

2022.06.30 — Modéliser les connaissances en SHS : pourquoi, comment, jusqu'où ?

THOMAS BOTTINI Institut de Recherche en Musicologie (IReMus)
UMR 8223 CNRS/SU/BnF/MC Centre d'expérimentation en méthodes
numériques pour les recherches en SHS, Faculté des Lettres, SU

Plan

- Du rôle de l'ingénieur · e dans un projet en SHS
- Quelques mots sur le Web sémantique comme milieu technique adapté à l'expression et à la diffusion des données de la recherche
- Les fondements de l'ontologie CIDOC-CRM
- Cas concrets et situations types

De l'activité de recherche à la donnée

Les étapes possibles d'un projet de recherche en SHS en rapport avec l'ingénierie des données :

1. ÉCOUTER les chercheur · se · s en situation de production de connaissances susceptibles d'être traduites en données informatiques structurées. En SHS, ces connaissances peuvent résulter d'une activité descriptive (sources) ou interprétative (choses dites à propos des 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, leur contexte de production et les sources auxquelles elles se rapportent. [← c'est l'objet de cette formation]
4. GÉNÉRER techniquement les données, en les rendant conformes au modèle. [← ce n'est pas l'objet de cette formation, mais je vous dirai tout de même deux mots du Web Sémantique et du RDF]
5. DIFFUSER techniquement les données :
 1. Accès par les machines (API, SPARQL Endpoint)
 2. Consultation par les humains (sites Web)

Une situation traditionnelle de saisie de données

Un exemple volontairement « sous-instrumenté » mais efficace :

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

Vocabulaire contrôlé :

https://docs.google.com/spreadsheets/d/1wS9punldFYlqZpkRgEsDXUSPRH1CDhdPKJYwjMXq_sw/edit#gid=0

Le Web sémantique en quelques notions clefs

- Web vs. Web sémantique, le fantasme d'un sens partagé (https://en.wikipedia.org/wiki/Semantic_Web)
- Le triplet RDF (sujet—prédicat—objet) comme élément de base pour représenter les données (<https://www.w3.org/TR/rdf11-primer/>)
- IRI, la promesse d'une addressabilité totale
- Graphe de connaissances (sources, contexte, annotations, ressources analytiques...)
- Sémantique partagée
- Et par rapport aux BDD « traditionnelles » (Drupal, FileMaker, Access, Excel...) ?
- Que faut-il avoir en tête pour utiliser les méthodes et techniques du Web sémantiques pour représenter les connaissances en SHS ?

Qu'est ce que modéliser ?

- Comprendre les choses de l'intérieur, leur logique d'organisation interne.
- Identifier :
 - les entités du domaine
 - leurs propriétés intrinsèques
 - les propriétés qui les connectent
- Toujours chercher à identifier les cas vicieux pour être certain que le modèle les prend bien en compte
- Trois outils : la pensée, la parole, le tableau blanc.
- Qu'est ce qu'une ontologie ?
- Mon approche : modéliser les choses de l'intérieur, dans leur singularité, permet de communiquer dessus avec une plus grande précision (notion d'adressabilité).

Réinventer la roue ?

- 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 CRM en quelques liens

- <https://www.cidoc-crm.org/>

Des exemples !

- Une peinture et son contexte de production
- Groupes sociaux
- Un cas vicieux : les institutions de l'Ancien Régime :
 - <https://opentheso3.mom.fr/opentheso3/index.xhtml>
 - https://miro.com/app/board/uXjVOz0NAKE=
 - <https://www.dropbox.com/sh/ylrbwswxvqbh30a/AACxjUOUVpHM2Km1LfvG-Xqla?dl=0>
- Annotations signées
-