

# Utilisation des LLM pour l'extraction automatique d'annotations conformes au modèle CIDOC-CRM

**HARIRI Ali**

ali.hariri@ensma.fr

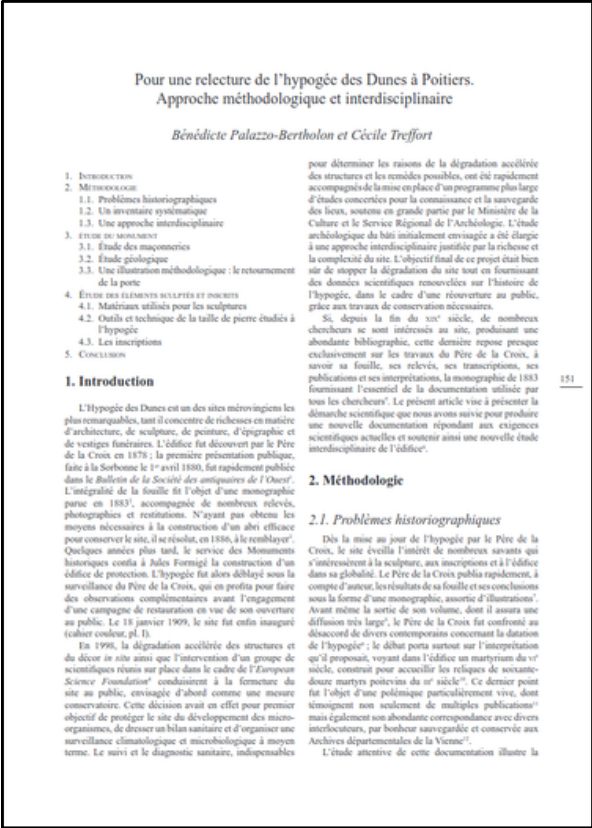
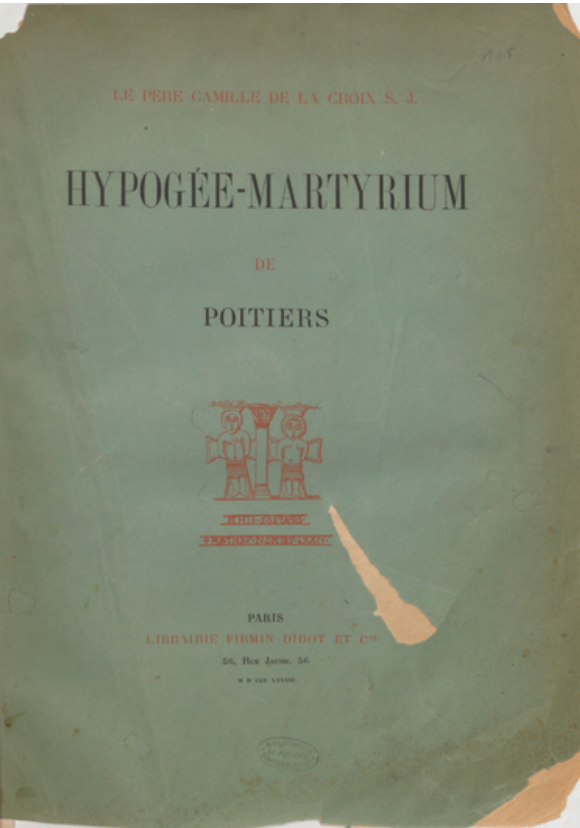
<https://www.lias-lab.fr/members/alihariri>

# RAPPEL DE L'ÉTUDE DE CAS

## Hypogée des Dunes

### Les documents

- Données non structurées
  - Articles scientifiques
  - Documents de travail
- Données structurées
  - 8 tableurs Excel
    - inventaires
    - résultats d'analyse
  - document Word (inventaire des outils de taille)



|    |                                                    |                           |              |                                                                                                                                                                                                                                                                                                                |                                                                         |             |                                                |
|----|----------------------------------------------------|---------------------------|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|-------------|------------------------------------------------|
| 1  | HYPOGEE DES DUNES : LISTE DES FAITS version 050215 |                           |              |                                                                                                                                                                                                                                                                                                                |                                                                         |             |                                                |
| 2  |                                                    |                           |              |                                                                                                                                                                                                                                                                                                                |                                                                         |             |                                                |
| 3  | FAIT                                               | n° US                     | Localisation | Description                                                                                                                                                                                                                                                                                                    | Etendue                                                                 | Antérieur à | Postérieur à                                   |
| 4  |                                                    |                           |              |                                                                                                                                                                                                                                                                                                                |                                                                         |             | Egal à                                         |
| 5  | FAIT 1                                             | Limite entre US03 et US05 | Mur est      | Décrochement de la maçonnerie entre les deux assises. L'assise supérieure est construite X cm en retrait par rapport à l'assise inférieure.                                                                                                                                                                    |                                                                         |             |                                                |
| 6  | FAIT 2                                             | US02                      | Mur est      | Construction d'une maçonnerie avec des moellons calcaires allongés portant des traces d'outil fraîche en surface. Trois moellons sont positionnés est-ouest comme des pierres d'attente.                                                                                                                       | Extrémité sud du mur est.                                               |             | US01, US03, US04, US05, US06, US07, US08       |
| 7  | FAIT 3                                             | US04 et US07              | Mur est      | Percement d'une baie à simple ébrasement : construction du glacis avec des petits moellons calcaires et des morceaux de travertin. La maçonnerie est surplombée par une dalle qui servirait d'appui extérieur de la baie.                                                                                      | Centre du mur est.                                                      | US02        |                                                |
| 8  | FAIT 4                                             | US05                      | Mur est      | Construction de la partie supérieure du mur, avec des moellons de travertin et des moellons calcaires. Cette maçonnerie est posée en retrait de X cm sur l'assise inférieure (US03).                                                                                                                           | Partie nord du mur est.                                                 |             | US03                                           |
| 9  | FAIT 5                                             | US01 et US05 ?            | Mur est      | Application d'un enduit de surface sur le parement du mur (US03). Il est couvert d'un badigeon blanc sur la totalité et conserve un décor peint (encadrement) dans le prolongement de l'ébrasement sud de la baie. De plus, il conserve en dessous du décor peint, une inscription peinte en rouge [monus...]. | Partie basse du mur est.                                                | US02        | US03 et US04 et US05 ? US06 ?                  |
| 10 | FAIT 6                                             | US03, US02, US04, US05    | Mur est      | Réfection des joints avec le mortier de type C (= restauration Formigé).                                                                                                                                                                                                                                       | Partie supérieure du mur est.                                           |             | US01, US03, US04, US05, US06, US07, US08, US02 |
| 11 | FAIT 7                                             | US18                      | Mur ouest    | Restauration XXe : enlèvement du bouchage US17 sur la partie haute pour mettre en place les barres métalliques qui fixent ensemble les parties haute et basse du jambage sud de la POR1.                                                                                                                       | Entre le jambage sud de la POR1 et la limite du mur ouest ancien (US12) |             | US17 US14                                      |

# RAPPEL

Objectif

Fournir une structure de représentation des données garantissant  
**la sûreté de l'information et l'interopérabilité**

# RAPPEL

Objectif

Fournir une structure de représentation des données garantissant  
**la sûreté de l'information et l'interopérabilité**

Les ontologies et les graphes de connaissances répondent à ce besoin

# RAPPEL

Objectif

Fournir une structure de représentation des données garantissant  
**la sûreté de l'information et l'interopérabilité**

Les ontologies et les graphes de connaissances répondent à ce besoin

**Mais**

Introduisent une couche de complexité technique pour l'utilisateur final (l'archéologue)

# RAPPEL

Objectif

Fournir une structure de représentation des données garantissant  
**la sûreté de l'information et l'interopérabilité**

Les ontologies et les graphes de connaissances répondent à ce besoin

**Mais**

Introduisent une couche de complexité technique pour l'utilisateur final (l'archéologue)

**La modélisation**

**Création de graphes de connaissances**

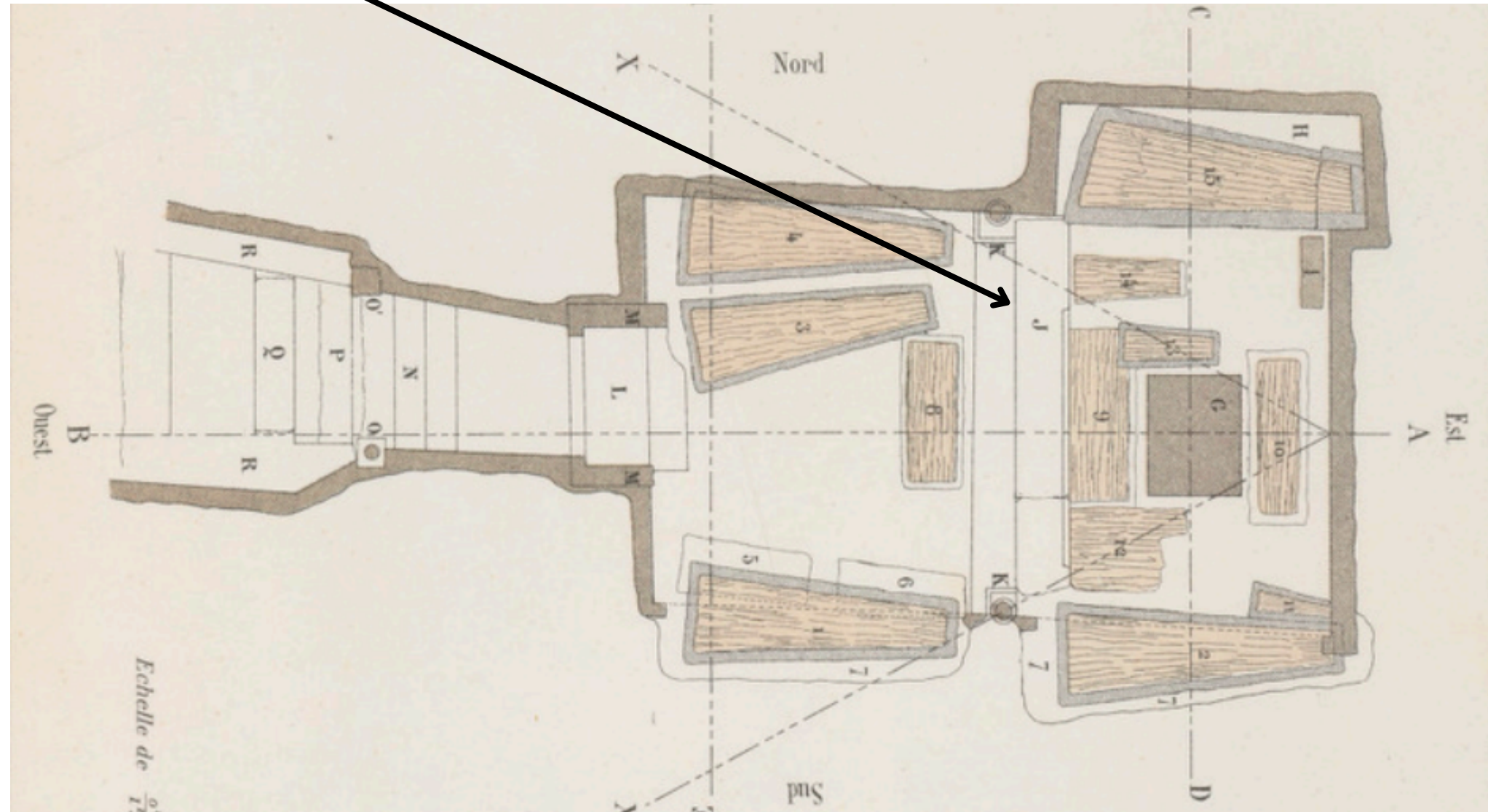
# **RAPPEL**

Problématique

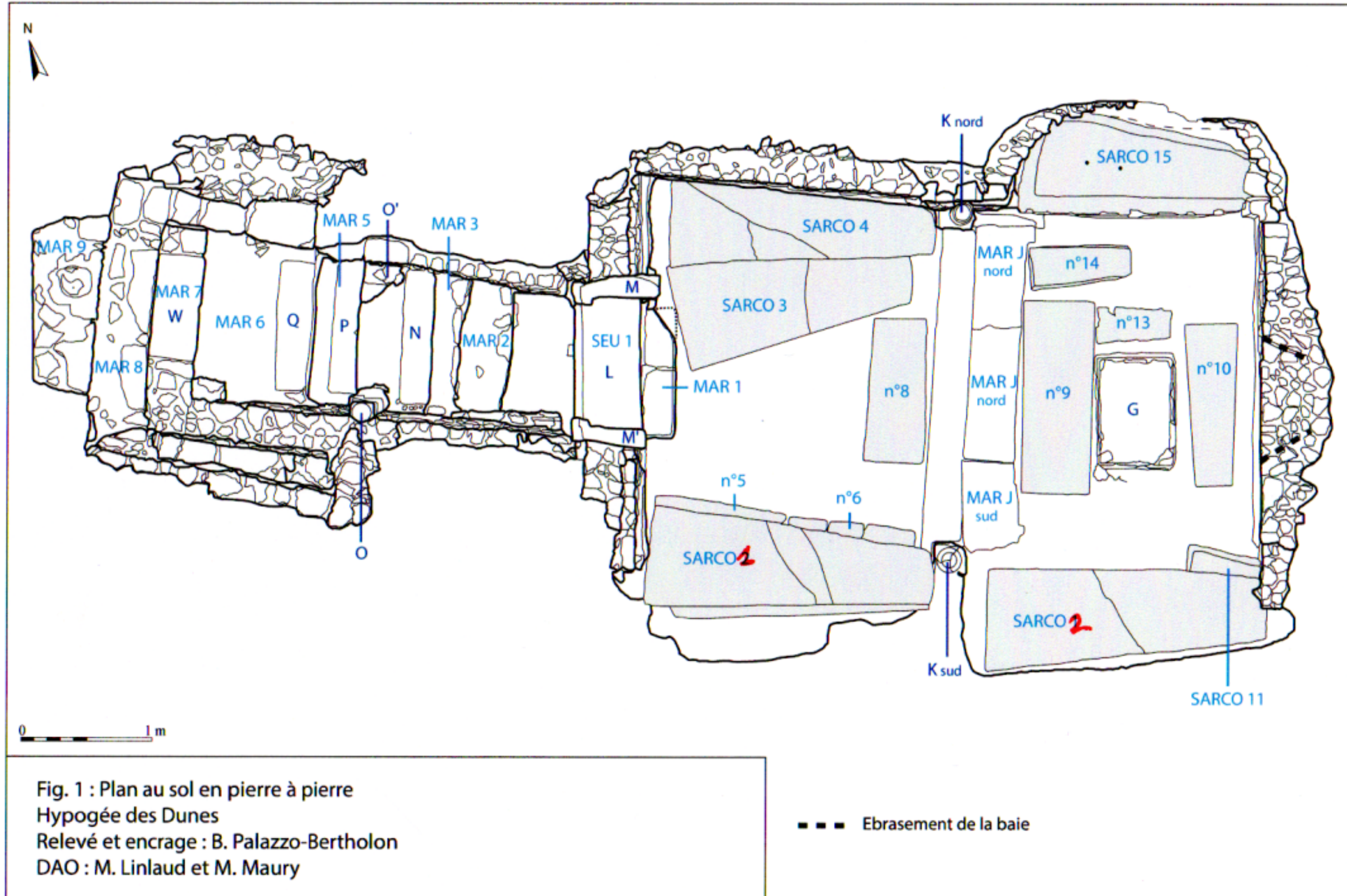
Comment automatiser l'étape de transformation des données  
(modélisation et création du graphe de connaissances) ?

# EXEMPLE

Marche J



# EXEMPLE



# EXEMPLE

## Peuplement de CIDOC CRM

| Champ        | Valeur                         |
|--------------|--------------------------------|
| n° US        | US29                           |
| Localisation | Entre la salle 1 et la salle 2 |
| Description  | Marche sculptée (marche J)     |

L'élément lapidaire portant le numéro d'inventaire **2008.0.6.171 (fragment a)** est une marche sculptée, inscrite et peinte, dite « **marche J** », en **calcaire Callovien inférieur**. Ses dimensions sont : **hauteur maximale 14 cm (minimale 11 cm), longueur 183 cm, profondeur maximale 45 cm**. Elle présente un **décor sculpté de rosaces, une inscription et des cavités à incrustations**. Elle est datée du **Haut Moyen Âge**. Sa provenance est située **entre la salle 1 et la salle 2**, partie nord, et correspond à l'unité stratigraphique **US29**. Elle est **conservée in situ**

| Champ                      | Valeur                                                          |
|----------------------------|-----------------------------------------------------------------|
| n° inv. Musées de France   | 2008.0.6.171                                                    |
| Fragment n°                | a                                                               |
| N° d'inv. Père de La Croix | J (partie nord)                                                 |
| Dénomination               | Marche sculptée inscrite et peinte dite "marche J"              |
| Matériaux                  | Calcaire Callovien inférieur                                    |
| Dimensions                 | H. 14 (max.) et 11 (min.) ; L. 183 ; Pr. 45 (max.) cm           |
| Présence d'inscription     | Décor sculpté : rosaces, inscription et cavités à incrustations |
| Datation                   | Haut Moyen Âge                                                  |
| Provenance                 | Entre la salle 1 et la salle 2, partie nord, <b>US29</b>        |
| Lieu de conservation       | In situ                                                         |

# EXEMPLE

## Peuplement de CIDOC CRM

|    |                                                                 |                           |              |                                                                                                                                                                                                                                                                                                                                 |                                                                         |             |                         |                        |      |         |
|----|-----------------------------------------------------------------|---------------------------|--------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------|-------------|-------------------------|------------------------|------|---------|
| 2  | HYPOGÉE DES DUNES : LISTE DES FAITS <span>version 050215</span> |                           |              |                                                                                                                                                                                                                                                                                                                                 |                                                                         |             |                         |                        |      |         |
| 3  |                                                                 |                           |              |                                                                                                                                                                                                                                                                                                                                 |                                                                         |             |                         |                        |      |         |
| 4  | FAIT                                                            | n° US                     | Localisation | Description                                                                                                                                                                                                                                                                                                                     | Etendue                                                                 | Antérieur à | Postérieur à            | Egal à                 | Etat | Mortier |
| 5  | FAIT 1                                                          | Limite entre US03 et US05 | Mur est      | Décrochement de la maçonnerie entre les deux assises. L'assise supérieure est construite X cm en <span>retrait</span> par rapport à l'assise inférieure.                                                                                                                                                                        |                                                                         |             |                         |                        |      |         |
| 6  | FAIT 2                                                          | US02                      | Mur est      | Construction d'une maçonnerie avec des moellons calcaires allongés portant des traces d'outil fraîche en surface. Trois moellons sont positionnés est-ouest comme des pierres d'attente.                                                                                                                                        | Extrémité sud du mur est.                                               |             | US01, US04, US06, US08  | US03, US05, US07       |      |         |
| 7  | FAIT 3                                                          | US04 et US07              | Mur est      | Percement d'une baie à simple ébrasement : construction du glacis avec des petits moellons calcaires et des morceaux de travertin. La maçonnerie est surplombée par une dalle qui servirait d'appui extérieur de la baie.                                                                                                       | Centre du mur est.                                                      | US02        |                         |                        |      |         |
| 8  | FAIT 4                                                          | US05                      | Mur est      | Construction de la partie supérieure du mur, avec des moellons de travertin et des moellons calcaires. Cette maçonnerie est posée en retrait de X cm sur l'assise inférieure (US03).                                                                                                                                            | Partie nord du mur est.                                                 |             | US03                    |                        |      |         |
| 9  | FAIT 5                                                          | US01 et US05 ?            | Mur est      | Application d'un enduit de surface sur le parement du mur (US03). Il est couvert d'un badigeon blanc sur la totalité et conserve un décor peint (encadrement) dans le prolongement de l'ébrasement sud de la baie. De plus, il conserve en dessous du décor peint, une inscription peinte en rouge [ <span>monetus</span> ...]. | Partie basse du mur est.                                                | US02        | US03 et US04 et US05 ?  | US06 ?                 |      |         |
| 10 | FAIT 6                                                          | US03, US02, US04, US05    | Mur est      | Réfection des joints avec le mortier de type C (= restauration <span>Formigé</span> ).                                                                                                                                                                                                                                          | Partie supérieure du mur est.                                           |             | US01, US04, US06, US08. | US03, US05, US07, US02 |      |         |
| 11 | FAIT 7                                                          | US18                      | Mur ouest    | <b>Restauration XXe</b> : enlèvement du bouchage US17 sur la partie haute pour mettre en place les barres métalliques qui fixent ensemble les parties haute et basse du jambage sud de la POR1.                                                                                                                                 | Entre le jambage sud de la POR1 et la limite du mur ouest ancien (US12) |             | US17                    | US14                   |      |         |

|    |                                           |              |                                                                                                     |                                                                                              |                          |              |        |                                                |         |                    |  |  |  |
|----|-------------------------------------------|--------------|-----------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|--------------------------|--------------|--------|------------------------------------------------|---------|--------------------|--|--|--|
| 1  | HYPOGÉE DES DUNES : ENREGISTREMENT DES US |              |                                                                                                     |                                                                                              |                          |              |        |                                                |         | VERSION 11:12:2015 |  |  |  |
| 2  |                                           |              |                                                                                                     |                                                                                              |                          |              |        |                                                |         |                    |  |  |  |
| 3  |                                           |              |                                                                                                     |                                                                                              |                          |              |        |                                                |         |                    |  |  |  |
| 4  | n° US                                     | Localisation | Description                                                                                         | Etendue                                                                                      | Antérieur à              | Postérieur à | Egal à | Etat                                           | Mortier |                    |  |  |  |
| 5  |                                           |              |                                                                                                     |                                                                                              |                          |              |        |                                                |         |                    |  |  |  |
| 6  | US01                                      | Mur est      | Enduit de surface + badigeon blanc + motif et inscription peints [monetus]                          | Depuis la base jusqu'à 50 à 120 cm de haut.                                                  |                          |              |        | Etat 2                                         |         |                    |  |  |  |
| 7  | US60                                      | Mur est      | Négatif de l'angle sud-est du mur (US03 et US136) détruit.                                          | Angle sud-est de l'hypogée.                                                                  | US59, US58, US02         | US01, US03   |        | Etat 1                                         |         |                    |  |  |  |
| 8  | US02                                      | Mur est      | Maçonnerie de moellons calcaires allongés + traces d'outil fraîches en surface.                     | extrémité sud du mur, partie haute.                                                          |                          |              |        | Restauration Formigé                           |         |                    |  |  |  |
| 9  | US03                                      | Mur est      | Maçonnerie de petits moellons calcaires liés au mortier                                             |                                                                                              | US01 / US02              |              |        | Etat 1                                         |         |                    |  |  |  |
| 10 | US04                                      | Mur est      | Glacis de la baie, maçonné avec des moellons calcaires et du travertin.                             | Glacis de la baie située au centre du mur.                                                   | US02                     |              |        | Etat 2                                         |         |                    |  |  |  |
| 11 | US05                                      | Mur est      | Remontage de la maçonnerie avec des moellons de travertin et de petits moellons calcaires mélangés. | Partie haute du mur côté nord et angle nord-est descendant jusqu'à l'enduit (US01).          | US02                     | US03         | US04   | Etat 2                                         |         |                    |  |  |  |
| 12 | US06                                      | Mur est      | Plaque d'enduit de surface qui recouvre l'angle nord-est du mur.                                    | Angle nord-est du mur.                                                                       | US02                     | US03         | US01 ? | Etat 2                                         |         |                    |  |  |  |
| 13 | US07                                      | Mur est      | Dalle surplombant le glacis de la baie percée au centre du mur. Supposée associée à la baie.        | Au centre du mur, surplombant le mur conservé.                                               | US02                     | US03         | US04   | Etat 1                                         |         |                    |  |  |  |
| 14 | n° US                                     | Localisation | Description                                                                                         | Etendue                                                                                      | Antérieur à              | Postérieur à | Egal à | Etat                                           | Mortier |                    |  |  |  |
| 15 | US08                                      | Mur ouest    | Enduit de surface + badigeon blanc.                                                                 | Côté nord, depuis la base du mur jusqu'à 80 cm de haut. Côté sud, depuis 12 à 90 cm de haut. | US16, US15, US14         | US12, US09   |        | Etat 2                                         |         |                    |  |  |  |
| 16 | US09                                      | Mur ouest    | Mortier (type A) qui recouvre la base du mur, côté sud.                                             | Base du mur ouest, au sud de la POR 1.                                                       | US08, US15, US16, US14   |              | US12   | Etat 1                                         |         |                    |  |  |  |
| 17 | US10                                      | Mur ouest    | Scellement de la marche d'accès dans l'hypogée (MAR1) à l'est du seuil (SEU1).                      | Mur ouest, sous le seuil (SEU1) donnant accès à la salle de l'hypogée.                       | US14, US15, US16, US08 ? |              | US17   | 2 pour le scellement de MAR1 issue de l'état 1 | Type B  |                    |  |  |  |

# EXEMPLE

## Peuplement de CIDOC CRM

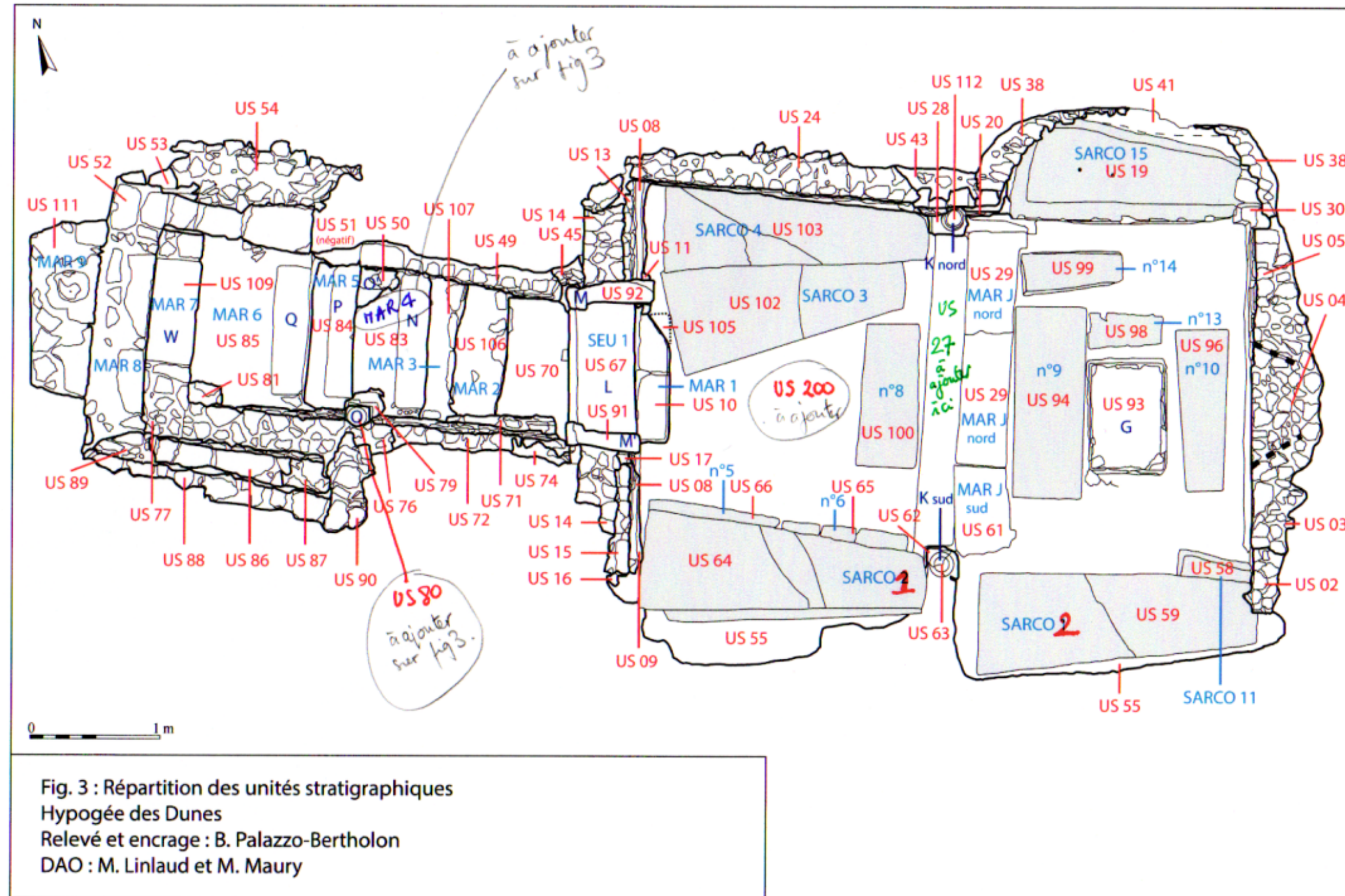
| Champ        | Valeur                         |
|--------------|--------------------------------|
| n° US        | US29                           |
| Localisation | Entre la salle 1 et la salle 2 |
| Description  | Marche sculptée (marche J)     |

L'élément lapidaire portant le numéro d'inventaire **2008.0.6.171 (fragment a)** est une marche sculptée, inscrite et peinte, dite « **marche J** », en **calcaire Callovien inférieur**. Ses dimensions sont : **hauteur maximale 14 cm (minimale 11 cm), longueur 183 cm, profondeur maximale 45 cm**. Elle présente un **décor sculpté de rosaces, une inscription et des cavités à incrustations**. Elle est datée du **Haut Moyen Âge**. Sa provenance est située **entre la salle 1 et la salle 2**, partie nord, et correspond à l'unité stratigraphique **US29**. Elle est **conservée in situ**

| Champ                      | Valeur                                                          |
|----------------------------|-----------------------------------------------------------------|
| n° inv. Musées de France   | 2008.0.6.171                                                    |
| Fragment n°                | a                                                               |
| N° d'inv. Père de La Croix | J (partie nord)                                                 |
| Dénomination               | Marche sculptée inscrite et peinte dite "marche J"              |
| Matériaux                  | Calcaire Callovien inférieur                                    |
| Dimensions                 | H. 14 (max.) et 11 (min.) ; L. 183 ; Pr. 45 (max.) cm           |
| Présence d'inscription     | Décor sculpté : rosaces, inscription et cavités à incrustations |
| Datation                   | Haut Moyen Âge                                                  |
| Provenance                 | Entre la salle 1 et la salle 2, partie nord, <b>US29</b>        |
| Lieu de conservation       | In situ                                                         |

# EXEMPLE

## Peuplement de CIDOC CRM



# EXEMPLE

## Peuplement de CIDOC CRM

| Champ        | Valeur                         |
|--------------|--------------------------------|
| n° US        | US29                           |
| Localisation | Entre la salle 1 et la salle 2 |
| Description  | Marche sculptée (marche J)     |

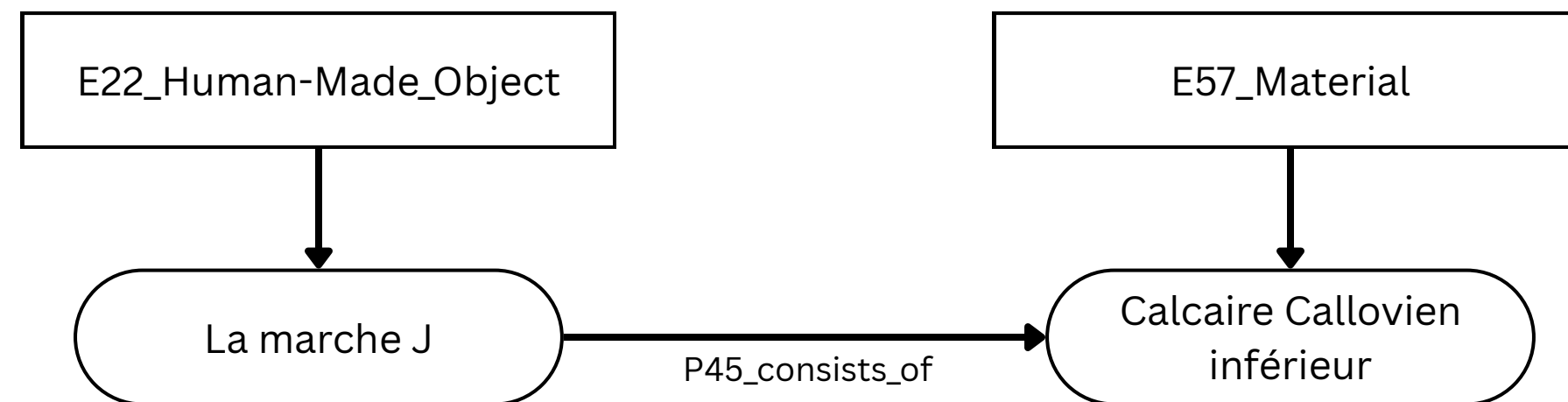
L'élément lapidaire portant le numéro d'inventaire **2008.0.6.171 (fragment a)** est une marche sculptée, inscrite et peinte, dite « **marche J** », en **calcaire Callovien inférieur**. Ses dimensions sont : **hauteur maximale 14 cm (minimale 11 cm), longueur 183 cm, profondeur maximale 45 cm**. Elle présente un **décor sculpté de rosaces, une inscription et des cavités à incrustations**. Elle est datée du **Haut Moyen Âge**. Sa provenance est située **entre la salle 1 et la salle 2**, partie nord, et correspond à l'unité stratigraphique **US29**. Elle est **conservée in situ**

| Champ                      | Valeur                                                          |
|----------------------------|-----------------------------------------------------------------|
| n° inv. Musées de France   | 2008.0.6.171                                                    |
| Fragment n°                | a                                                               |
| N° d'inv. Père de La Croix | J (partie nord)                                                 |
| Dénomination               | Marche sculptée inscrite et peinte dite "marche J"              |
| Matériaux                  | Calcaire Callovien inférieur                                    |
| Dimensions                 | H. 14 (max.) et 11 (min.) ; L. 183 ; Pr. 45 (max.) cm           |
| Présence d'inscription     | Décor sculpté : rosaces, inscription et cavités à incrustations |
| Datation                   | Haut Moyen Âge                                                  |
| Provenance                 | Entre la salle 1 et la salle 2, partie nord, <b>US29</b>        |
| Lieu de conservation       | In situ                                                         |

# EXEMPLE

Peuplement de CIDOC CRM

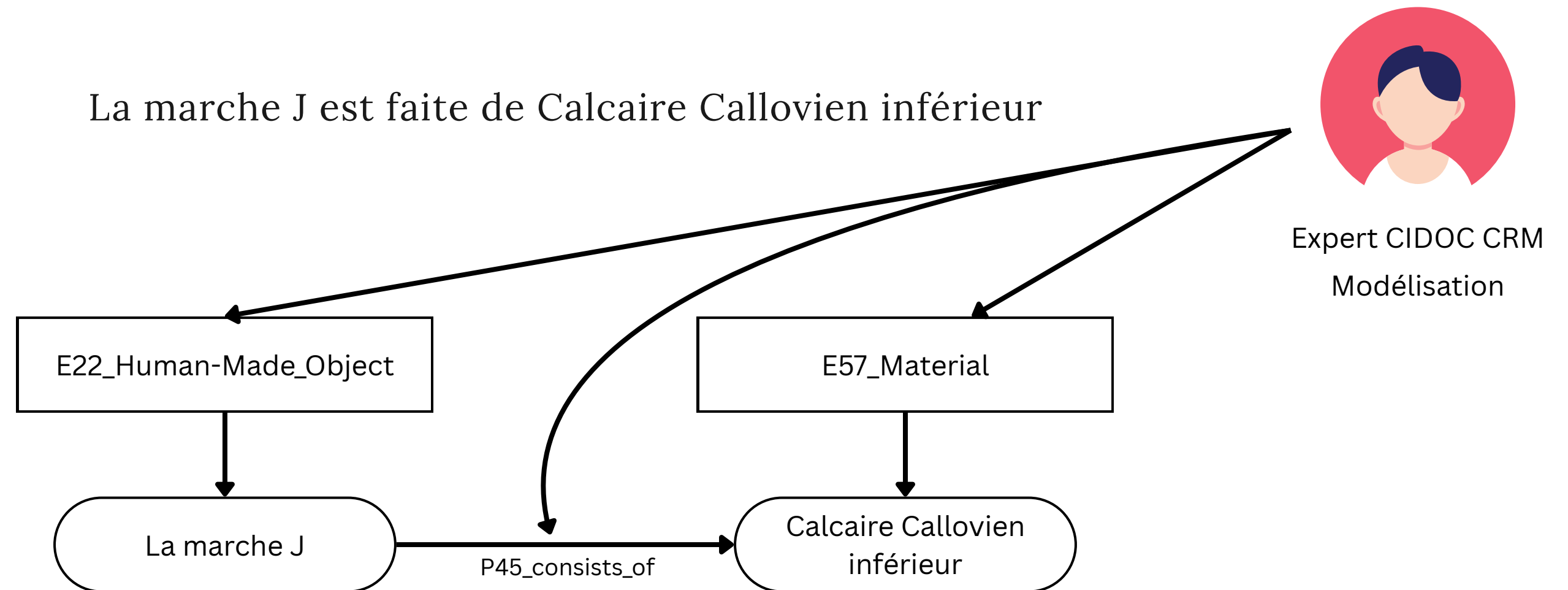
La marche J est faite de Calcaire Callovien inférieur



# EXEMPLE

Acteurs

La marche J est faite de Calcaire Callovien inférieur



Expert en archéologie  
Cohérence des données

# EXEMPLE

Acteurs

La marche J est faite de Calcaire Callovien inférieur



Expert CIDOC CRM  
Modélisation

E22\_Human-Made\_Object

E57\_Material

La marche J

Calcaire Callovien  
inférieur

P45\_consists\_of



Expert en archéologie  
Cohérence des données

# EXEMPLE

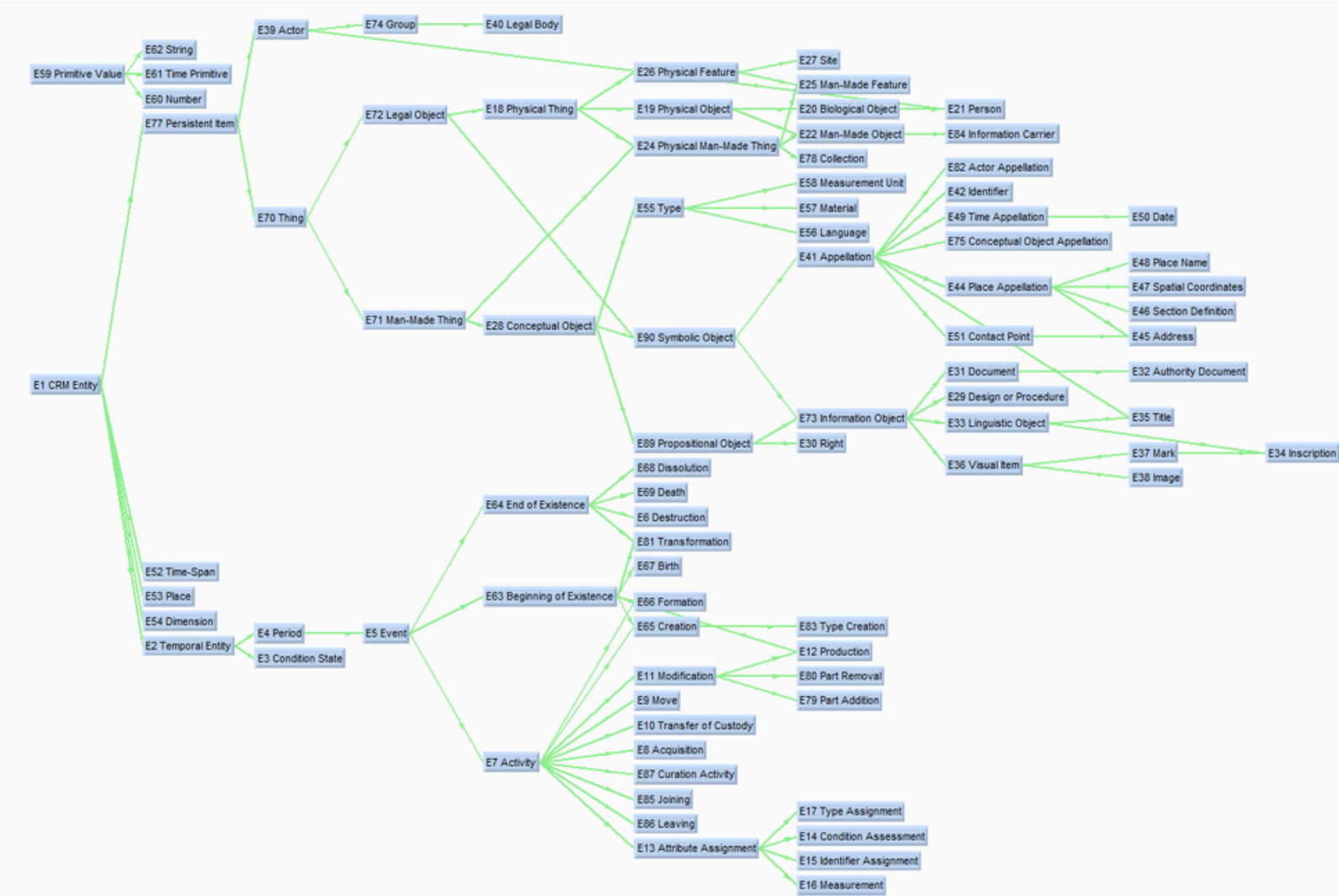
Peuplement de CIDOC CRM

L'élément lapidaire portant le numéro d'inventaire **2008.0.6.171 (fragment a)** est une marche sculptée, inscrite et peinte, dite « **marche J** », en **calcaire Callovien inférieur**. Ses dimensions sont : **hauteur maximale 14 cm (minimale 11 cm), longueur 183 cm, profondeur maximale 45 cm**. Elle présente un **décor sculpté de rosaces, une inscription et des cavités à incrustations**. Elle est datée du **Haut Moyen Âge**. Sa provenance est située **entre la salle 1 et la salle 2**, partie nord, et correspond à l'unité stratigraphique **US29**. Elle est **conservée in situ**

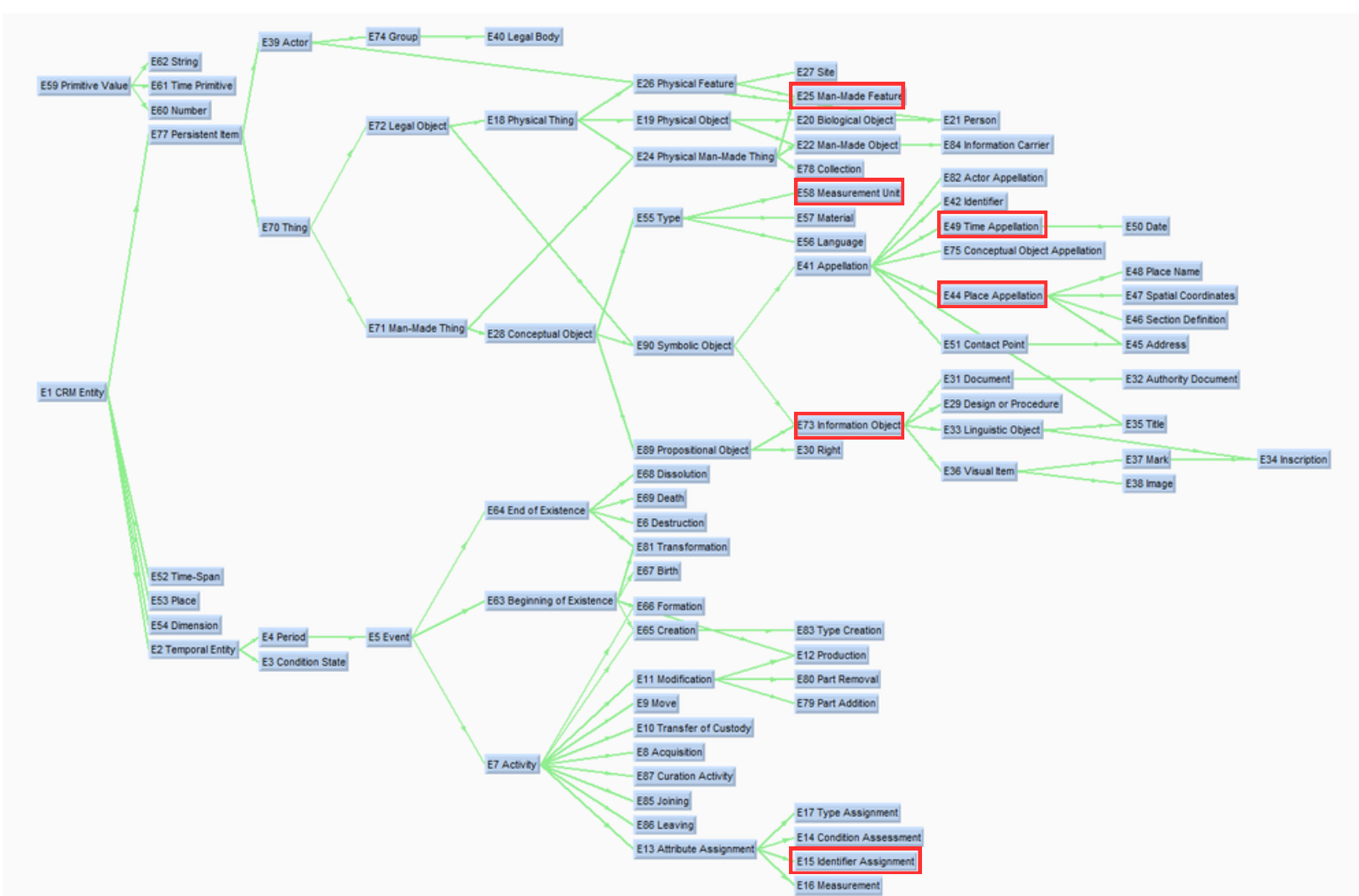
# SÉLECTION AUTOMATIQUE DU SOUS-ENSEMBLE

CIDOC CRM

CIDOC CRM

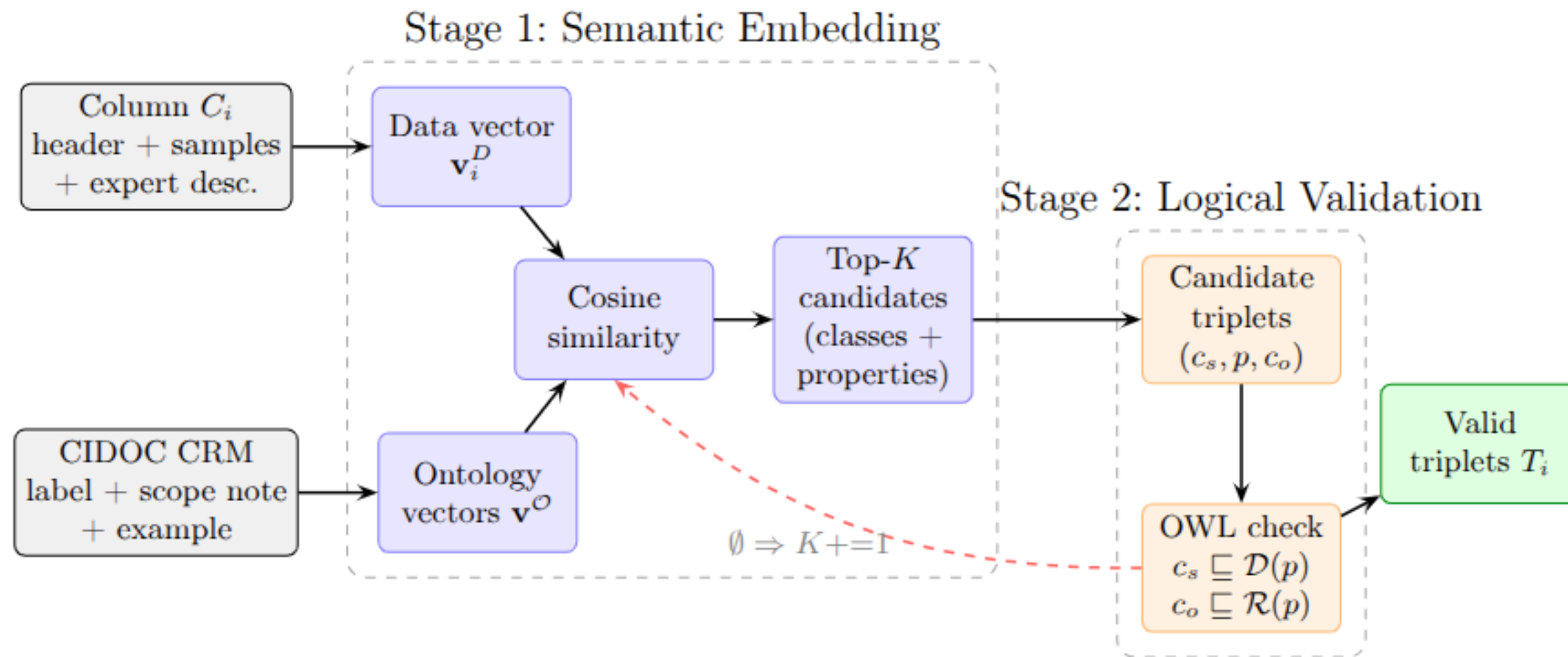


Sous-ensemble de CIDOC CRM



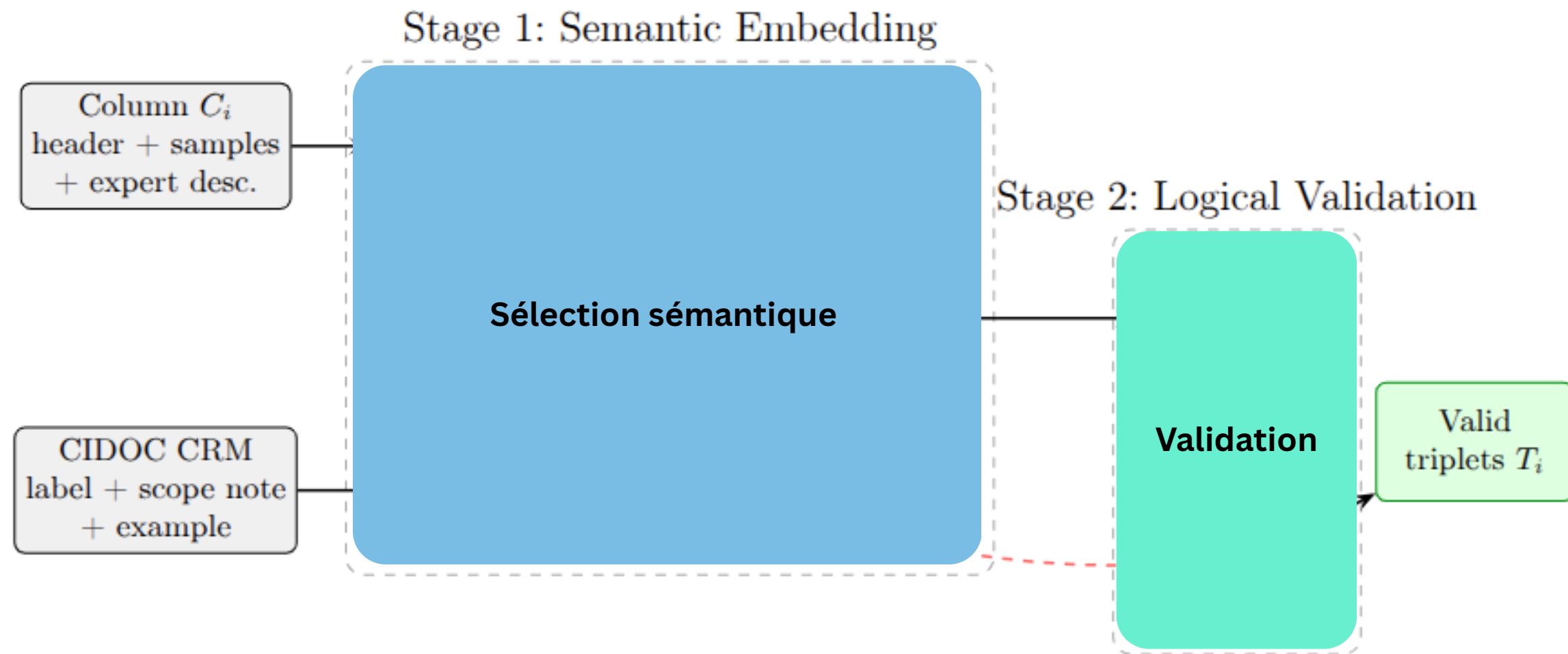
# SÉLECTION AUTOMATIQUE DU SOUS-ENSEMBLE

CIDOC CRM



# SÉLECTION AUTOMATIQUE DU SOUS-ENSEMBLE

CIDOC CRM



# SÉLECTION AUTOMATIQUE DU SOUS-ENSEMBLE

Article accepté



La 28ème Conférence Internationale sur l'Analyse des  
Mégadonnées et la Découverte de Connaissances

11 - 13 Août, 2026

## ASOS-CRM: Automated Semantic Scoping for CIDOC CRM Population

Ali Hariri<sup>[0009–0004–9380–5138]</sup>, Stéphane Jean<sup>[0000–0002–9434–9320]</sup>, and Mickaël  
Baron<sup>[0000–0002–3356–0835]</sup>

ISAE-ENSMA, Université de Poitiers, LIAS, France  
{ali.hariri, stephane.jean, mickael.baron}@ensma.fr  
<https://www.lias-lab.fr>

**Abstract.** Mapping structured data to large ontologies such as CIDOC CRM remains a challenging task for Large Language Models (LLMs). Previous studies have shown that providing the LLM with a small, domain-relevant subset of the ontology rather than the full specification significantly improves annotation quality. However, constructing this subset has so far required manual expert work, limiting the Automation of semantic integration approaches. In this paper, we present **ASOS-CRM** (Automated Semantic Ontological Scoping for CRM), a two-stage approach that automatically selects the relevant CIDOC CRM classes and properties for a given dataset. ASOS-CRM combines multilingual sentence embeddings with formal OWL constraints checking to select logically valid ontological classes and properties. Evaluated on four real-world archaeological datasets, our approach yields annotation results that match and marginally surpass those of a manually curated subset, while requiring significantly less preparation time.

**Keywords:** Knowledge Graph · CIDOC CRM · Embeddings · OWL Validation

### 1 Introduction

The CIDOC Conceptual Reference Model (CIDOC CRM) is a formal event-centred ontology for cultural heritage information exchange. Applying CIDOC CRM correctly requires significant ontological expertise; building a compliant knowledge graph (KG) typically requires days of work per dataset. While LLMs offer promise, prior work showed that feeding the full CIDOC CRM specification produces poor results; models fall back on generic classes and generate non-existent properties. Instead, restricting LLMs to manually curated subsets of relevant classes and properties consistently improves annotation quality. However, constructing this subset for each new dataset is time-consuming and requires expert judgment. This paper investigates automating subset selection while maintaining logical correctness. We propose ASOS-CRM, a two-stage approach: Stage 1 uses multilingual embeddings to retrieve semantically similar CIDOC CRM classes and properties; Stage 2 verifies candidates against OWL

# COMMUNICATION

# Huma-Num MASA plus

**08/06/2026 PAR AUGUSTIN BRAUD**

Journée d'étude inter-CHN : « Cas d'usage :  
Le partage des données des SHS sur le web  
sémantique »

Le 4 juin, le consortium Musica\* et MASA+ ont accueilli des collègues pour la journée d'étude « **Cas d'usage : Le partage des données des SHS sur le web sémantique** » au Centre d'Études Supérieures de la Renaissance, à Tours. Cette journée faisait suite à une journée introductive et didactique sur le CIDOC-CRM à la BnF.



**Journée inter-consortiums MASA+ et MUSICA\* :**

## « PARTAGER LES DONNÉES DES SHS SUR LE WEB SÉMANTIQUE »

**Intervenants :**

- Thomas Bottini (UMR 8223 IReMUS, Paris)
- Augustin Braud (UMR 7323 CESR, Tours)
- Florian Hivert (Expert web sémantique indépendant, Tours)
- Olivier Marlet (UMR 7324 CITERES, Tours)

|| Salle Colette, BnF Quai François Mauriac, 75706 Paris



→ 12 mars 2025 →

10h-12h

## Introduction au CIDOC CRM

Cette intervention démontre pourquoi le CIDOC CRM est le bon choix pour faciliter l'interopérabilité et la pérennisation des données de la recherche.

### Utiliser le CIDOC CRM

Nous aborderons ici les principes de modélisation avec le CIDOC CRM, notamment à travers le jeu *OntoMatchGame*.

Pause déjeuner

13h30 - 17h

## Exploiter et publier des données RDF/CIDOC CRM

Ces mises en situation concrètes présenteront différents processus de génération de triplets RDF.

### Cas d'usage

Les intervenants présenteront divers cas d'usages au travers de plusieurs disciplines dans le champ des sciences humaines.



citeres



(BnF



Merci