

Maitriser une technologie de gestion des ontologies et vocabulaires en France : défis et enjeux.

Clement Jonquet, LIRMM, Université de Montpellier & CNRS

Texte

Les principes FAIR [1] ont établi l'importance d'utiliser des vocabulaires ou des ontologies pour décrire les données et faciliter l'interopérabilité et la réutilisation. Comme toutes autres données, les ontologies, vocabulaires, thésaurus et terminologies doivent eux-mêmes être FAIR. Cependant, il est fastidieux pour les experts du domaine d'identifier les ontologies dont ils ont besoin et de gérer leur hétérogénéité, et leur chevauchement. Bien qu'il y ait de multiples façons de rendre les ontologies FAIR, on s'accordera à dire que développer des portails ouverts d'ontologies en fait certainement partie. Les fonctionnalités des portails d'ontologies vont d'une simple liste avec plus ou moins de descriptions de métadonnées (e.g., OBO Foundry, WebProtégé, FAIRsharing, BARTOC), à des plateformes qui indexent le contenu et offrent des services avancés : navigation, recherche, visualisation, métriques, annotation, recommandation, etc. (e.g., LOV, OntoBee, CISMeF HeTOP, EBI OLS, NCBO BioPortal, MMI ORR, OntoHub). Elles sont parfois thématiques, et le domaine biomédical a initié plusieurs d'entre elles [2].

BioPortal [3], le portail d'ontologies de référence en biomédecine, construit par le US *National Center for Biomedical Ontology* s'appuie sur une technologie ouverte, indépendante du domaine, régulièrement réutilisée. Par exemples, par l'ESIP Portal, MMI ORR, BiblioPortal ou lorsqu'il faut traiter des données en interne (e.g., hôpitaux).

Il existe plusieurs défis pour les portails d'ontologies [4] : métadonnées et sélection, multilinguisme, alignement, fournir des nouveaux services génériques, les annotations et données liées, l'interopérabilité et le passage à l'échelle. Au LIRMM, nous essayons de relever certains de ces défis dans le cadre des projets SIFR BioPortal (<http://biportal.lirmm.fr>) [5] et AgroPortal (<http://agroportal.lirmm.fr>) [6] en réutilisant, la technologie du NCBO au service des applications biomédicales (en français) et à l'agronomie et ses domaines connexes -agriculture, science des plantes, nutrition et biodiversité.

En partenariat avec l'Université de Stanford, nous développons et maintenons ces deux portails d'ontologies et de services à base d'ontologies qui nous permettent d'attaquer des problèmes de recherche complexes tels que le traitement de la langue, l'alignement d'ontologies, l'annotation sémantique, l'ingénierie ontologique, tout en étant pilotés par des scénarios d'applications concrets dont les impacts en santé et agronomie sont mesurables. Ces projets sont particulièrement fédérateurs pour plusieurs institutions de recherche (INRA, IRSTEA, CIRAD, IRD, CGIAR) et ont un impact national et international reconnu.

Aujourd'hui, le besoin pour des plateformes équivalentes se fait ressentir dans différents domaines e.g., l'écologie, les sciences sociales, les géosciences. En outre, avec l'adoption du Web sémantique dans le monde professionnel et l'impact de standards de ressources sémantiques plus souples et

accessibles comme SKOS, les utilisateurs intéressés ne sont plus seulement les scientifiques mais inclus tous les acteurs de la donnée en général.

Dans toutes les infrastructures de recherche, la question de la pérennisation se pose. Quels sont les valorisations qui permettent de garder un modèle ouvert et catalyseur de découvertes scientifiques ? Comment encourager le partage des données suivant les principes FAIR tout en pérennisant et consolidant les outils impliqués ? Quels sont les enjeux pour la France ? - qui initie des projets d'étude comme VisaTM [7] ou nous nous intéressons à la production de services de *text & data mining* pour les scientifiques en France. Quid des enjeux de l'adoption de SNOMED-CT par la France dans la santé ? Comment répondre au besoin de sémantique dans le futur paysage Européen de la science ouverte (EOSC - *European Open Science Cloud*) ?

Références

- [1] M. D. Wilkinson, M. Dumontier, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3.
- [2] M. D'Aquin and N. F Noy. 2012. Where to Publish and Find Ontologies? A Survey of Ontology Libraries. *Web semantics*, 11(August):96–111.
- [3] N. F Noy, N.H Shah, et al.. 2009. BioPortal: ontologies and integrated data resources at the click of a mouse. In B Smith, editor, *Nucleic Acids Research*, volume 37, pages 170–173.
- [4] C. Jonquet. 2017. Challenges for ontology repositories and applications to biomedicine & agronomy. *Keynote SIMBig 2017*, Lima, Peru.
- [5] C. Jonquet, A. Annane, et al. 2016. SIFR BioPortal : Un portail ouvert et générique d'ontologies et de terminologies biomédicales françaises au service de l'annotation sémantique. *16th Journées Francophones d'Informatique Médicale, JFIM'16*, Geneva, Switzerland.
- [6] C. Jonquet, A. Toulet, et al. 2018. AgroPortal: A vocabulary and ontology repository for agronomy. *Computers and Electronics in Agriculture*, 144, pp.126-143.
- [7] F. Kettani, S. Schneider, et al. 2018. Projet VisaTM : l'interconnexion OpenMinTeD – AgroPortal – ISTEEX, un exemple de service de Text et Data Mining pour les scientifiques français. 29emes journées francophones d'Ingénierie des Connaissances (IC'18). Session poster. Nancy, France.

Biographie

Clément Jonquet (LIRMM, Université de Montpellier), docteur en informatique, Maître de Conférences et chercheur invité à l'Université de Stanford, a 10 ans d'expérience dans les ontologies et le Web sémantique appliqués à la biomédecine et à l'agronomie. Il travaille sur la conception et le développement de portail d'ontologies/vocabulaires et d'outil d'annotation sémantique en tant que porteur du projet SIFR (ANR-JCJC & H2020-MSCA), et co-porteur de l'ANR PraktikPharma. Depuis 2015, il rassemble la communauté agronomique nationale et internationale autour de l'initiative AgroPortal. C. Jonquet est (co)auteur de +70 publications, qui cumulent plus de 2000 citations. Il est membre de plusieurs comités de programmes de conférence en informatique et applications à la santé. Il a organisé ESWC 2013, et plusieurs workshops dans le domaine. Il pilote également les groupes Meetup Web Science Montpellier et AgroHackathon. Depuis 2010, il est enseignant à Polytech Montpellier dans le département Informatique et Gestion.

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 701771.