

OpenArchaeo : une plateforme Web sémantique pour interroger les données archéologiques.

En collaboration avec Xavier Rodier, Béatrice
Markhoff et Olivier Marlet

Florian HIVERT

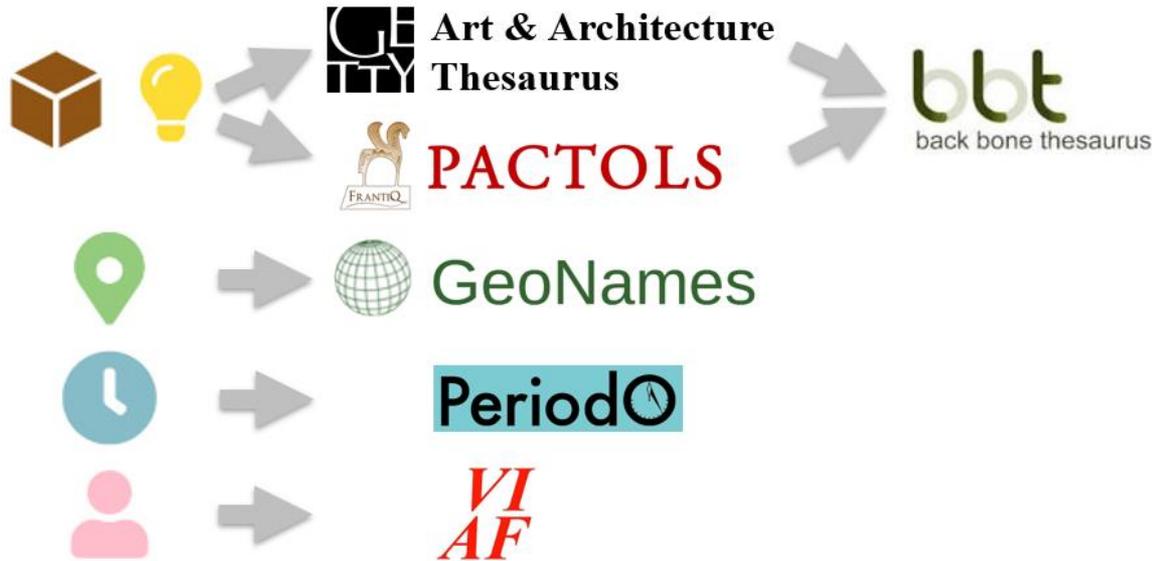
MSH Val de Loire, Tours/Consortium MASA

Le Consortium MASA



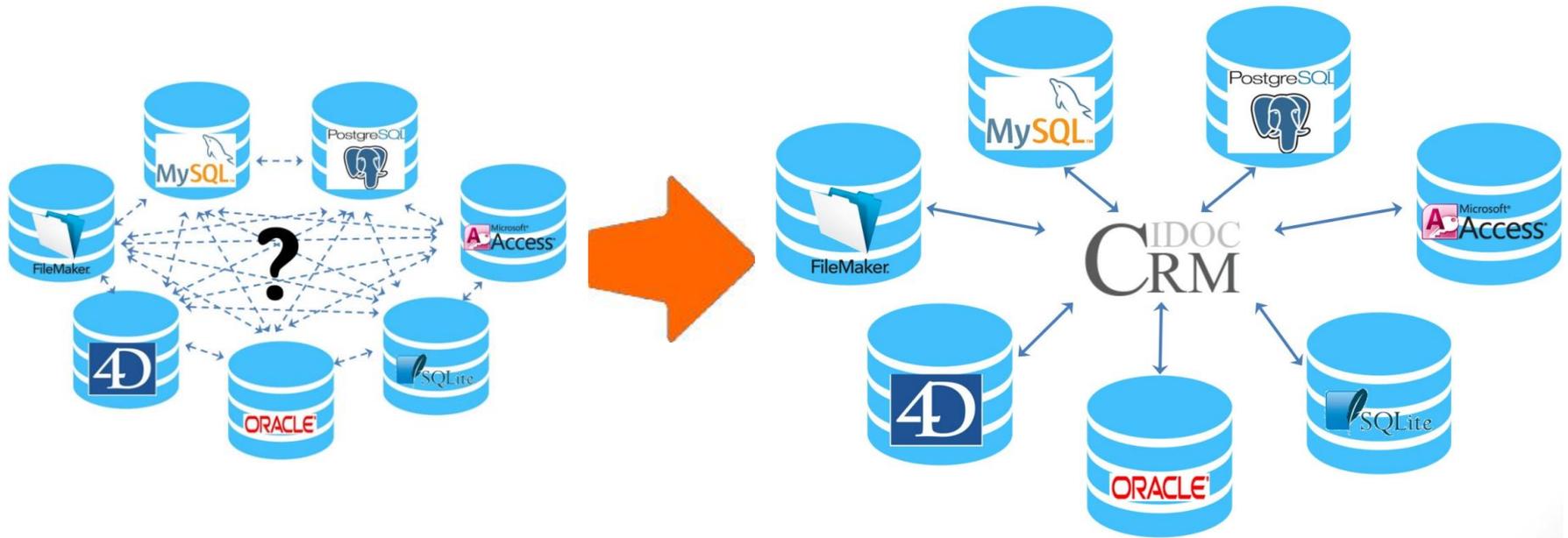
- Depuis 2013, le Consortium MASA de la TGIR Huma-Num œuvre pour aider la communauté archéologique française à adopter les principes FAIR.
- L'objectif est d'aider les archéologues à rendre leur données interopérables et à ouvrir leur jeux de données vers le web sémantique.
- L'hétérogénéité des jeux de données à travers la France crée un besoin de standardisation.

Les référentiels



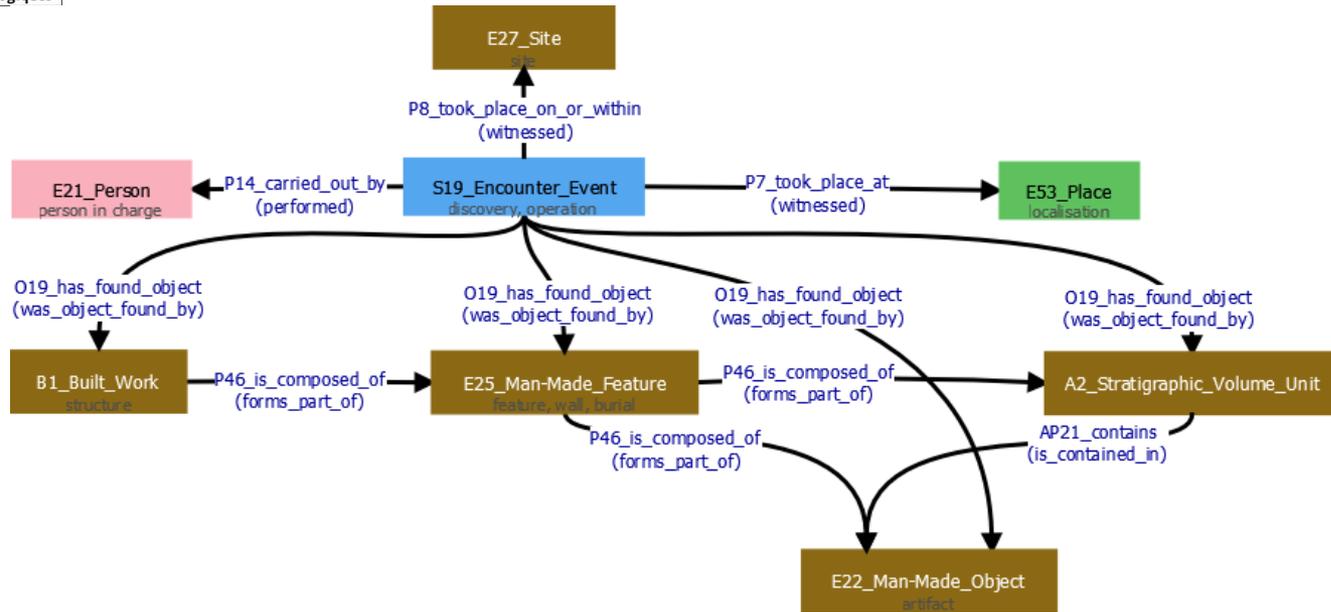
L'expression terminologique, temporelle et spatiale des jeux de données sont alignées sur des vocabulaires standards au maximum afin d'apporter une interopérabilité des données. Les personnes doivent quant à elle être représentés par des URI standards provenant de VIAF, ORCID ou HAL.

L'interopérabilité



En plus de l'interopérabilité des données, une structure commune était nécessaire afin que les jeux de données hétérogènes en format et en structure, puissent être interrogeable en même temps.

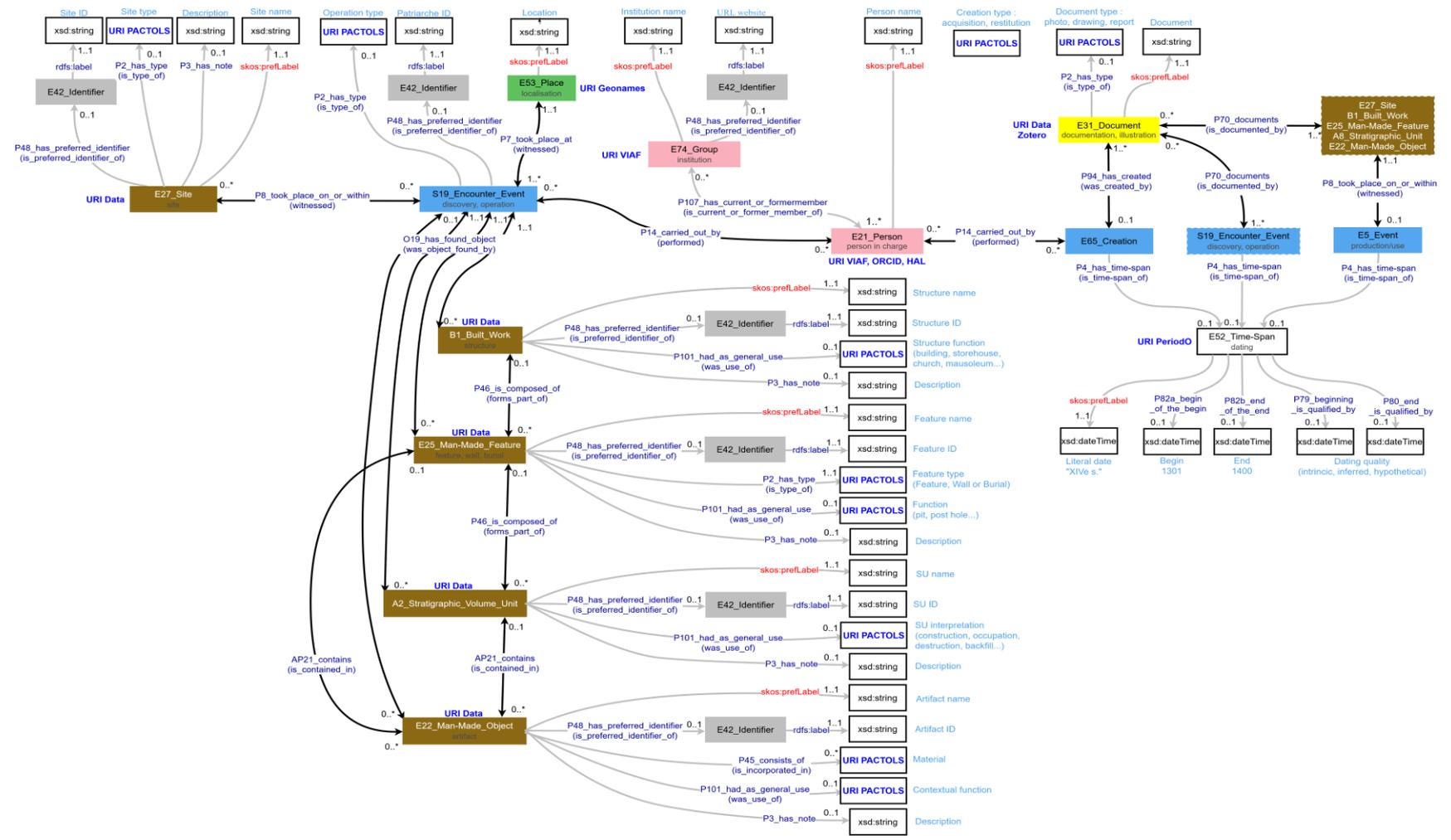
La solution du Consortium MASA se porta sur l'utilisation de l'ontologie standard du CIDOC-CRM et de ses extensions, pour obtenir une structure partagée par tous les jeux de données.



Le modèle générique de OpenArcheo utilise donc des entités et des propriétés du CIDOC-CRM (et des extensions CRMba, CRMsci et CRMarchaeo) pour représenter les concepts archéologiques que les jeux de données ont le plus de probabilité d'avoir en commun (site, découverte, artefact, fait, stratigraphie.).

L'objectif n'est pas de représenter les concepts les plus complexes et spécifiques de l'archéologie en utilisant le CIDOC-CRM, mais de faciliter le croisement des différents jeux de données et des informations décrites par ces données.

Le modèle générique



Les outils de mapping avec le CIDOC



XML dataset ▶ 3M



DBMS ▶ Ontop

In Progress
HRID: 134305 OUTAGR_Mapping_OpenArc...

Mapping #1 - Named Graph (Optionally add one)

Link	Source Node	Source Relation	Target Entity	Target Relation
Link #1	/scopecontent/localisation/@geomname exists	@id	E27_Site site	P48_has_preferred_identifier
Link #2	/ead/archdesc/desc/c01/c02	@id	E42_Identifier	
Link #3	/ead/archdesc/desc/c01/c02		P80_witnessed S19_Inventory_Event decouverte	
Link #4	./biblio	./biblio	P70i_is_documented_in E31_Document document	
Link #5	scopecontent/subject	@normal exists AND @source equals contente de site.	P2_has_type E55_Type	
Link #6	scopecontent/(descendant::date[1])		P80_witnessed	

cidoc-crm (http://www.cidoc-crm.org/cidoc-crm/) : [C:\CITERES\MASA\Atelier_ontologie\ontologieCIDOC.ttl]

Active Ontology: cidoc-crm (http://www.cidoc-crm.org/cidoc-crm/)

Class hierarchy: Thing, Entité CRM, I5_Inference_Making, S4_Observasion

Mapping editor: Mapping Assistant - BETA

Datasource selection: Select datasource: testarcheo

Mapping manager: Create, Remove, Copy, Select all, Select none

MOB-G_usage-01: arsol:imobi{CoSIUSObj} :P2_has_type arsol:imobi{CoSIUSObj}#usage .
SELECT CoSIUSObj FROM mobilier

MOB-G_usage-02: arsol:imobi{CoSIUSObj}#usage a :E55_Type .
SELECT CoSIUSObj FROM mobilier

MOB-G_usage-03: arsol:imobi{CoSIUSObj}#usage :P1_is_identified_by arsol:imobi{CoSIUSObj}#usageid .
SELECT CoSIUSObj FROM mobilier

MOB-G_usage-04: arsol:imobi{CoSIUSObj}#usageid a :E41_Appellation .
SELECT CoSIUSObj FROM mobilier

MOB-G_usage-05: arsol:imobi{CoSIUSObj}#usageid :P3_has_note {G_usage}^^rdfs:Literal .
SELECT CoSIUSObj, G_usage FROM mobilier



<http://139.91.183.3/3M/>
<https://demos.isl.ics.forth.gr/3m/>



Validation results of 13 shapes

Download validation report in [CSV](#) [Turtle](#) [RDF/XML](#)

663 Violations

- Violation** 592 "Property needs to have at least 1 values, but found 0" (crm:P8_took_place_on_or_within|^crm:P8i_witnessed in shape oash:P80) [see details](#)
- Violation** 30 "Property needs to have at least 1 values, but found 0" (crm:P4_has_time-span in shape oash:P81) [see details](#)
- Violation** 19 "Property may only have 1 value, but found 2" (skos:prefLabel in shape oash:P32) [see details](#)
- Violation** 11 "Property may only have 1 value, but found 2" (skos:prefLabel in shape oash:P35) [see details](#)
- Violation** 8 "Property needs to have at least 1 values, but found 0" (crmsci:O19i_was_object_found_by|^crmsci:O19_has_found_object in shape oash:P51) [see details](#)
- Violation** 1 "Property may only have 1 value, but found 13" (skos:prefLabel in shape oash:P35) [see details](#)
- Violation** 1 "Property may only have 1 value, but found 3" (skos:prefLabel in shape oash:P32) [see details](#)
- Violation** 1 "Property may only have 1 value, but found 3" (skos:prefLabel in shape oash:P35) [see details](#)

SHACL Play! by Thomas Francart **SPARNA** | version : 0.2 (timestamp) | SHACL Play! embeds [TobBraid SHACL API](#) from [TopQuadrant](#)

Chaque graphe RDF produit, avec les outils d'appariement 3M/Ontop, est ensuite vérifié via SHACL.

SHACL utilise un fichier représentant les « formes » du modèle générique (soit tous les triplets possibles utilisés dans le modèle) et vérifie si le graphe de données RDF représente correctement ces « formes ».

OpenArcheo Fédération SPARQL API Administration fr

Requêtes

- ARSOL : Liste des sépultures trouvées dans un site
- ARSOL : Nombre de sites découverts par découvreur
- ARSOL : Faits contenant du mobilier en fer
- ARSOL : Monnaies trouvées dans le cadre d'une fouille
- ARSOL : Responsables qui ont trouvé des sépultures
- AERBA : Responsables ayant trouvés un site par prospection aérienne daté avec fin postérieure à 400
- AERBA : Site trouvé dans la commune de Cormainville
- ARSOL/AERBA : Sites où un fossé à été mis à jour
- ARSOL/AERBA : Liste de tous les E27_Site
- Explorer une URI connue
- Afficher une frise chronologique

```

1 PREFIX wficheweb: <http://arsol.univ-tours.fr/4DACTION/WFICHEWEB/>
2 PREFIX crmsci: <http://www.ics.forth.gr/is1/CRMsCi/>
3 PREFIX crmba: <http://www.ics.forth.gr/is1/CRMba/>
4 PREFIX crmarch: <http://www.ics.forth.gr/is1/CRMarchaeo/>
5 PREFIX crm: <http://www.cidoc-crm.org/cidoc-crm/>
6 PREFIX frantiq: <https://ark.frantiq.fr/ark:/26678/>
7 PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
8 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
9 PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
10 SELECT DISTINCT ?person ?personLabel
11 FROM NAMED <http://openarchaeo.huma-num.fr/federation/sources/arsol>
12 WHERE {
13   ?person a crm:E21_Person .
14   ?person skos:prefLabel ?personLabel .
15   ?person crm:P14i_performed/crm:P8_took_place_on_or_within ?site .
16   ?site crm:P8i_witnessed/crmsci:019_has_found_object ?sepulture .
17   ?sepulture a crm:E25_Man-Made_Feature .
18   ?sepulture crm:P2_has_type <https://ark.frantiq.fr/ark:/26678/pcrt795b632nlw> .
19 }
                
```

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

Showing 1 to 2 of 2 entries Search: Show 50 entries

Un endpoint SPARQL est mis à disposition sur la plateforme pour pouvoir interroger les graphes de données de manière traditionnelle.

 Choisir les sources de données à interroger

Recherche

Rechercher...

Filter par mots-clé

Algérie (1) Beauce (1) Belgique (2) Centre (1)
 Cherchell (1) Chinon (1) Chypre (2) EpiCherchell (1)
 France (3) Gallo-Romain (1)

Filter par période

-6500 / 2020

Filter par lieu

Africa (1) Algeria (1) Arrondissement de Tours (1)
 Beauce (1) Centre-Val de Loire (4) Cherchell (1)
 Cyprus (1) Eure-et-Loir (1) Europe (8) France (6)

Filter par collection

Enregistrement de terrain (3) Inventaire de sites (7)

AERBA

Atlas des établissements ruraux de la Beauce antique

Détails... ▾

ArSoL

Archives du Sol - Enregistrement des fouilles archéologiques menées par le Laboratoire Archéologie et Territoire de Tours

Détails... ▾

Chronique des fouilles en ligne

La Chronique des fouilles en ligne est une base de données richement illustrée, disposant d'un outil cartographique pour localiser les opérations de terrain ayant eu lieu en Grèce et à Chypre.

Détails... ▾

EpiCherchell

Corpus des inscriptions antiques de Césarée de Maurétanie (Cherchell, Algérie)

Détails... ▾

ICERAMM

Information sur la CÉRAmique Médiévale et Moderne - Lots céramiques typochronologiques du réseau national

Détails... ▾

Kition-Pervolia

Données de fouilles du site de Kition-Pervolia

Détails... ▾

OUTAGR

Outillage Agricole Gallo-Romain - Inventaire de l'outillage agricole Gallo-Romain

Détails... ▾

RITA

Recherche en Image - Tours Archéologique, base de données des différentes sources visuelles archéologiques de la ville de Tours

Détails... ▾

SOLiDAR

Étude de la dynamique de l'occupation humaine dans les massifs de Blois, Russy, Boulogne et Chambord (Loir-et-Cher, France) des origines à nos jours, à partir d'une campagne de télédétection LIDAR (Light Detection And Ranging) de 270 km2

Détails... ▾

VALIDER

L'originalité de l'instance Sparnatural de OpenArcheo, développée en 2018, repose sur l'interrogation d'un ou plusieurs graphes de données. L'utilisateur peut utiliser des filtres pour mieux choisir le ou les jeux de données que l'on souhaite interroger.



✓ Query successful! - View/hide SPARQL query

```

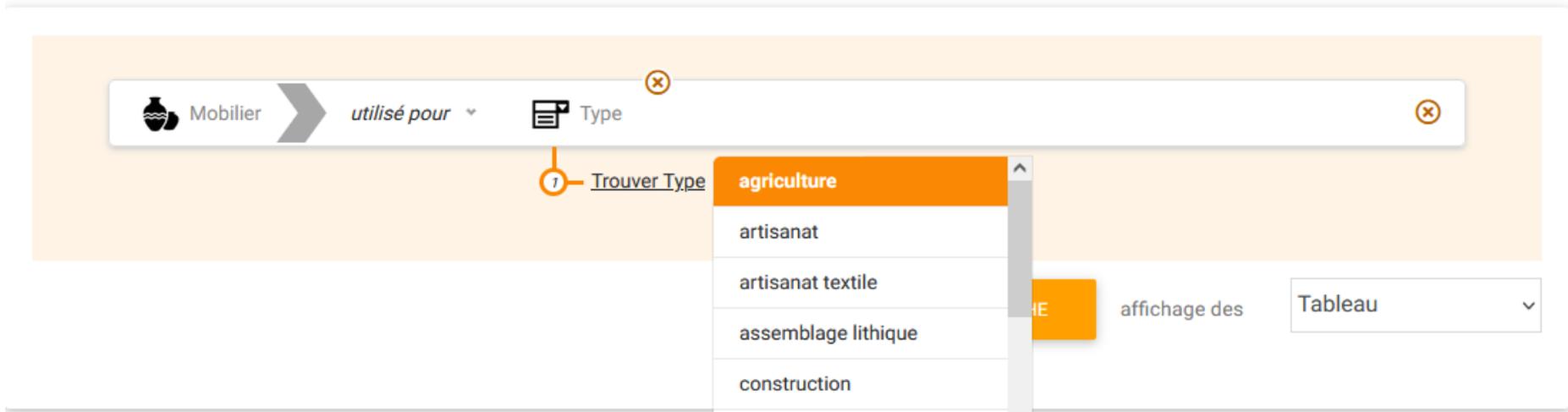
1 SELECT DISTINCT ?this ?thisLabel
2 FROM NAMED <http://openarchaeo.huma-num.fr/federation/sources/arsol>
3 WHERE
4 {
5   ?this a
6     <http://www.cidoc-crm.org/cidoc-crm/E25_Man-Made_Feature> ;
7     <http://www.cidoc-crm.org/cidoc-crm/P2_has_type> <https://ark.frantiqu.fr/ark:/26678/pcrt795b632nMw> .
8   ?this <http://www.ics.forth.gr/isl/CRMsci/019i_was_object_found_by>/<http://www.cidoc-crm.org/cidoc-crm/P8_took_place>
9     ?Site1 a
10      <http://www.cidoc-crm.org/cidoc-crm/E27_Site> .
11   ?Site1 <http://www.cidoc-crm.org/cidoc-crm/P8i_witnessed>/<http://www.cidoc-crm.org/cidoc-crm/P14_carried_out_by> <
12   OPTIONAL
13     { ?this <http://www.w3.org/2004/02/skos/core#prefLabel> ?thisLabel }
14 }
    
```

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

Showing 1 to 50 of 722 entries Search: Show 50 entries

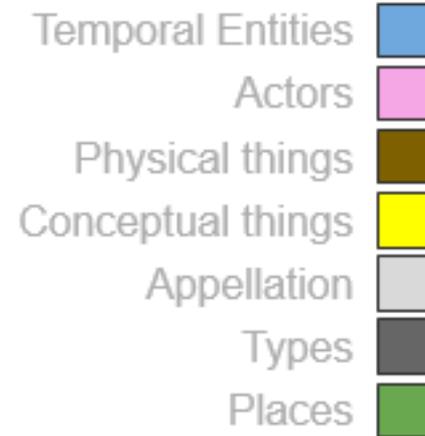
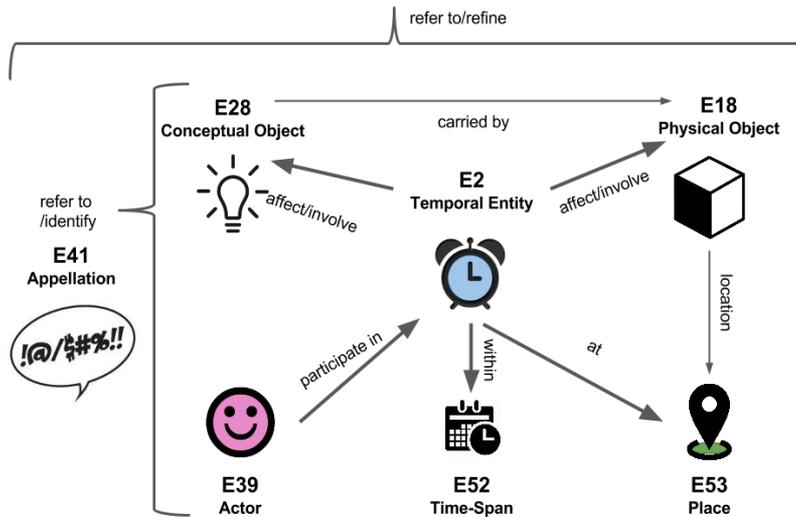
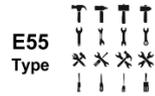
this	thisLabel
1 http://arsol.univ-tours.fr/4DACTION/WFICHEWEB/isepuAJ000001	Sepulture AJ000001
2 http://arsol.univ-tours.fr/4DACTION/WFICHEWEB/isepuAJ000002	Sepulture AJ000002
3 http://arsol.univ-tours.fr/4DACTION/WFICHEWEB/isepuAJ000003	Sepulture AJ000003

- La configuration du Sparnatural utilisée par OpenArcheo répond à un besoin d'édition du CIDOC-CRM afin de faire comprendre la philosophie de modélisation de cette ontologie à l'utilisateur du système de requête.
- Par exemple : les instances de E22_Man-Made_Object sont désignés comme des mobiliers dans OpenArcheo.



L'utilisation de référentiel commun comme PACTOLS ou VIAF permet à la plateforme d'utiliser les URIs dans des listes déroulantes et ainsi rechercher à l'aide de la même expression sur plusieurs jeux de données.

Apprendre en jouant : jeu de carte CIDOC-CRM



L'apprentissage du CIDOC-CRM demande un investissement important pour comprendre les subtilités et mécanisme de l'ontologie. Pour que les chercheurs comprennent les enjeux de l'utilisation de cette ontologie, George Brusseker et Anaïs Guilhem ont créé un jeu de cartes pour aider les chercheur à la comprendre.

Apprendre en jouant : jeu de carte CIDOC-CRM

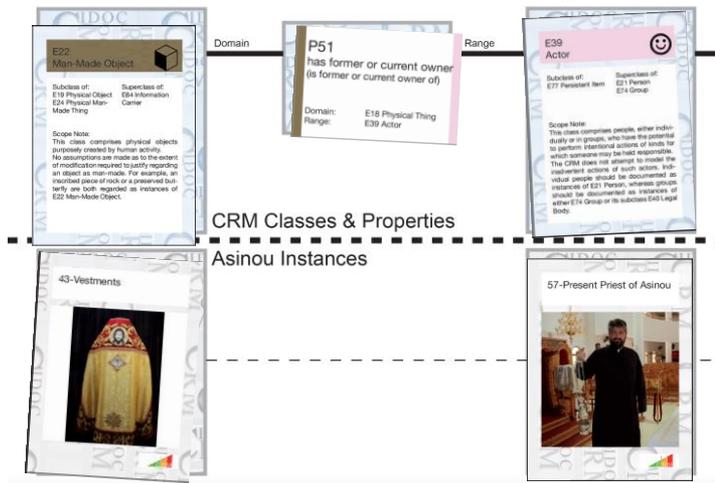
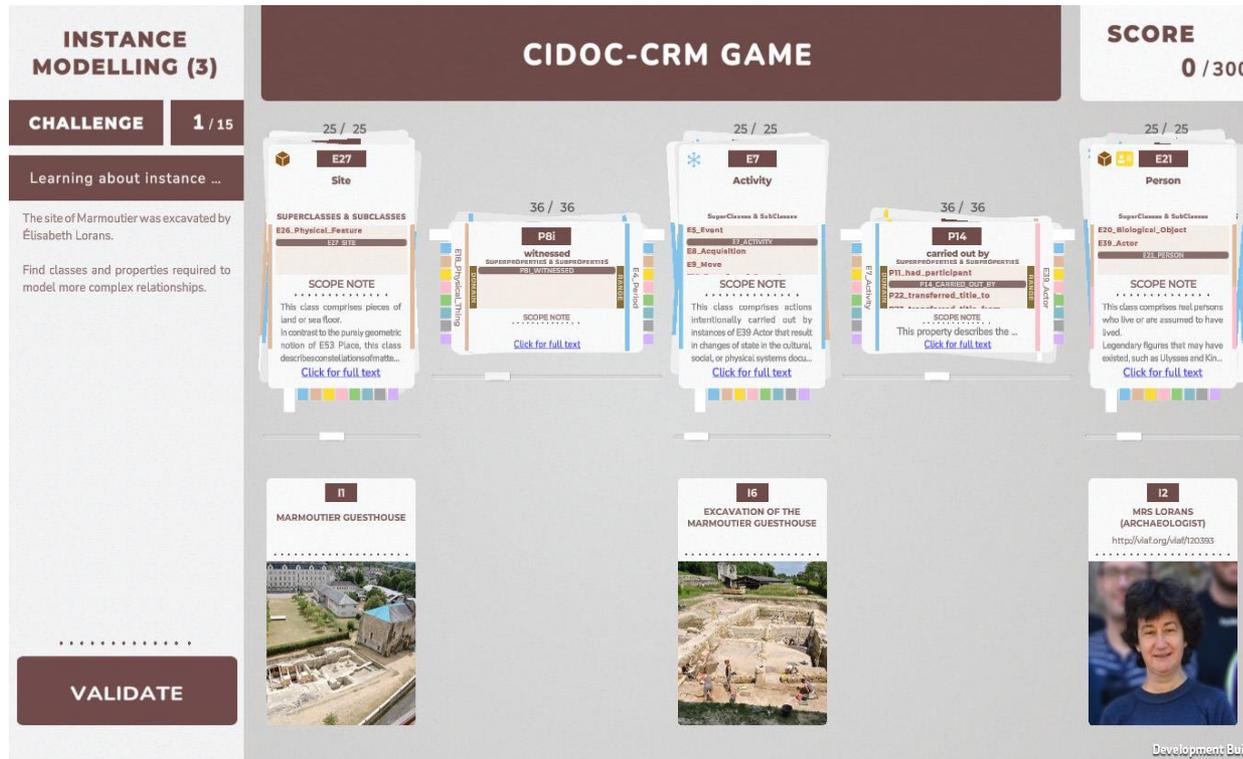


Photo from ITN-DCH workshop game play

Les cartes proposent un système de couleurs et d'icônes associées aux différentes branches du CIDOC.

On peut lire un minimum d'informations sur les cartes pour vérifier la validité de l'utilisation d'une propriété entre deux entités, ou la validité d'un appariement entre un individu et une entité.

Apprendre en jouant : jeu de carte CIDOC-CRM



Le Consortium MASA a également participé au développement d'une version virtuelle du jeu de carte CIDOC-CRM avec George Bruseker et Anaïs Guillem, afin de faciliter l'utilisation et l'accès du jeu de carte. La version virtuelle est entièrement paramétrable pour y incorporer une autre ontologie que le CIDOC-CRM. Une version en ligne devrait bientôt être disponible.

Merci

<https://masa.hypotheses.org/>

florian.hivert@univ-tours.fr