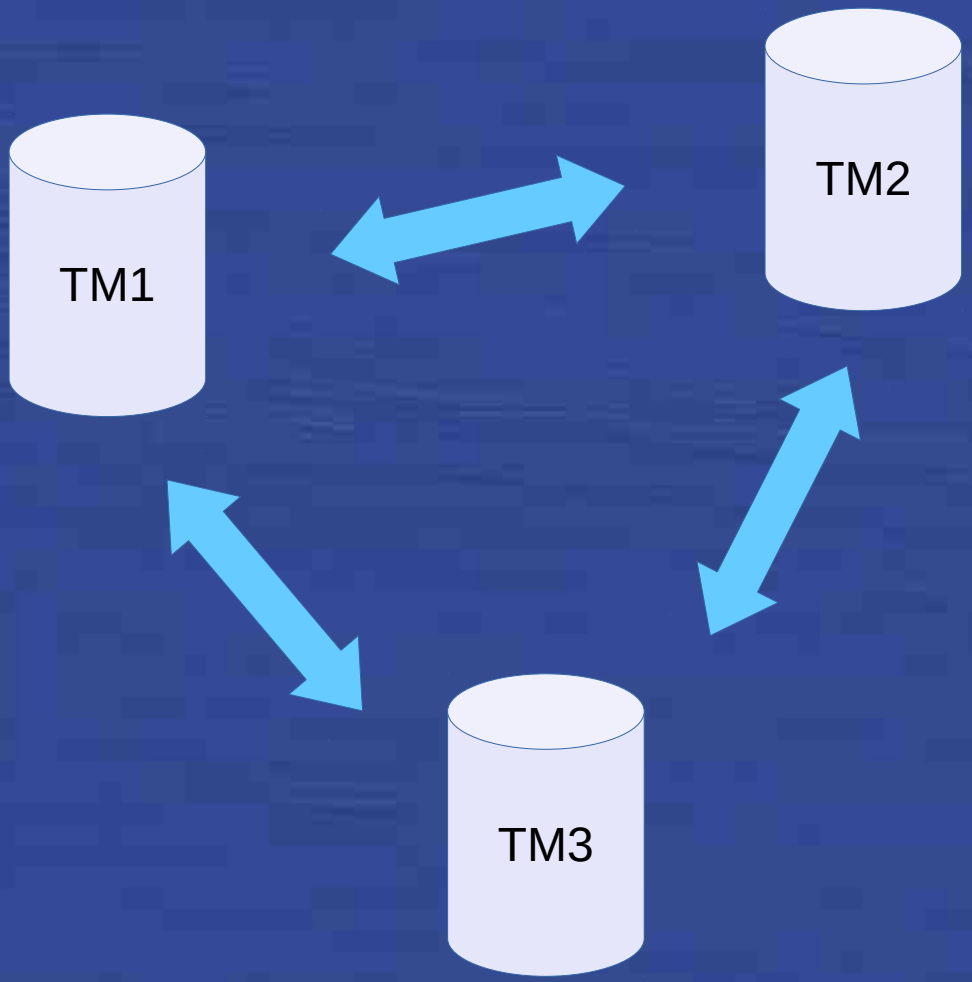
- Expertise
- Theory
- Proof

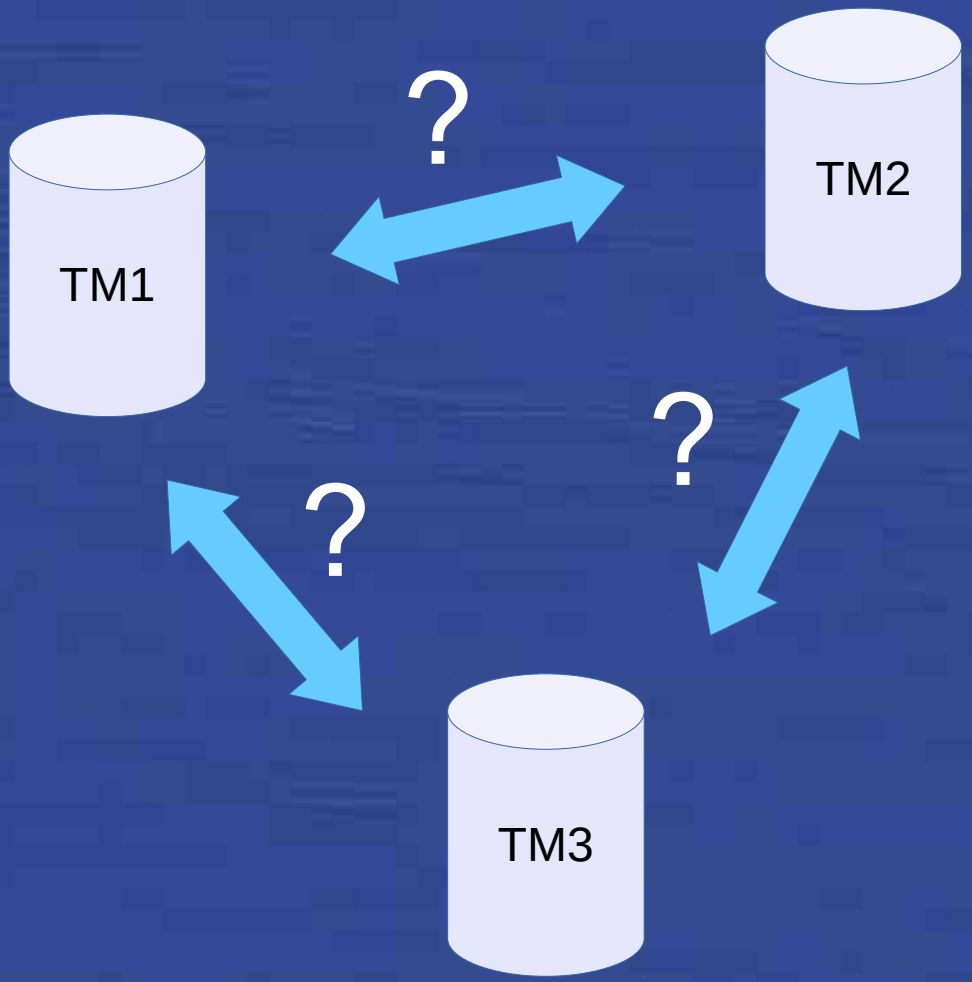
Analysis
Visualisation
Storytelling

Research agenda : gather relevant information

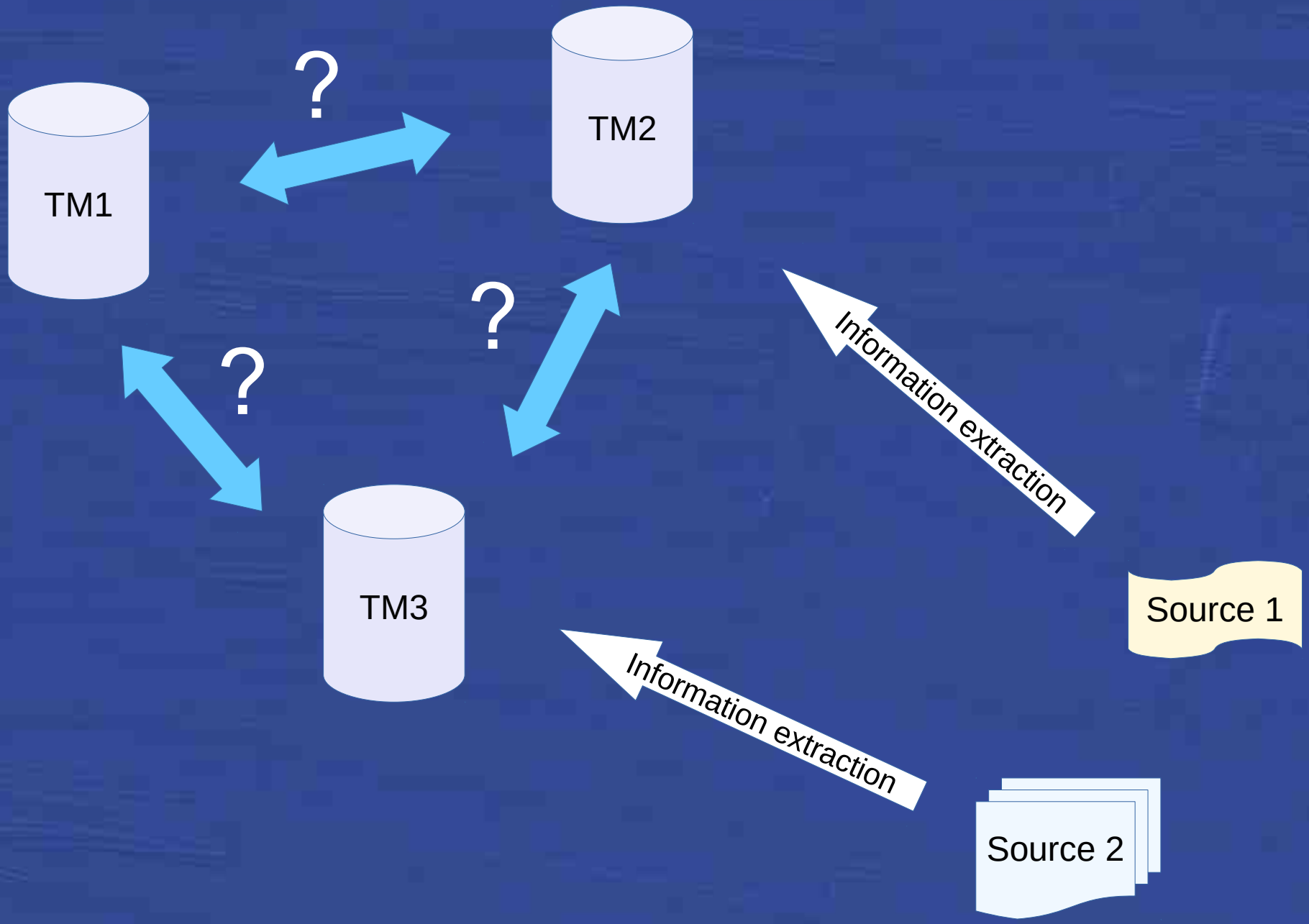
- Context
- Method
- Model

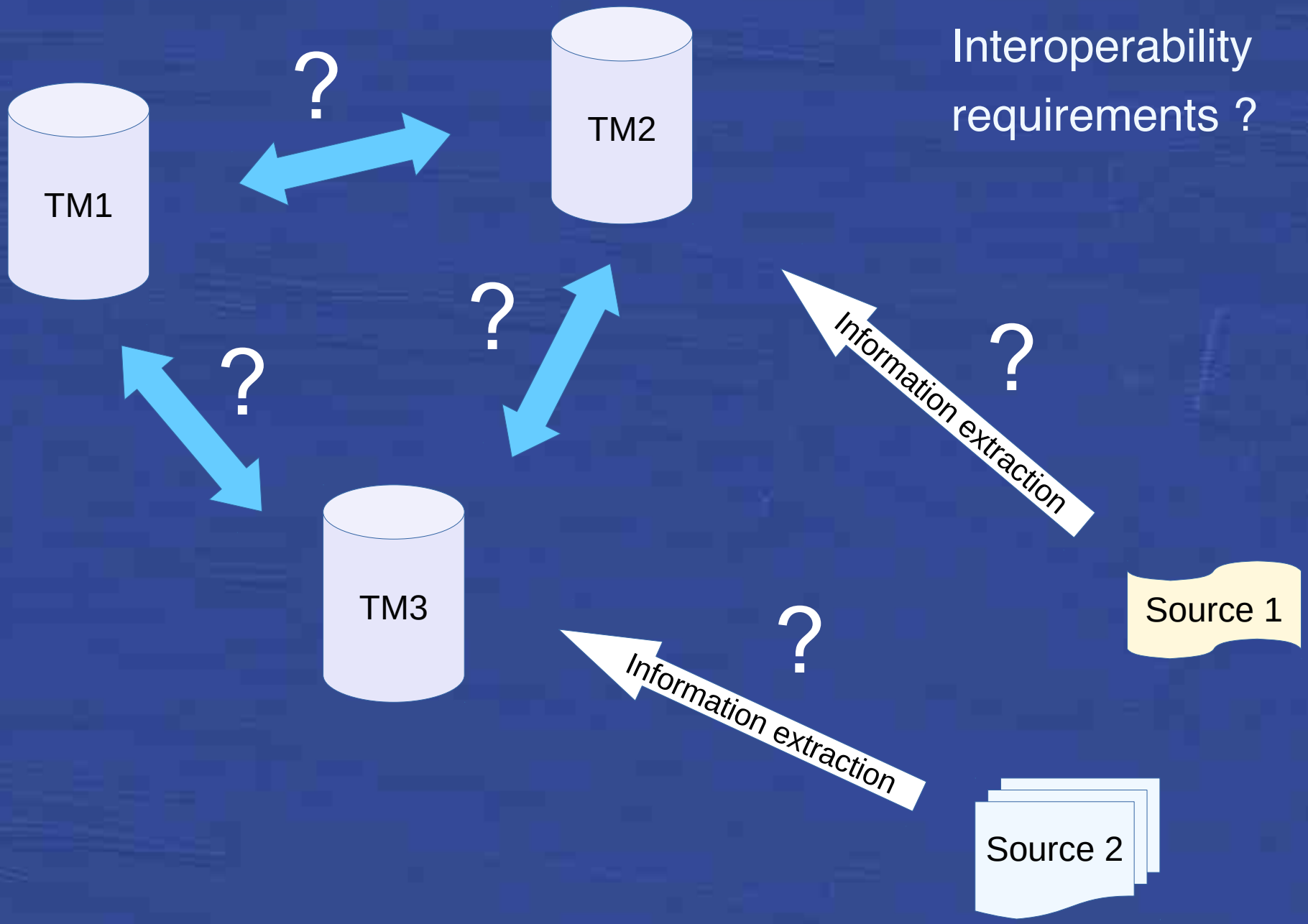
Data production, cleansing and alignment


TM1
TM2
TM3



Interoperability requirements ?


TM1
TM2
TM3
?
?
?
Information extraction
Information extraction
Source 1
Source 2


Interoperability requirements ?
TM1
TM2
TM3
?
?
?
?
?
Information extraction
Information extraction
Source 1
Source 2

Requirements for structured data interoperability :

Requirements for structured data interoperability :

* Share and interlink dictionaries of identified entities (authority files or gazetteers) : people, places, organisations, books, …

« In 1592, he [Galileo Galilei] moved to the University of Padua where he taught [mathematics] until 1610»

[https://en.wikipedia.org/wiki/Galileo_Galilei#Reference-Sharratt-1994]

« In 1592, he [Galileo Galilei] moved to the University of Padua where he taught [mathematics] until 1610»

[https://en.wikipedia.org/wiki/Galileo_Galilei#Reference-Sharratt-1994]

Record linkage : is my Galilei your Galilei ?

« In 1592, he [Galileo Galilei] moved to the University of Padua where he taught [mathematics] until 1610»

[https://en.wikipedia.org/wiki/Galileo_Galilei#Reference-Sharratt-1994]

Record linkage : is my Galilei your Galilei ?

Galileo Galilei :

- *1564, mathematician and astronomer

« In 1592, he [Galileo Galilei] moved to the University of Padua where he taught [mathematics] until 1610»

[https://en.wikipedia.org/wiki/Galileo_Galilei#Reference-Sharratt-1994]

Record linkage : is my Galilei your Galilei ?

Galileo Galilei :

- *1564, mathematician and astronomer
- *1370, medicine lecturer at Florence *studio*

« In 1592, he [Galileo Galilei] moved to the University of Padua where he taught [mathematics] until 1610»

[https://en.wikipedia.org/wiki/Galileo_Galilei#Reference-Sharratt-1994]

Record linkage : is my Galilei your Galilei ?

Galileo Galilei : **http://symogih.org/resource/Actr161**

- *1564, mathematician and astronomer
- *1370, medicine lecturer at Florence *studio*

http://symogih.org/resource/Actr646

« In 1592, he [Galileo Galilei] moved to the University of Padua where he taught [mathematics] until 1610»

[https://en.wikipedia.org/wiki/Galileo_Galilei#Reference-Sharratt-1994]

Record linkage : is my Galilei your Galilei ?

Galileo Galilei :

- *1564, mathematician and astronomer
- *1370, medicine lecturer at Florence *studio*

Who was Johannes Teutonicus ?

« In 1592, he [Galileo Galilei] moved to the University of Padua where he taught [mathematics] until 1610»

[https://en.wikipedia.org/wiki/Galileo_Galilei#Reference-Sharratt-1994]

Record linkage : is my Galilei your Galilei ?

Galileo Galilei :

- *1564, mathematician and astronomer
- *1370, medicine lecturer at Florence *studio*

Who was Johannes Teutonicus ?

What about places, organizations, books ?

Requirements for structured data interoperability :

* Share and interlink dictionaries of identified entities (authority files or gazetteers) : people, places, organisations, books, …

* Share controlled vocabulaires (thesauri and taxonomies of ‘types’) that are used to classify entities (and provide their identity)

Galilei → Mathematician and astronomer

Galilei → Mathematician and astronomer

Mathematicians :

- Louis de Broglie (1892-1987), French mathematician and physicist, Nobel Prize for Physics in 1929

Galilei → Mathematician and astronomer

Mathematicians :

- Louis de Broglie (1892-1987), French mathematician and physicist, Nobel Prize for Physics in 1929
- Stephen Smale (*1930), American mathematician. His research concerns topology, dynamical systems and mathematical economics.

Galilei → Mathematician and astronomer

Mathematicians :

- Louis de Broglie (1892-1987), French mathematician and physicist, Nobel Prize for Physics in 1929
- Stephen Smale (*1930), American mathematician. His research concerns topology, dynamical systems and mathematical economics.

Mathematics (17th century)

Galilei → Mathematician and astronomer

Mathematicians :

- Louis de Broglie (1892-1987), French mathematician and physicist, Nobel Prize for Physics in 1929
- Stephen Smale (*1930), American mathematician. His research concerns topology, dynamical systems and mathematical economics.

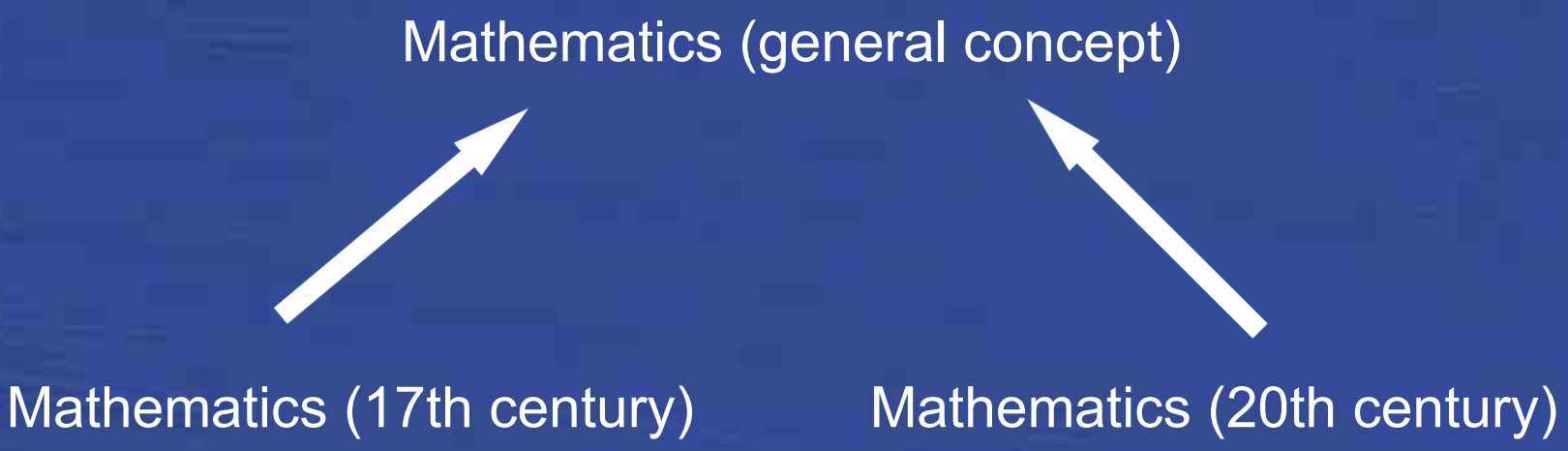
Mathematics (17th century)

Mathematics (20th century)

Galilei → Mathematician and astronomer

Mathematicians :

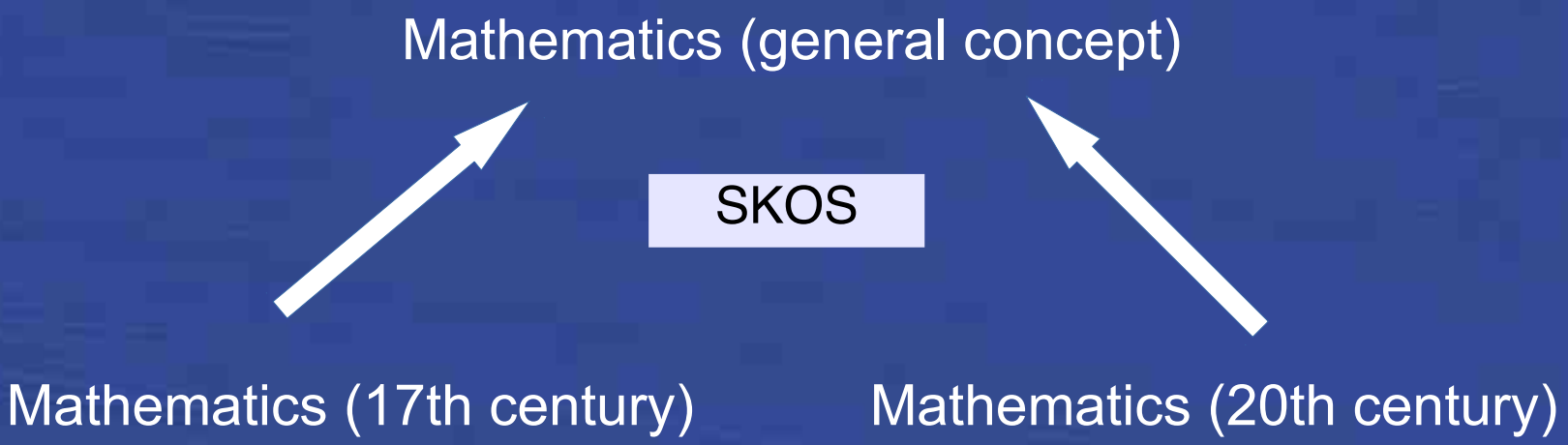
- Louis de Broglie (1892-1987), French mathematician and physicist, Nobel Prize for Physics in 1929
- Stephen Smale (*1930), American mathematician. His research concerns topology, dynamical systems and mathematical economics.



Galilei → Mathematician and astronomer

Mathematicians :

- Louis de Broglie (1892-1987), French mathematician and physicist, Nobel Prize for Physics in 1929
- Stephen Smale (*1930), American mathematician. His research concerns topology, dynamical systems and mathematical economics.



Requirements for structured data interoperability :

* Share and interlink dictionaries of identified entities (authority files or gazetteers) : people, places, organisations, books, …

* Share controlled vocabulaires (thesauri and taxonomies of ‘types’ that are used to classify entities (and provide their identity)

* Share information about entities :
  a common, multi-tiered and extensible data model, i.e.
  an ontology as a shared conceptualization of the ‘world’ and its geo-historical development (cultural heritage domain)

« In 1592, he [Galileo Galilei] moved to the University of Padua where he taught [mathematics] until 1610»

[https://en.wikipedia.org/wiki/Galileo_Galilei#Reference-Sharratt-1994]

« In 1592, he [Galileo Galilei] moved to the University of Padua where he taught [mathematics] until 1610»

[https://en.wikipedia.org/wiki/Galileo_Galilei#Reference-Sharratt-1994]

Galilei moves to Padua

« In 1592, he [Galileo Galilei] moved to the University of Padua where he taught [mathematics] until 1610»

[https://en.wikipedia.org/wiki/Galileo_Galilei#Reference-Sharratt-1994]

Galilei moves to Padua

Galilei is hired by the University of Padua

« In 1592, he [Galileo Galilei] moved to the University of Padua where he taught [mathematics] until 1610»

[https://en.wikipedia.org/wiki/Galileo_Galilei#Reference-Sharratt-1994]

Galilei moves to Padua

Galilei is hired by the University of Padua

Galilei teaches mathematics...

« In 1592, he [Galileo Galilei] moved to the University of Padua where he taught [mathematics] until 1610»

[https://en.wikipedia.org/wiki/Galileo_Galilei#Reference-Sharratt-1994]

Galilei moves to Padua

Galilei is hired by the University of Padua

Galilei teaches mathematics...

- Galilei has the status of professor of mathematics

« In 1592, he [Galileo Galilei] moved to the University of Padua where he taught [mathematics] until 1610»

[https://en.wikipedia.org/wiki/Galileo_Galilei#Reference-Sharratt-1994]

Galilei moves to Padua

Galilei is hired by the University of Padua

Galilei teaches mathematics...

- Galilei has the status of professor of mathematics
- Galilei teaches mathematics courses but he has a sabbatical year

« In 1592, he [Galileo Galilei] moved to the University of Padua where he taught [mathematics] until 1610»

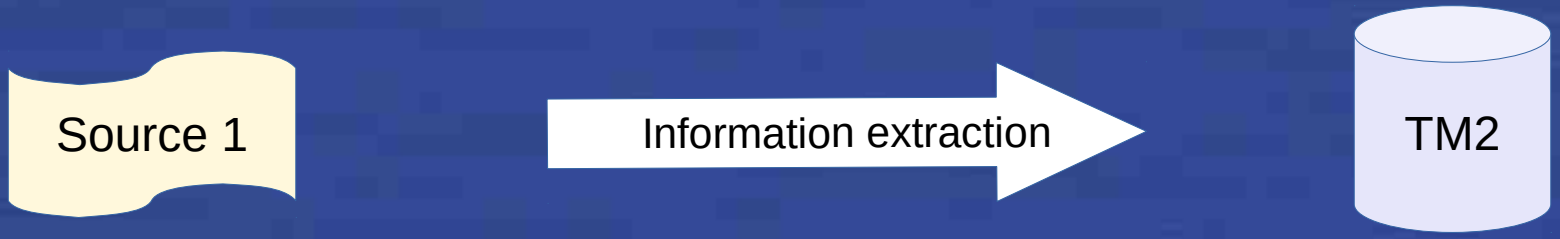
[https://en.wikipedia.org/wiki/Galileo_Galilei#Reference-Sharratt-1994]

Galilei moves to Padua

Galilei is hired by the University of Padua

Galilei teaches mathematics...

- Galilei has the status of professor
- Galilei teaches mathematics courses but he has a sabbatical year
- During this semester Galilei regularly teaches mathematics



Johannes Kepler was born on December 27 1571.

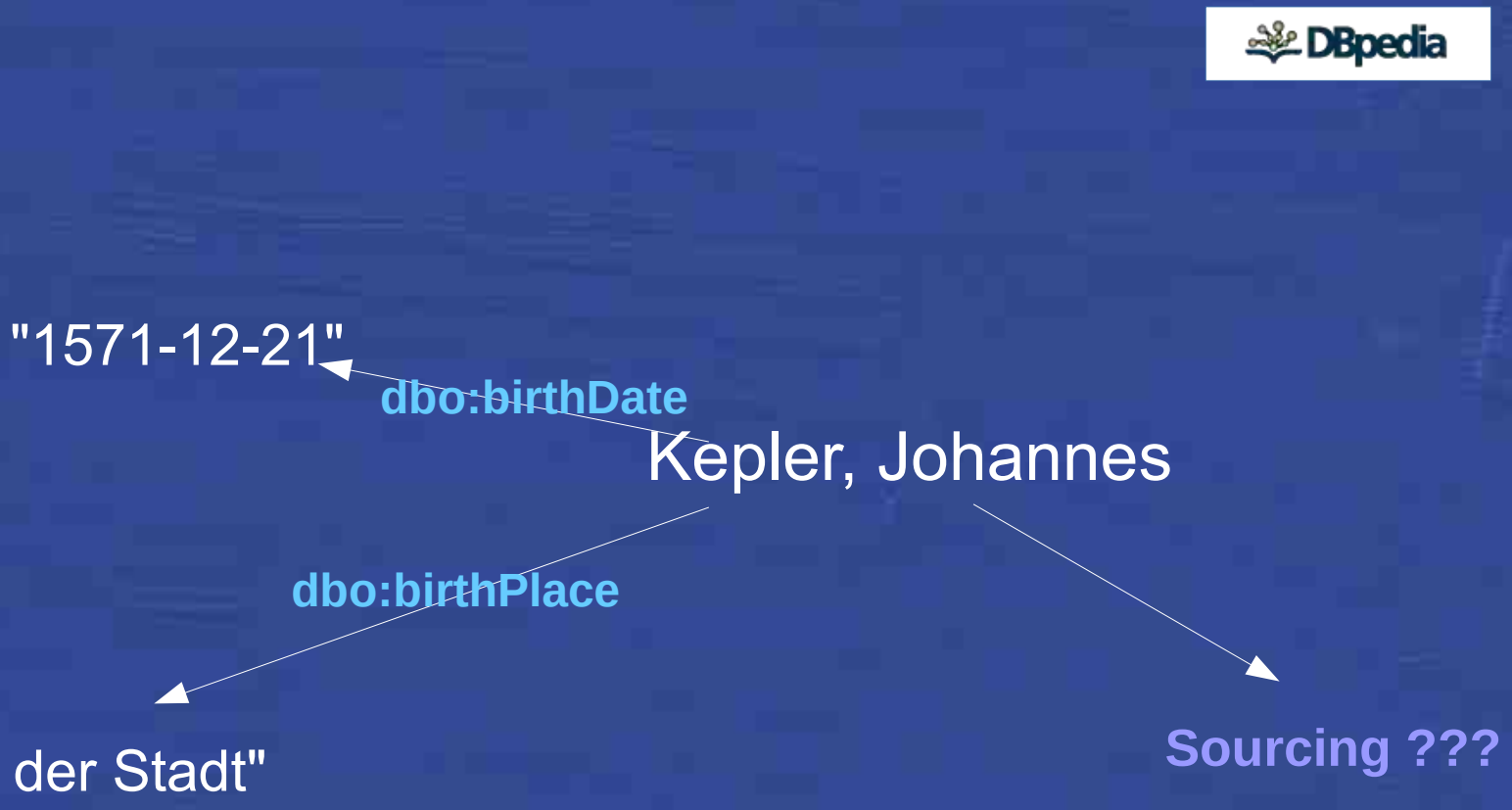
{"Tokens" : [
{"doc" : 1, "sentence" : 6, "token_id":1, "token" : Kepler}, was
{"doc" : 1, "sentence" : 6, "token_id":2, "token" : **was**},
{"doc" : 1, "sentence" : 6, "token_id":3, "token" : **born**},
{"doc" : 1, "sentence" : 6, "token_id":4, "token" : on}, was
{"doc" : 1, "sentence" : 6, "token_id":5, "token" : December},
{"doc" : 1, "sentence" : 6, "token_id":6, "token" : 27},
{"doc" : 1, "sentence" : 6, "token_id":7, "token" : 1571},
…]}

Chunk [Tokens ids]	Entity	Value
1	dbr:Johannes_Kepler	
5,6,7		1571-12-27

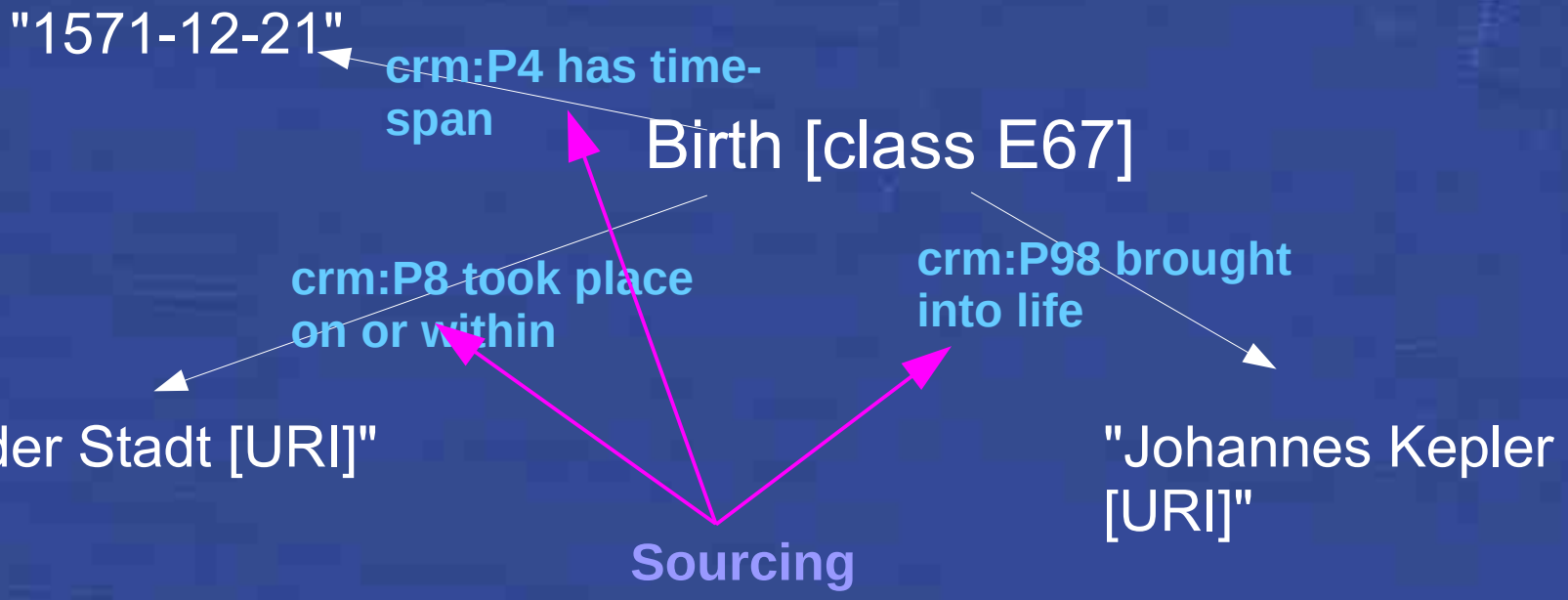
The most important challenges at this stage :

- how to cope with data redundancy and duplication ?
- how to transform *mentions* of people and situations to data that contain consistent information about *identified* persons ?
- how to reformat the collected data and produce information that can be used for systematic processing ?

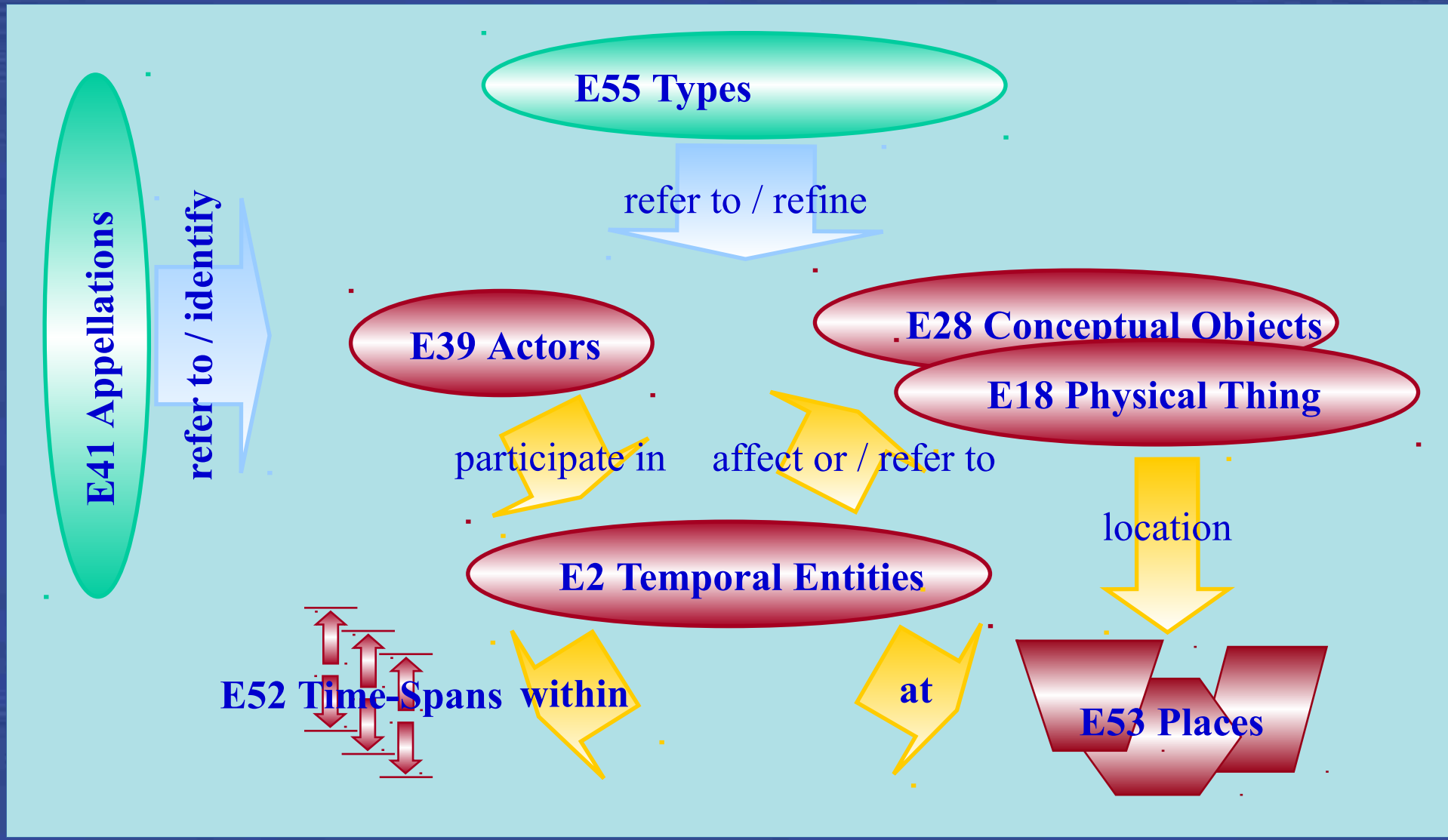
Adopt a resource-centered model...



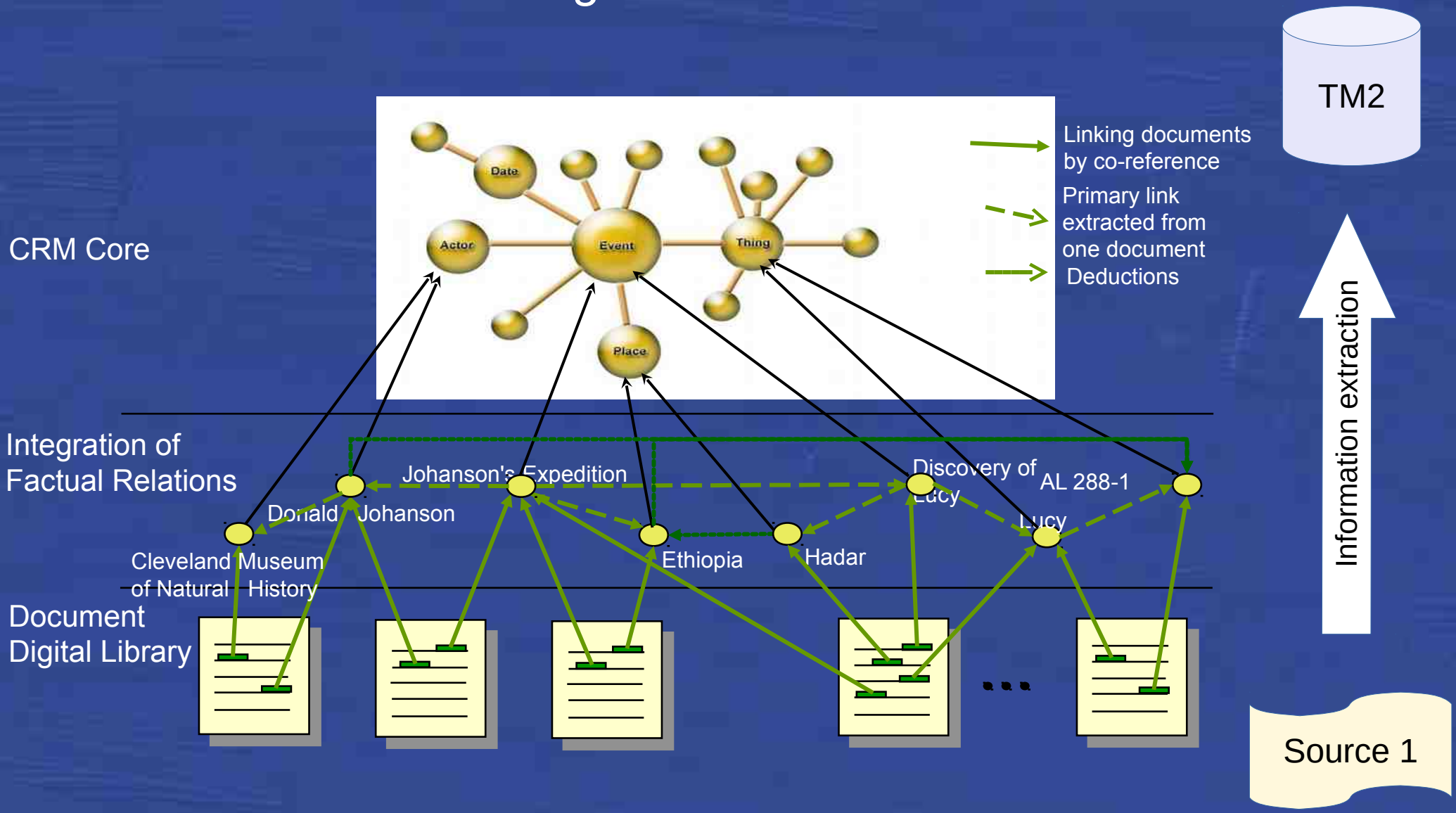
… or prefer a more robust event-centered model !

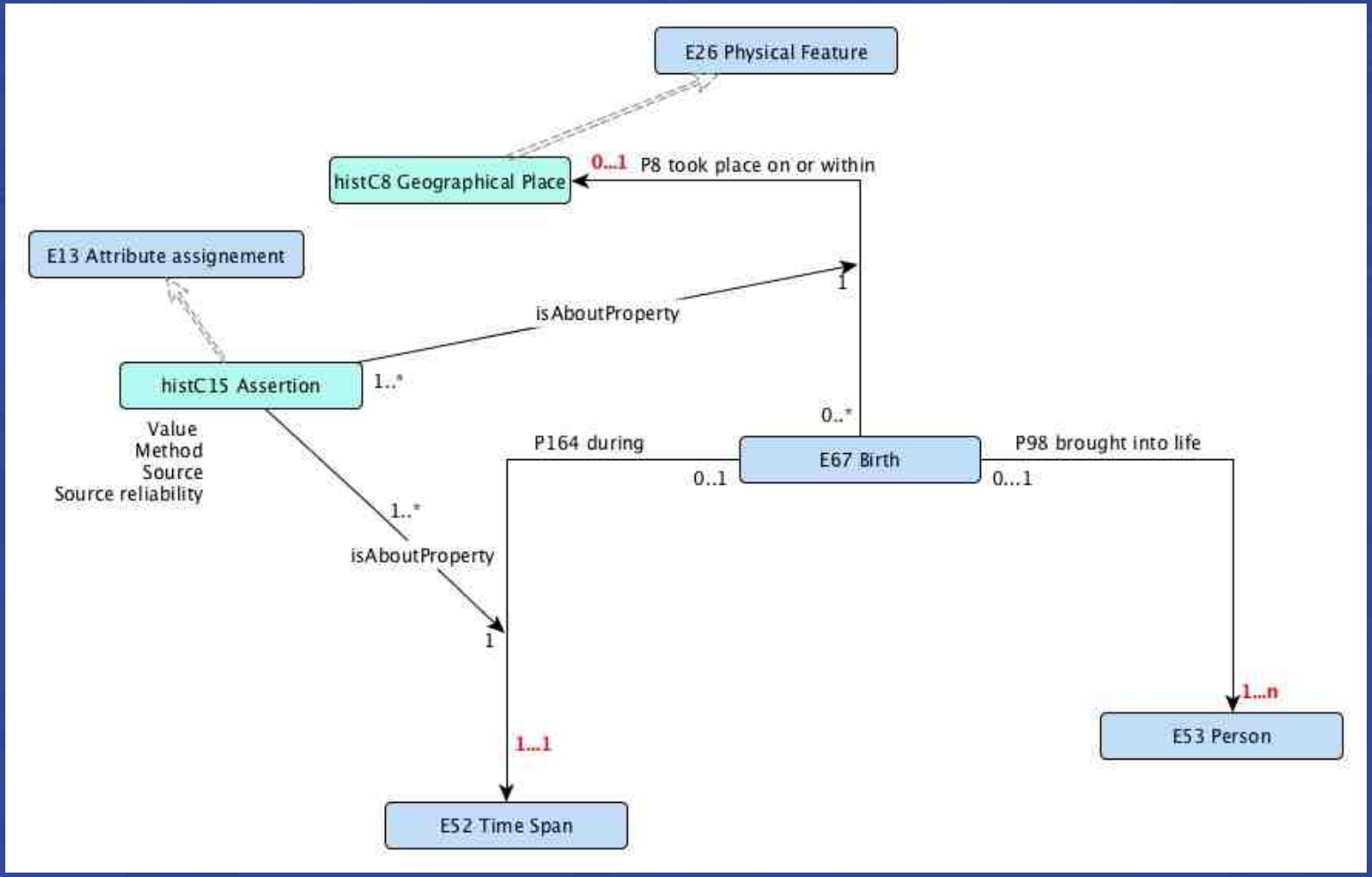


The CIDOC CRM (ISO21127:2006)
A semantic framework that provides interoperability between different sources of cultural heritage information



Integration of information extracted from documents using the CIDOC CRM

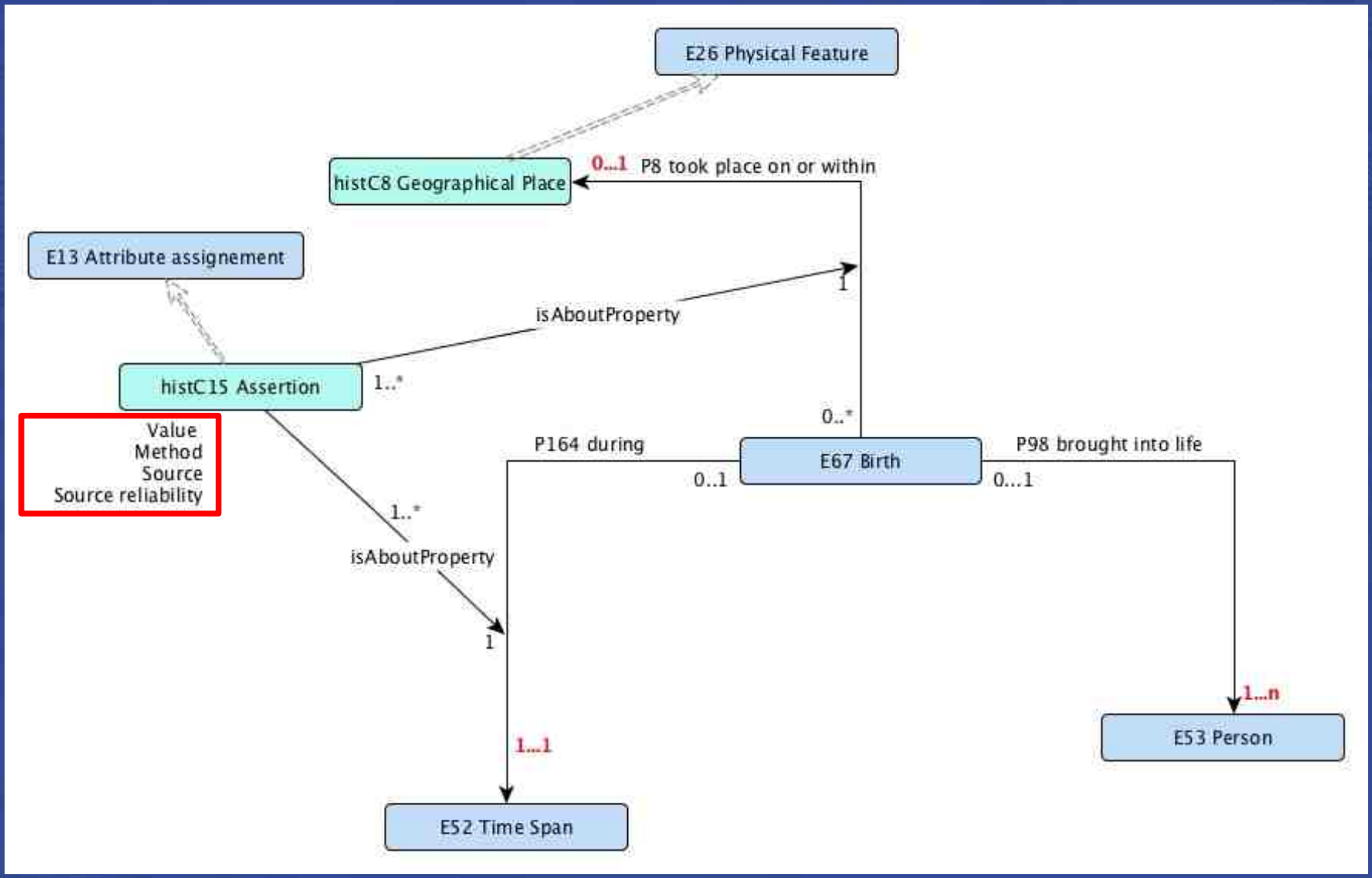


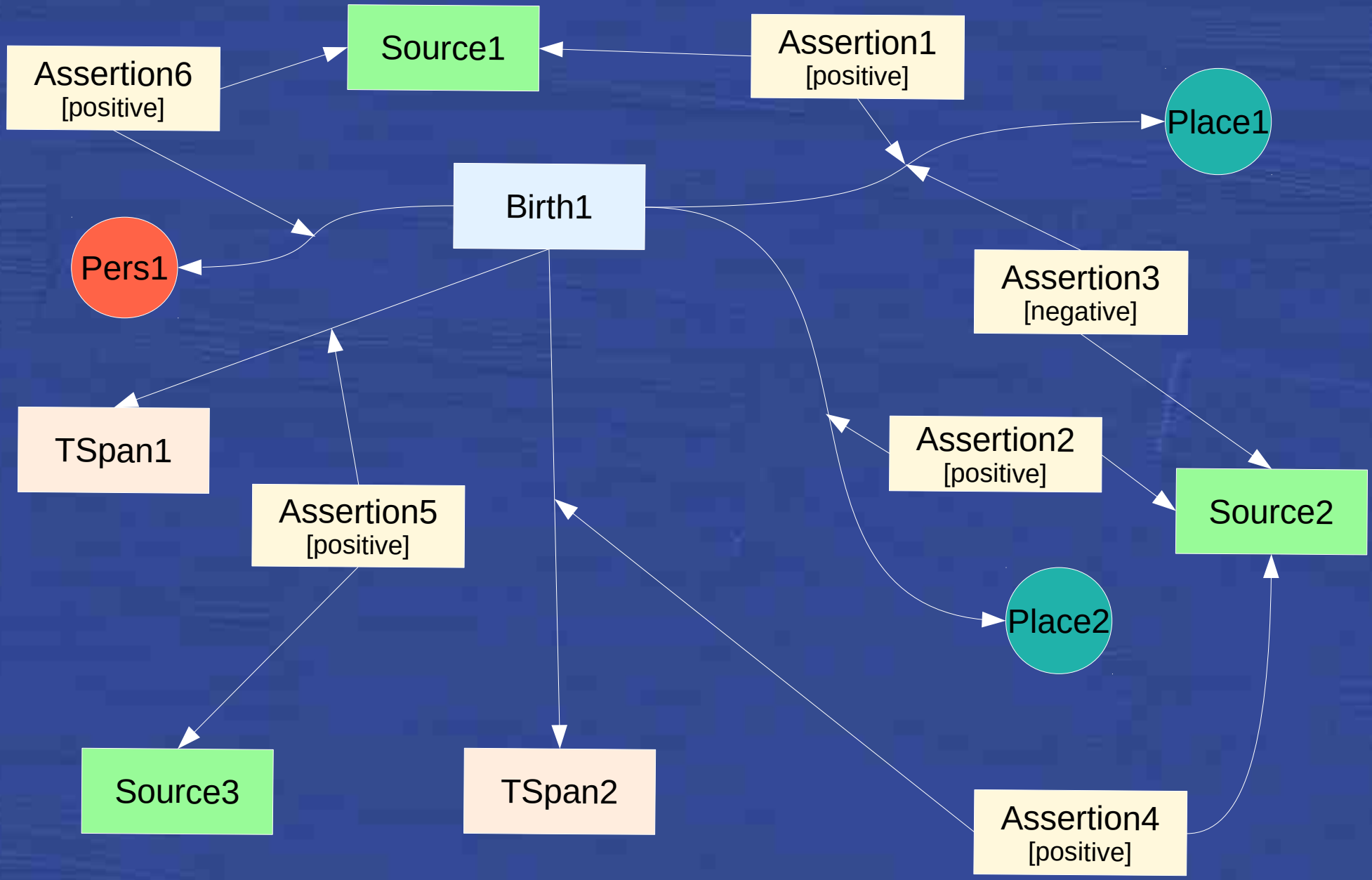


« Quantifiers for properties are provided for the purpose of semantic clarification only, and should **not** be treated as implementation recommendations.

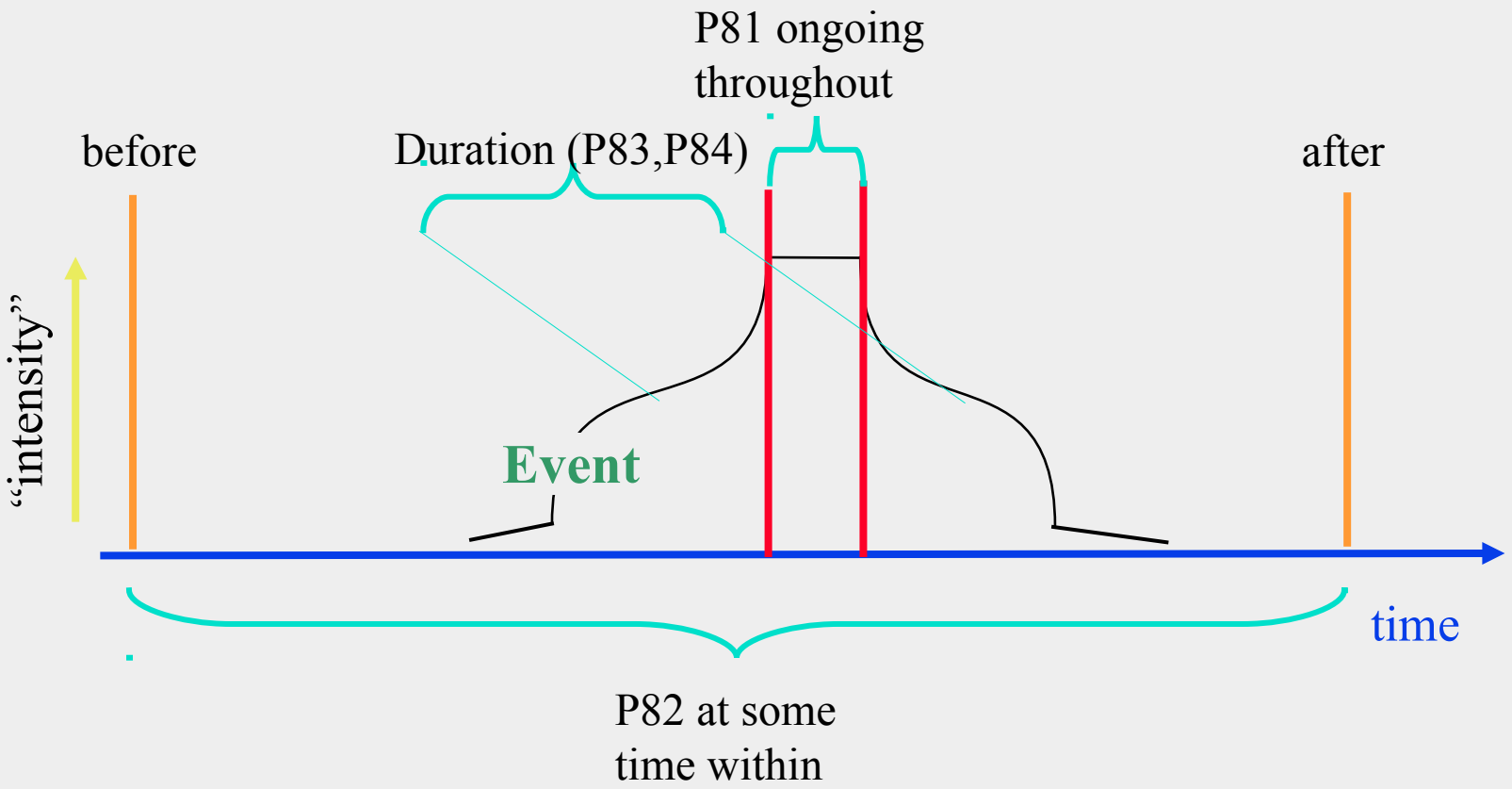
The CRM **has been designed to accommodate alternative opinions** and **incomplete information**, and therefore **all** properties should be implemented as optional and repeatable for their domain and range ("many to many (0,n:0,n)").

Therefore the term "cardinality constraints" is avoided here, as it typically pertains to implementations. » (CRM 6.2)


E26 Physical Feature
histC8 Geographical Place
0...1
P8 took place on or within
E13 Attribute assignement
1
isAboutProperty
histC15 Assertion
1..*
Value
Method
Source
Source reliability
0..*
P164 during
E67 Birth
P98 brought into life
0..1
0...1
1..*
isAboutProperty
1
1...n
E53 Person
1...1
E52 Time Span


Source1
Assertion1
[positive]
Assertion6
[positive]
Place1
Birth1
Pers1
Assertion3
[negative]
TSpan1
Assertion2
[positive]
Source2
Assertion5
[positive]
Place2
Source3
TSpan2
Assertion4
[positive]

The CIDOC CRM
Time Uncertainty, Certainty and Duration



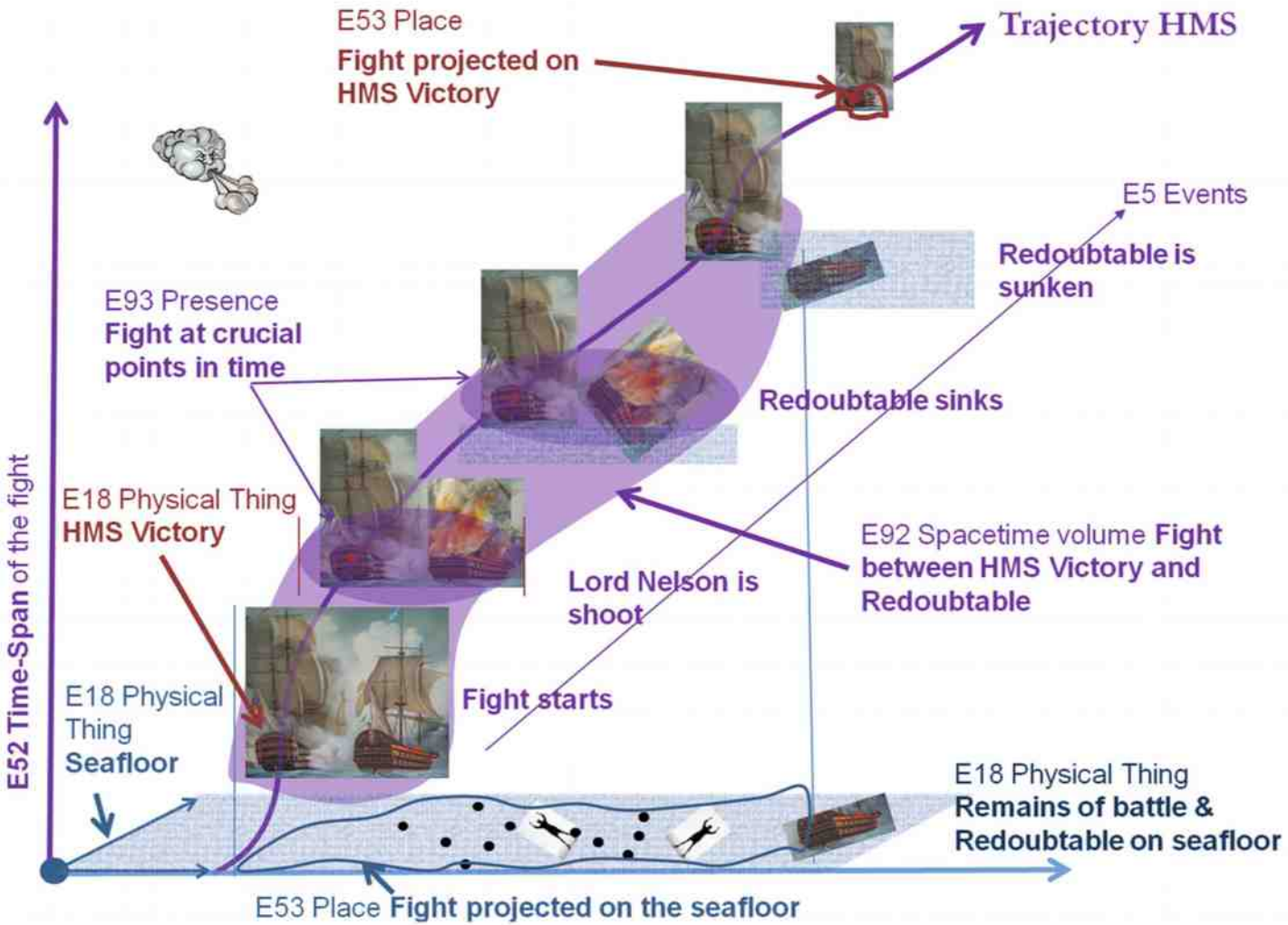


Fig. 2 The fight of the HMS Victory and the Redoubtable in the Battle of Trafalgar illustrating Spacetime volume (E92) and Presence (E93) and their projection to different Places (E53)

DOI 10.1007/s00799-016-0192-4

CRMgeo: A spatiotemporal extension of CIDOC-CRM Gerald Hiebel1 · Martin Doerr2 · Øyvind Eide3

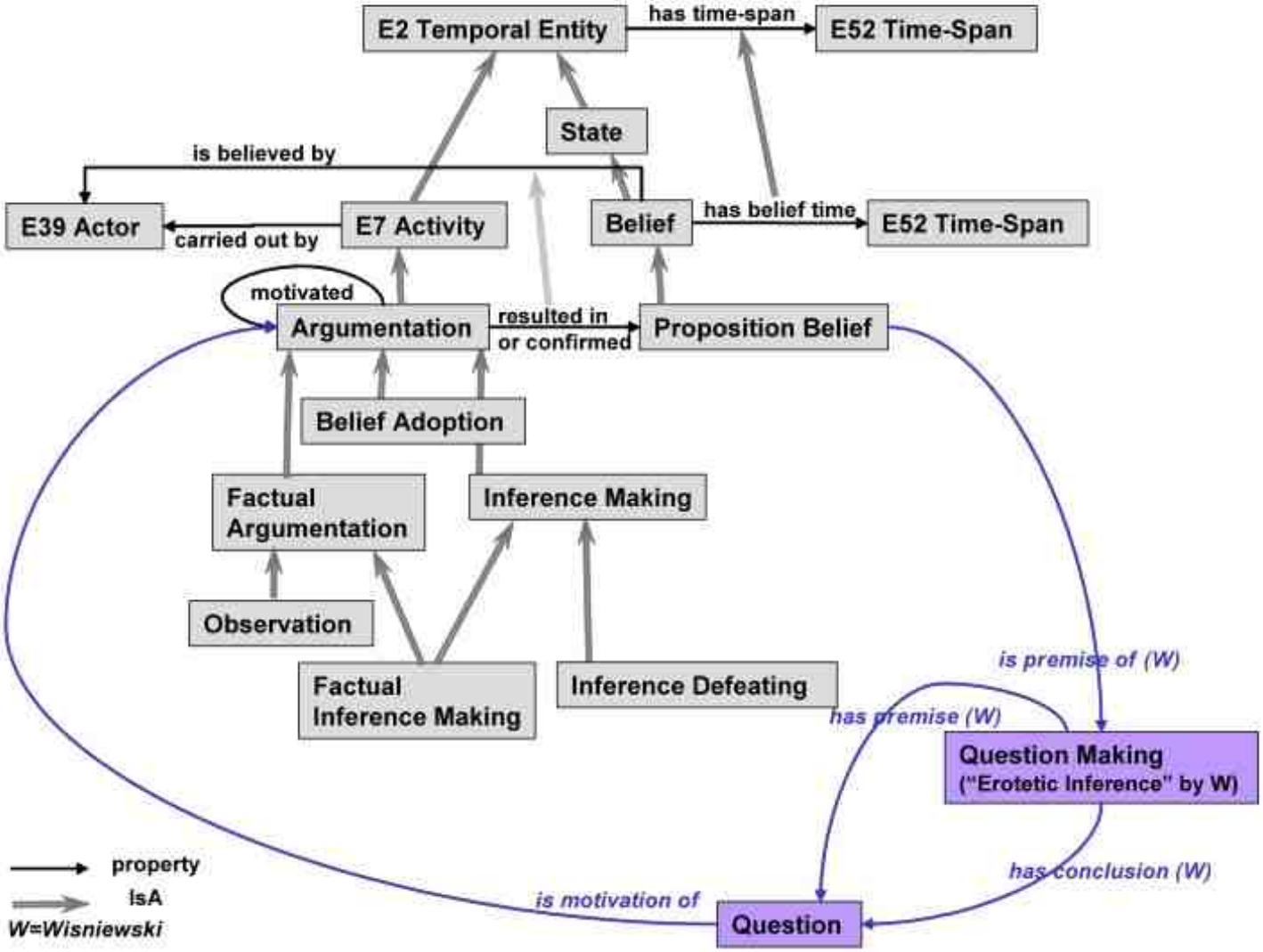
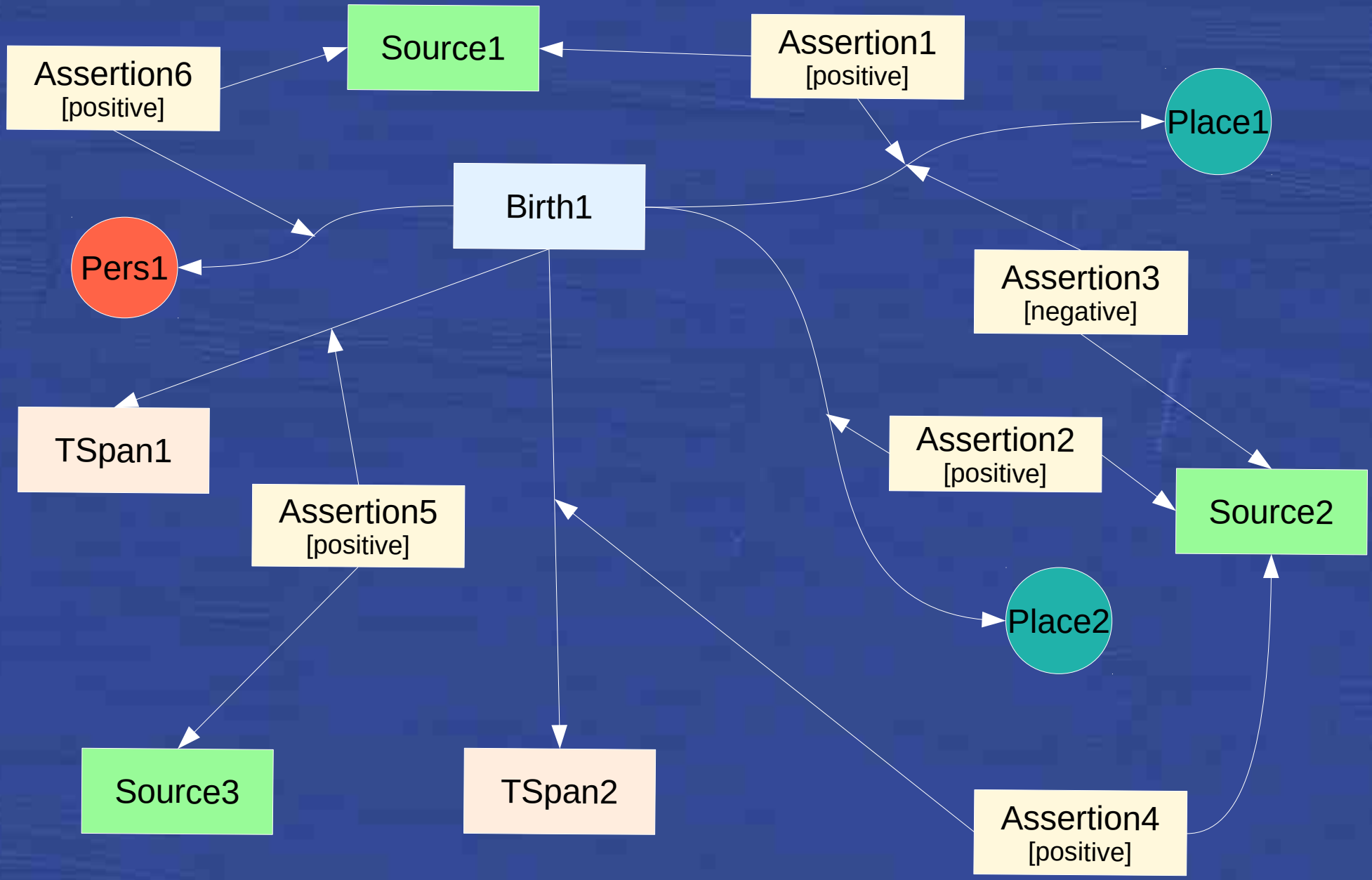
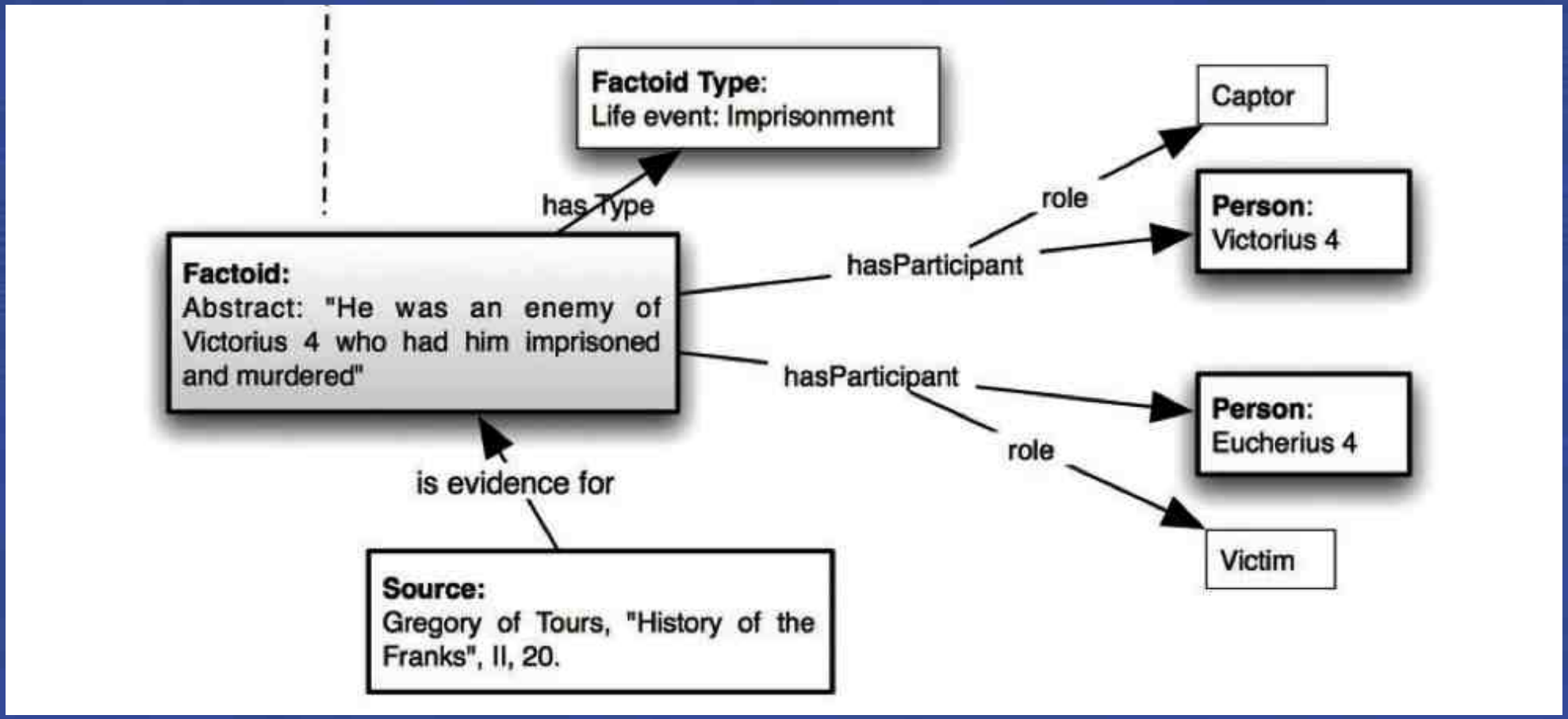


Figure 5. Types of argumentation: Factual Argumentation, Inference making, and Belief Adoption.

CRMInf

Martin DOERR, Athina KRITSOTAKI and Katerina BOUTSIKA,
Factual Argumentation – A Core Model for Assertions Making, Journal on Computing and Cultural Heritage
Volume 3 Issue 3, March 2011 Article No. 8


Source1
Assertion1
[positive]
Assertion6
[positive]
Place1
Birth1
Pers1
Assertion3
[negative]
TSpan1
Assertion2
[positive]
Source2
Assertion5
[positive]
Place2
Source3
TSpan2
Assertion4
[positive]



Factoïd ontology

KCL's Department of Digital Humanities (DDH : Prosopographies of the Byzantine World (PBE and PBW), Anglo-Saxon England (PASE), Medieval Scotland (PoMS), Anglo-Scottish cross-border society ('Breaking of Britain': BoB)

Michele Pasin and John Bradley, Factoid-based prosopography and computer ontologies: Towards an integrated approach, Literary and Linguistic Computing Advance Access published June 29, 2013

M. Pasin and J. Bradley

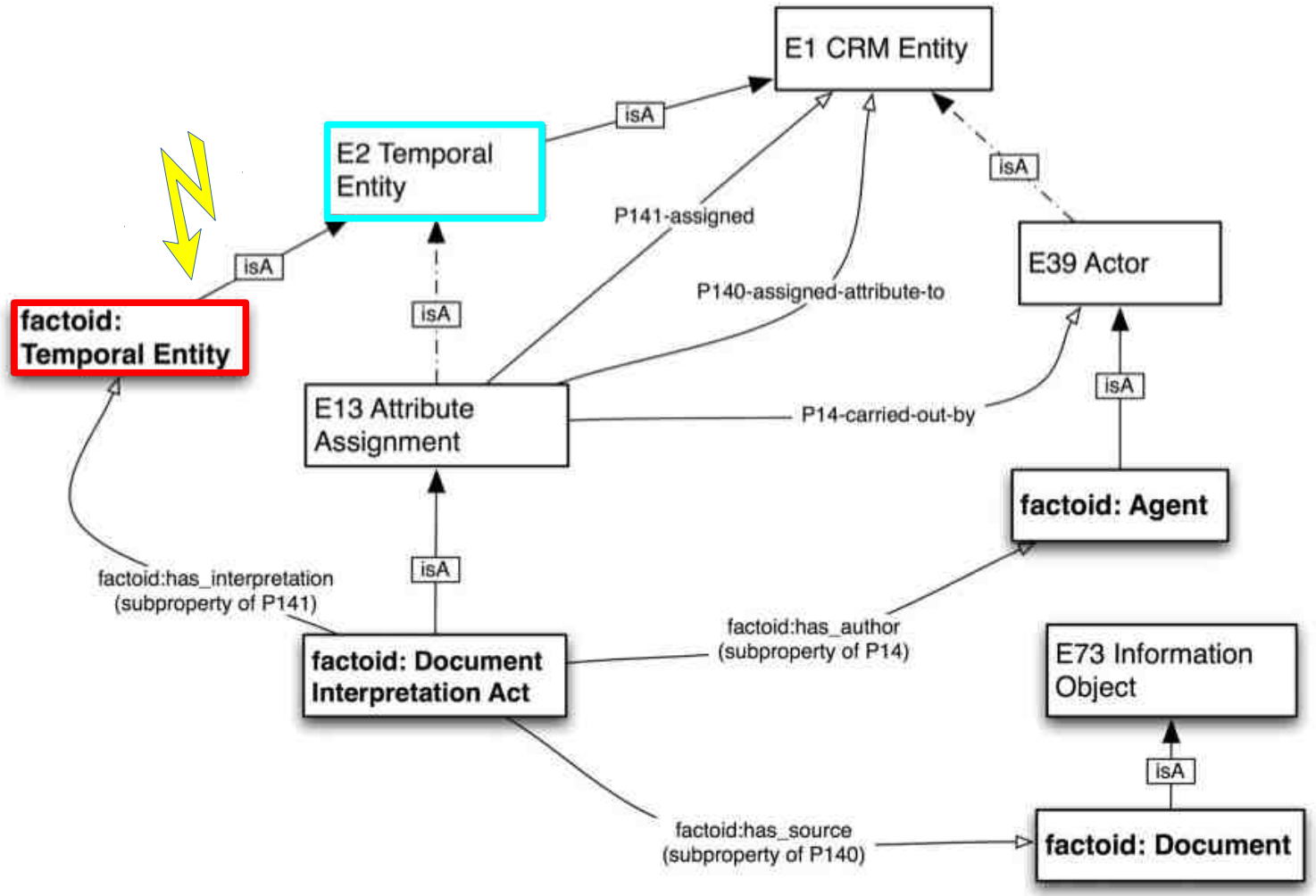


Fig. 4 Mapping the factoid model to CIDOC-CRM primitives

« The factoid approach prioritizes the sources, rather than our historians' reading of them » (Bradley / Pasin).

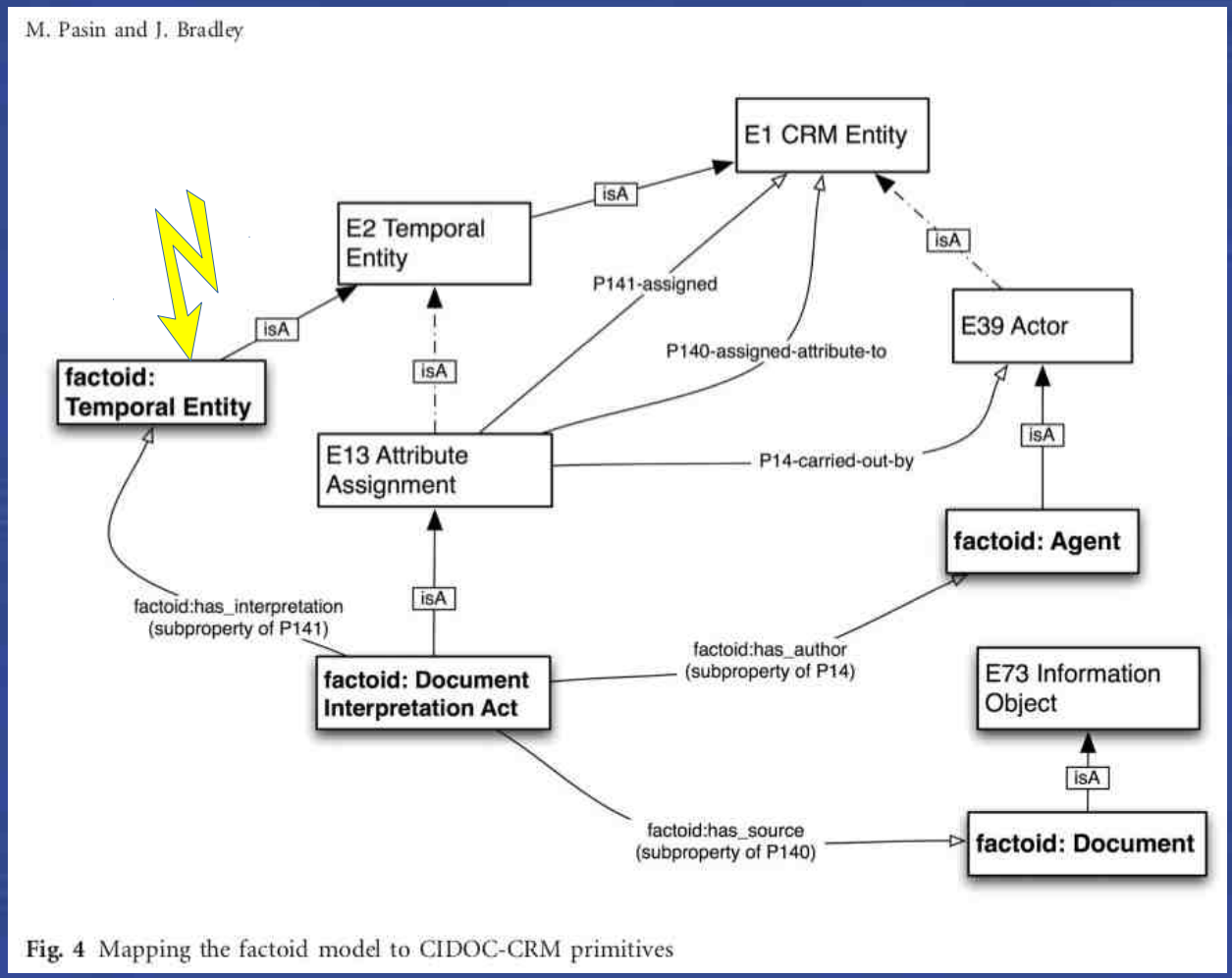
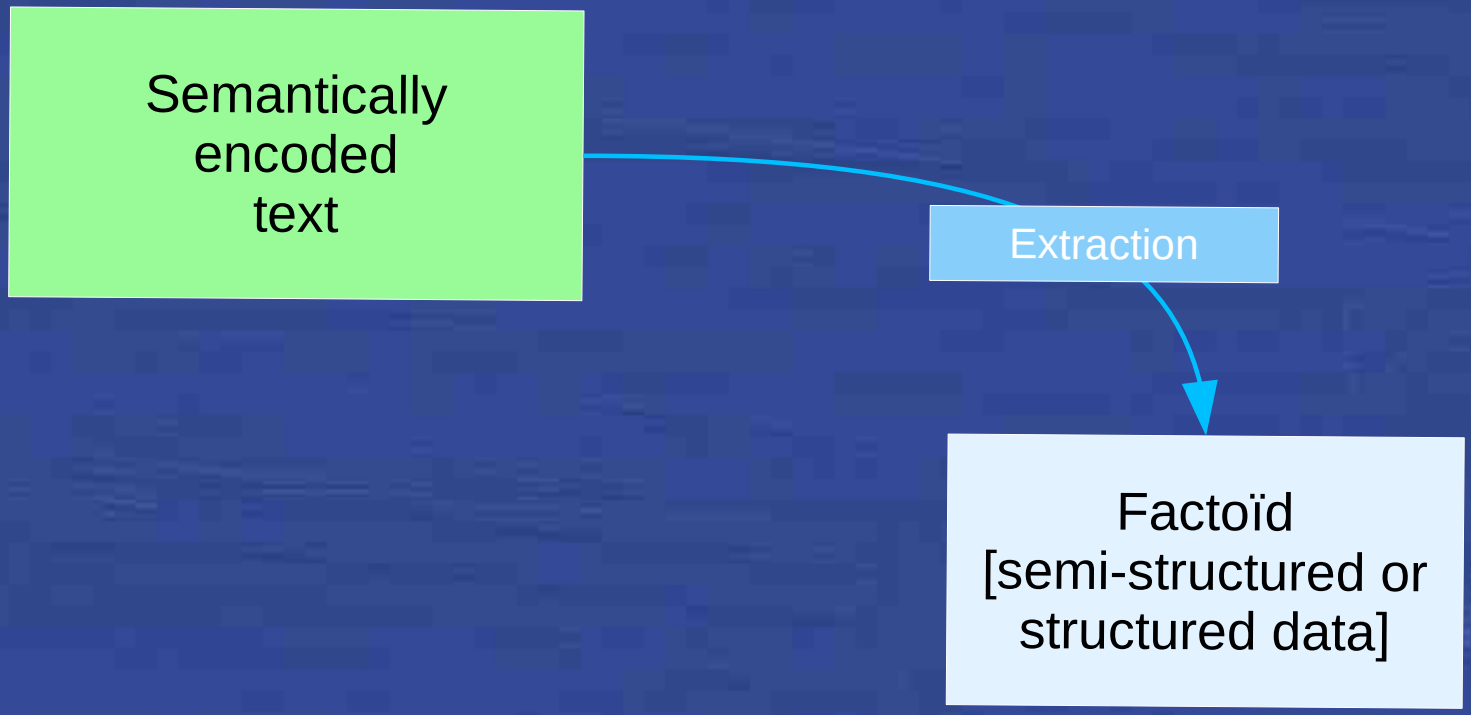
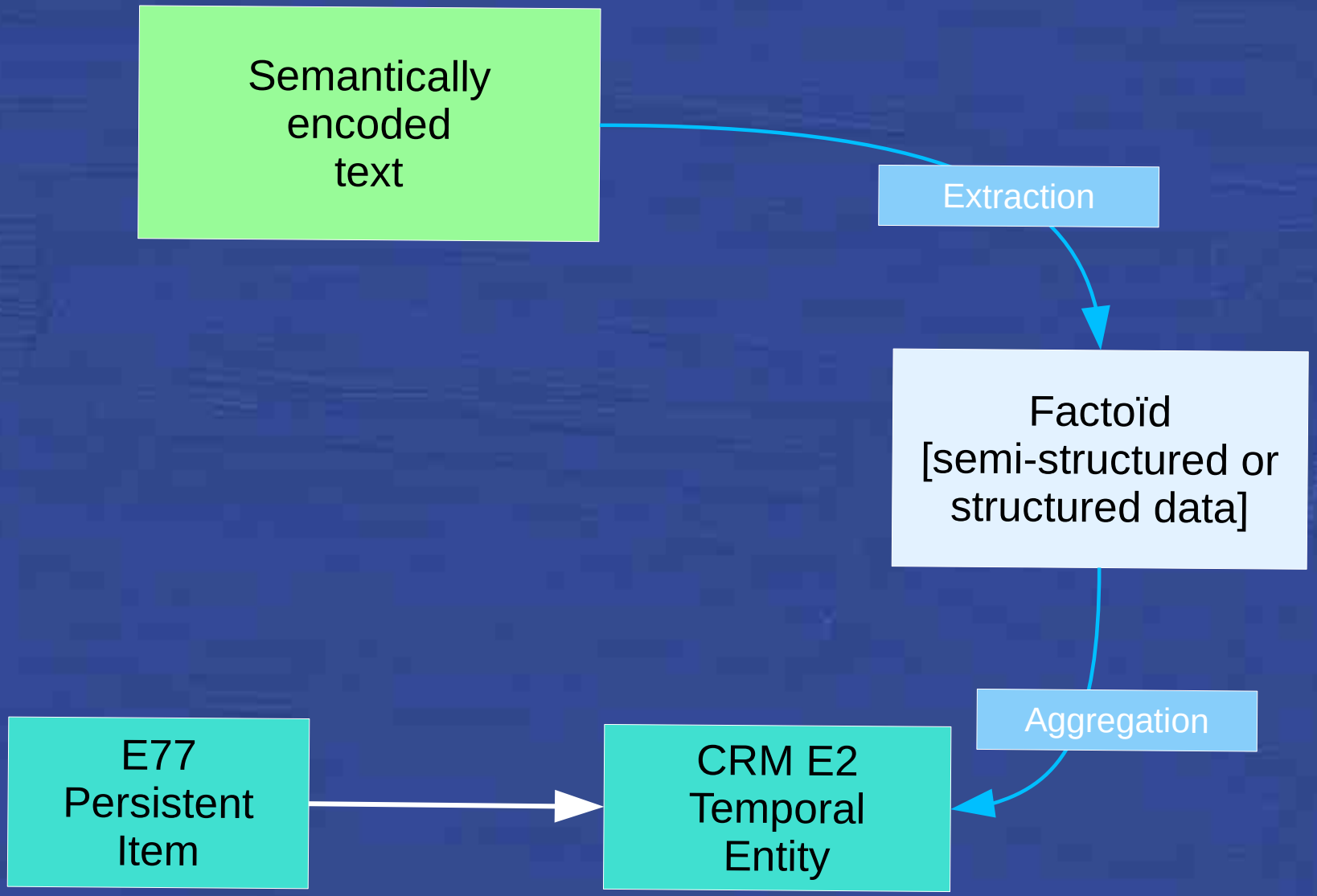
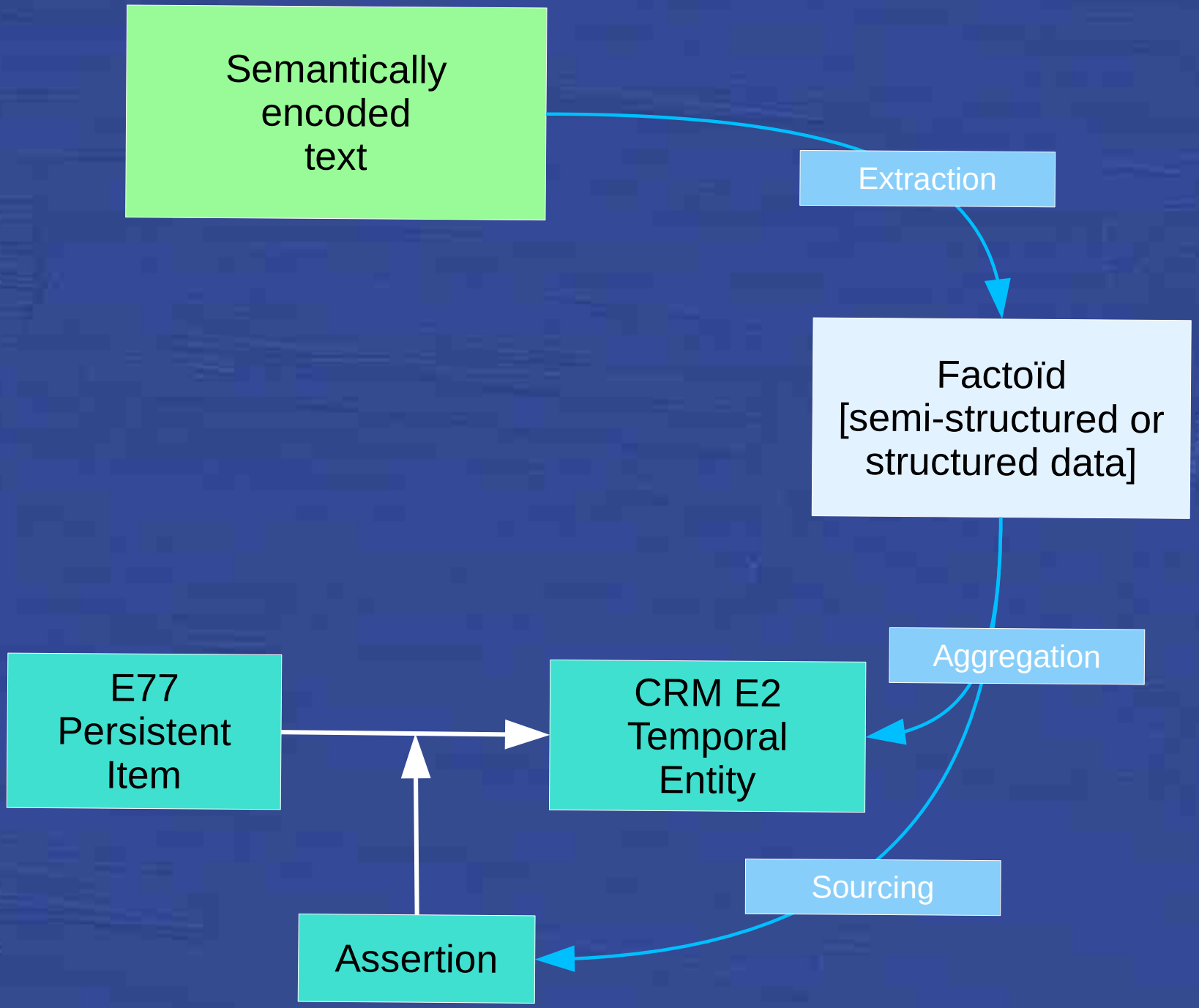


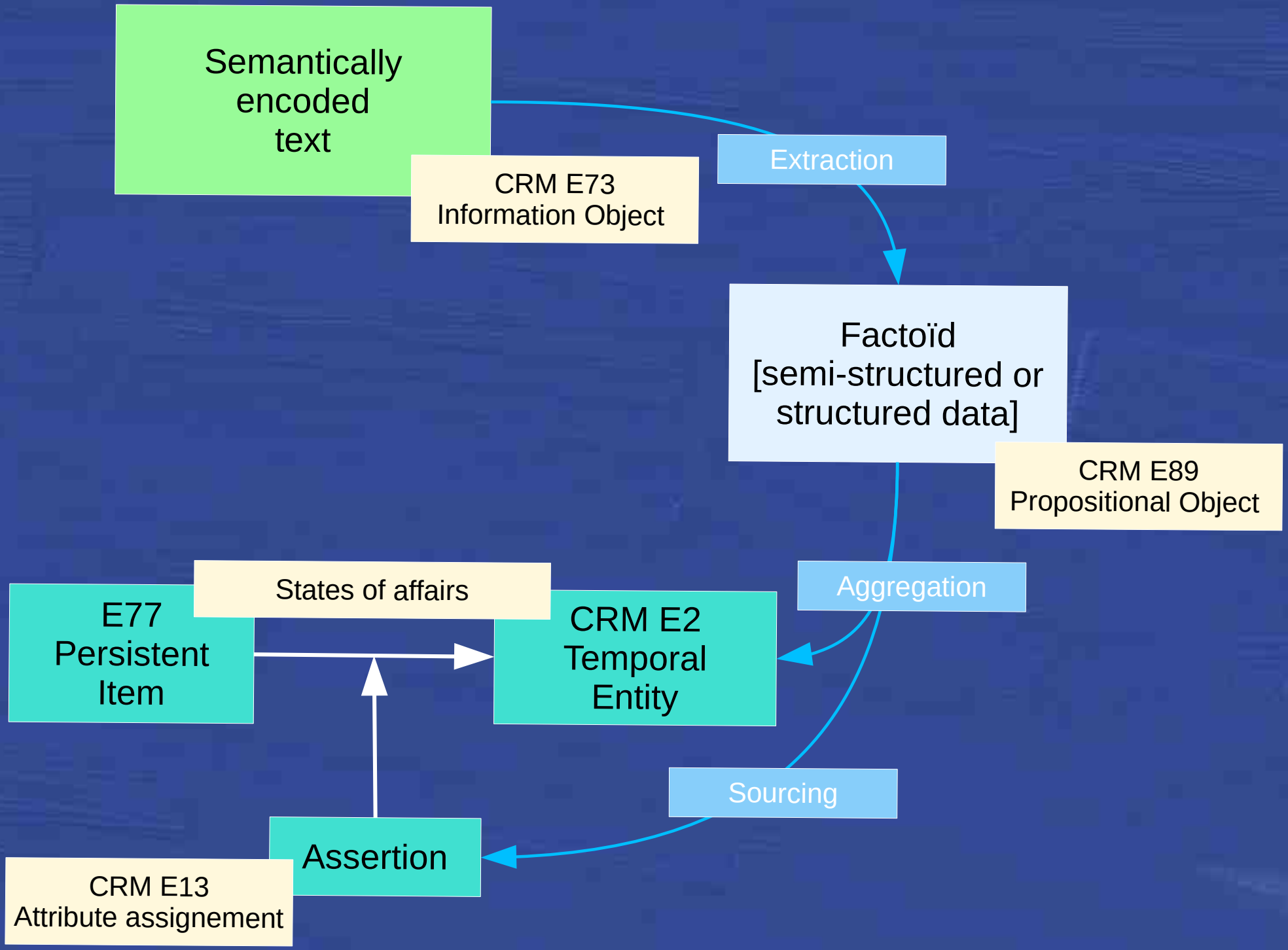
Fig. 4 Mapping the factoid model to CIDOC-CRM primitives

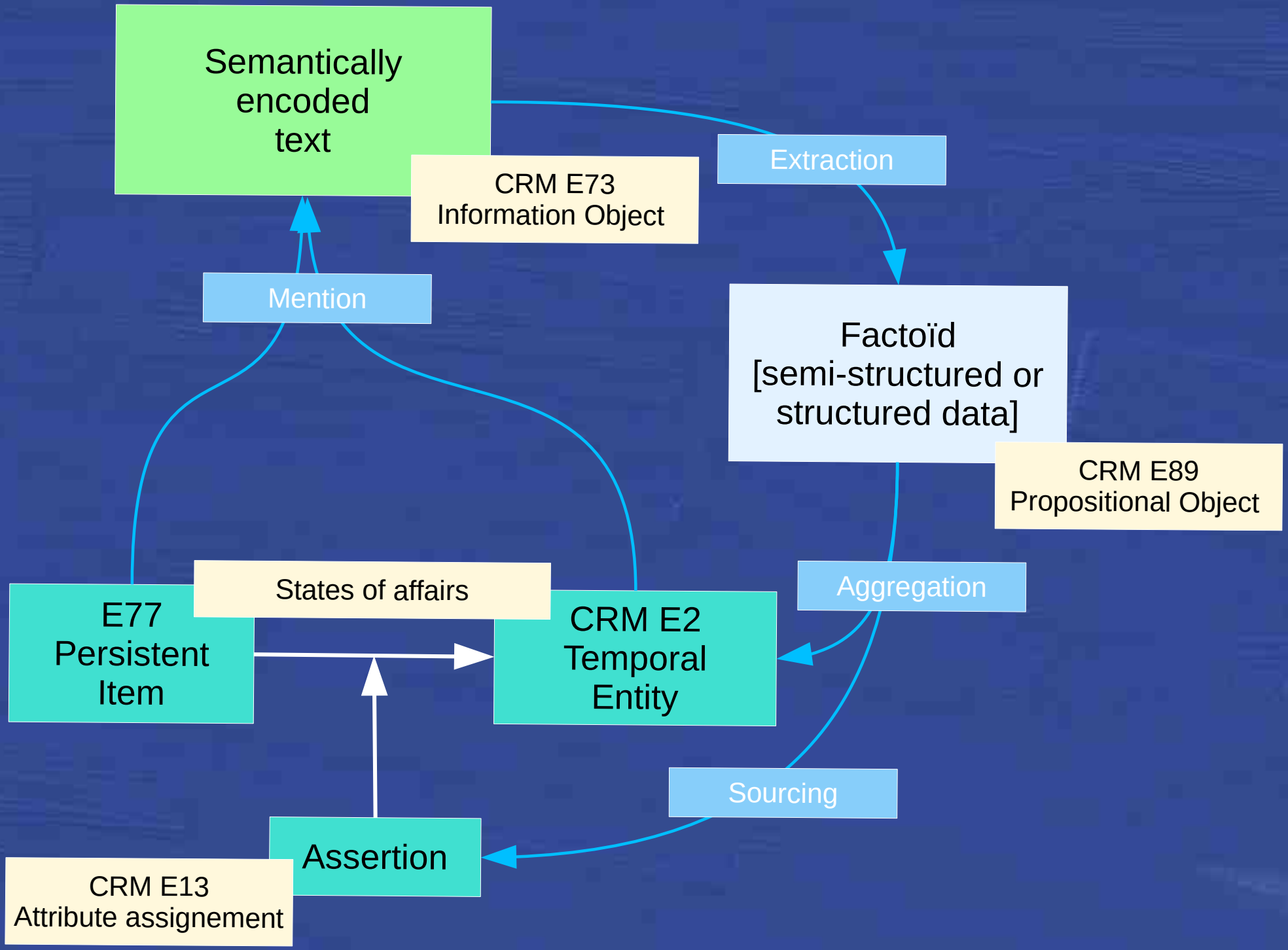
The essential issue is to distinguish what we want to model:
the *content of the source* as is,
or
states of affairs, as the historian reconstructs them using
the methods of source criticism.


Semantically encoded text
Extraction
Factoïd [semi-structured or structured data]

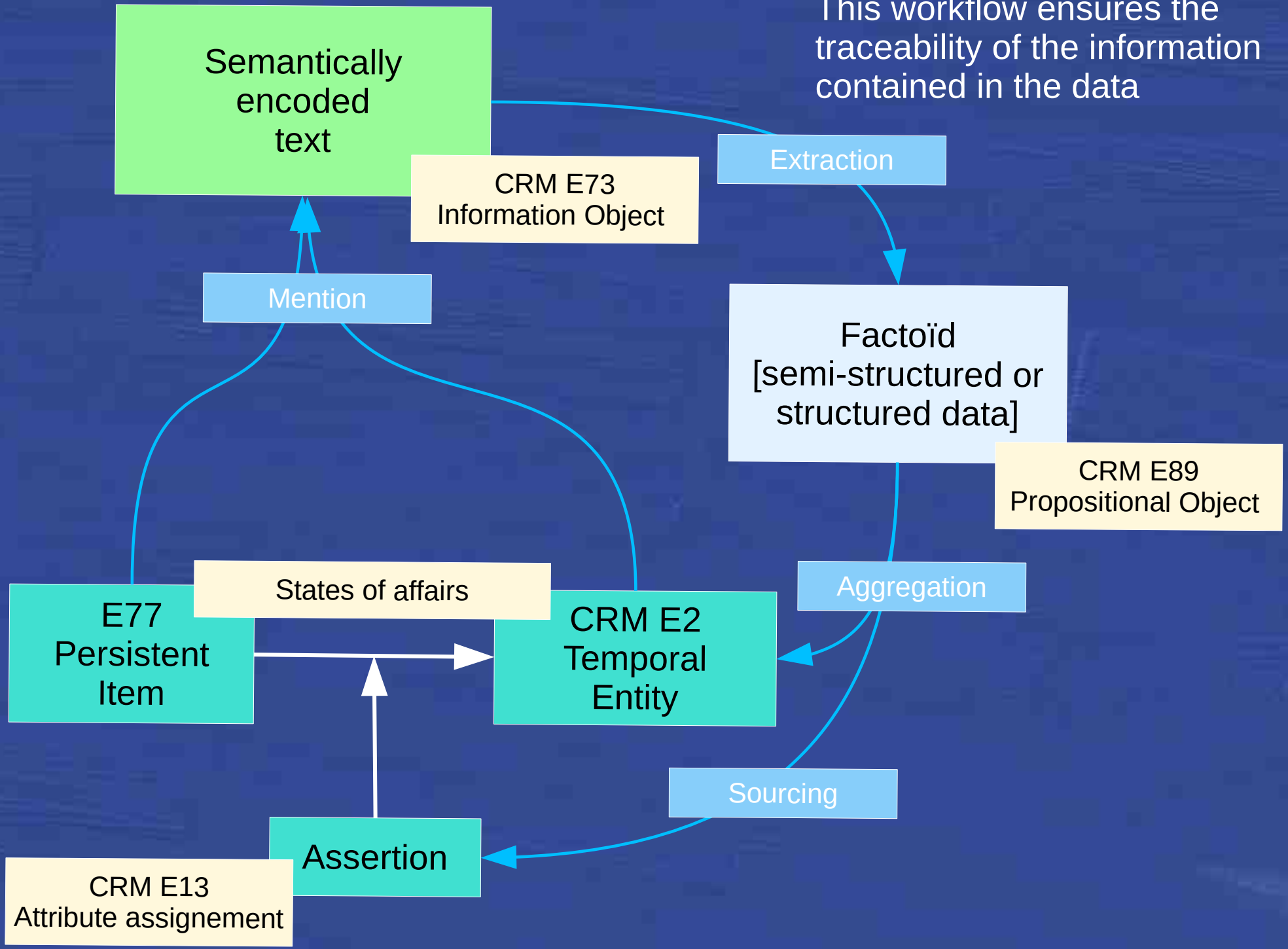

Semantically encoded text
Extraction
Factoïd [semi-structured or structured data]
Aggregation
E77 Persistent Item
CRM E2 Temporal Entity

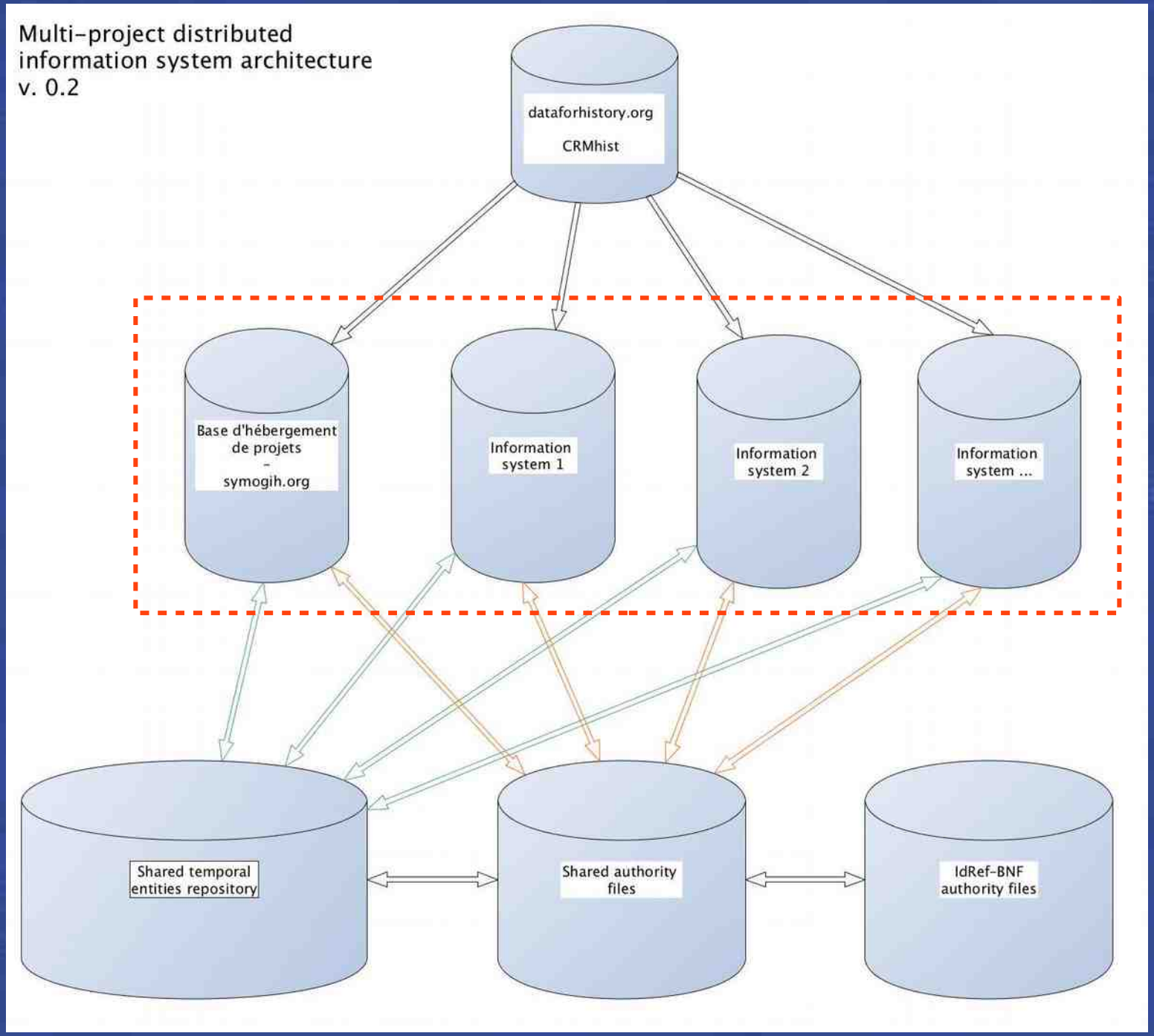

Semantically encoded text
Extraction
Factoïd [semi-structured or structured data]
Aggregation
E77 Persistent Item
CRM E2 Temporal Entity
Sourcing
Assertion


Semantically encoded text
CRM E73 Information Object
Extraction
Factoïd [semi-structured or structured data]
CRM E89 Propositional Object
Aggregation
States of affairs
E77 Persistent Item
CRM E2 Temporal Entity
Sourcing
Assertion
CRM E13 Attribute assignement

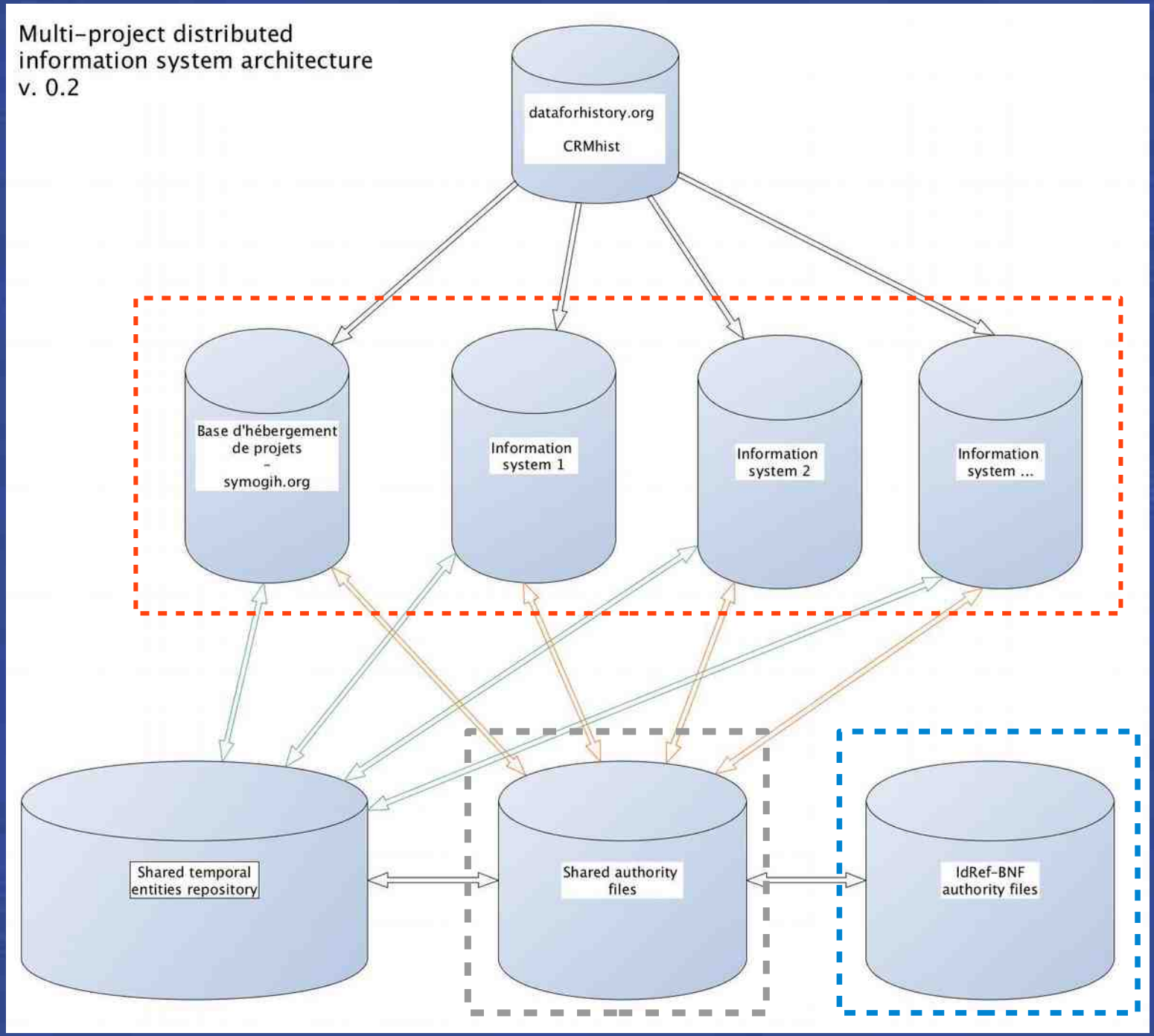

Semantically encoded text
CRM E73 Information Object
Extraction
Mention
Factoïd [semi-structured or structured data]
CRM E89 Propositional Object
States of affairs
E77 Persistent Item
CRM E2 Temporal Entity
Aggregation
Sourcing
Assertion
CRM E13 Attribute assignement

This workflow ensures the traceability of the information contained in the data



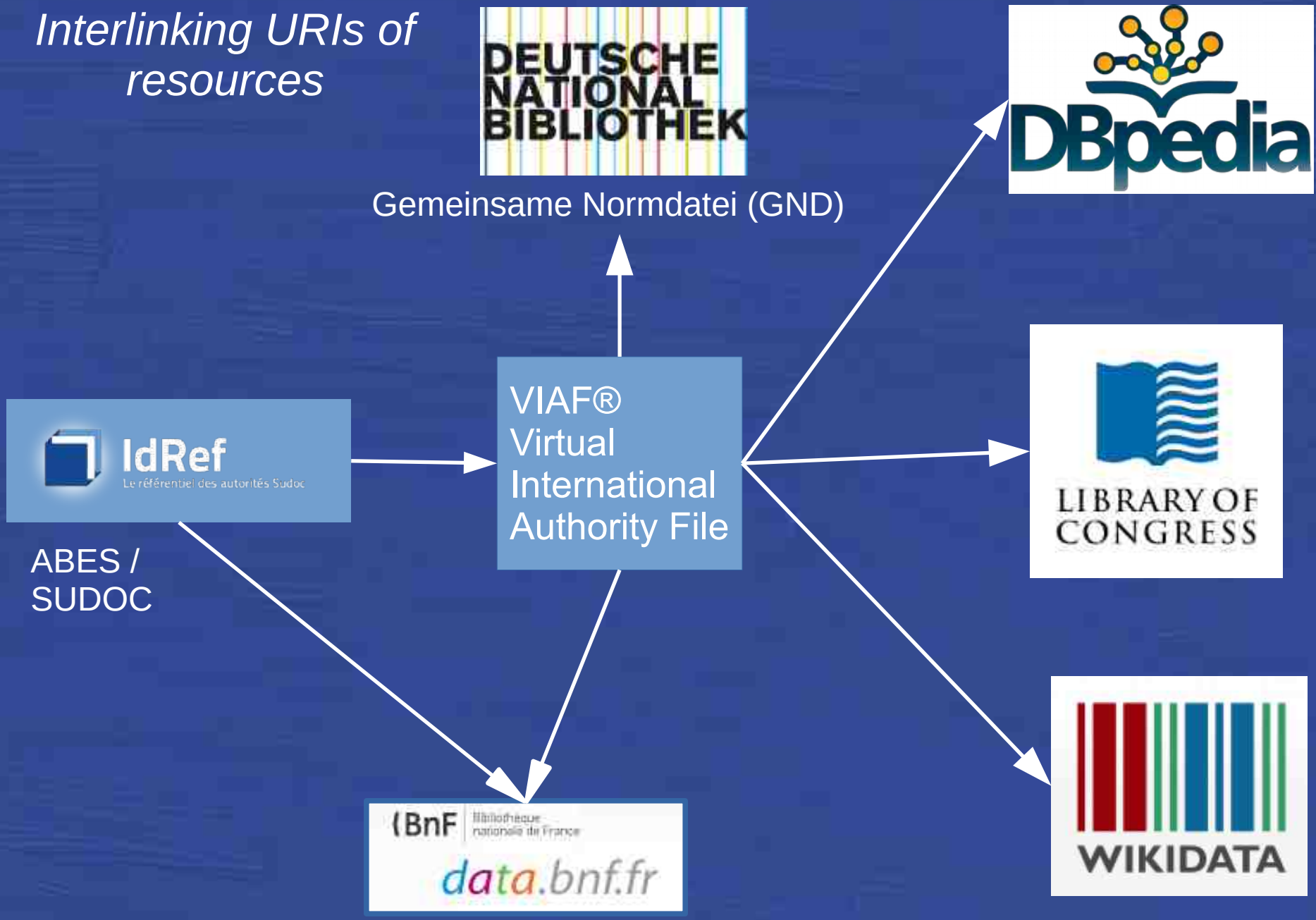


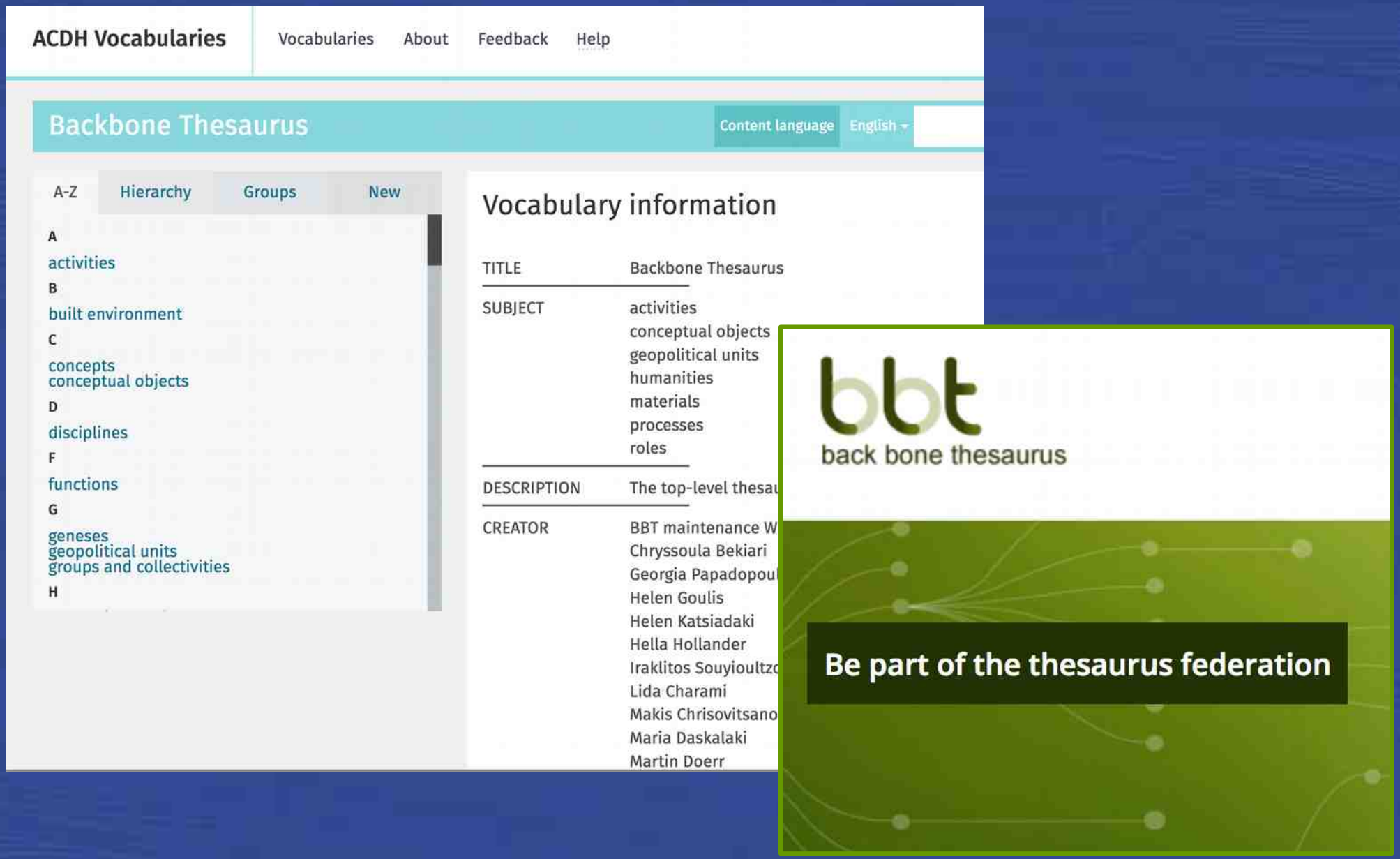
dataforhistory.org



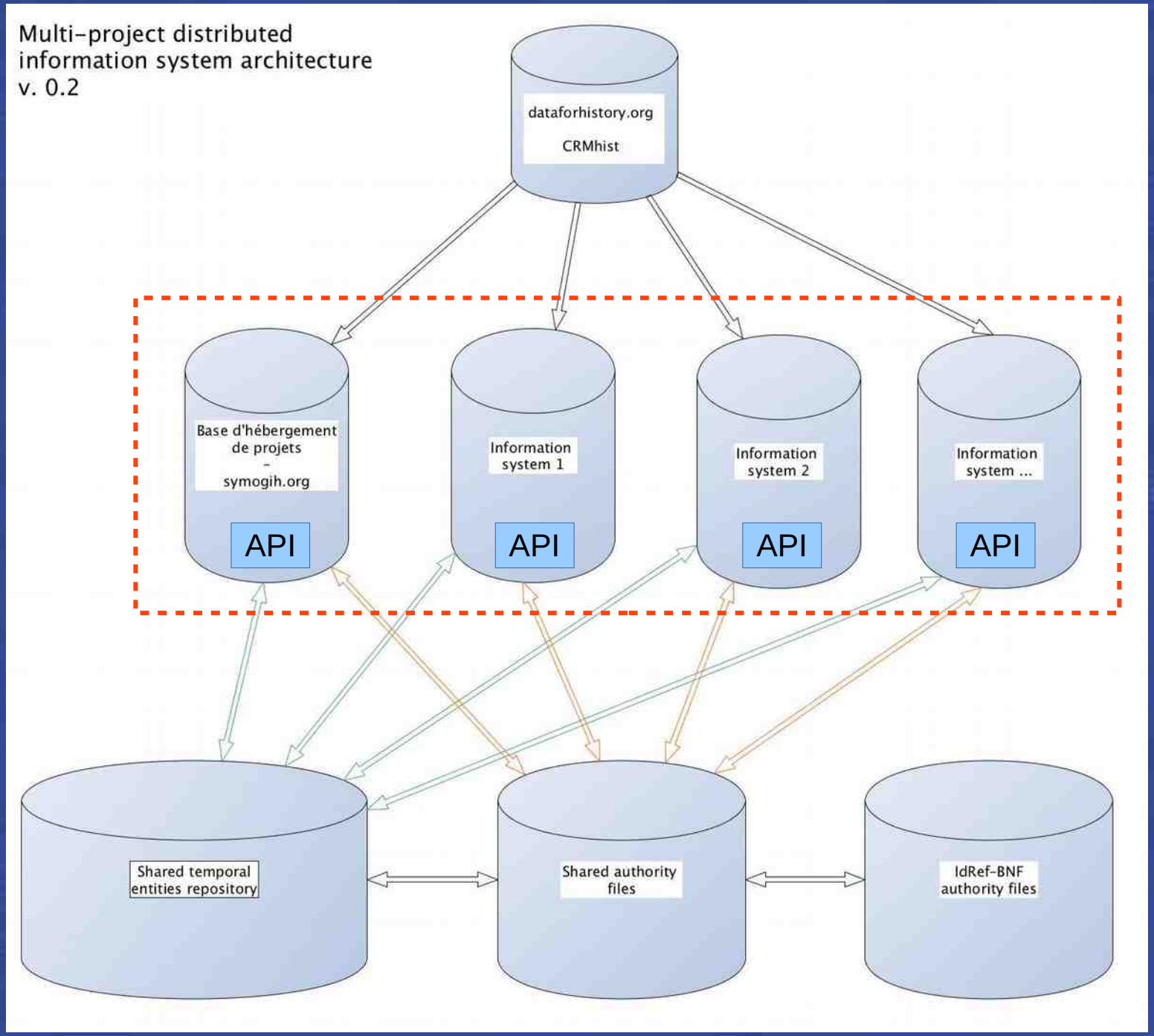
dataforhistory.org

Interlinking URIs of resources

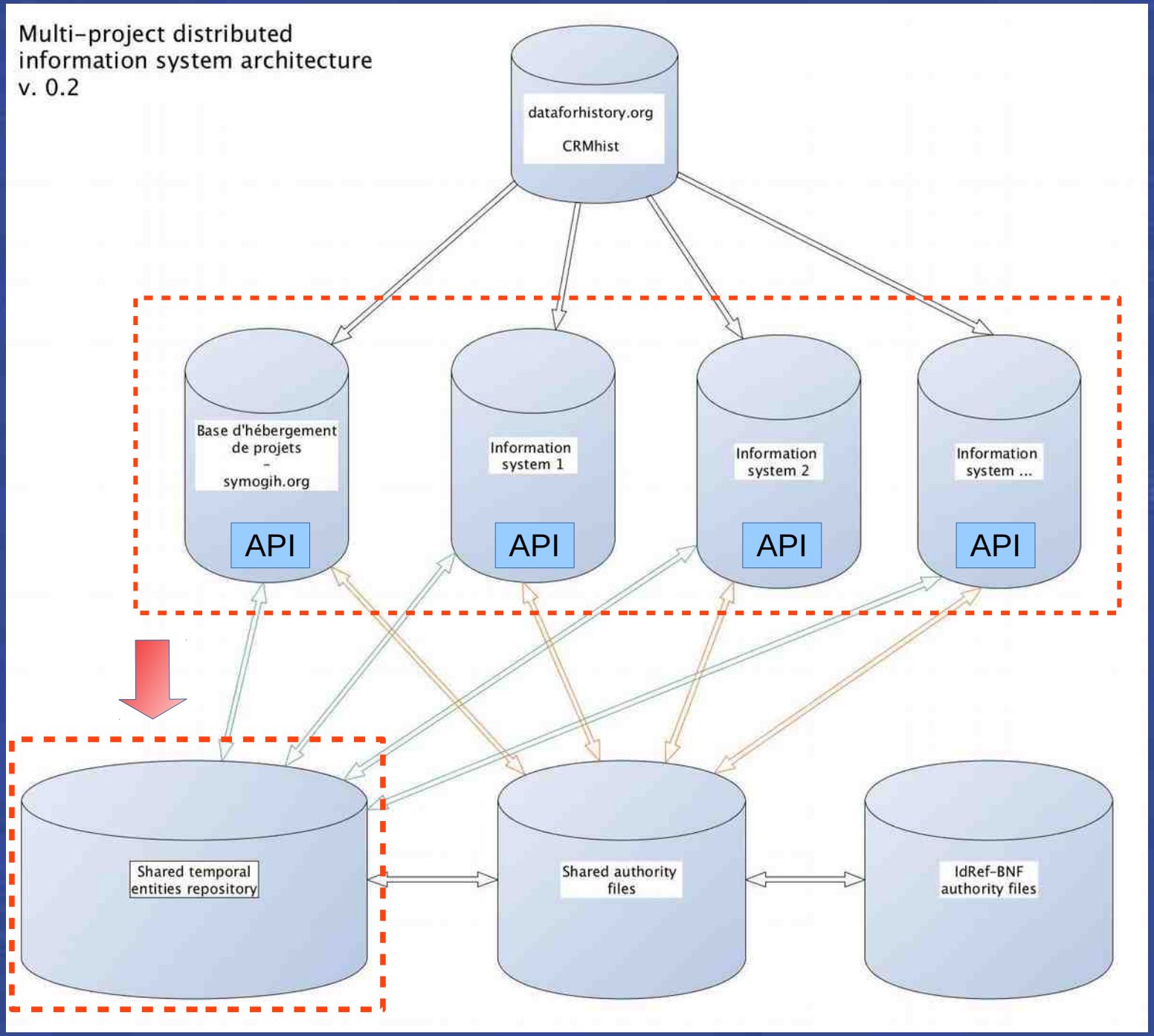




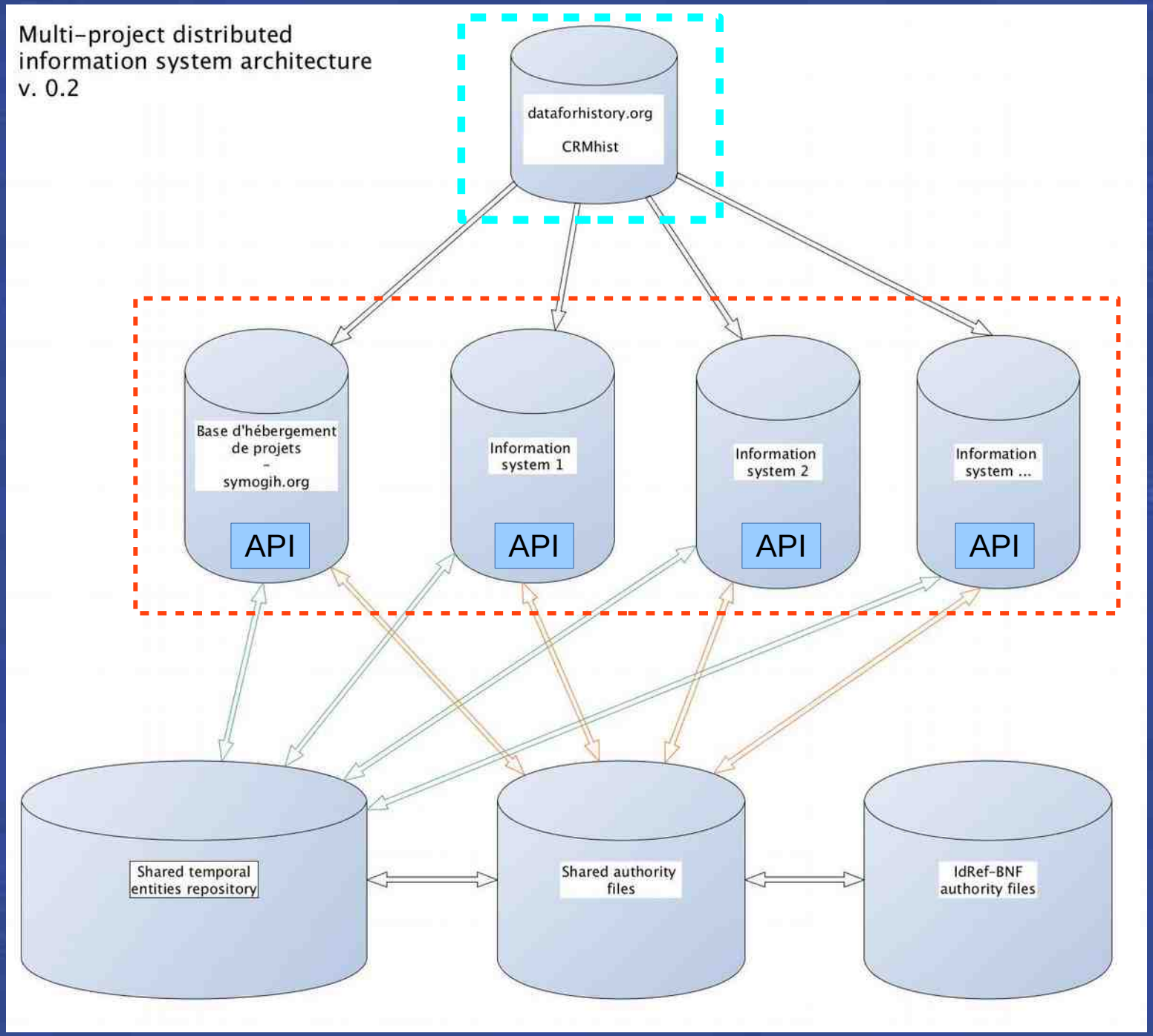
backbonethesaurus.eu



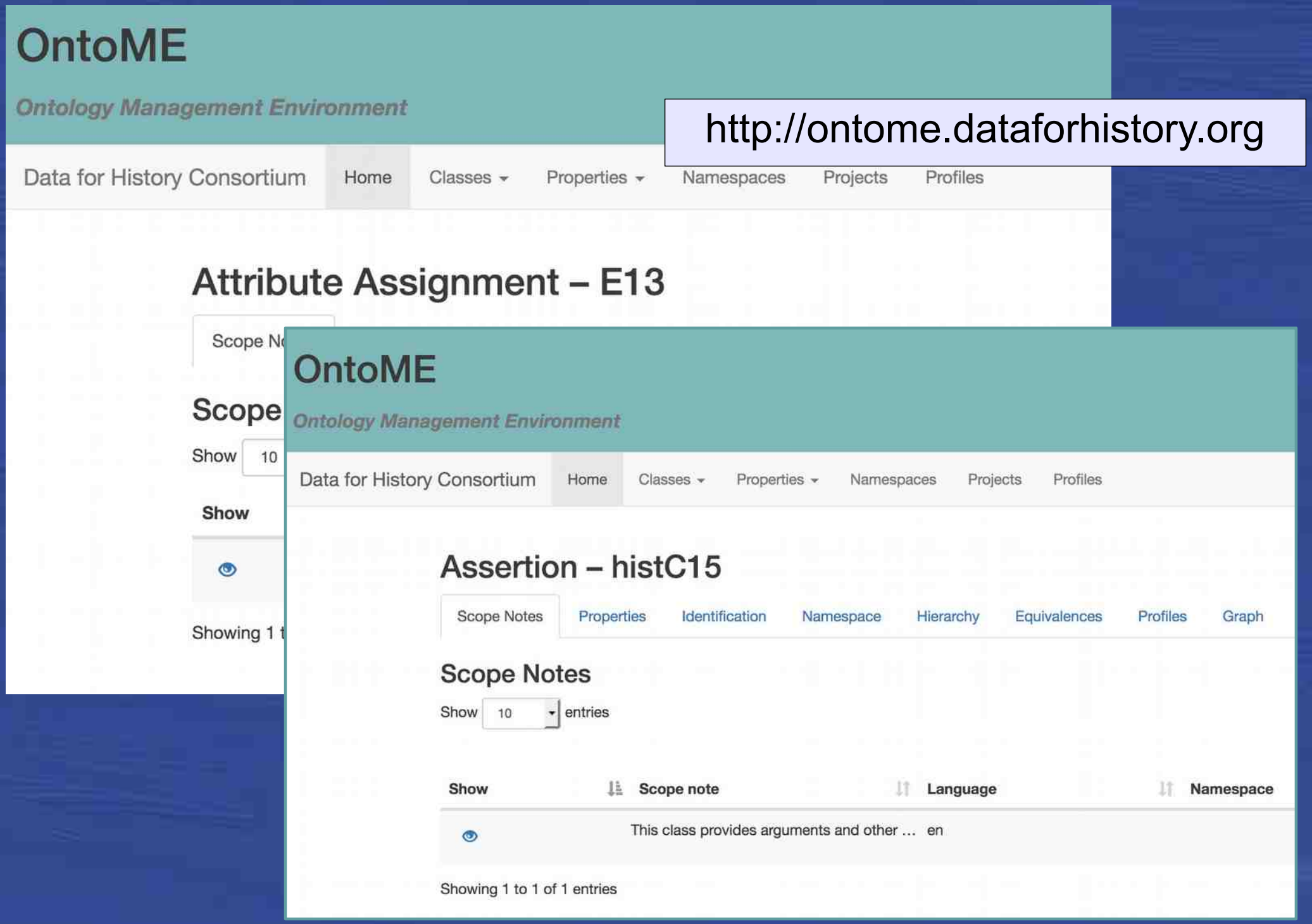
dataforhistory.org

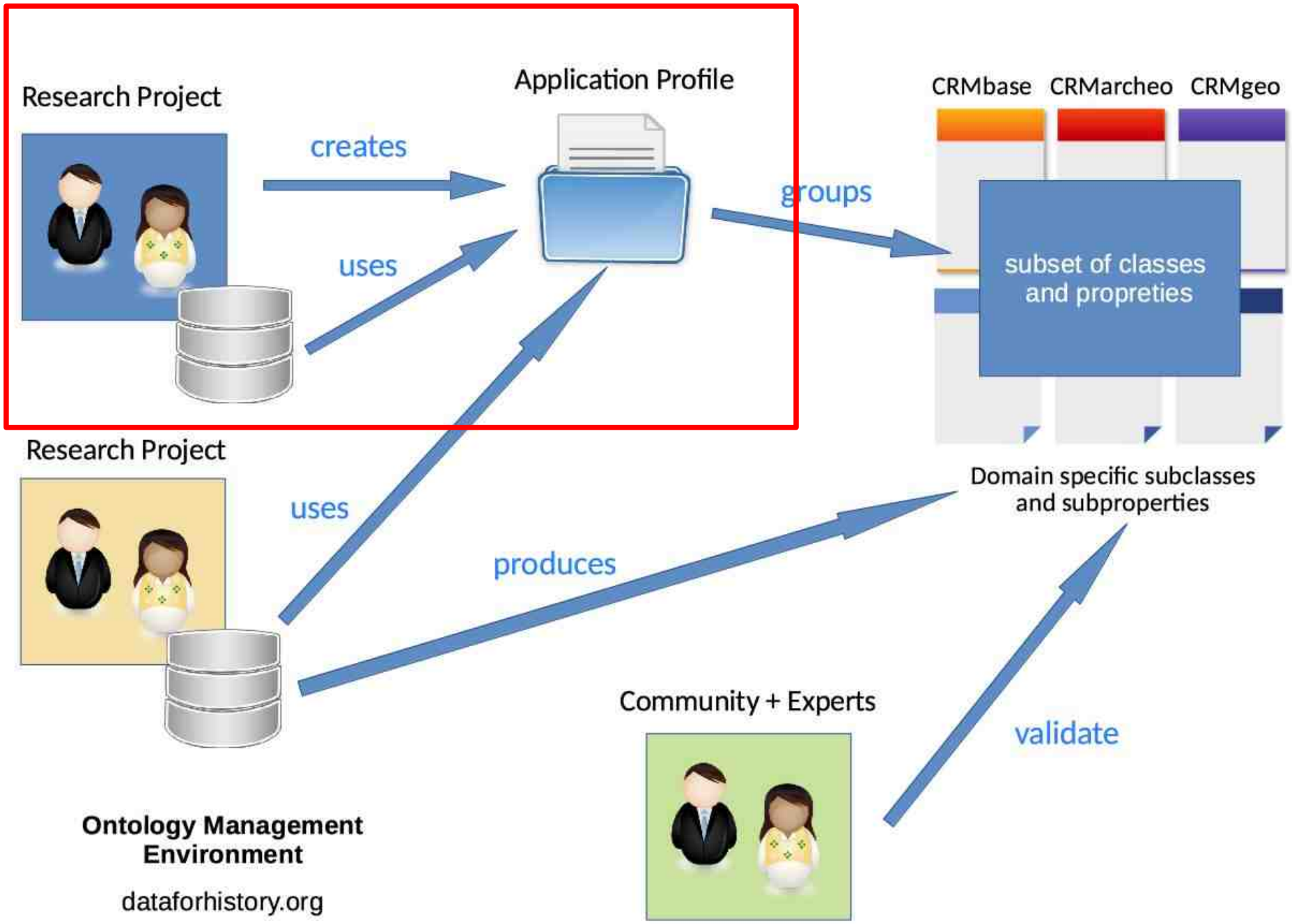


dataforhistory.org



dataforhistory.org

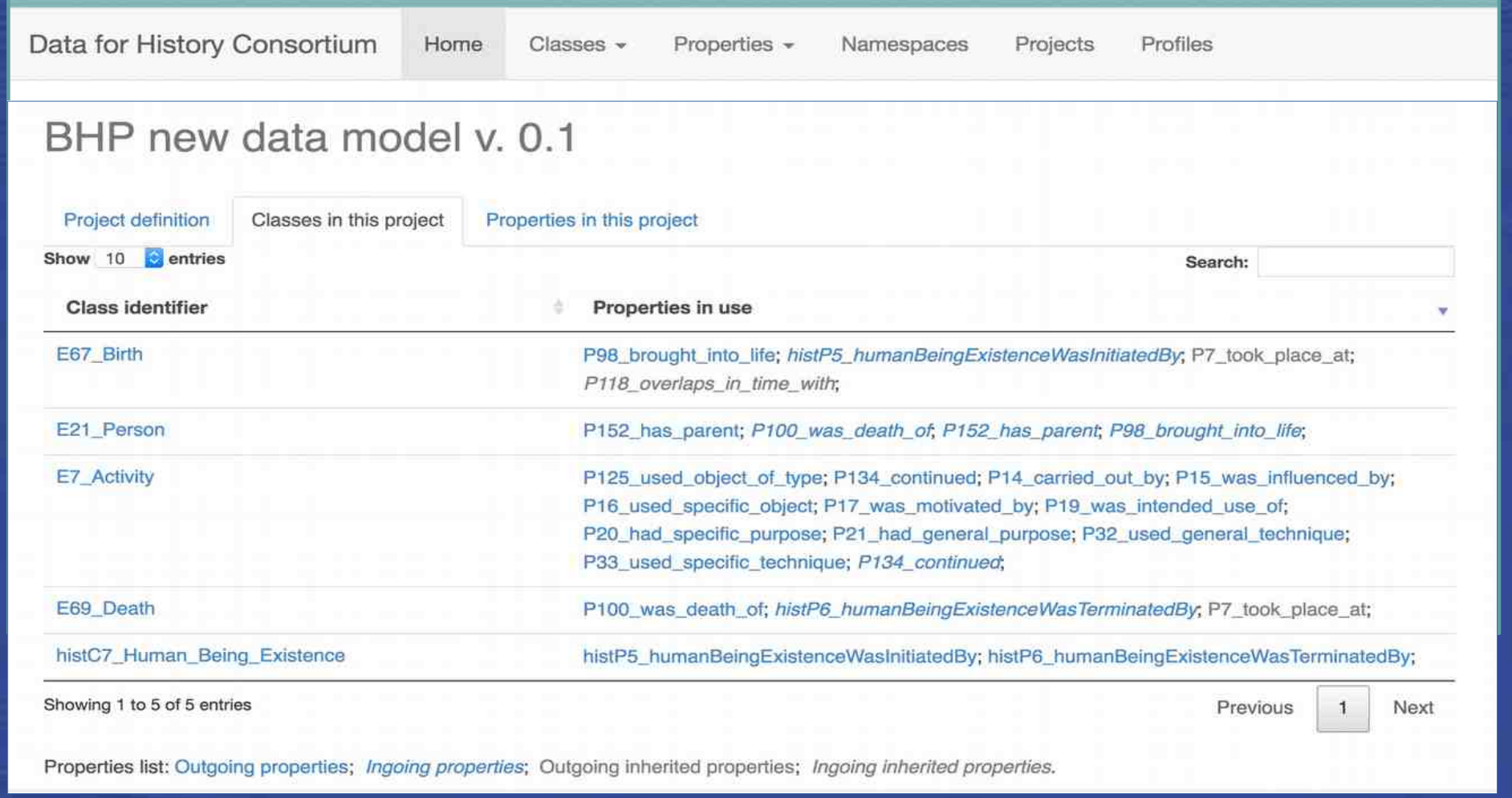

OntoME
Ontology Management Environment
http://ontome.dataforhistory.org
Data for History Consortium
Home
Classes
Properties
Namespaces
Projects
Profiles
Attribute Assignment – E13
Scope N
Scope
Show 10
Show
Showing 1 t
OntoME
Ontology Management Environment
Data for History Consortium
Home
Classes
Properties
Namespaces
Projects
Profiles
Assertion – histC15
Scope Notes
Properties
Identification
Namespace
Hierarchy
Equivalences
Profiles
Graph
Scope Notes
Show 10 entries
Show
Scope note
Language
Namespace
This class provides arguments and other … en
Showing 1 to 1 of 1 entries


Research Project
creates
Application Profile
uses
groups
CRMbase
CRMarcheo
CRMgeo
subset of classes
and propreties
Research Project
uses
produces
Domain specific subclasses
and subproperties
Community + Experts
validate
Ontology Management
Environment
dataforhistory.org

OntoME

Ontology Management Environment

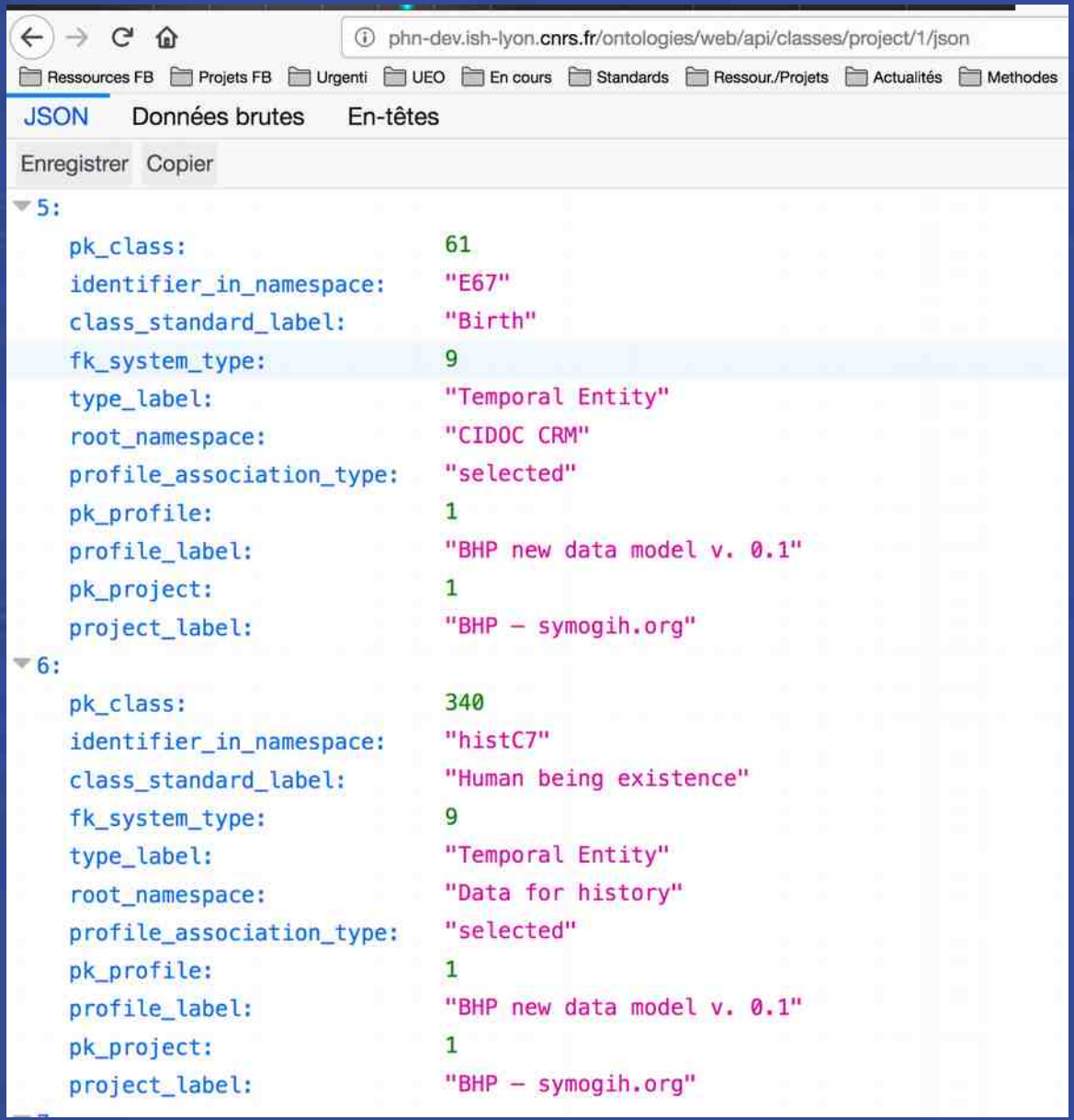
http://ontome.dataforhistory.org

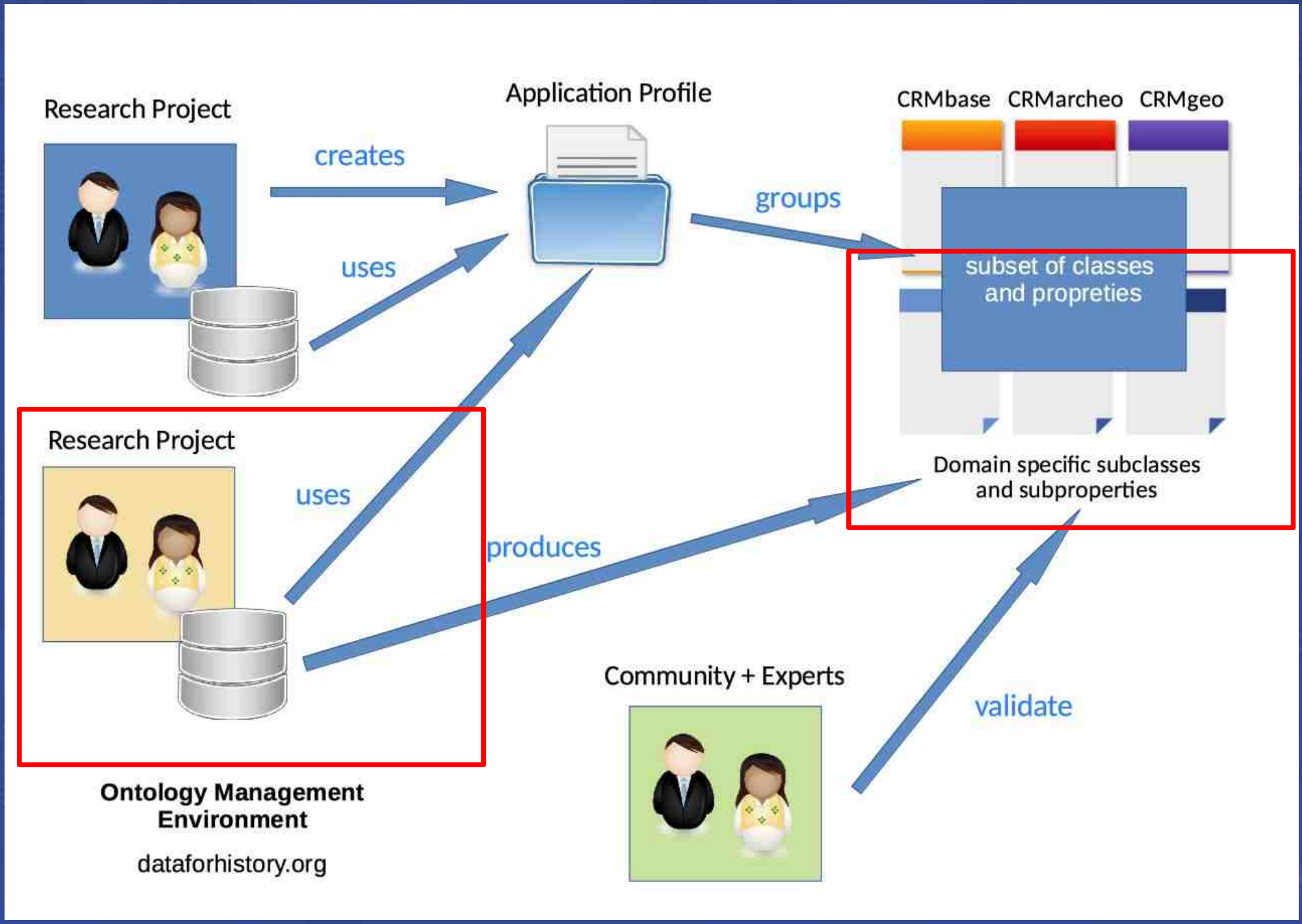


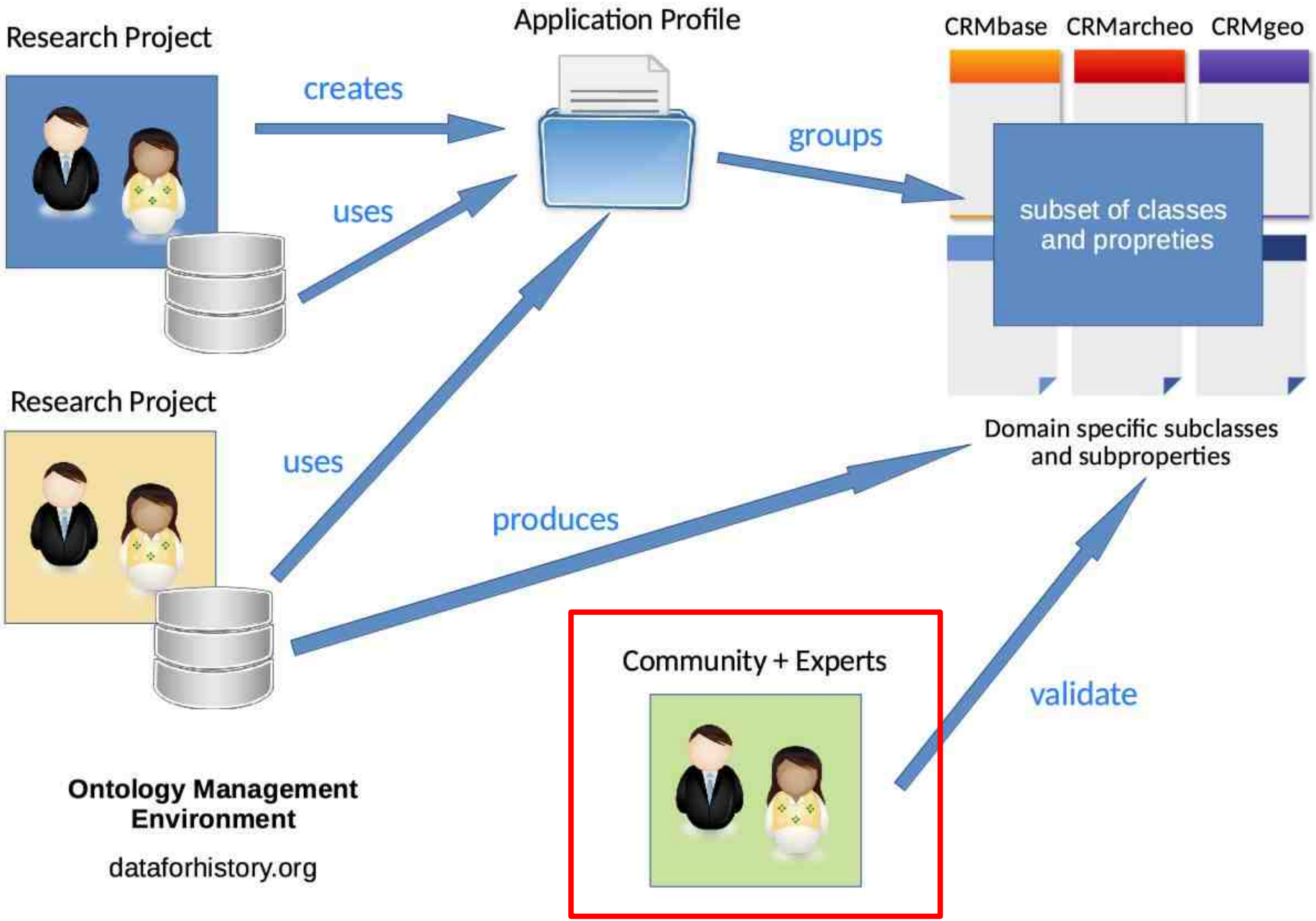
Define and customize the application profile of your own project

Retrieve your project's application profiles from an API

http://ontologies.dataforhistory.org/api/classes/project/1/json



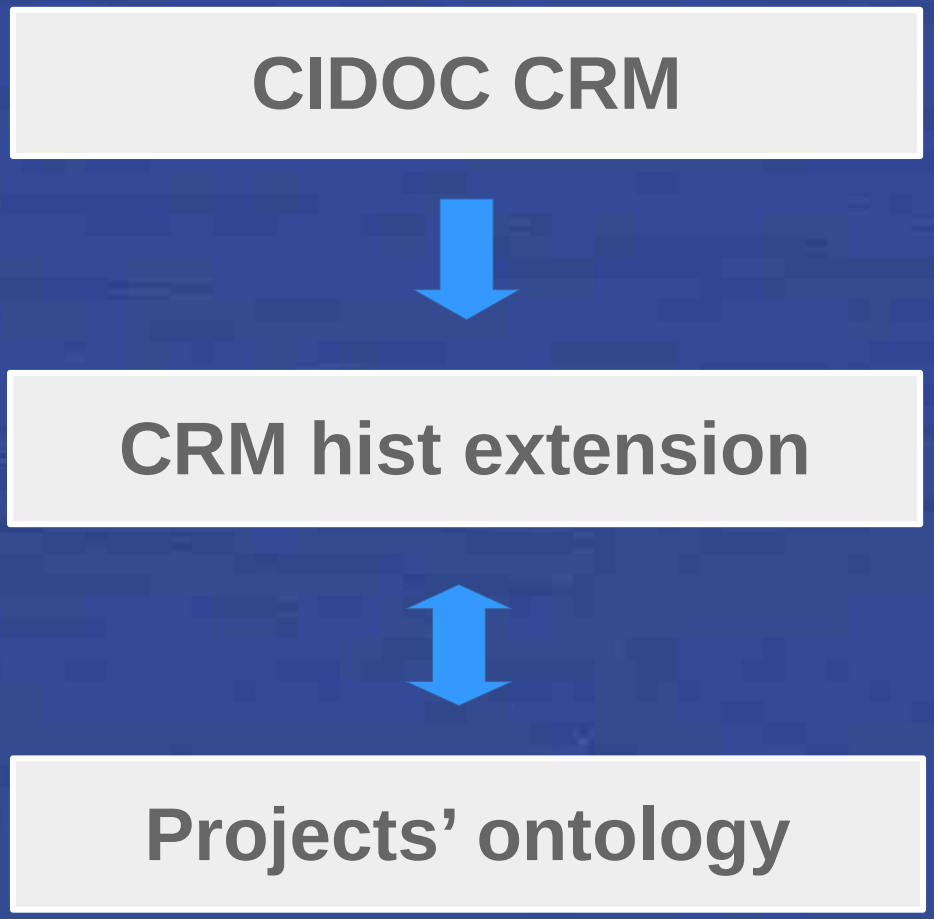

Research Project
creates
Application Profile
CRMbase
CRMarcheo
CRMgeo
groups
uses
subset of classes
and propreties
Research Project
uses
produces
Domain specific subclasses
and subproperties
Community + Experts
validate
Ontology Management
Environment
dataforhistory.org

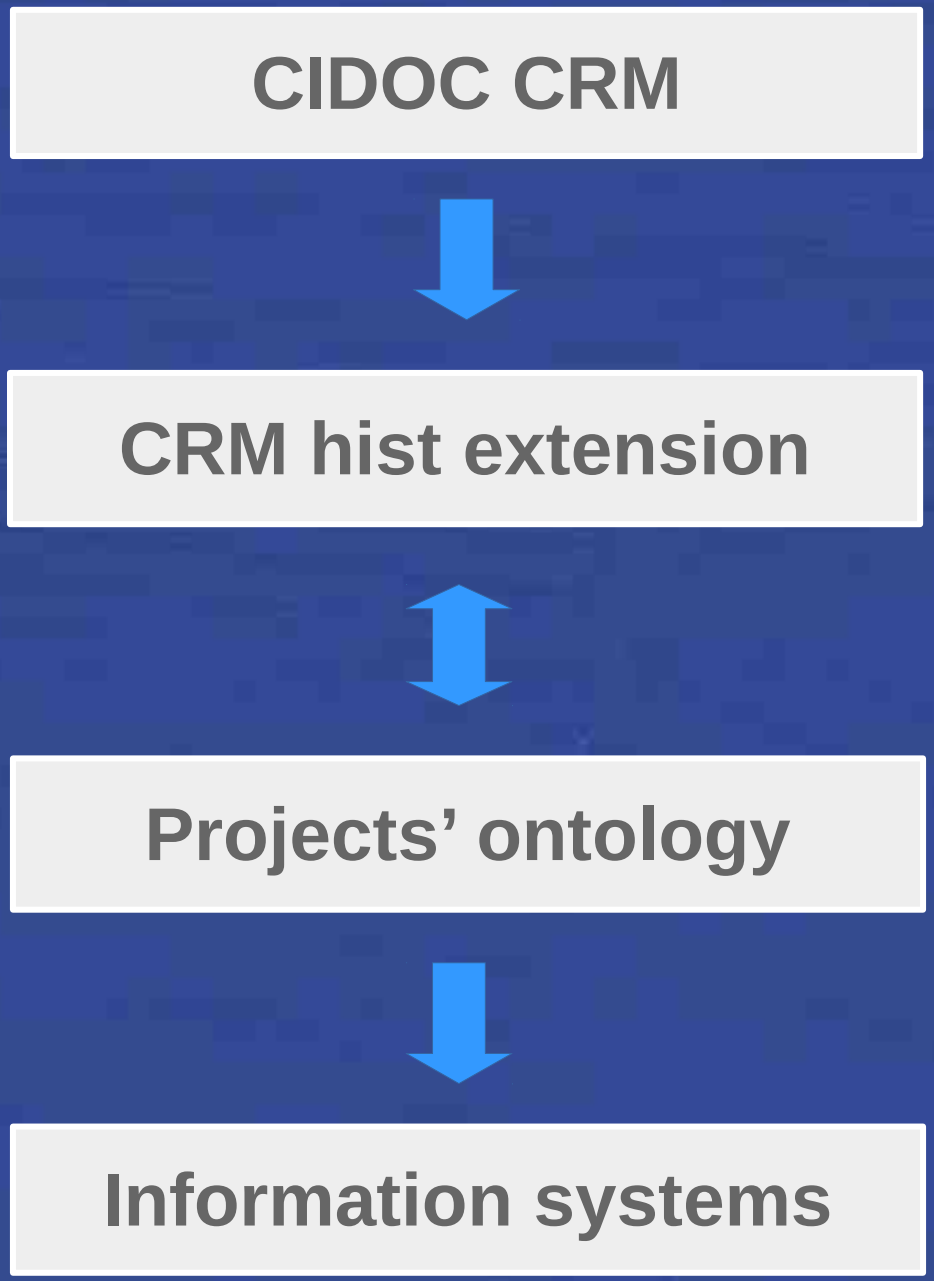

Research Project
creates
Application Profile
CRMbase
CRMarcheo
CRMgeo
groups
uses
subset of classes
and propreties
Research Project
uses
Domain specific subclasses
and subproperties
produces
Community + Experts
validate
Ontology Management
Environment
dataforhistory.org

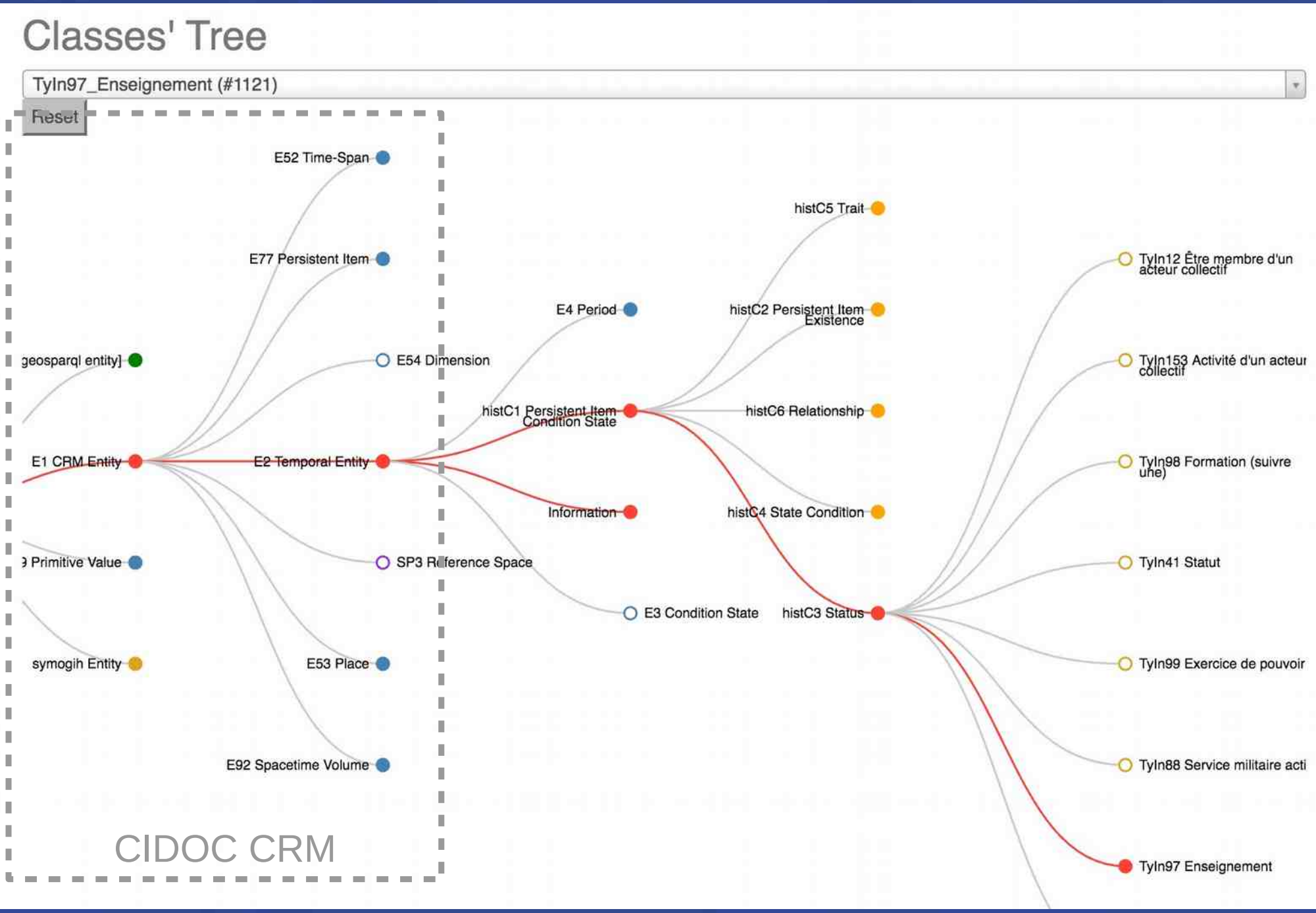
CIDOC CRM

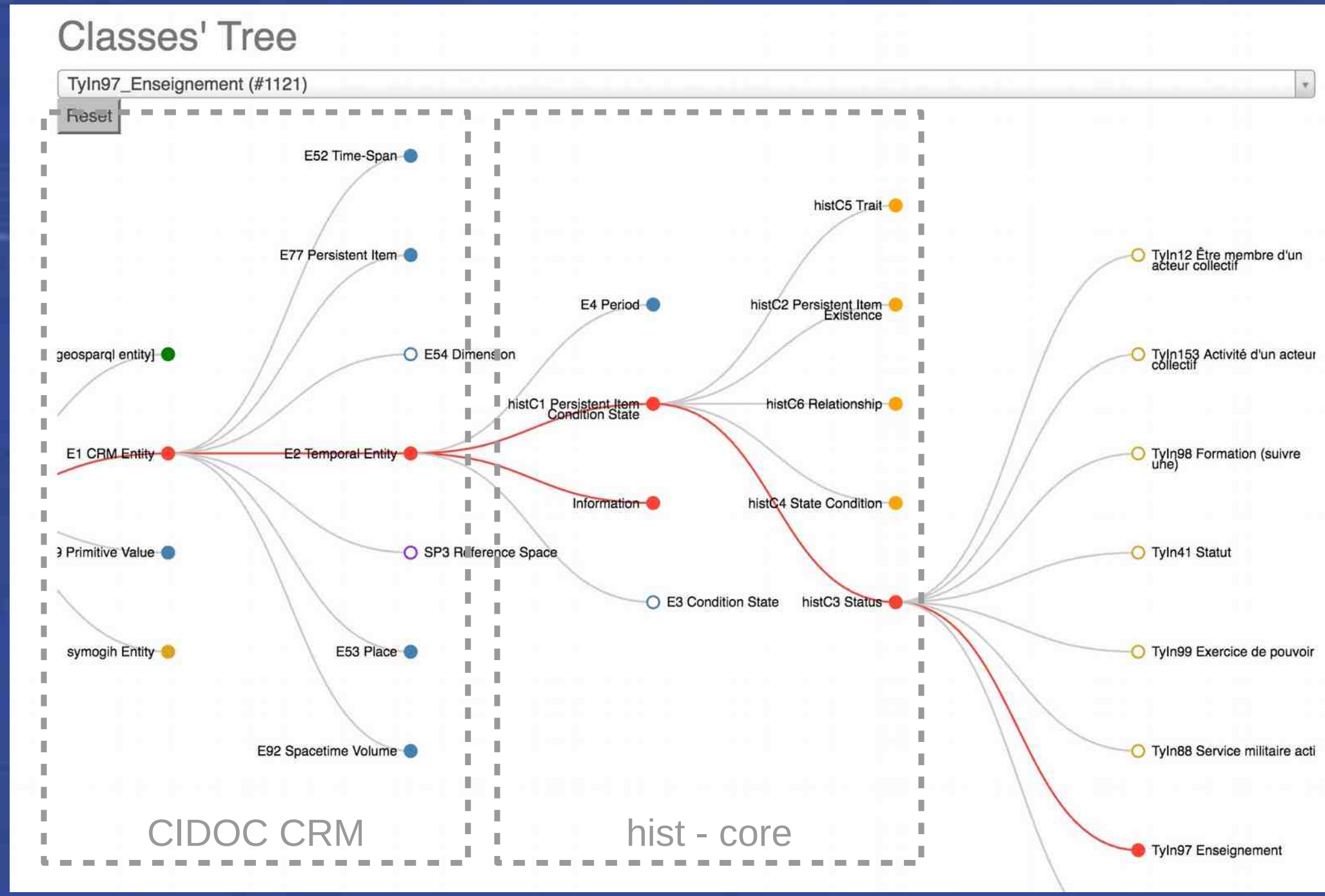
CIDOC CRM

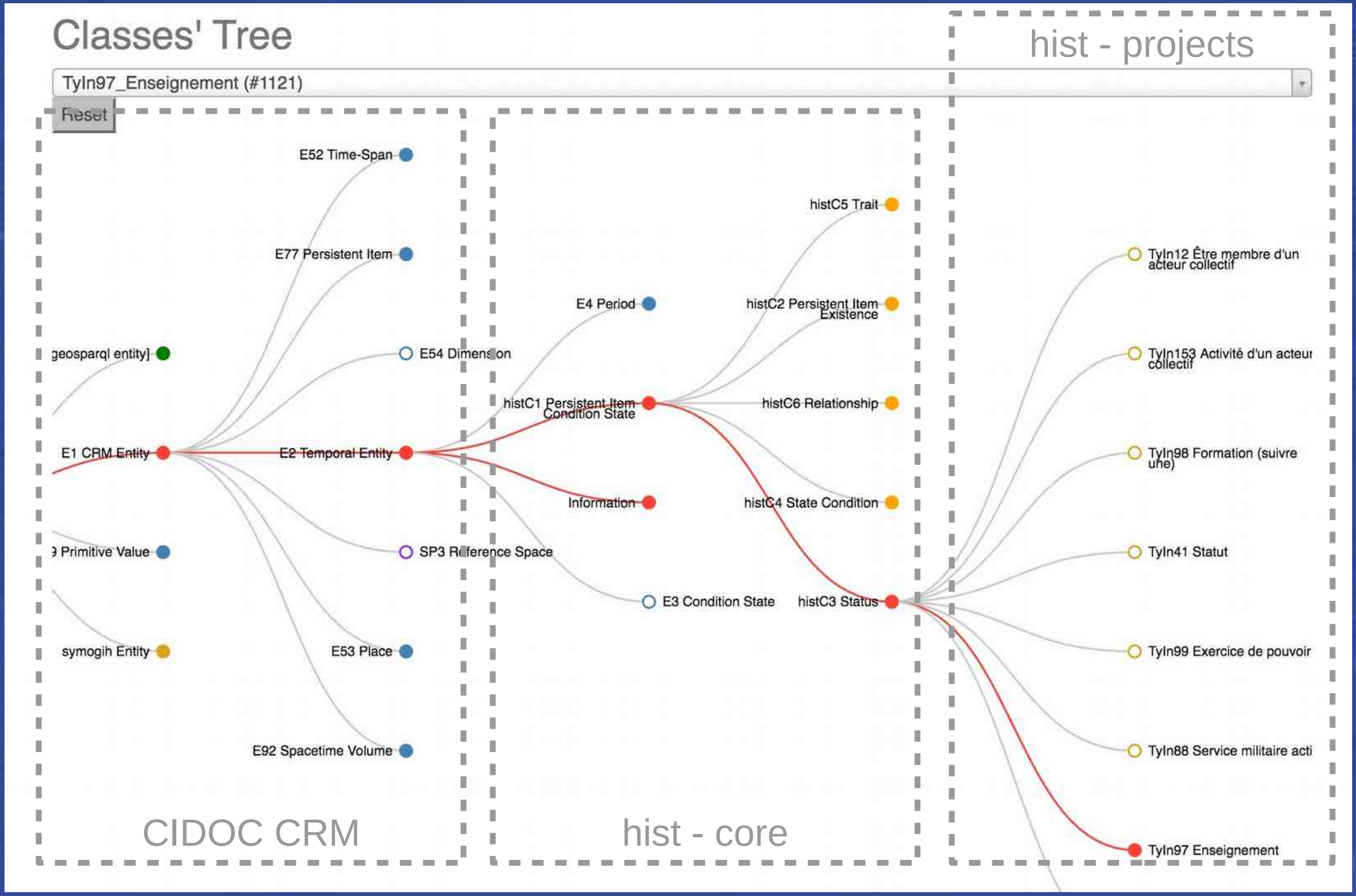
CRM hist extension


CIDOC CRM
CRM hist extension
Projects' ontology


CIDOC CRM
CRM hist extension
Projects' ontology
Information systems


Classes' Tree
TyIn97_Enseignement (#1121)
Reset
E52 Time-Span
E77 Persistent Item
geosparql entity]
E54 Dimension
E1 CRM Entity
E2 Temporal Entity
Primitive Value
SP3 Reference Space
symogih Entity
E53 Place
E92 Spacetime Volume
CIDOC CRM
E4 Period
histC1 Persistent Item Condition State
Information
E3 Condition State
histC5 Trait
histC2 Persistent Item Existence
histC6 Relationship
histC4 State Condition
histC3 Status
TyIn12 Être membre d'un acteur collectif
TyIn153 Activité d'un acteur collectif
TyIn98 Formation (suivre une)
TyIn41 Statut
TyIn99 Exercice de pouvoir
TyIn88 Service militaire acti
TyIn97 Enseignement


Classes' Tree
TyIn97_Enseignement (#1121)
Reset
E52 Time-Span
E77 Persistent Item
geosparql entity]
E54 Dimension
E1 CRM Entity
E2 Temporal Entity
Primitive Value
SP3 Reference Space
symogih Entity
E53 Place
E92 Spacetime Volume
CIDOC CRM
E4 Period
histC1 Persistent Item Condition State
Information
E3 Condition State
histC5 Trait
histC2 Persistent Item Existence
histC6 Relationship
histC4 State Condition
histC3 Status
hist - core
TyIn12 Être membre d'un acteur collectif
TyIn153 Activité d'un acteur collectif
TyIn98 Formation (suivre une)
TyIn41 Statut
TyIn99 Exercice de pouvoir
TyIn88 Service militaire acti
TyIn97 Enseignement


Classes' Tree
TyIn97_Enseignement (#1121)
Reset
hist - projects
E52 Time-Span
E77 Persistent Item
geosparql entity]
E54 Dimension
E1 CRM Entity
E2 Temporal Entity
Primitive Value
SP3 Reference Space
symogih Entity
E53 Place
E92 Spacetime Volume
CIDOC CRM
E4 Period
histC1 Persistent Item Condition State
Information
E3 Condition State
histC5 Trait
histC2 Persistent Item Existence
histC6 Relationship
histC4 State Condition
histC3 Status
hist - core
TyIn12 Être membre d'un acteur collectif
TyIn153 Activité d'un acteur collectif
TyIn98 Formation (suivre une)
TyIn41 Statut
TyIn99 Exercice de pouvoir
TyIn88 Service militaire acti
TyIn97 Enseignement

Data for History Consortium

Meetings

- Lyon, November 2017 (founding meeting)
- Lyon, Mai 2018
- Panel presenting the Data for History vision and tools
  at the EADH 2018 conference in Galway (December 2018)
- Leipzig, March 2019 (planned)

The Data for History consortium

is in the process of being formally established

and is **open to all interested institutions and researchers**

with the aim of sharing FAIR data, best practices, tools and platforms

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.

dataforhistory.org