



HAL
open science

Une ontologie pour Henri et ses amis

Pierre Willaime

► **To cite this version:**

Pierre Willaime. Une ontologie pour Henri et ses amis. 10 ans avec CAHIER. Des corpus d'auteurs pour les humanités à leur exploitation numérique, Jun 2021, Bordeaux, France. halshs-03583092

HAL Id: halshs-03583092

<https://shs.hal.science/halshs-03583092>

Submitted on 21 Feb 2022

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.

Pierre Willaime

"Une ontologie pour Henri et ses amis"

Pierre Willaime

(Archives Henri-Poincaré - Nancy)

Le projet

- projet collectif depuis 2002
- édition numérique de la correspondance de Henri Poincaré (1854-1912)
- ~2100 lettres
- un apparat critique pour chaque lettre

une méthodologie un peu particulière basée sur le Web sémantique :

Représenter des connaissances sans encoder un texte en TEI.

Modéliser plutôt que décrire

S'extraire du document (voire du corpus) pour formuler des interrogations plus globales (SPARQL)

Omeka S

- pas de barrière entre publier et exploiter
- adapté à un corpus dont la description se fait au fil de l'eau
- gestion native des relations entre contenus (RDF)
- utilisation native de schémas de métadonnées interopérables (DC, FOAF, BIBO, BIO, ...)

Au prix d'un aplatissement de la structuration de l'ontologie (ex : dcterms:created sous-propriété de dcterms:date, perte des domaines d'application des métadonnées, etc.)

(démon)

Web sémantique

1- des triplets RDF (Resource Description Framework) : "sujet - prédicat - objet"

2- une structuration plus complète avec RDFS : ajout de classes et sous-classes, de domaine d'application, de type de données attendu (range)

3- encore raffinée par OWL (Web Ontology Language) : équivalent des

propriétés et des classes, (a)symétrie des relations, ...

-> F.A.I.R : on permet la réutilisation d'un univers.

Problème : les ontologies sont de cohérences très variables et peu adaptées aux besoins historiques.

Ontologie(s)

- Terminologie : notions + définition
- Taxonomie : terminologie + une relation hiérarchisée d'un seul type (celui de la subsumption)
- Thésaurus : taxonomie + d'autres informations (TG, TS, TA, ...) et des relations non hiérarchiques plus lâches
- Ontologie : terminologie + des relations structurées pas seulement hiérarchiquement.

Une ontologie est le degré le plus avancé d'une structuration représentationnelle

parenthèse : Le cas du Dublin Core (version DCMI terms)

- DCMI Metadata Terms sont publiés sous forme de schémas RDF (<http://purl.org/dc/terms/>),
- mais ils ne sont pas directement exprimables en OWL.

Définir des relations structurées particulières demande souvent de délimiter un domaine où ces relations s'exercent. Des relations ou des objets peuvent cesser d'exister lorsqu'on sort du domaine.

Dublin Core n'a pas de domaine d'application (mais décrit un type de ressources, les ressources numériques). Dès lors faut-il le considérer comme une ontologie ? Le peut-on ?

Pour moi, il s'agit plutôt d'un vocabulaire de base, exprimable sémantiquement, et utilisé par la plupart des ontologies.

AHPo

Problème : les ontologies sont de cohérences très variables et peu adaptées aux besoins historiques. Solution : développer une ontologie maison (alignée sur d'autres lorsque c'est possible).

<file:ahpo_v2.owl>

(démon)

Réseau de personnes

Des choix de modélisation :

1- les liens généalogiques : rel:spouseOf, rel:childOf 2- les liens d'amitiés, 3- les liens entre collègues, 4- les professions, 5- la genèse des relations : "connaît par le réseau Polytechnicien", "connaît par telle académie", "connaît par le réseau familial"

Risque d'une inflation ontologique.

Conclusion

Pas d'opposition entre deux technologies (XML-TEI et le Web sémantique) mais des focus différents.

- TEI : centré sur le texte avec une granularité souvent fine -> objectif d'édition numérique
- Web sémantique : modélisation des informations extraites du corpus -> objectif de formulation de requêtes et de fouille.

Les deux technologies peuvent conduire à des visions complémentaires d'un même corpus. Il serait intéressant de coupler ces usages, présentés souvent (à tort) comme deux approches incompatibles et comme deux directions opposées pour un projet d'humanités numériques.