



HAL
open science

Modélisation sémantique des annotations pour l'exploitation critique des sources iconographiques

Thomas Bottini, Rebecca Bristow

► **To cite this version:**

Thomas Bottini, Rebecca Bristow. Modélisation sémantique des annotations pour l'exploitation critique des sources iconographiques. L'IReMus Présente, IReMus, Jun 2022, Paris, France. halshs-03950395

HAL Id: halshs-03950395

<https://shs.hal.science/halshs-03950395v1>

Submitted on 21 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

2022.06.16 — IReMus Présente

Figure 1: téléchargement.gif

Modélisation sémantique des annotations pour l'exploitation critique des sources iconographiques

Thomas Bottini — IReMus UMR 8223 CNRS Rebecca Bristow — IReMus UMR 8223 CNRS + CERES

Le projet SHERLOCK

- Émergence SU 2019-2021, porté par Achille Davy-Rigaux & TB
- Projet de refonte méthodologique, scientifique et technique de l'infrastructure de gestion des données scientifiques numériques de l'IReMus, de la production à la publication.
- Trois compétences : musicologie, humanités numériques, informatique.

SHERLOCK, c'est :

1. Écouter les chercheur•se•s en situation de production données informatiques structurées représentant des connaissances à propos de sources.
2. Accompagner méthodologiquement les chercheur•se•s dans la saisie des données scientifiques :
 1. Aider à révéler la structure interne des sources et des phénomènes par lesquels est appréhendé l'objet de recherche (dimension maïeutique + heuristique).
 2. Mettre en place des interfaces de saisie ergonomiques.
3. Modéliser les **données scientifiques** de l'IReMus, leur **contexte de production** et les **sources** auxquelles elles se rapportent (pensée + dessin). Nous nous appuyons sur une ontologie sémantique, le CIDOC-CRM, sur laquelle nous reviendrons un peu plus tard.
4. Nourrir le graphe de données RDF de l'IReMus.
 - Le web sémantique en quelques notions clefs :
 - Web vs Web sémantique
 - Triplet RDF (sujet — prédicat — objet)
 - IRI
 - Graphe de connaissances (sources, contexte, annotations, ressources analytiques...)
 - Sémantique partagée
 - Par rapport aux BDD « traditionnelles » (Drupal, FileMaker, Access...) ?
 - Exemples marquants : DOREMUS, data.bnf.fr, BML, l'INA, Wikidata...

- Comment utiliser les méthodes et techniques du Web sémantiques pour la recherche ?
- 5. Développer des applications Web :
 1. pour publier, explorer et valoriser les données scientifiques en lien avec les sources
 2. pour annoter les sources de manière coopérative

Aujourd'hui, nous vous parlons plus spécifiquement du travail iconographique...

Des opérations mentales & techniques

Produire des savoirs sur des corpus iconographiques renvoie à un ensemble d'opérations à la fois mentales et matérielles :

- Constituer un corpus d'images
- Documenter une image et son contexte de production
- Accéder au contexte éditorial de l'image
- Indexer une image avec des catégories analytiques
- Annoter une image
- Explorer un corpus d'images
- ...

Notion de « Scholarly primitives » dans la littérature DH anglo-saxonne (Unsworth, 2000)

Un cas : l'étude du corpus d'estampes du *Mer-cure Galant*

- périodique mensuel de l'époque de Louis XIV (1672-1710)
- fondé en 1672 à l'initiative de l'homme de lettres Jean Donneau de Visé
- actualité sociale, littéraire et artistique
- près de 700 airs gravés et 476 estampes

Une situation de travail

- Regarder les estampes du corpus
- Les analyser
- Reporter et formaliser ces analyses dans une feuille de saisie



Figure 2: https://www.iremus.cnrs.fr/sites/default/files/mercure_lecture_dair.jpeg

Propriétés génériques

- date d'édition
- titres
- acteurs de l'édition
- techniques
- dimensions

Éléments indexés

- lieux et personnes
- objets
- thématiques
- types de représentation

https://docs.google.com/spreadsheets/d/1xI4XzA_PTOOz1rsHCpJMwxj5sNZQsJtDGf2ZcGEht4g/edit?skip_itp2_check=true&pli=1#gid=0

Créer et utiliser des référentiels

- Vocabulaire d'indexation des estampes

https://docs.google.com/spreadsheets/d/1wS9punldFYlqZpkRgEsDXUSPRH1CDhdPKJYwjMXq_

EXTRAORDINAIRE
DU MERCURE
GALANT.

QUARTIER D'OCTOBRE 1678.

DEDIE' AU ROY.

TOME IV.



G. 1563. A PARIS.
28.
AV PALAIS



Source gallica.bnf.fr / Bibliothèque nationale de France

Figure 3: Extraordinaire_du_Mercure_galant__bpt6k6373165g_5.jpeg



Figure 4: 1689-05a_254.jpg



Source gallica.bnf.fr / Bibliothèque nationale de France

Figure 6: Mercure_galant__bpt6k63706517_143.jpeg

sw/edit#gid=0

- S'aligner sur Iconclass
- Les référentiels de l'Ancien Régime

Modélisation sémantique

Choix méthodologique

- Le choix du CIDOC-CRM (ICOM, <https://www.cidoc-crm.org/>)
- Pourquoi est-ce rassurant ?
- Fonction heuristique (par rapport à un modèle ad-hoc)

Le CIDOC-CRM en trente secondes

- Entités & propriétés
- Modèle centré événement

CRM Class Hierarchy

Martin Doerr
2/16/2020

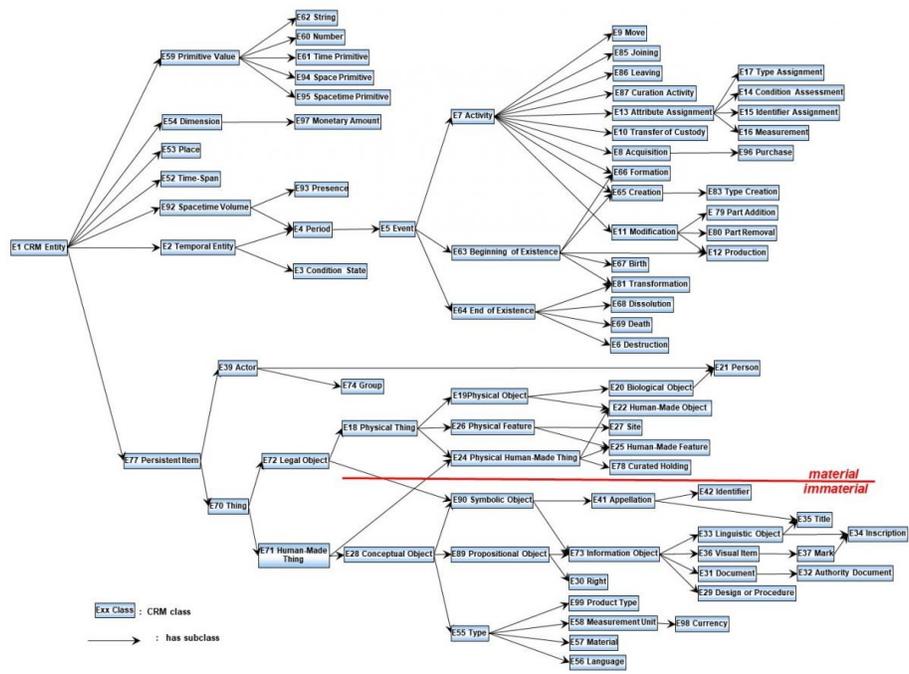


Figure 7: cidoc-crm.JPG

Quelques fragments du modèle pour le travail sur les images

Représenter la structure d'une annotation

- Pour nous, toute donnée résulte d'une annotation, qui se définit par :
 - Une **cible**, c'est-à-dire ce à quoi se rapporte l'annotation. Ici, une image ou une zone d'une image.
 - Un **corps**, un contenu ajouté (ce qu'on dit de la cible)
 - Une propriété qui connecte la cible et le corps
 - Un **• e auteur • rice**
 - Toute autre information sur le contexte de production de l'annotation (horodatage, justification scientifique, niveau de certitude, liens vers des preuves, etc.)
- Une donnée scientifique est une donnée contextualisée.
- Toute cellule d'un tableur peut être vue comme une annotation dont le contexte serait implicite.

https://docs.google.com/spreadsheets/d/1xI4XzA_PTOOz1rsHCpJMwxj5sNZQsJtDGr2ZcGEht4g/edit?skip_itp2_check=true&pli=1#gid=0

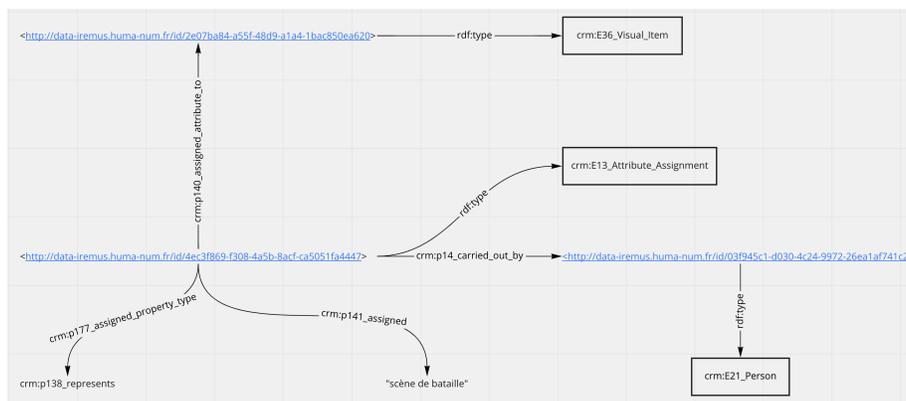


Figure 8: Capture d'écran 2022-06-14 à 16.43.41.png

Le CRM encourage à représenter une connaissance comme l'événement duquel émane cette connaissance, et donc favorise l'expression du contexte.

Analyser le contenu d'une image

« Cette image contient une zone qui représente une médaille »

Exprimer le contexte de production et faire des attributions

- Il est intéressant de le représenter autour d'événements
- *Ex : invenit, sculpsit...*

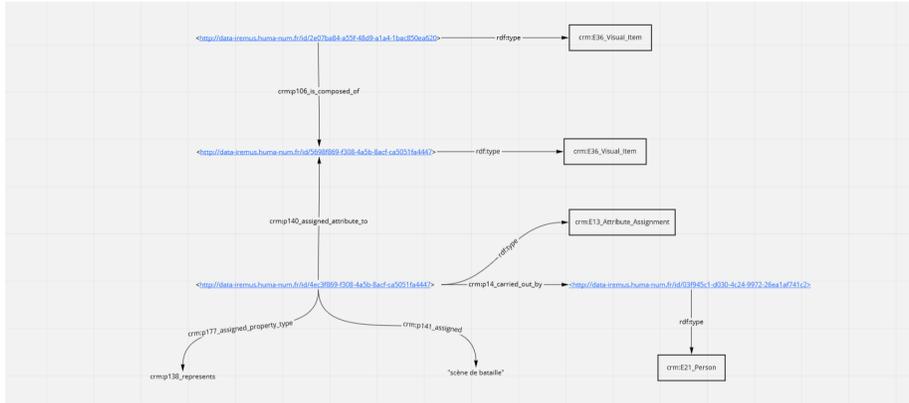


Figure 9: mg-graph-E36-P106.png

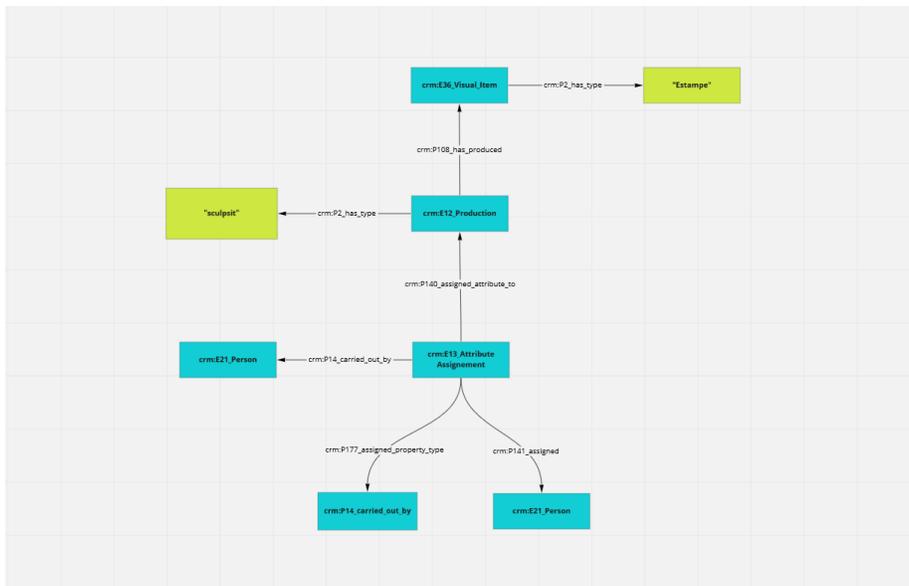


Figure 10: mg-graph-E36-E12.png

Le contexte éditorial des estampes

- Un périodique, des livraisons, des articles
- Utilisation de l'ontologie LRMoo (fille du CRM et de FRBR, qui nous vient du monde des bibliothèques) pour représenter les œuvres et leur structure.



Figure 11: mg-graph-E36-F2.png

Représenter une image et la publier sur un serveur IIIF

https://ceres.huma-num.fr/iiif/3/mercure-galant-estampes-1684-01_067/full/max/0/default.jpg

https://ceres.huma-num.fr/iiif/3/mercure-galant-estampes-1684-01_067/info.json

https://ceres.huma-num.fr/iiif/3/mercure-galant-estampes-1684-01_067/550,1100,200,300/max/0/default.jpg

Vers une interface d'exploration & d'annotation sur le Web

- Notre approche : dégager les invariants fonctionnels entre différents corpus et pratiques scientifiques ayant pour point commun le fait de travailler sur des images.
- Plusieurs projets pilotes (nous recherchons des corpus différents pour éprouver notre hypothèse de généricité)
- Trois niveaux :

- Fonctions génériques non spécifiques aux corpus iconographiques
 - * Visualiser les annotations attachées à une ressource
 - * Effectuer des recherche sémantiques
 - * Annoter une ressource
 - * Créer des corpus et des projets analytiques
 - * Constituer des sélections de ressources ou de fragments de ressources dans un panier et les transformer en projet/corpus
 - * Explorer le corpus par les projets analytiques
 - * Obtenir des statistiques et les convoquer pour bâtir des stratégies d’exploration des corpus
 - * Elastic Search (sources + contexte + annotations)
 - * Naviguer dans les triplets
- Fonctions génériques aux corpus iconographiques
 - * Afficher une image et zoomer
 - * Délimiter et identifier des zones d’une image
 - * Bâtir des vues synoptiques de type « galerie »
- Fonctions spécifique par projet
 - * Accès au contexte éditorial