

Linked Open Vocabularies, un écosystème encore fragile

Pierre-Yves Vandenbussche^{1,2}, Bernard Vatant¹, Jean Charlet^{2,3}

¹ MONDECA, Paris, France

`pierre-yves.vandenbussche@mondeca.com`, `bernard.vatant@mondeca.com`

² INSERM UMRS 872 ÉQ.20, Ingénierie des connaissances en santé, Paris, France
`jean.charlet@upmc.fr`

³ AP-HP – Assistance Publique Hôpitaux de Paris, Paris, France.

Résumé : La publication de données formalisées, liées et ouvertes sur le Web est en plein essor. Pour les décrire, des ontologies et des vocabulaires sont utilisés. L'initiative Linked Open Vocabularies (LOV) se positionne comme un observatoire de cet écosystème de vocabulaires réutilisables. Cet observatoire ne se contente pas de récolter et de rendre visibles les vocabulaires et leurs métadonnées, il joue aussi un rôle social prépondérant dans la promotion de pratiques visant à l'amélioration de la qualité globale de l'écosystème, et à sa gouvernance responsable et durable. Après un an, nous proposons une analyse fine de cet écosystème et sensibilisons la communauté du Web Sémantique sur la nécessité d'une politique de gestion durable des vocabulaires.

Mots-clés : Vocabulaires, Web Sémantique, Open data, Gestion durable.

1 Introduction

Le Web des données ouvertes (Linked Open Data ou LOD) s'appuie sur des vocabulaires formels (RDFS, OWL) pour définir les types et propriétés des ressources décrites (Heath & Bizer, 2011; Ferrara *et al.*, 2011). Le projet Linked Open Vocabularies (Vandenbussche *et al.*, 2011) (LOV) a été développé à partir de ce qui peut paraître une évidence : les données ouvertes liées devraient utiliser des vocabulaires ouverts liés. Notre implication préalable dans l'univers de ces vocabulaires nous a permis de savoir que ce principe séduisant par sa simplicité était encore loin d'être une règle générale. Ce projet nous permet de mettre en place et d'évaluer de façon

plus systématique les pratiques réelles, sur la base d'un certain nombre d'indicateurs de la qualité des vocabulaires. Démarré au printemps 2011 dans le cadre du projet Datalift¹, le projet LOV se positionne désormais comme un observatoire privilégié et novateur de l'écosystème des vocabulaires, entre autres par la prise en compte d'indicateurs quantitatifs et qualitatifs encore très peu évalués auparavant, comme l'interconnexion entre vocabulaires. Les résultats d'ores et déjà acquis montrent la diversité des pratiques, tant techniques que sociales intervenant dans le cycle de vie des vocabulaires. Ils mettent en évidence un certain nombre de points fragiles de l'écosystème mais en même temps des signes encourageants d'une prise de conscience croissante de l'importance de le maintenir en bonne santé, tant technique que sociale. Enfin cet observatoire ne se contente pas de récolter et de rendre visible des données et métadonnées, il joue aussi un rôle social dans la promotion de pratiques visant à l'amélioration de la qualité globale de l'écosystème, et à sa gouvernance responsable et durable. Dans la suite de cet article, nous présentons le projet LOV et le positionnons par rapport aux projets existants. Nous analysons ensuite l'état de l'écosystème des vocabulaires liés dans le périmètre du LOV. Enfin, de cette analyse, nous dégagons certains enseignements dont la gestion durable des vocabulaires qui constitue un enjeu majeur pour le futur du Web Sémantique.

2 LOV, un observatoire des vocabulaires

L'initiative LOV a pour objectif de favoriser et faciliter la réutilisation de vocabulaires pour la description de jeux de données sémantiques liés. Le LOV met à disposition un jeu de données exprimé en langage RDF, qui inventorie les vocabulaires de description des jeux de données du Web sémantique mais également les relations qui lient ces vocabulaires. Le travail de recensement des vocabulaires dans le LOV n'est pas exhaustif. Le Web de données est par définition non contrôlé et par nature en constante évolution. De même que le Web de Données, le LOV a comme caractéristique intrinsèque d'être inachevé et incomplet.

2.1 Travaux connexes et positionnement

Notre initiative complète d'autres projets existants tels que Schemape-
dia, Open Metadata Registry, Swoogle, Falcon-S, Watson ou encore Sin-

1. Voir <http://datalift.org>

dice :

- Schemapedia² est un projet qui inventorie 200 vocabulaires sur la base d'un site Web communautaire. La description d'exemples d'utilisation permet notamment à un utilisateur de le comprendre et de l'utiliser plus efficacement. Ce projet ne met pas à disposition toutes les métadonnées déclarées dans un vocabulaire comme ses créateurs ou les liens qu'il entretient avec le reste de l'écosystème des vocabulaires ;
- Open Metadata Registry (Phipps & Hillmann, 2011) (OMR) est un projet qui permet de référencer, partager et publier des vocabulaires contrôlés, c'est-à-dire des listes de valeurs ou des schémas de concepts (également appelés « Value Vocabularies »³) exprimés principalement en SKOS. Les concepts de ces listes sont ensuite utilisés comme instances dans les données d'un utilisateur. OMR identifie dans une moindre mesure et principalement dans le domaine bibliothécaire des vocabulaires (également appelés « Metadata Element Sets ») qui ont le même sens que ceux utilisés par le LOV ;
- Swoogle(Finin *et al.*, 2005), Falcon-S(Wu *et al.*, 2006) et Watson(D'Aquin *et al.*, 2007) sont tous trois des moteurs de recherche centrés sur la recherche d'éléments d'ontologies dont les vocabulaires constituent un sous-ensemble. Toutefois ces projets ne mettent pas à disposition l'ensemble des métadonnées concernant les vocabulaires tels que la dernière date de mise à jour ou encore sa description ;
- Sindice(Tummarello *et al.*, 2007) est un catalogue et un service de recherche de documents prenant en compte les métadonnées RDF mais aussi microdata, microformats dans les documents HTML. Ces documents représentent une petite partie de l'utilisation des vocabulaires au regard des données liées.

Notre projet s'inscrit dans cet ensemble de projets existants et se focalise sur les vocabulaires dédiés à être réutilisés pour la description de données du Web sémantique. Plus que répertorier les vocabulaires, nous complétons les métadonnées disponibles explicitement dans ces vocabulaires par des informations que nous pouvons recueillir (via une documentation, par échanges avec les auteurs, etc.). Un autre point innovant de notre projet est l'explicitation des relations qu'entretiennent ces vocabulaires.

2. Voir <http://schemapedia.com/>

3. Voir Library Linked Data Incubator Group Report.

2.2 Le vocabulaire VOAF pour la description de vocabulaires

Notre initiative adopte une approche pour les vocabulaires, similaire à ce que LOD est pour les données (Bizer *et al.*, 2009). De même que le vocabulaire Vocabulary of Interlinked Datasets (VoiD) sert à décrire les relations entre jeux de données (Alexander *et al.*, 2009), nous avons étudié les vocabulaires qui permettent de décrire des vocabulaires et les relations qu'ils entretiennent. Afin de respecter les bonnes pratiques du Web de données, nous avons cherché à réutiliser, quand cela était possible, des classes et propriétés existantes. Nous avons créé le vocabulaire Vocabulary Of A Friend⁴ (VOAF) pour permettre la description de relations entre vocabulaires. Certaines propriétés de notre vocabulaire sont donc extraites de vocabulaires de référence du Web de données :

- Dublin Core Terms (dcterms) est un vocabulaire pour la description de métadonnées sur des documents électroniques (Weibel, 1997). Nous utilisons entre autres les propriétés `dcterms:title` pour décrire le titre ; `dcterms:modified` pour indiquer la date de dernière mise à jour ou encore `dcterms:creator` pour préciser le ou les créateurs d'un vocabulaire ;
- Vocabulary Of Interlinked Datasets (void) est un vocabulaire pour décrire les jeux de données. Il apporte des informations de statistiques ou encore la localisation du fichier de contenu d'un jeu de données (Alexander *et al.*, 2009) ;
- Vocabulary for ANNotating vocabulary (vann) est un vocabulaire pour la description de vocabulaires⁵. Nous utilisons ses propriétés `vann:preferredNamespacePrefix` et `vann:preferredNamespaceUri` pour représenter le préfixe à utiliser et l'espace de nom d'un vocabulaire ;
- BIBliographic Ontology (bibo) est un vocabulaire pour la description de données bibliographiques. Nous utilisons sa propriété `bibo:shortTitle` afin de décrire le titre abrégé d'un vocabulaire (D'Arcus & Giasson, 2010).

En ce qui concerne l'interconnexion entre vocabulaires, nous avons défini huit relations qui capturent les caractéristiques de spécialisation, extension, etc. D'autres propriétés spécifiques telles que l'indication du nombre de

4. Voir <http://labs.mondeca.com/vocab/voaf>

5. Voir <http://vocab.org/vann/.html>

classes et de propriétés d'un vocabulaire constituent le vocabulaire VOAF⁶.

3 État de l'écosystème

Le LOV peut être vu comme un écosystème où chaque vocabulaire est maintenu et publié de manière indépendante tout en entretenant des liens de dépendance vis-à-vis des autres vocabulaires. L'aspect social de ce réseau est donc très important et favorise la confrontation de différents points de vue, mais souligne également la fragilité du réseau de vocabulaires lorsqu'une ressource n'est plus accessible.

3.1 Périmètre d'observation

A ce jour⁷ plus de 250 vocabulaires ont été identifiés et intégrés au LOV. Ces vocabulaires peuvent avoir une portée aussi bien générale et servir à la description de métadonnées que spécifique et servir à la description de données d'un domaine particulier (e.g. géo-localisation, sciences de la vie ou encore e-business). Dans le cadre de l'initiative LOV, nous ne nous arrêtons pas sur la définition linguistique de « vocabulaire » mais l'abordons sous l'angle de son utilisation dans le Web sémantique⁸. Dans ce contexte, un vocabulaire se confond avec une ontologie. Toutefois, il est courant de différencier un vocabulaire d'une ontologie par sa nature et son utilisation. Parmi les systèmes d'organisation de la connaissance, les ontologies ont l'engagement formel le plus fort. La réutilisation d'une ontologie implique l'acceptation et le respect des définitions et des contraintes formelles qui sont déclarées dans l'ontologie. Les fournisseurs de données réutilisent plus facilement des structures de données possédant une plus grande souplesse formelle. Les vocabulaires se limitent souvent à RDFS et utilisent peu les contraintes formelles d'OWL. De par sa taille généralement petite et ses faibles contraintes formelles, un vocabulaire est plus enclin à être compris et réutilisé. La valeur médiane du nombre de classes et de propriétés par vocabulaire dans le LOV est de respectivement 7 et 16. Un vocabulaire ne contient généralement pas d'instances si ce n'est pour l'exemple. Un vocabulaire peut réutiliser tout ou partie d'un vocabulaire

6. Les huit relations sont détaillées dans un précédent papier (Vandenbussche *et al.*, 2011).

7. En date du 25 avril 2012.

8. Voir <http://www.w3.org/standards/semanticweb/ontology>

existant. Nous ne nous intéressons qu'aux schémas de description de données (autrement appelés « Metadata Element Sets »⁹) et non aux schémas de concepts (« Value Vocabularies ») généralement représentés en SKOS. L'entrée d'un vocabulaire dans le LOV est aujourd'hui conditionnée par le respect de normes et de bonnes pratiques. Les principaux critères conditionnant l'entrée d'un vocabulaire dans le LOV sont les suivants :

- Le vocabulaire doit être dédié à la description des ressources en RDF : certaines ontologies ont été adaptées d'un contexte « pré-Web ». Nous privilégions les vocabulaires développés spécifiquement dans le contexte des données liées sur le Web ;
- Le vocabulaire doit être exprimé dans un standard du Web sémantique : RDFS ou OWL¹⁰ ;
- Le vocabulaire doit être accessible sur le Web via son espace de noms : L'espace de noms (URI principale) du vocabulaire doit permettre d'accéder par négociation de contenu au vocabulaire formel et à sa documentation.

3.2 Acteurs

Les vocabulaires actuels sont édités par 32 organismes et 129 personnes appartenant à différents types :

- les organismes de normalisation (e.g. W3C ou DCMI) ;
- les conservateurs du patrimoine (e.g. IFLA, OCLC, Europeana, Library of Congress ou bibliothèques nationales) ;
- les groupes de presse et de média (e.g. New York Times, BBC ou IPTC) ;
- les entreprises du Web (promotion de vocabulaires comme schema.org) ;
- les gouvernements et les fournisseurs de données institutionnels (e.g. data.gov, Ordnance Survey ou INSEE) ;
- les universités et les centres de recherche (e.g. DERI ou INRIA) ;
- les sociétés de conseil, les PME, les éditeurs de logiciels (e.g. Talis, TopQuadrant ou Mondeca) ;
- les initiatives individuelles, plus ou moins soutenues par un ou plusieurs des organismes qui précèdent (e.g. taxonconcept, geospecies ou geonames).

9. Voir Library Linked Data Incubator Group Report.

10. Les vocabulaires exprimés en SKOS sont souvent des listes de valeurs (schémas de concepts) et ne rentrent pas dans le périmètre de notre projet.

La pérennité de ces acteurs et donc des vocabulaires qu'ils publient n'est actuellement pas garantie. Par exemple un vocabulaire développé dans le cadre d'un projet de recherche est susceptible de ne pas avoir de financement pour être maintenu au-delà de la date de fin de projet. Nous reviendrons sur cet aspect en section 4.2.

3.3 Indicateurs et résultats

3.3.1 Accessibilité

Pour qu'un vocabulaire prétende à la réutilisation, son espace de noms doit permettre d'accéder par négociation de contenu au vocabulaire formel et à sa documentation (Berrueta & Phipps, 2008). Malgré nos sollicitations auprès des acteurs de l'écosystème LOV, 3% des vocabulaires ne sont pas déréférencables. Cela signifie que la sémantique des données les utilisant ne peut pas être connue.

3.3.2 Métadonnées

La nature récursive de RDF peut être utilisée pour décrire les vocabulaires eux-mêmes. Il existe un large éventail de vocabulaires généraux pour la description de métadonnées relatives aux vocabulaires et de leurs éléments. La publication d'un vocabulaire devrait a minima respecter les bonnes pratiques en matière de métadonnées (Vandenbussche & Vatant, 2011). Seuls 18% des vocabulaires du LOV possèdent la métadonnée de date de création du vocabulaire. Grâce au travail de maintenance du LOV nous avons complété cette information, amenant à 85% des vocabulaires avec une date de création explicitement et formellement formulée. De même 54% des vocabulaires possèdent une métadonnée de créateur/contributeur/éditeur. Grâce au travail de maintenance du LOV nous avons complété cette information, amenant à 76% les vocabulaires possédant une de ces données explicitement et formellement formulée.

3.3.3 Multilinguisme

La recherche et la compréhension des vocabulaires est facilitée grâce aux termes qui dénotent les éléments (classes, propriétés) de ceux-ci. L'information de langue et le multilinguisme sont donc importants. Dans le LOV seulement 12% des vocabulaires possèdent des informations multilingues. Parmi les libellés décrits dans les vocabulaires, 35% n'ont pas

d'information de langue ; 41% sont exprimés en anglais ; 13 autres langues apparaissent moins fréquemment.

3.3.4 Utilisation dans les données

L'information d'utilisation d'un vocabulaire par des jeux de données est pertinente à plusieurs titres. Elle permet à un fournisseur de données de s'inspirer des vocabulaires utilisés par des données externes proches des siennes. Elle permet également de faciliter la compréhension d'un jeu de données en utilisant des vocabulaires déjà connus. Enfin, elle permet de faciliter la réutilisation des données qui seront exprimés dans un même schéma que d'autres données liées. Nous utilisons actuellement des métriques calculées par l'interrogation du serveur cache du LOD¹¹. Le résultat présenté en figure 1 présente le nombre d'occurrences d'un élément d'un vocabulaire dans les jeux de données du LOD. Le projet LOD2 stats¹² donne cette même information avec une technique de mise à jour automatique hebdomadaire (notre méthode précédente utilisait une méthode semi-automatique). Nous étudions actuellement avec cette équipe la manière dont nous pourrions automatiquement utiliser leurs résultats.

LOV distribution ↕	LOV popularity ↕	LOD popularity ↕	Vocabulary Prefix ↕	Element URI ↕
32	198	37690602	foaf	http://xmains.com/foaf/0.1/Person
2	3	32496151	foaf	http://xmains.com/foaf/0.1/nick
0	0	32371723	unpc	http://purl.uniprot.org/core/Sequence
0	0	31558903	foaf	http://xmains.com/foaf/0.1/weblog
0	0	30924404	unpc	http://purl.uniprot.org/core/Domain_Assignment_Statement
0	0	30745633	unpc	http://purl.uniprot.org/core/Resource
0	0	29411616	unpc	http://purl.uniprot.org/core/reviewed
0	0	28469887	unpc	http://purl.uniprot.org/core/identity
0	0	26914075	unpc	http://purl.uniprot.org/core/length
0	0	26884281	unpc	http://purl.uniprot.org/core/obsolete
21	54	26843510	dce	http://purl.org/dcelements/1.1/identifier
0	0	26731600	unpc	http://purl.uniprot.org/core/commonName
1	1	26485921	geo	http://www.w3.org/2003/01/geo/wgs84_pos#long

FIGURE 1 – Métriques d'utilisation des vocabulaires et de leurs éléments dans le LOD et dans le LOV.

3.3.5 Réutilisation par les vocabulaires

La réutilisation de vocabulaires pour la construction de son propre vocabulaire ou ontologie permet de bénéficier des bonnes pratiques de construc-

11. Voir <http://lod.openlinksw.com>

12. Voir <http://stats.lod2.eu/vocabularies>

tion de ces premiers. En effet les vocabulaires dans le LOV ont été éprouvés sur des données et discutés au sein de la communauté. Un vocabulaire est également une source d'idées ou d'enrichissements pour son propre vocabulaire. Dans le LOV, chaque vocabulaire est lié en moyenne à 4 autres formant ainsi le maillage fin de l'écosystème. Parmi ces liens on peut distinguer les vocabulaires utilisés pour décrire des métadonnées sur un vocabulaire (en moyenne 2 vocabulaires distincts sont utilisés) et les vocabulaires dont un autre étend, spécialise, se lie d'équivalences, etc. (en moyenne 3 vocabulaires distincts sont utilisés).

3.4 Rétroactions positives

Comme nous l'avons vu dans les sections précédentes, notre action dans le projet LOV dépasse l'observation et inclut l'interaction avec les acteurs pour améliorer la qualité générale de l'écosystème. A ce jour, toutes les sollicitations de modification ont été prises en compte par les auteurs. Nous continuons donc cette tâche et proposons d'autres services comme la possibilité de suggérer un nouveau vocabulaire pour son intégration dans le LOV¹³ (19 vocabulaires ont été suggérés). Une perspective intéressante est l'intégration d'une composante communautaire aux interfaces du LOV dans laquelle les utilisateurs peuvent commenter, recommander un vocabulaire, argumenter leur utilisation, etc. Le service de recherche¹⁴ au sein de l'écosystème bénéficie également d'un retour positif avec plus de 1200 requêtes par mois. Notre projet reçoit le soutien institutionnel de OKFN¹⁵ (Open Knowledge Foundation) qui se propose d'héberger le site Web de manière durable ; mais également d'autres organismes comme le DCMI, le W3C ou encore la Library of Congress¹⁶.

4 Les vocabulaires, patrimoine commun

4.1 Un patrimoine vivant à conserver et à valoriser

Nous avons jusqu'ici montré l'importance technique des vocabulaires et la position critique qu'ils occupent dans la représentation des données

13. Voir <http://labs.mondeca.com/dataset/lov/suggest/>

14. Voir <http://labs.mondeca.com/dataset/lov/search/>

15. Voir <http://okfn.org/>

16. Voir <http://www.loc.gov>

ouvertes et les applications qui les exploitent. Mais les vocabulaires représentent également un patrimoine précieux par le savoir accumulé qu'ils contiennent, connaissances techniques aussi bien que terminologiques. Même si un vocabulaire n'est plus maintenu ni utilisé aujourd'hui, il représente une mine de connaissance qui peut être réexploitée à tout moment dans le futur. On peut comparer à cet égard les vocabulaires à l'édifice des mathématiques, où certains résultats peuvent être en sommeil pendant des décennies, voire des siècles, et brusquement être réveillés par des applications nouvelles et insoupçonnables au temps de leur découverte. Les vocabulaires ouverts liés représentent donc un patrimoine commun dont la valeur technique et culturelle mérite qu'on le gère et le préserve durablement au même titre que toute autre production de la création intellectuelle humaine. Cette gestion doit permettre de façon pérenne d'accéder facilement aux vocabulaires publiés et d'évaluer leur qualité selon tous les indicateurs cités dans les sections précédentes. Elle doit permettre également aux gestionnaires d'un vocabulaire de connaître et mesurer l'utilisation qui en est faite, de répondre aux retours de la communauté des utilisateurs et de le faire évoluer, de l'interconnecter et l'aligner avec d'autres vocabulaires semblables ou complémentaires, et d'éviter de réinventer ce qui existe déjà. Les leçons apprises à travers le projet LOV nous permettent d'ores et déjà d'envisager des solutions pour une telle gestion, durable et responsable, de l'écosystème des vocabulaires. Ces leçons doivent être consolidées et complétées. L'idée d'une grande enquête dirigée vers les producteurs de vocabulaires est actuellement à l'étude. Elle permettrait d'une part de rassembler des informations introuvables dans la publication des vocabulaires, et d'autre part de sensibiliser la communauté des producteurs de vocabulaires à l'importance de publier et de maintenir à jour de telles informations.

4.2 Vers des solutions de gestion durable

4.2.1 Visibilité, sélection et découverte

Pour être utilisé, un vocabulaire doit d'abord être connu ou facile à découvrir, et des critères de sélection doivent être mis en place de façon à ce qu'un gestionnaire de données ou de vocabulaires puisse juger rapidement de la pertinence de tel ou tel vocabulaire pour organiser ses données ou compléter ses propres vocabulaires. LOV ouvre la voie à des outils de sélection des vocabulaires, offrant à la fois des fonctionnalités de recherche basées sur des métadonnées riches, et un support social à la sélection, par la définition de responsabilités visibles. Ces outils doivent être complétés

et consolidés. La migration du projet LOV dans le cadre de l'Open Knowledge Foundation est un pas dans cette direction.

4.2.2 Maintenance et responsabilité

Les réactions de la communauté montrent l'importance pour les responsables de vocabulaires d'exposer de façon explicite leur politique de maintenance : pérennité ou non du service et des URI, stabilité ou évolution, fréquence des mises à jour, rétro-compatibilité ou non avec les versions antérieures, politique de dépréciation et de versioning. L'établissement de critères partagés, dans l'esprit des niveaux de droits définis par Creative Commons, permettrait une clarification de la multiplicité actuelle des pratiques, à la fois pour les producteurs et pour les utilisateurs de vocabulaires.

4.2.3 Interconnexion et résilience de l'écosystème

On a vu l'importance de l'interconnexion des vocabulaires, tant du point de vue technique (meilleure interopérabilité des données), qu'économique (éviter de réinventer la roue) et social (travail commun d'ajustement, construction modulaire de l'écosystème). Mais l'interconnexion génère également des dépendances. Réutiliser un vocabulaire suppose une confiance dans sa fiabilité à moyen, voire long terme, c'est-à-dire en particulier la permanence des URI, de leur définition, et la pérennité des acteurs qui les maintiennent. Des solutions doivent être imaginées en cas de défaillance des ressources techniques ou humaines.

4.2.4 Conservation à long terme

Quelle que soit son implication et son investissement pour maintenir ses vocabulaires, un responsable individuel ou institutionnel, doit prévoir qu'il risque un jour de ne plus être en mesure de le faire. Une politique d'héritage de la gestion des DNS est donc souhaitable. En dernier ressort, il semble que le meilleur légataire final d'un vocabulaire qui n'est plus maintenu soit une grande bibliothèque. Une synergie entre la communauté des éditeurs de vocabulaires et celle des bibliothèques doit donc être mise en place¹⁷. C'est entre autres l'une des tâches auxquelles le groupe de travail DCMI Vocabulary Management Community devrait s'intéresser dans les mois et années à venir.

17. Voir Library Linked Data Incubator Group Final Report.

5 Conclusion

Quel modèle de gouvernance peut-on imaginer pour un avenir responsable et durable de l'écosystème des vocabulaires, que nous avons tendance à désigner maintenant sous le terme « Vocabulary Commons » ? Cette gouvernance doit trouver un équilibre entre d'une part la dispersion extrême à travers la multiplication de jargons spécifiques à de petites communautés et à des domaines limités, et d'autre part une approche centralisée et contrôlée, où les « bons » vocabulaires seraient imposés par des autorités de droit (organismes de standardisation) ou de fait (grandes multinationales du Web). Autrement dit, il s'agit de combiner la richesse et l'agilité du modèle décentralisé, multilingue, multiculturel et social du Web, et le savoir-faire millénaire des bibliothèques dans l'organisation des vocabulaires et des connaissances. Un premier pas est la reconnaissance par chacune des parties de la valeur ajoutée par l'autre. La dynamique dans cette direction est en bonne voie.

Remerciements

Les auteurs remercient L'Agence Nationale de la Recherche qui a en partie financée ces travaux : ANR-10-CORD-009.

Références

- ALEXANDER K., CYGANIAK R., HAUSENBLAS M. & ZHAO J. (2009). Describing linked datasets. In *Proceedings of the 2nd Workshop on Linked Data on the Web (LDOW2009)*.
- BERRUETA D. & PHIPPS J. (2008). Best practice recipes for publishing rdf vocabularies. *Working draft, W3C*.
- BIZER C., HEATH T. & BERNERS-LEE T. (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, **5**(3), 1–22.
- D'AQUIN M., BALDASSARRE C., GRIDINOC L., SABOU M., ANGELETOU S. & MOTTA E. (2007). Watson : Supporting next generation semantic web applications.
- D'ARCUS B. & GIASSON F. (2010). Bibliographic ontology specification. *Retrieved October, 8*.
- FERRARA A., NIKOLOV A. & SCHARFFE F. (2011). Data linking for the semantic web. *International Journal on Semantic Web and Information Systems (IJSWIS)*, **7**(3), 46–76.

- FININ T., DING L., PAN R., JOSHI A., KOLARI P., JAVA A. & PENG Y. (2005). Swoogle : Searching for knowledge on the semantic web. In *PROCEEDINGS OF THE NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE*, volume 20, p. 1682 : Menlo Park, CA ; Cambridge, MA ; London ; AAAI Press ; MIT Press ; 1999.
- HEATH T. & BIZER C. (2011). *Linked Data : Evolving the Web into a Global Data Space : Theory and Technology*, volume 1. Morgan & Claypool Publishers.
- PHIPPS J. & HILLMANN D. (2011). The open metadata registry : An update. *Bulletin of the American Society for Information Science and Technology*, **37**(4), 35–37.
- TUMMARELLO G., DELBRU R. & OREN E. (2007). Sindice. com : Weaving the open linked data. In *Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference*, p. 552–565 : Springer-Verlag.
- VANDENBUSSCHE P.-Y. & VATANT B. (2011). *Metadata Recommendations For Linked Open Data Vocabularies*. Rapport interne, Mondeca.
- VANDENBUSSCHE P.-Y., VATANT B. & ROZAT L. (2011). Qualité et robustesse dans le web de données : Linked open vocabularies. In *Atelier Qualité et Robustesse pour le Web de Données IC 2011*.
- WEIBEL S. (1997). The dublin core : a simple content description model for electronic resources. *Bulletin of the American Society for Information Science and Technology*, **24**(1), 9–11.
- WU H., CHENG G. & QU Y. (2006). Falcon-s : An ontology-based approach to searching objects and images in the soccer domain. *Supplemental Proceedings of ISWC*.