

Editorial

L'émergence de données ouvertes et les particularités des données liées font naître des problématiques nouvelles pour les systèmes d'information qui les exploitent ou les génèrent. Que ce soit dans leur construction ou leur exploitation, ces types de données affectent les SI dans toute leur chaîne de traitement. Pour leur production, de leur genèse à leur mise à disposition, les données passent par des séries de transformations, de validation, d'intégration, d'agrégation, de publication puis, éventuellement, de visualisation. Pour leur consommation, il convient de les sélectionner, les interroger, les indexer, les fusionner. De manière transversale, la question de la gestion de telles données est prééminente puisque par essence, les données ouvertes et liées sont distribués, changeantes, variées et abondantes.

Il se pose donc des problématiques fortes d'intégration de sources multiples, de gestion de l'hétérogénéité des données, de traitement de données dynamiques, de qualité de données, de la provenance des données, de passage à l'échelle, et enfin de valorisation.

Les articles soumis à ce numéro spécial ont été évalués par au moins trois relecteurs experts. Et, lorsque cela était nécessaire, une évaluation en deux temps a été mise en place, c'est-à-dire que nous avons proposé aux auteurs de renvoyer une nouvelle version de leur article après commentaires des relecteurs afin que celle-ci soit à nouveau évaluée et que décision soit prise. Pour ce numéro spécial, nous avons au final retenu trois articles.

1) « A survey of the first 20 years of research on semantic Web and linked data », de Fabien Gandon est une étude des sujets de recherche dans le domaine du Web sémantique, des données liées et du Web des données. Cette étude se penche sur les contributions de cette communauté de recherche au cours de ses vingt premières années d'existence. En compilant plusieurs sources bibliographiques et indicateurs bibliométriques, Fabien Gandon identifie les principales tendances de la recherche et il référence certaines de leurs publications majeures pour donner un aperçu de cette période initiale. Une conclusion avec une discussion sur les tendances et perspectives de recherche est proposée.

2) « Impact des données ouvertes et liées sur les catalogues bibliographiques » de Fabien Duchateau, Nicolas Lumineau et Trond Aalberg aborde la problématique des systèmes d'information des bibliothèques. Ces systèmes incluent en général un élément central, le catalogue, dont les notices bibliographiques décrivent chacune des œuvres disponibles (e.g., livre, enregistrement vidéo). La gestion de ce

catalogue a peu évolué au cours des dernières décennies et rencontre de nombreux problèmes : un modèle plat, centré sur l'œuvre et qui ne permet pas de représenter d'autres entités et leurs relations, de nombreuses incohérences et ambiguïtés dans les données, des pratiques locales de catalogage limitant les échanges, etc. Une transition vers le web sémantique implique une transformation du catalogue en profondeur : définition d'un nouveau modèle, migration de données garantissant un haut degré de qualité, enrichissement des données originales et amélioration des services associés au catalogue (e.g., recherche sémantique). Pour réussir ces différentes étapes, l'exploitation des données ouvertes et liées est nécessaire, que ce soit pour faciliter l'appariement d'ontologies, pour lever des ambiguïtés avec l'aide de référentiels ou pour lier des entités vers d'autres sources. Cet article dresse donc un panorama de l'impact des données ouvertes et liées dans le domaine bibliographique, en particulier sur la migration de son catalogue vers le web sémantique.

3) « Détection de liens d'identité erronés en utilisant la détection de communautés dans les graphes d'identité » Joe Raad, Wouter Beek, Nathalie Pernelle, Fatima Saïs, Frank Van Harmelen. En l'absence d'une autorité de nommage centralisée dans le web sémantique, il est courant que différents jeux de données fassent référence à la même entité avec différents identifiants de ressources internationalisés (IRI). Des liens owl:sameAs sont alors nécessaires pour relier ces données. Cependant, des études menées depuis 2009 ont relevé que la propriété owl:sameAs est parfois utilisée de façon incorrecte. Dans cet article, les auteurs montrent comment la topologie du réseau des liens d'identité peut être utilisée pour détecter des liens probablement erronés en se basant sur des approches de détection de communauté. L'un des intérêts d'une telle méthode est qu'elle ne nécessite pas de connaissances autre que le réseau lui-même. Les auteurs ont évalué notre approche sur une large collection comportant 558 millions de liens d'identités issus du LOD. Cette évaluation a montré les capacités de leur approche à passer à l'échelle et son efficacité dans la détection de liens incorrects.

Max CHEVALIER

IRIT, Université Paul Sabatier Toulouse 3 - IUT

Sébastien LABORIE

LIUPPA, Université de Pau et des Pays de l'Adour

Cassia TROJAHN

IRIT, Université Toulouse 2 Jean-Jaurès

Antoine ZIMMERMANN

LaHC, École des mines de Saint-Étienne

Comité de lecture

Nathalie Aussenac-Gilles – IRIT, Université Paul Sabatier Toulouse 3
Sylvie Calabretto – LIRIS, Université Claude Bernard Lyon 1
Pierre-Antoine Champin – LIRIS, Université Claude Bernard Lyon 1
Mathieu d’Aquin – Insight Centre for Data Analytics, National University of Ireland Galway, RU
Jérôme David – INRIA/LIG, Université de Grenoble
Catherine Faron-Zucker – I3S, Université Nice Sophia Antipolis
Stéphane Frenot – CITI-INRIA, INSA Lyon Telecom
Antoine Isaac – Europeana & VU University Amsterdam
Maxime Lefrançois – LaHC, École des Mines de Saint-Étienne
Lionel Médini – LIRIS, Université Claude Bernard Lyon 1
Imen Megdiche – IRIT, Institut National Universitaire Champollion
Philippe Ramadour – LSIS, Aix-Marseille Université
Philippe Roose – LIUPPA, Université de Pau et des Pays de l’Adour
Franck Ravat – IRIT, Université Toulouse 1 Capitole
Christian Sallaberry – LIUPPA, Université de Pau et des Pays de l’Adour
Olivier Teste – IRIT, Université Toulouse 2 Jean-Jaurès
Élodie Thiéblin – IRIT, Université Toulouse 2 Jean-Jaurès
Nathalie Valles-Parlangeau – IRIT, Université Toulouse 1 Capitole

Ressources liées aux thématiques de ce numéro spécial

Les ressources sont essentiellement disponibles en ligne. Veuillez nous excuser par avance si certaines d’entre elles sont modifiées/indisponibles après la publication de ce numéro spécial.

- Best Practices for Publishing Linked Data: W3C Working Group Note 09 January 2014, <https://www.w3.org/TR/ld-bp/>
- Linked Data (W3C): <https://www.w3.org/standards/semanticweb/data>
- Linked Data - Connect Distributed Data across the Web (Tom Heath) : <http://linkeddata.org/>
- OpenData France : <http://www.opendatafrance.net/>
- W3C study of practices and tooling for Web data standardisation: Dave Raggett dsr@w3.org, W3C Data Activity Lead, Dec. 2017: <https://www.w3.org/2017/12/odi-study/>
- RDF 1.1 Concepts and Abstract Syntax: W3C Recommendation 25/02/2014 : <https://www.w3.org/TR/rdf11-concepts/>
- RDF 1.1 Semantics: W3C Recommendation 25/02/2014: <https://www.w3.org/TR/rdf11-mt/>

- OWL 2 Primer: OWL 2 Web Ontology Language Primer (2nd Edition); W3C Recommendation 11/12/2012 : <https://www.w3.org/TR/owl2-primer/>
- SPARQL 1.1 Query Language: W3C Recommendation 21/03/2013 : <https://www.w3.org/TR/sparql11-query/>
- Compléments sur les ontologies: https://www.w3.org/wiki/Ontology_Dowsing