

TTProfiler

Computing Types and Terms Profiles of Assertional Knowledge Graphs

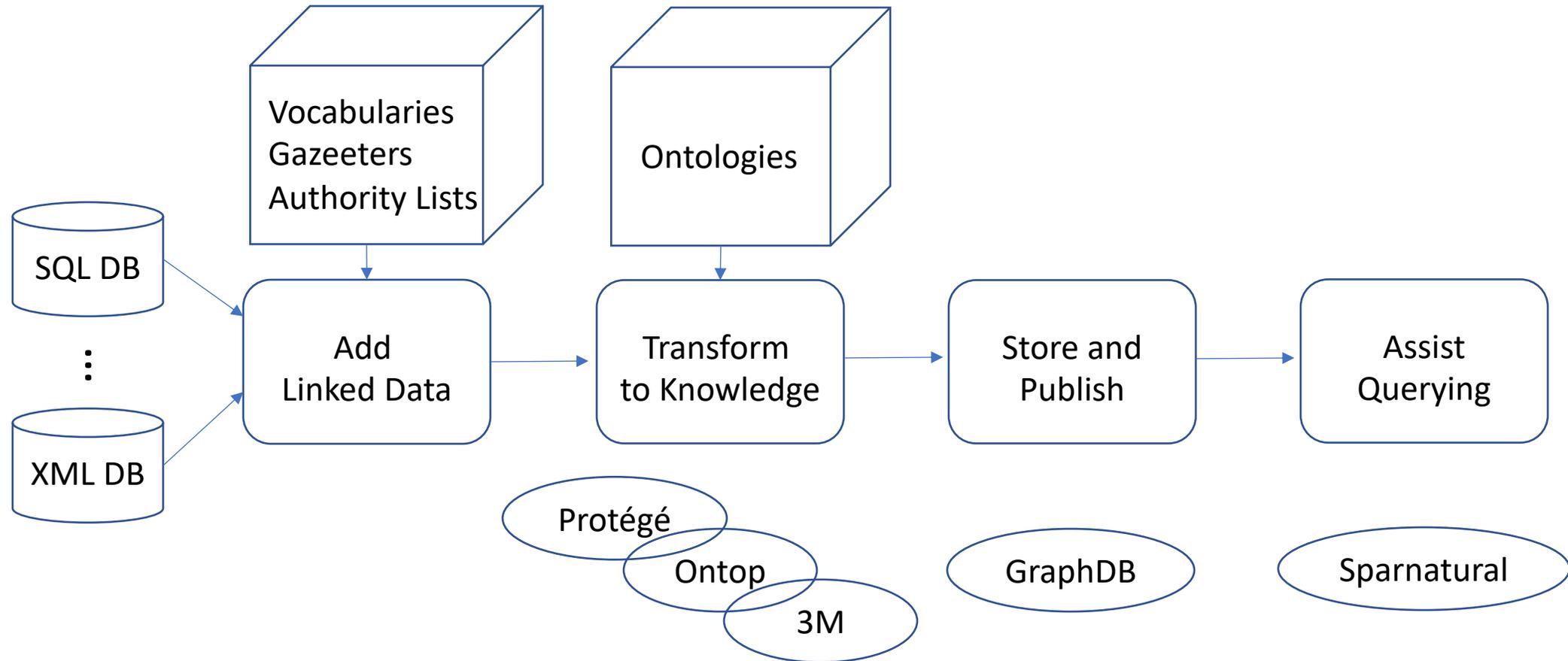
Lamine Diop, Arnaud Giacometti, Béatrice Markhoff, and Arnaud Soulet

Agenda

- Context
- Issues
- Related Works
- Algorithm
- Experiments
- Conclusion & Future Works

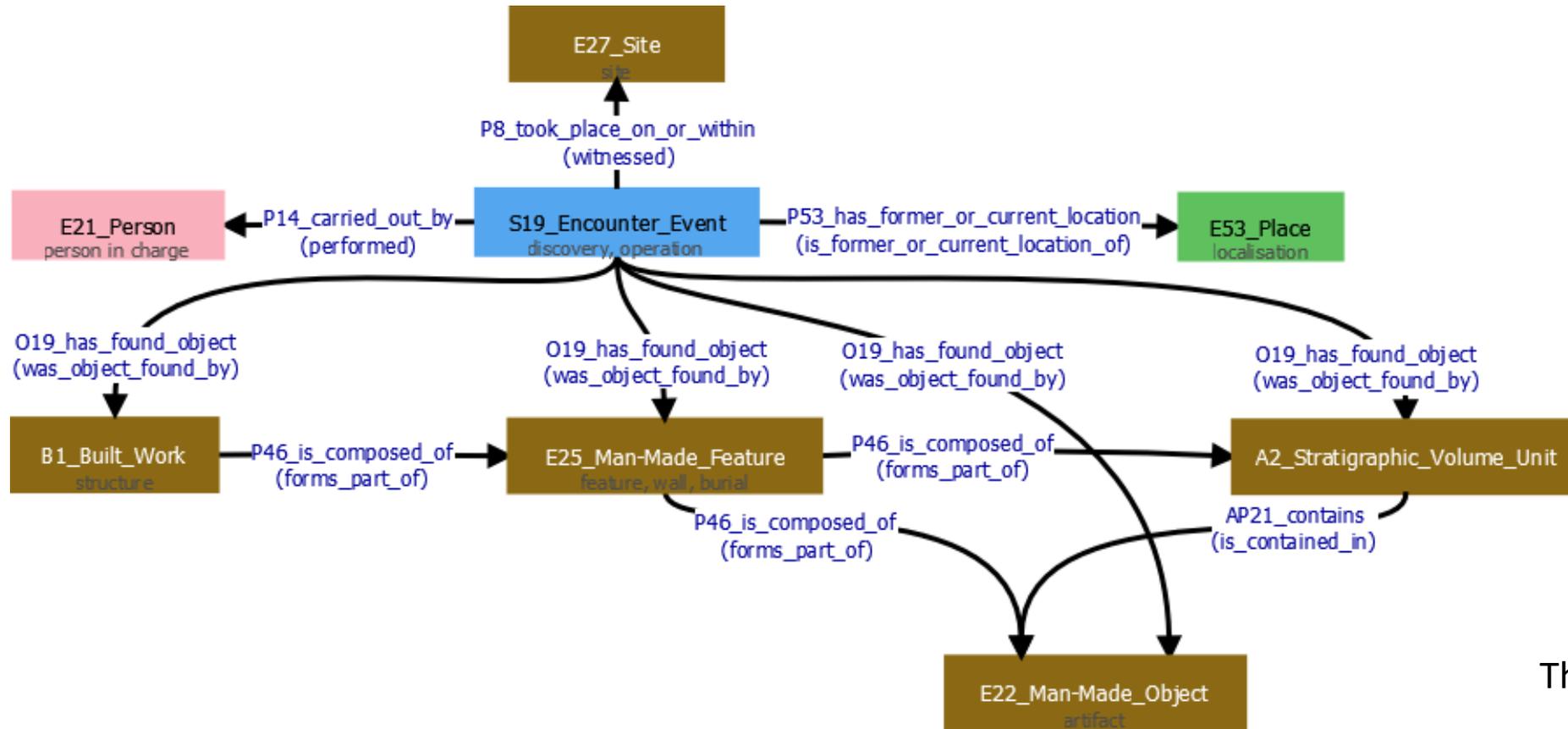


Context: KGs generation, publication and querying



Generate interoperable graphs: shared model

Extracted from the CIDOC CRM and its extensions



This is an excerpt!

Generate interoperable graphs: vocabularies

Compte : miled.rousset

Thésaurus PACTOLS (TH_1) français mes Theso

Thésaurus Candidats Configurations Edition Statistiques

Terme(s) générique(s)
bijou (BT)
fixation de vêtement (BT)

Terme(s) associé(s)
broche
Caché:

Terme(s) synonyme(s)

fibule
Nombre de notices: 2456
id : 14737
ark: 26678/pcrt1tSmMf0Tw8
Domaine : Sujets

definition: Agrafe ou broche articulée par un ressort ou une charnière et dont la pointe (ardillon) se fixait à l'extrémité d'un arc de forme variée, servant à attacher les vêtements flottants et non cousus. (Musée d'Argentomagus) (fr)

Créé le : 08/02/2007
Modifié le : 01/07/2014

Alignement
exactMatch :
Wikidata - Fibule

Coordonnées Gps

©Opentheso 4.3.7 PACTOLS

	uriTerm	term
1	https://ark.frantiq.fr/ark:/26678/cryuEo4S4GyD	"fait archéologique"@fr
2	https://ark.frantiq.fr/ark:/26678/pcrt795b632nWw	"sépulture"@fr
3	https://ark.frantiq.fr/ark:/26678/pcrtms2OAv82PY	"fosse"@fr
4	https://ark.frantiq.fr/ark:/26678/pcrthQAINOX0GB	"sol"@fr
5	https://ark.frantiq.fr/ark:/26678/pcrtb1E0Dz7cSX	"fouille"@fr
6	https://ark.frantiq.fr/ark:/26678/pcrtUDM5NvfJZ1	"coquillage"@fr
7	https://chypre.mom.fr	Site de Kition-Pervolia
8	https://n2t.net/ark:/99152/p0m63njd3g2	"Cypro Classical II"@fr
9	https://orcid.org/0000-0001-8552-1283	Sabine Fourier
10	https://n2t.net/ark:/99152/p0dg76fwn27	"Archaïc"@fr

Publication and querying: OpenArcheo [3]

Semantic web platform for archaeology

MASA (Mémoire des Archéologues et des Sites Archéologiques), a consortium of the TGIH in semantic web, and the Laboratoire d'Informatique Fondamentale et Appliquée de Tours (LIFA) Exploiting the archaeological triplestores paired with the CIDOC in a federated way, OpenArcheo

User Interface : Explorer

ACCESS TO THE EXPLORER

The Explorer, resolutely user-friendly, is based on a system of icons and concepts well understood by archaeologists (Stratigraphic Unit, Artifact, Feature, Site...). In order to produce a SPARQL query from the natural language, OpenArcheo has been the subject of the development of the SPARNATURAL component, which is modular and reusable. This system enables to generate a SPARQL query intuitively, without knowledge of the SPARQL language, but with knowledge of the main concepts of archaeology.

The OpenArcheo data model

To implement the MASA Triplestore, the datasets are mapped to the CIDOC Conceptual Reference Model (CRM), which is a standard for cultural heritage, coordinated by an international consortium (CRM GIS). The role of the CIDOC CRM



The screenshot shows the OpenArcheo Explorer interface. At the top, a query builder allows users to construct queries using natural language concepts. The query shown is: "Burial found in Site studied by Elisabeth Lorans". Below the query builder is an "EXECUTE QUERY" button and a dropdown menu to select the output format, currently set to "Table".

A green notification bar indicates: "Query successful! - View/hide SPARQL query".

The SPARQL query is displayed in a text area:

```
1 SELECT DISTINCT ?this ?thisLabel
2 FROM NAMED <http://openarchaeo.huma-num.fr/federation/sources/arsol>
3 WHERE
4 { ?this a <http://www.cidoc-crm.org/cidoc-crm/E25_Man-Made_Feature> ;
5   <http://www.cidoc-crm.org/cidoc-crm/P2_has_type> <https://ark.frantiq.fr/ark:/26678/pcrt795b632nww> .
6   ?this <http://www.ics.forth.gr/isl/CRMsci/019i_was_object_found_by>/<http://www.cidoc-crm.org/cidoc-crm/P8_took_place>
7   ?Site1 a <http://www.cidoc-crm.org/cidoc-crm/E27_Site> .
8   ?Site1 <http://www.cidoc-crm.org/cidoc-crm/P8i_witnessed>/<http://www.cidoc-crm.org/cidoc-crm/P14_carried_out_by> <
9   OPTIONAL
10  { ?this <http://www.w3.org/2004/02/skos/core#prefLabel> ?thisLabel}
11 }
12
```

Below the query, there are tabs for "Response", "Table", "Pivot Table", "Geo", "Google Chart", and "Timeline". The "Table" tab is selected. The results are displayed as a table with 3 entries:

this	thisLabel
1 http://arsol.univ-tours.fr/4DACTION/WFICHEWEB/isepuAJ000001	Sepulture AJ000001
2 http://arsol.univ-tours.fr/4DACTION/WFICHEWEB/isepuAJ000002	Sepulture AJ000002
3 http://arsol.univ-tours.fr/4DACTION/WFICHEWEB/isepuAJ000003	Sepulture AJ000003

At the bottom, there is a search bar and a "Show 50 entries" dropdown menu.

KGs in OpenArcheo

\mathcal{A}	nb triples	nb nodes
Aerba	3,318	1,695
Epicherchell	3,488	1,372
Kition	26,773	9,165
Iceramm	32,687	9,325
Rita	40,479	10,769
Outagr	79,420	39,573
Arsol	670,757	212,143

Number of Predicates

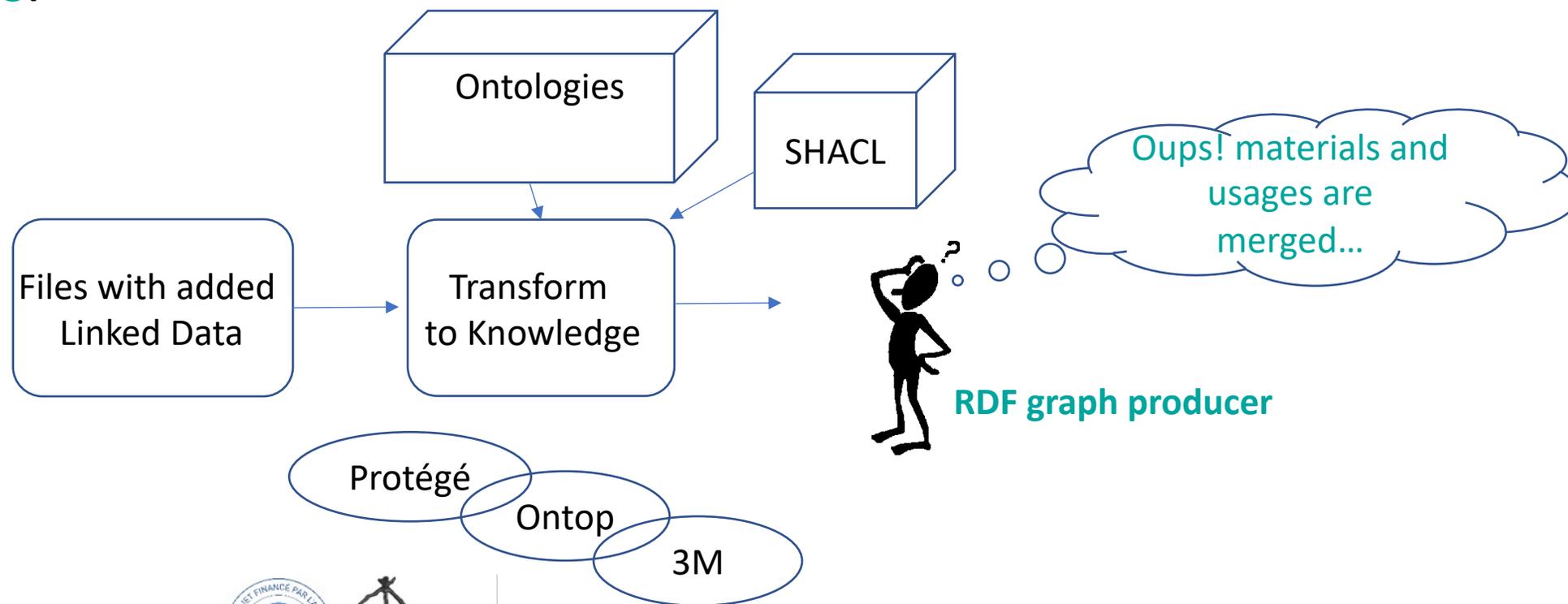
\mathcal{K}	crm	crmsci	crmarch
Aerba	3	0	0
Epicherchell	13	2	0
Kition	26	3	1
Iceramm	19	2	0
Rita	5	0	0
Outagr	7	1	0
Arsol	29	3	1

Number of Types and Terms

\mathcal{K}	crm	crmsci	crmarch	crmba	PACTOLS
Aerba	4	0	0	0	0
Epicherchell	9	1	0	0	21
Kition	12	1	1	0	57
Iceramm	11	1	0	0	1
Rita	5	0	0	0	178
Outagr	6	1	0	0	1
Arsol	11	1	1	1	79

Issue 1: KG generation errors

- Difficulties for **RDF graph producers** to check the results of their automatic generation processes, especially when it comes to deal with **terms**.



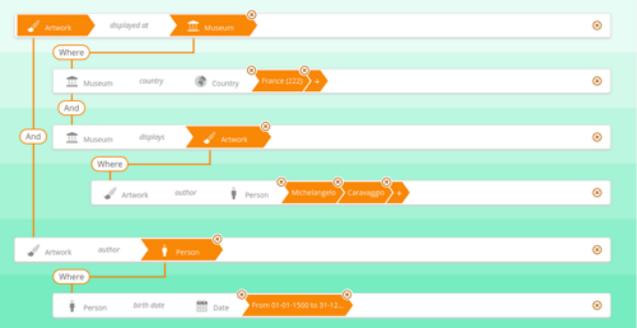
Issue 2: query tool reusability

- Difficulties to improve the reusability of Sparnatural [4] by automating its generation from any given knowledge graph.

- Need of extracting a “profile” from the knowledge graph

About Sparnatural

Sparnatural is a Javascript component that allows users to explore an RDF Knowledge Graph by building SPARQL queries intuitively. It is configurable to adapt to your knowledge graph ontology.



How do **you** explore your knowledge graph ?

Maybe you have lists of entities. Or a search engine. Maybe you have facets. Maybe you even have data-visualisation. But can you explore your knowledge graph *really as a graph* ? This is what Sparnatural does.

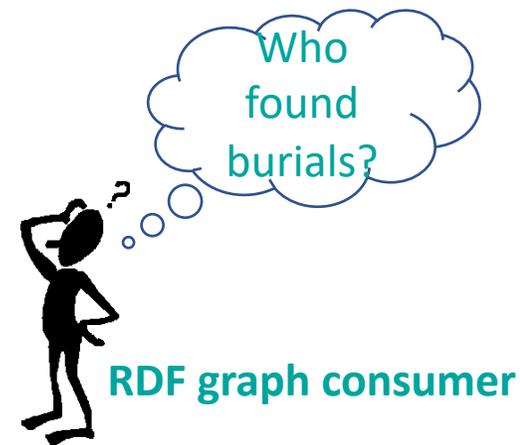
Have you ever been frustrated by not being able to show - and to sell - all this nice structured data you have aggregated in your graph ? Sparnatural can demonstrate the value of heterogeneous data integration.

Are you looking for a solution that can *bring your knowledge graph to your end-users* ? Sparnatural is the right tool for this : it allows try-and-error interactions, it is visual, and it "gamifies" the knowledge graph experience.

Issue 3: user understanding of the KG's content

- Difficulties for a computer scientist user to know what can be queried, even with such a simple shared ontology

```
SELECT DISTINCT ?person ?personLabel
FROM NAMED <http://openarchaeo.huma-num.fr/federation/sources/arsol>
WHERE {
  ?person a crm:E21_Person .
  ?person skos:prefLabel ?personLabel .
  ?person crm:P14i_performed/crm:P8_took_place_on_or_within ?site .
  ?site crm:P8i_witnessed/crmsci:019_has_found_object ?sepulture .
  ?sepulture a crm:E25_Man-Made_Feature .
  ?sepulture crm:P2_has_type <https://ark.frantiq.fr/ark:/26678
/pcrt795b632nWw> .
}
```



Related Works

- Huge literature on computing *graph summaries*
- We don't want to *infer a schema* from an Assertional KG
 - it has been generated using ontologies
 - in general **not only one**, often **not accessible**
- We want to *compute and show the Predicates, Types and Terms* used in an Assertional KG
 - ABSTAT partly does it
 - Evidence of its helpfulness for users when querying an unknown KG

ABSTAT [\(http://abstat.disco.unimib.it/\)](http://abstat.disco.unimib.it/) [Principe et al., 2017]

Subject Type (occurrences)	Predicate (occurrences)	Object Type (occurrences)	Frequency
subject	https://dbpedia.org/ontology/	object	
dbo:Location (839987)	dbo:country (770950)	dbo:Country (13169)	560532
schema:Place (1665807)	dbo:country (770950)	dbo:Country (13169)	560532
dbo:Location (839987)	dbo:country (770950)	wgspos:SpatialThing (1965986)	546131
schema:Place (1665807)	dbo:country (770950)	wgspos:SpatialThing (1965986)	546131

ABSTAT Abstract Pattern $C_1 \xrightarrow{P(n)} C_2$

A **profile** is a set of Abstract Patterns

Basic Graph Pattern



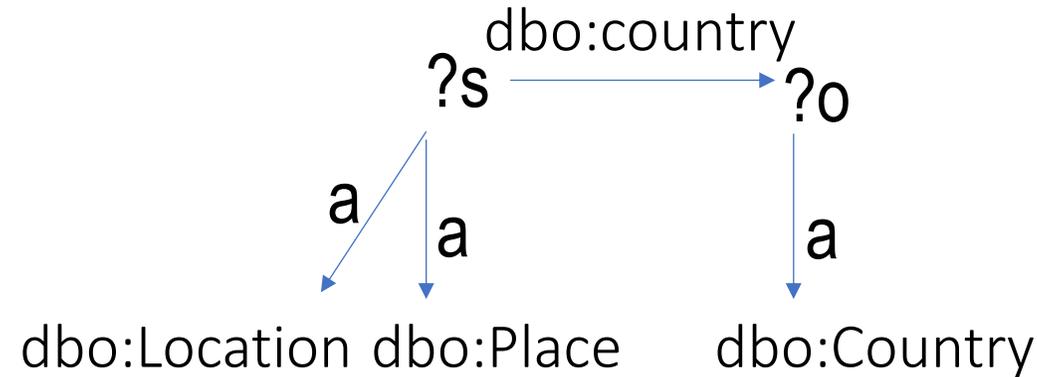
Subject Type (occurrences)	Predicate (occurrences)	Object Type (occurrences)	Frequency
dbo:Location (839987)	dbo:country (770950)	dbo:Country (13169)	560532

dbo:Location dbo:Country

dbo:Location $\xrightarrow{\text{dbo:country (560532)}}$ dbo:Country

Problem 1

Basic Graph Pattern

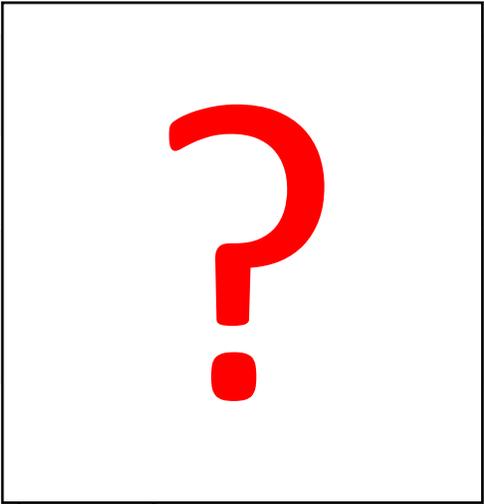


Subject Type (occurrences)	Predicate (occurrences)	Object Type (occurrences)	Frequency
<code>dbo:Location</code> (839987)	<code>dbo:country</code> (770950)	<code>dbo:Country</code> (13169)	560532
<code>schema:Place</code> (1665807)	<code>dbo:country</code> (770950)	<code>dbo:Country</code> (13169)	560532
<code>dbo:Location</code> (839987)	<code>dbo:country</code> (770950)	<code>wgspos:SpatialThing</code> (1965986)	546131
<code>schema:Place</code> (1665807)	<code>dbo:country</code> (770950)	<code>wgspos:SpatialThing</code> (1965986)	546131

same instances of (s,o)

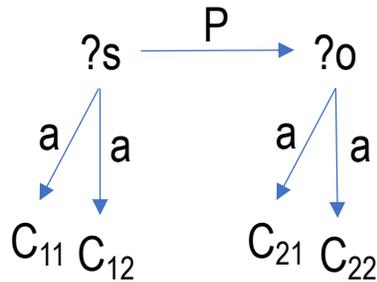
Problem 2

Nothing appears about the terms used

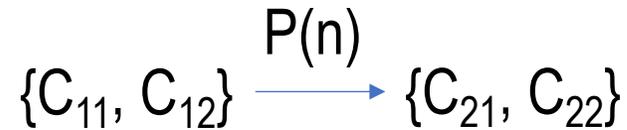
Subject Type (occurrences)	Predicate (occurrences)	Term (occurrences)	Frequency
<input type="text" value="subject"/>	<input type="text" value="https://dbpedia.org/ontology,"/>	<input type="text" value="object"/>	
dbo:Location (839987)	dbo:country (770950)		
schema:Place (1665807)	dbo:country (770950)		
dbo:Location (839987)	dbo:country (770950)		
schema:Place (1665807)	dbo:country (770950)		

Solutions

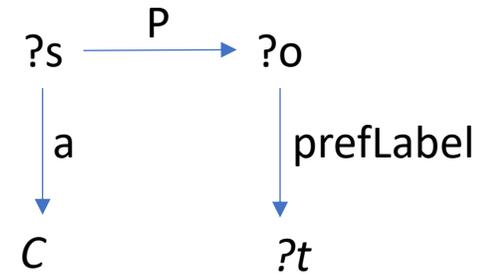
Basic Graph Pattern



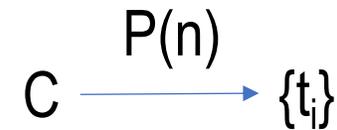
Types Abstract Pattern



Basic Graph Pattern



Terms Abstract Pattern



Goal:

Given the ABox \mathcal{A} of a knowledge graph $\mathcal{KB} = (\mathcal{A}, \mathcal{T})$

find a profile of \mathcal{A} taking into account **Types** and **Terms** without counting multiple times some instances of \mathcal{A}

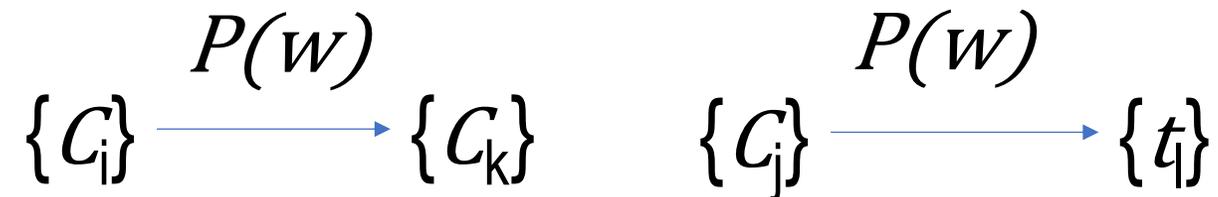
Algorithm TTPProfiler

1. Retrieve all **Abstract Patterns (one Type and/or one Term)** and their **weights**:

$$\mathcal{R} = \{((C, P, D), w) : ((\exists P(a, b), C(a), D(b) \in ABox) \vee (\exists P(a, b), C(a), \text{skos:prefLabel}(b) = D))\}$$

where w is the number of instances $P(a, b)$ of (C, P, D) in the Abox

2. Compute all Types and Terms (TT) Patterns and store them in \mathbb{P}



3. For visualisation: group all TT Patterns of \mathbb{P} that share a type/term

TTPProfiler: retrieval of Abstract Patterns (step 1)

```
1 SELECT *
2 FROM NAMED <http://openarchaeo.huma-num.fr/federation/sources/kition-
pervolia>
3 WHERE {
4   {
5     SELECT ?C ?P ?D (COUNT(?P) AS ?w) {
6       ?subj ?P ?obj ; a ?C .
7       ?obj a ?D .
8       filter(!isBlank(?C) && !isBlank(?D) )
9     } GROUP BY ?C ?P ?D
10  }
11 }
12
```



Query: Find all weighted pattern **(C, P, D):w** where C and D are types and w the number of times this pattern is instantiated in the knowledge graph.

TProfiler: retrieval of Abstract Patterns

Response **Table** Pivot Table Geo Google Chart Timeline  

Showing 1 to 42 of 42 entries Search: Show entries

	C	P	D	w
1	http://www.cidoc-crm.org/cidoc-crm/E21_Person	http://www.cidoc-crm.org/cidoc-crm/P14i_performed	http://www.cidoc-crm.org/cidoc-crm/E65_Creation	"3270" ^{^^xsd:integer}
2	http://www.cidoc-crm.org/cidoc-crm/E31_Document	http://www.cidoc-crm.org/cidoc-crm/P70_documents	http://www.cidoc-crm.org/cidoc-crm/E25_Man-Made_Feature	"1340" ^{^^xsd:integer}
3	http://www.cidoc-crm.org/cidoc-crm/E31_Document	http://www.cidoc-crm.org/cidoc-crm/P2_has_type	http://www.cidoc-crm.org/cidoc-crm/E55_Type	"1090" ^{^^xsd:integer}
4	http://www.cidoc-crm.org/cidoc-crm/E31_Document	http://www.cidoc-crm.org/cidoc-crm/P94i_was_created_by	http://www.cidoc-crm.org/cidoc-crm/E65_Creation	"1090" ^{^^xsd:integer}
5	http://www.cidoc-crm.org/cidoc-crm/E65_Creation	http://www.cidoc-crm.org/cidoc-crm/P14_carried_out_by	http://www.cidoc-crm.org/cidoc-crm/E21_Person	"1090" ^{^^xsd:integer}
6	http://www.cidoc-crm.org/cidoc-crm/E65_Creation	http://www.cidoc-crm.org/cidoc-crm/P94_has_created	http://www.cidoc-crm.org/cidoc-crm/E31_Document	"1090" ^{^^xsd:integer}

TTPProfiler: retrieval of One Term Patterns (step 1)

```
1 SELECT *
2 FROM NAMED <http://openarchaeo.huma-num.fr/federation/sources/kition-
  pervolia>
3 WHERE {
4   {
5     SELECT ?C ?P ?D (COUNT(?P) AS ?w) {
6       ?subj ?P ?obj ; a ?C .
7       FILTER(CONTAINS(STR(?obj), "frantiq")).
8       ?obj <http://www.w3.org/2004/02/skos/core#prefLabel> ?D .
9       FILTER(!isBlank(?C))
10      FILTER(lang(?D)='fr')
11    } GROUP BY ?C ?P ?D
12  }
13 }
14
```



Query: Find all weighted pattern **(C, P, D):w** where **C** is a type, **D** a term and **w** the number of times this pattern is instantiated in the knowledge graph.

TTPProfiler: retrieval of One Term Patterns

Response **Table** Pivot Table Geo Google Chart Timeline  

Showing 1 to 58 of 58 entries Search: Show entries

	C	P	D	w
1	http://www.cidoc-crm.org/cidoc-crm/E31_Document	http://www.cidoc-crm.org/cidoc-crm/P2_has_type	"Photographie"@fr	"895"^^xsd:integer
2	http://www.cidoc-crm.org/cidoc-crm/E31_Document	http://www.cidoc-crm.org/cidoc-crm/P2_has_type	"Graphisme"@fr	"186"^^xsd:integer
3	http://www.cidoc-crm.org/cidoc-crm/E22_Man-Made_Object	http://www.cidoc-crm.org/cidoc-crm/P101_had_as_general_use	"cruche"@fr	"124"^^xsd:integer
4	http://www.cidoc-crm.org/cidoc-crm/E25_Man-Made_Feature	http://www.cidoc-crm.org/cidoc-crm/P2_has_type	"fait archéologique"@fr	"71"^^xsd:integer
5	http://www.cidoc-crm.org/cidoc-crm/E22_Man-Made_Object	http://www.cidoc-crm.org/cidoc-crm/P101_had_as_general_use	"bol"@fr	"66"^^xsd:integer

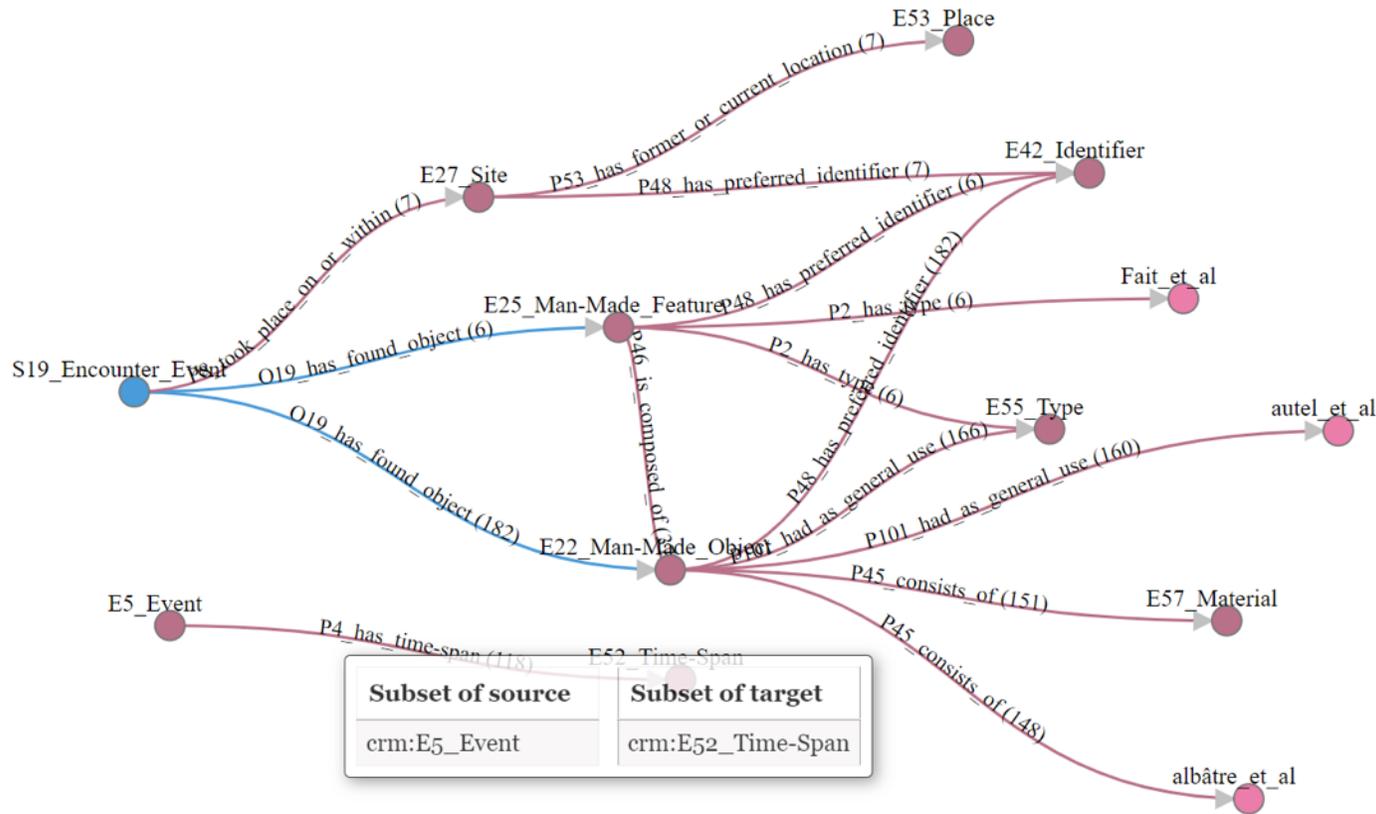
Experiments

- **TTPProfiler** is implemented [1] in Java using Jena
- Input
 - TTPProfiler queries the public **online** SPARQL endpoints (Abox part)
- Output
 - TTPProfiler outputs a JSON file containing the TT profile
- The visualisation is implemented [2] in Javascript using D3.js

TT Profiles of Cultural Heritage Knowledge Graphs

A	nb triples	nb nodes	statistics for TT profile		
			nb types & terms	nb AP	nb nodes
Aerba	3,318	1,695	5	3	5
Epicherchell	3,488	1,372	31	15	13
Kition	26,773	9,165	72	31	19
Iceramm	32,687	9,325	13	21	13
Rita	40,479	10,769	184	6	7
Outagr	79,420	39,573	8	8	8
Arsol	670,757	21,2143	94	34	17
Smithsonian	2,542,142	969,172	18	35	18
Doremus	91,093,326	24,141,972	599	678	146

Graph EPICHERCHELL Visualiser



Subset of source	Subset of target
crm:E5_Event	crm:E52_Time-Span

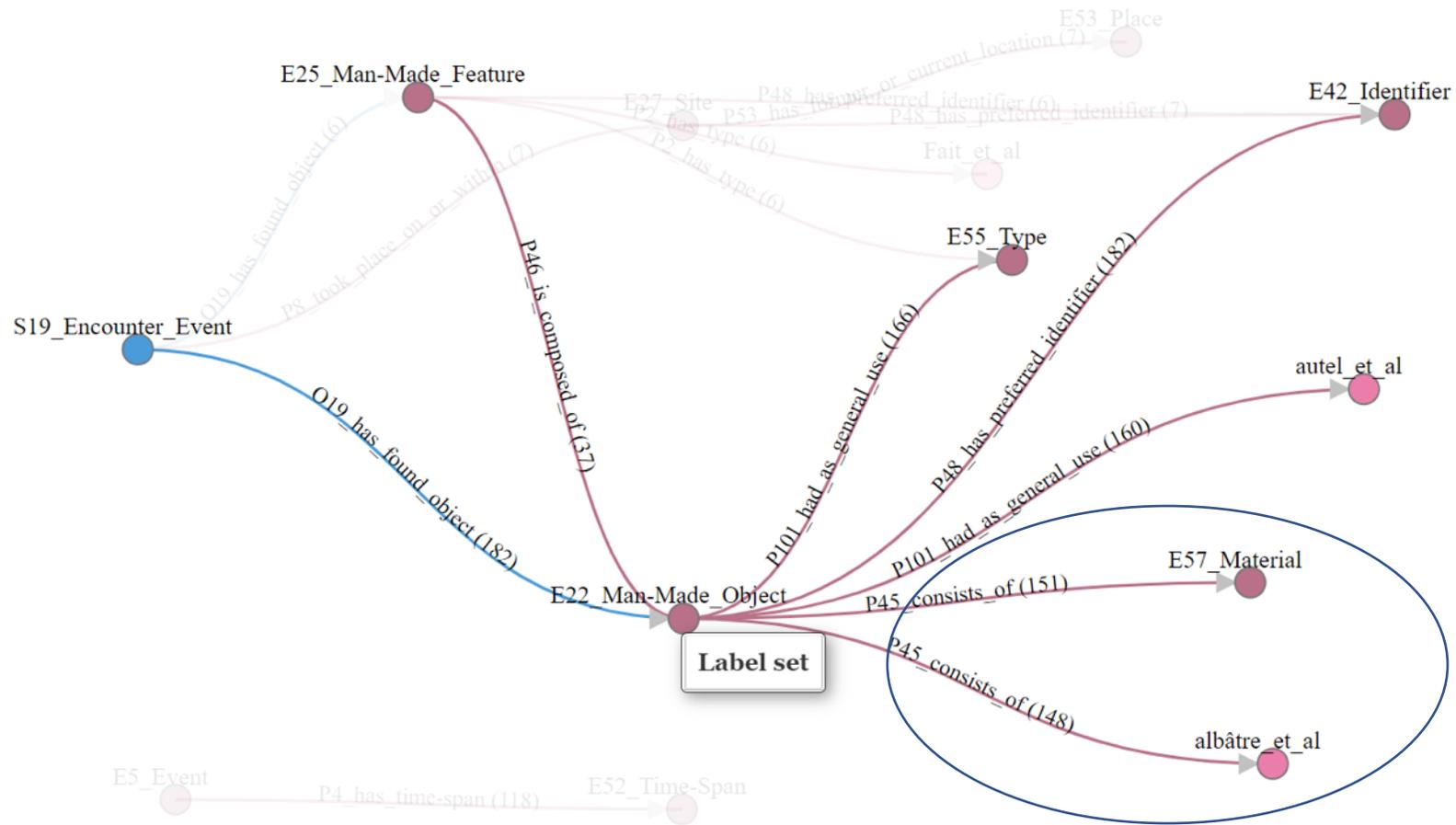
Prefix by:

- crm: <http://www.cidoc-crm.org/cidoc-crm/>
- crmsci: <http://www.ics.forth.gr/isl/CRMsci/>
- frantiq: <https://ark.frantiq.fr/ark:/26678/>

Link *P4_has_time-span*

Description

- Subset of source:**
crm:E5_Event
- Subset of target:**
crm:E52_Time-Span
- Prefix:** crm
- Frequency:** 118
- Uri:** http://www.cidoc-crm.org/cidoc-crm/P4_has_time-span



Prefix by:

- crm: <http://www.cidoc-crm.org/cidoc-crm/>
- crmsci: <http://www.ics.forth.gr/isl/CRMsci/>
- frantiq: <https://ark.frantiq.fr/ark:/26678/>

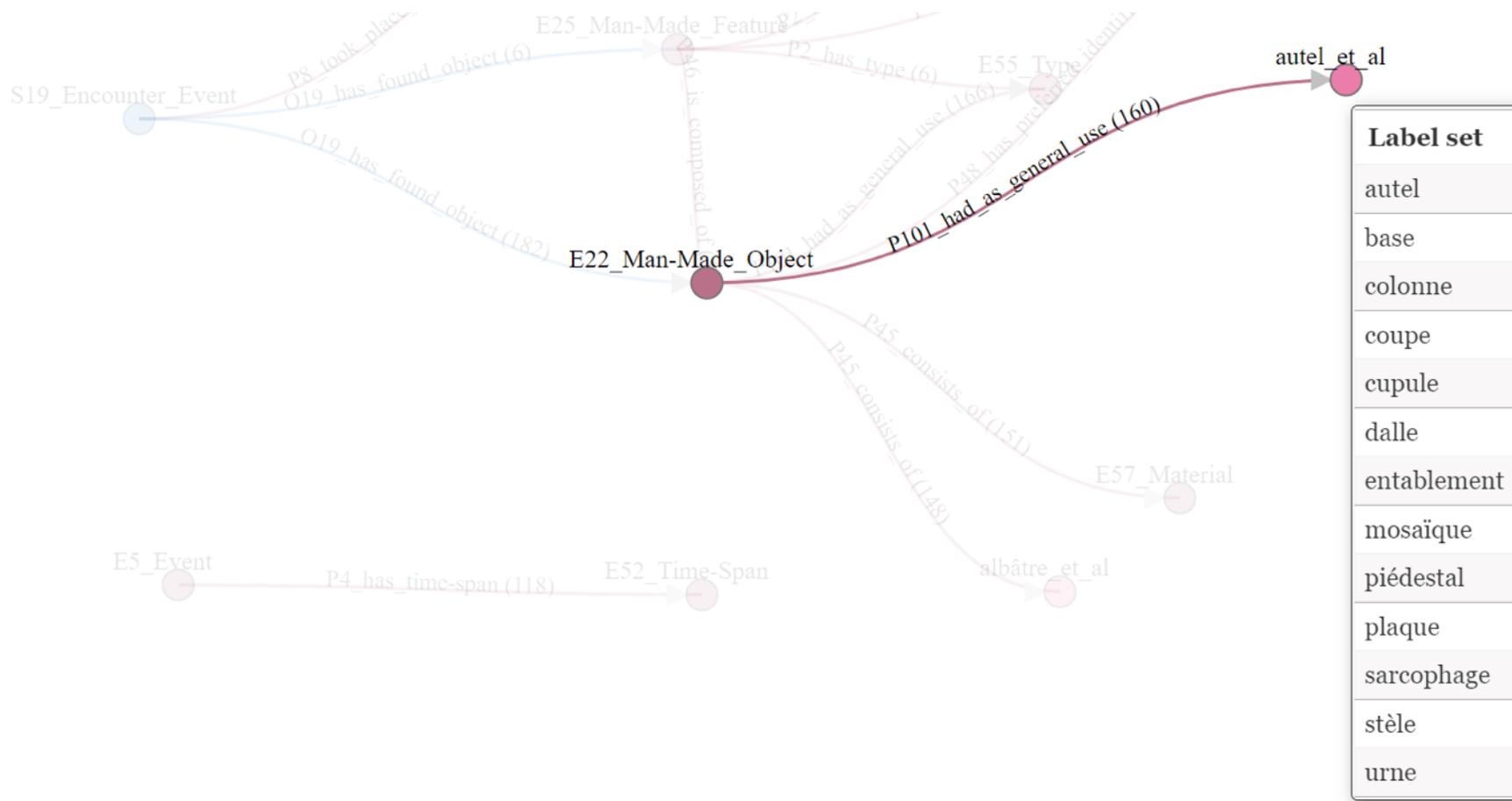
Node *E22_Man-Made_Object*

Description

Data properties

- *E22_Man-Made_Object*, prefLabel (182), string
- *E22_Man-Made_Object*, P3_has_note (182), string

Query



Label set
autel
base
colonne
coupe
cupule
dalle
entablement
mosaïque
piédestal
plaque
sarcophage
stèle
urne

Node *autel_et_al*

Description

- **Prefix:** frantiq
- **Label link set:**
[autel](#)
[base](#)
[colonne](#)
[coupe](#)
[cupule](#)
[dalle](#)
[entablement](#)
[mosaïque](#)
[piédestal](#)
[plaque](#)
[sarcophage](#)
[stèle](#)
[urne](#)

Data properties

Queries

Conclusion

- A profile builder taking into account terms of thesaurus
 - Help the **graph producers** to check their result (Issue 1)
 - Step towards automating Sparnatural's configuration (Issue 2)
 - Provides to **graph consumers** a precise picture of the KG's content (Issue 3)
- Direct use of online SPARQL endpoints instead of loading graphs locally
- Interactive Types and Terms Profile visualizer, developed by **Zilu YANG**

- Future Works:
 - Compute a **KG Summary of size k** from the TT profile
 - Complete with ontologies: from the TT Profile to the ontology of that KG could be with a “forgetting-based approach for computing views of ontologies” [Li et al, The Web Conference 2021]

THANK YOU FOR YOUR ATTENTION!

(beatrice.markhoff@univ-tours.fr)

[1] TTPProfiler Github: <https://github.com/DTTProfiler/DTTProfiler>

[2] TT Profile Visualiser (work in progress): <https://kgsumviz.univ-tours.fr>

[3] OpenArchaeo: <http://openarchaeo.huma-num.fr/explorateur/home>

[4] Sparnatural: <http://sparnatural.eu/>