



HAL
open science

Ce qui fragilise l'intéropérabilité dans la mise en oeuvre des technologies du web sémantique et les prémices d'un outil pour la renforcer

Loic Jeanson

► To cite this version:

Loic Jeanson. Ce qui fragilise l'intéropérabilité dans la mise en oeuvre des technologies du web sémantique et les prémices d'un outil pour la renforcer. Journée nationale DAHLIA " Informatique et Humanités numériques: quelles problématiques pour quels domaines? ", Jun 2019, Nantes, France. hal-02498768

HAL Id: hal-02498768

<https://hal.science/hal-02498768v1>

Submitted on 4 Mar 2020

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.

Ce qui fragilise l'interopérabilité dans la mise en oeuvre des technologies du web sémantique et les prémices d'un outil pour la renforcer

Loïc Jeanson*

*LS2N

1, rue de la Noë - BP 92101
44321 NANTES Cedex 3
<http://ls2n.fr>

**Centre François Viète

UFR Sciences et Techniques
2 rue de la Houssinière - BP 92208
44322 NANTES CEDEX 3
<http://www.cfv.univ-nantes.fr>
loic.jeanson@univ-nantes.fr

Résumé. Ce document présente une faiblesse d'interopérabilité dans les usages des technologies du web sémantique. Il présente quelques exemples de la même question du point de vue des humanités classiques et propose une assistance pour renforcer cette interopérabilité sémantique.

1 L'utilisation répandue des technologies du web sémantique

La nébuleuse de travaux à laquelle on associe le nom d'*humanités numériques* peut s'observer de multiples points de vue. Une analyse par les technologies utilisées permet un contact avec plusieurs tensions : la diffusion des pratiques ou l'évolution de certaines tendances. Parmi les technologies en usage, le web sémantique constitue un cas intéressant : comme ces standards et technologies portent sur le stockage d'informations, ils sont assez répandus. À l'appui de cette affirmation, on peut mettre en avant de nombreux projets : en histoire et en histoire des sciences et des techniques (e.g. Laubé et al. (2014); Zacklad (2005); de Boer et al. (2013)), ou des projets patrimoniaux (Hyvönen (2012); Hyvönen et al. (2009); Pauwels et al. (2013)). Cette dynamique a même produit certaines communautés scientifiques constituées autour de l'usage de ces technologies (e.g. Linked pasts, Pelagios, etc.).

Afin de ne pas présenter inutilement ces technologies à une audience initiée, nous resterons brefs. L'objectif est le passage d'un *web de documents* à un *web de données* : des données ouvertes liées, structurées par des vocabulaires, eux aussi, ouverts et liés. Ces technologies autorisent des modélisations d'objets (au sens large) en garantissant l'interopérabilité la plus grande possible par 1. une utilisation extensive d'identifiants uniques (permettant de réduire les problèmes d'ambiguïté) et 2. un accès le plus facile possible et un partage (utilisation et réutilisation) le plus large possible au modèle ontologique derrière la modélisation.

2 L'interopérabilité visée rencontre des limites

L'interopérabilité nous semble n'en présenter que le titre quand elle n'est qu'une acception partagée d'éléments d'une ontologie. Cette acception partagée n'est d'ailleurs souvent pas satisfaisante, preuve en est le grand nombre d'ontologies de domaine ou spécialisées, d'extensions à l'usage plus ou moins local que l'on trouve dans tous les domaines.

2.1 Diversité et évolution sémantique et lexicale

Les *humanités*, bien avant qu'elles ne s'outillent d'informatique pour devenir *numériques* rencontrent ces problèmes. Les études de sémiotique Eco (1972) ou de sémantique Benveniste (1966) montrent que ces problèmes sont ardues. En histoire, le courant sur la sémantique historique (entre autres Koselleck et al. (1990)) travaille aux changements d'acceptions à l'interprétation des informations historiographiques. Les pratiques anciennes de philologie, d'herméneutique ou d'exégèse sont d'autres réponses à ces besoins d'interprétation et de réinterprétation. Le numérique aussi porte ce besoin d'interprétation et de réinterprétation pour permettre pérennité d'accès et de fonctionnement (Alaoui (2019)). Plus généralement, la modélisation produit des classifications dont Percey affirme avec humour et sincérité Percey (1985) :

Mon problème avec les classements, c'est qu'ils ne durent pas ; à peine ai-je fini de mettre de l'ordre que cet ordre est déjà caduc.

Nous sommes dans un environnement conceptuel de désaccord ténus et en mouvement.

2.2 Les limites de la pratique

Les cadres conceptuels fixes ou les ontologies écrites une fois pour toutes semblent assez incompatibles avec la souplesse et l'évolution inhérente au lexique, à la sémantique et à la sémiotique si chères aux sciences humaines. Cette contradiction mène à des mésusages de deux ordres. Puisque toute modélisation répond à un objectif premier et que les objectifs peuvent être très différents, il se peut que les classes et propriétés des ontologies soient utilisées à mauvais titre, ou dans une acception approximative (c'est à dire non plus strictement dans le cadre de leur définition). Après tout, pourquoi redéfinir une classe ou une propriété locale si une ontologie existante en propose une qui correspond à peu près à ce qui m'intéresse ? De plus, la frontière entre web sémantique et web lexical semble parfois ténue tant ces technologies sont parfois sous-utilisées (affranchissement des règles de logique de premier ordre, des domaines de validité de ces règles,...).

Les numériciens s'en sont cependant déjà rendu compte depuis longtemps, et les travaux sur les systèmes permettant des alignements entre ontologies (Noy et al. (2000) ; Euzenat et al. (2004) ; Otero-Cerdeira et al. (2015) ainsi que ceux portant sur le suivi de l'évolution des ontologies (Zablith et al. (2015) ; Troullinou et al. (2016) montrent bien ce travail entamé.

3 Penser l'exégèse des bases de données

Ces initiatives permettent de suivre l'évolution d'une ontologie dans le temps et de la tenir à jour. Elles permettent aussi de systématiser des équivalences entre ontologies. Le besoin se fait cependant sentir d'être en mesure que ce choix soit réalisé au cas par cas, ou au moins

d'une session à l'autre. Le besoin d'une sorte d'exégèse de bases de données où l'utilisateur peut choisir les alignements de façon délibérée à partir d'indicateurs guidant ces choix. L'interopérabilité sémantique reste sujette à discussion (on peut ne pas être d'accord avec le choix d'interprétation/d'alignement). Cependant, elle permet de renforcer l'interopérabilité effective au sein du cadre conceptuel qui l'a guidée en premier lieu. Ainsi, un outil permettant une forme de *distant reading*, produisant des indicateurs basés sur des calculs de distances serait un bon guide à ces décisions. On retrouve alors le champ classique des outils d'alignement automatisés : calcul de distances à l'intérieur du graphe (des données ou du modèle de données) et calcul de distances par lexicométrie (similarité ou étude du contexte). Notons qu'en premier lieu, pour réaliser ces calculs, le prérequis le plus fort est le cadre conceptuel ou ontologique de référence duquel on calculera l'éloignement des autres. Fondamentalement, on le voit bien, cette proposition n'apporte rien de très nouveau aux systèmes classiques d'alignement des classes et propriétés, si ce n'est de restituer la décision d'alignement à l'utilisateur, qui rejoignent Dragisic et al. (2016); Lambrix et Kaliyaperumal (2017). De plus, que l'alignement soit réalisé ou non, le calcul de la distance peut être renouvelé en changeant de méthode ou être simplement conservé.

Références

- Alaoui, S. (2019). Patrimoine et numérique : technique et politique de la mémoire, par bruno bachimont. *Canadian Journal of Academic Librarianship* 4, 1–3.
- Benveniste, E. (1966). Problèmes de linguistique générale, 1 vol.
- de Boer, V., J. van Doornik, L. Buitinck, M. Marx, T. Veken, et K. Ribbens (2013). Linking the kingdom : enriched access to a historiographical text. In *Proceedings of the seventh international conference on Knowledge capture*, pp. 17–24. ACM.
- Dragisic, Z., V. Ivanova, P. Lambrix, D. Faria, E. Jiménez-Ruiz, et C. Pesquita (2016). User validation in ontology alignment. In *International Semantic Web Conference*, pp. 200–217. Springer.
- Eco, U. (1972). La structure absente introduction à la recherche sémiotique.
- Euzenat, J., D. Loup, M. Touzani, et P. Valtchev (2004). Ontology alignment with ola. In *Proc. 3rd ISWC2004 workshop on Evaluation of Ontology-based tools (EON)*, pp. 59–68. No commercial editor.
- Hyvönen, E. (2012). Publishing and using cultural heritage linked data on the semantic web. *Synthesis Lectures on the Semantic Web : Theory and Technology* 2(1), 1–159.
- Hyvönen, E., E. Mäkelä, T. Kauppinen, O. Alm, J. Kurki, T. Ruotsalo, K. Seppälä, J. Takala, K. Puputti, H. Kuittinen, et al. (2009). Culturesampo : A national publication system of cultural heritage on the semantic web 2.0. In *European Semantic Web Conference*, pp. 851–856. Springer.
- Koselleck, R., M.-C. Hoock-Demarle, et J. Hoock (1990). Le futur passé contribution à la sémantique des temps historiques.
- Lambrix, P. et R. Kaliyaperumal (2017). A session-based ontology alignment approach enabling user involvement 1. *Semantic Web* 8(2), 225–251.

Renforcer l'interopérabilité dans l'utilisation du web sémantique

- Laubé, S., B. Rohou, et S. Garlatti (2014). Humanités numériques et web sémantique : de l'intérêt de la modélisation des connaissances en histoire des sciences et des techniques pour une histoire comparée des ports de brest (france) et mar del plata (argentine). In *DI 2014 : 1st international conference on Digital Intelligence*.
- Noy, N. F., M. A. Musen, et al. (2000). Algorithm and tool for automated ontology merging and alignment. In *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI-00)*. Available as SMI technical report SMI-2000-0831, Volume 115. sn.
- Otero-Cerdeira, L., F. J. Rodríguez-Martínez, et A. Gómez-Rodríguez (2015). Ontology matching : A literature review. *Expert Systems with Applications* 42(2), 949–971.
- Pauwels, P., R. Bod, D. Di Mascio, et R. De Meyer (2013). Integrating building information modelling and semantic web technologies for the management of built heritage information. In *2013 Digital Heritage International Congress (DigitalHeritage)*, Volume 1, pp. 481–488. IEEE.
- Perec, G. (1985). *Penser, classer*. Hachette Paris.
- Troullinou, G., G. Roussakis, H. Kondylakis, K. Stefanidis, et G. Flouris (2016). Understanding ontology evolution beyond deltas. In *EDBT/ICDT Workshops*.
- Zablith, F., G. Antoniou, M. d'Aquin, G. Flouris, H. Kondylakis, E. Motta, D. Plexousakis, et M. Sabou (2015). Ontology evolution : a process-centric survey. *The knowledge engineering review* 30(1), 45–75.
- Zacklad, M. (2005). Vers le web socio sémantique : introduction aux ontologies sémiotiques. In *Ingénierie des Connaissance 2005*.

Summary

This work presents a flaw in interoperability in the use of semantic web technologies. It gives a few examples of how the humanities deal with it and ends up with a tool proposal to reinforce such semantic interoperability.