Interopérabilité et web sémantique : le cas d'une plateforme de publication d'articles scientifiques en sciences de gestion

Eric Thivant

▶ **To cite this version:**

Eric Thivant. Interopérabilité et web sémantique : le cas d'une plateforme de publication d'articles scientifiques en sciences de gestion. 8e Colloque International de l'ISKO-France, Jun 2011, Lille, France. hal-00671892

HAL Id: hal-00671892

https://univ-lyon3.hal.science/hal-00671892

Submitted on 19 Feb 2012

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.

8e Colloque International de l'ISKO-France, 27 -28 juin 2011, Lille

Interopérabilité et web sémantique : le cas d'une plateforme de publication d'articles scientifiques en sciences de gestion

Thivant Eric*

*SICOMOR, CR MAGELLAN Université Jean Moulin, Lyon 3,
6 Cours Albert Thomas, 69008 Lyon, France
04 78 78 72 85 / 04 78 78 77 50
thivant@univ-lyon3.fr

Résumé. Cet article s'intéresse à la question de la question essentielle de l'interopérabilité entre systèmes d'organisation des connaissances en prenant le cas d'une plateforme de publication d'articles scientifiques en sciences de gestion. L'utilisation du Dublin Core simple et/ou étendue avec un profil d'application spécifique à une discipline est largement utilisée par des entrepôts de données pour l'exportation des métadonnées. L'arrivée du Dublin Core Qualifié, exprimé sous le format RDF (Resource Description Framework) est une réponse immédiate au moissonnage des entrepôts de données, notamment avec les Coins (ContextObjects in Spans). Mais il faut aller plus loin selon la méthode SAFARI qui conseille de rédiger une taxonomie, de développer un langage standardisé SKOS, voir d'utiliser des vocabulaires contrôlés pour intégrer le web sémantique. Et dans cet article nous nous intéresserons donc à la question délicate des classifications et des vocabulaires contrôlés qui pourraient être définis dans le cadre spécifique des sciences de gestion.

Mots-clés : Dublin Core, RDF-XML, Web Sémantique, Plateforme de publications, sciences de gestion

Abstract. This article deals with the question of interoperability between Knowledge Organization Systems, taking into example a small repository of scientific article in Management Science. The use of simple Dublin Core and additional Dublin Core Elements with specific and local application profile is more and more used by database system. Nowadays, qualified Dublin Core is encoded in terms of RDF, the Resource Description Framework, is an answer to harvest metadata from repository with for example COINS (ContextObjects in Spans). But according to SAFARI methodology, we should go to the semantic web and be able to write better taxonomy, to develop really standardized language such as SKOS and to use standardized vocabularies. And in this article we will ask some questions about the development of theses standardized vocabularies used in management sciences.

Mots-clés : Dublin Core, RDF-XML, Sémantic Web, Repository, Management Sciences

1 Introduction

Cette communication s'intéresse à la question essentielle de l'interopérabilité entre des systèmes d'organisation des connaissances hétérogènes en prenant le cas d'une plate-forme de publication d'articles scientifiques en sciences de gestion. Comme le rappelle Ducloy, Nicolas, Le Hénaff, Foulonneau, Grivel et Ducasse (2006), les chercheurs sont à la recherche de la reconnaissance et de la visibilité de leurs travaux scientifiques. Si les publications des chercheurs sont référencées par les revues scientifiques qui intègrent souvent l'exportation de métadonnées liées aux références bibliographiques (Ducloy, 2004), il n'en est pas de même pour les communications scientifiques qui sont sous exploitées. En effet, nous ne pouvons que constater le peu d'intérêt de certains comités d'organisation de colloques qui ne mettent pas assez en valeur sur internet les communications acceptées et ne les associent pas nécessairement à des métadonnées adaptées. Donc, il y a là un important gisement sur lequel nous pouvons travailler pour faire connaître les travaux des chercheurs.

2 Problématique

Mais avant de commencer, nous rappellerons qu'une métadonnée est une « donnée qui définit et décrit une autre donnée » (ISO/IEC 11179-3, 2003) ou en d'autres termes « c'est de l'information structurée qui décrit, explique localise la ressource et en facilite la recherche, l'usage et la gestion » (NISO, Understanding Metadata, 2004). Ces métadonnées sont essentielles pour permettre l'échange et l'interopérabilité des données. C'est dans ce cadre qu'il faut choisir d'utiliser un protocole simple et interopérable pour le partage des données. Nos travaux s'inscrivent donc dans la lignée des travaux sur le protocole Open Archive Initiative (OAI) comme pour (Foulonneau, 2004).Comme le soulignent Claerr et Westeel (2010), l'Open Archive Initiative née en 1999, est « la création d'un protocole d'échange de métadonnées basé sur le Dublin Core, dont l'objectif est de faciliter le signalement des documents et la création de collection interopérable ». (2010 : 115). Le protocole qui lui est associé appelé « Open Archive Initiative Protocol for Metadata Harvesting » insiste sur l'utilisation du Dublin Core pour décrire les métadonnées exposées dans ces entrepôts de données D'autres formats de métadonnées plus complexes peuvent être utilisés mais à condition d'être exprimé en XML suivant un schéma accessible en ligne… Mais en réalité dans les faits et dans notre cas, le Dublin Core comportant seulement quinze éléments est très vite limité et le Dublin Core non qualifié permet aux fournisseurs de services de moissonner nos documents numériques.

L'arrivée du web sémantique du W3C et des nouveaux langages tels que le RDF/XML (Ressources Description Framework) permettrait-il d'améliorer l'interopérabilité et les échanges de données ? L'initiative des données liées (Linked Data) doit-il nous pousser à passer toutes nos ressources en RDF ? Devons nous développer une nouvelle représentation formelle de système d'organisation des connaissances avec un vocabulaire contrôlé, tels que nous pouvons le rencontrer avec le SKOS (Simple Knowledge Organisation System) ? L'existence d'un triplet « sujet », « prédicat » et « objet » nous permet-il d'envisager l'élimination de toute ambiguïté ? Pouvons-nous clarifier les relations entre usages et structures ? Comment faire émerger de nouvelles structures en fonction des usages ? Au cours de cette présentation, nous allons vous démontrer la difficulté de rendre interopérables des données en prenant donc l'exemple des publications scientifiques en sciences de gestion. Après un bref rappel sur les bouleversements technologiques à venir avec ce web sémantique, nous vous présenterons dans une première partie la plate-forme de publication des articles scientifiques du centre de recherche en gestion, puis dans une deuxième partie nous présenterons les solutions mises en place actuellement pour rendre interopérable nos métadonnées et nous poser la question d'une solution plus pérenne dans le futur.

3 Présentation de la plate-forme de publication

Le Centre de Recherche en gestion répertorie l'ensemble des publications et des communications des chercheurs de son centre de recherche depuis cinq ans et met en ligne les articles avec résumés, des métadonnées associées, et parfois des communications intégrales en accord avec les auteurs. À ce jour plus de mille références bibliographiques sont présentes sur ce site. C'est un processus long et complexe qui nécessite une implication de la part de tous les chercheurs du centre de recherche qui nous signalent leurs publications. Ce processus est parallèle à celui la mise en ligne volontaire par les chercheurs du centre de leurs articles sur la plate-forme de publication scientifique commune HAL en vertu des accords signés entre le CNRS et la CPU. En effet, notre université a effectivement choisi de créer une unité de dépôt sur le serveur HAL du CCSD, afin d'éviter les problèmes de mise à jour et d'indexation à long à terme. Néanmoins HAL ne remplit pas toutes nos exigences notamment en matière de reporting au quotidien. En effet, ce dépôt concerne l'ensemble des centres de recherche de l'Université. Et donc plusieurs opérations d'agrégation et de compilation des références bibliographiques des travaux des différents groupes de recherche sont nécessaires pour arriver à des résultats satisfaisants. La solution de création

d'une unité de dépôt supplémentaire pour une seule équipe d'accueil sur le serveur HAL pourrait être intéressante, mais elle n'est pas envisageable. Donc pour gérer les besoins ponctuels du centre (liste de publications et de communications à diffuser en interne ou en externe), nous avons donc choisi de développer une plate-forme spécifique. En revanche nous avons développé solution-maison avec une gestion un peu plus optimale des métadonnées associés à ces articles, nous avons pris le parti de développer d'abord des Coins (ContextObjects in Spans) sont des spécifications simples qui permettent de publier des références OpenURL en HTML. Ces COinS permettent une réutilisation des données avec le logiciel de gestion bibliographique Zotero. Sa facilité d'utilisation par les chercheurs est très appréciée. Plusieurs séminaires ont été organisés par le centre de recherche et/ou par l'URFIST à ce sujet. L'exportation sous le format Zotero RDF ou DC non qualifié RDF est intéressante et permet de limiter le temps de codage des métadonnées associés à ces articles pour leur exploitation ultérieure. Le développement des COinS ne résout pas cependant le problème de l'interopérabilité des données. Le choix d'une plate-forme open source comme Fedora Commons[1] ou D-Space[2] permettrait de faciliter l'indexation et renforcerait aussi l'interopérabilité des données, mais cela serait-il suffisant ?

4 L'interopérabilité en question à l'ère du web sémantique

Baser son développement uniquement sur des métadonnées associées au Dublin Core simple pose des problèmes de moissonnage aux fournisseurs de services. Notre projet s'est affiné avec, pour chaque article présenté, l'affichage des métadonnées de type Dublin Core étendu qui lui sont associées. Il a été en effet, proposé de rédiger un profil d'application spécifique aux articles du centre de recherche en gestion. Un profil d'application du Dublin Core est « *une déclaration spécifiant quelles métadonnées une organisation, un fournisseur de services ou une communauté d'utilisateurs utilisent dans ses métadonnées et comment ces termes ont été adaptés à une application particulière* ».[3i] Le profil

[1] **http://fedora-commons.org/**

[2] http://www.dspace.org/

[3] Dublin Core Application Profile (DCAP) A DCAP is a declaration specifying which metadata terms an organization, information provider or user community uses in its metadata and how those terms have been customized or adapted to a particular application. Par definition a DCAP is based on Dublin Core and follows DCMI Grammatical Principles [DCMI-PRINCIPLES]. A DCAP consists of a Descriptive Header and one or more Term Usages.

d'application peut aider à surmonter les problèmes d'interopérabilité liée à la spécificité des métadonnées associées à cette discipline. Cependant, nous sommes encore loin avec ces techniques de créer le web sémantique du W3C. Pour Ron (2010) la méthodologie SAFARI doit nous guider pour atteindre ce but. Cette méthode s'appuie sur six étapes :

S pour Start	**A pour Add**	**F pour Facets**	**A pour Adopt**	**R pour RDF**	**I pour Integrate**
15 éléments du Dublin Core	Ajout des éléments locaux et des éléments DCMI additionnels	Passer des champs de textes libres à la rédaction d'une taxonomie	Adopter un langage standardisé – type skos	Passer des URL aux URI et aux vocabulaires contrôlés	Lier les informations à l'initiative linked data et les intégrer au web sémantique.

TAB. 1: les six étapes de la démarche SAFARI

5 Futur développement et importance de la spécificité de la discipline « sciences de gestion »

Comme vous pouvez le constater, nous sommes seulement à la phase deux de cette marche intégratrice vers le web sémantique. Il n'y a plus la notion de validation comme dans le XML, mais la notion de vocabulaire est pourtant essentielle. Pour pouvoir donc avancer, il faudra définir des vocabulaires spécifiques aux sciences de gestion et réduire au maximum les ambiguïtés linguistiques inhérentes à cette discipline. Le thésaurus utilisé par la base de données EBSCO relie certains mots comme l'exemple de « carrières » ou « careers » en anglais qui renvoie vers les termes « occupational » ou « vocational guidance ». Cependant, il faudra dépasser les concepts généraux pour préciser des termes spécifiques. La présence d'une liste de termes validés et de relations existants entre ces termes scientifiquement devra être réalisée auprès des enseignants chercheurs volontaires dans cette démarche et devra être couplée à une analyse des mots-clés des articles scientifiques déjà rédigés. Cependant, cela

ne sera pas simple, car souvent les mêmes termes sont repris dans les sous-disciplines des sciences de gestion et elles peuvent alors parfois prêter à confusion. Par exemple, le terme « carrière » est très utilisé en « Gestion des ressources humaines » comme nous pouvons le voir dans ces deux communications présentées ci-dessous. Lorsque nous parlons de « carrières nomades » et « ancre de carrière », quelles relations existent entre ces deux termes ?

LACROUX, Alain. DEPREZ, Anne (2008). Peut-on faire carrière dans l'intérim ? Les apports du concept de carrières nomades. 5ème journées sur les carrières : Carrières et Contextes. Lyon, FRANCE. 2008-05-23
Mots-clés : Intérim, insécurité perçue, carrières nomades

Figure 1: détail d'une notice bibliographique sur le site web

ROGER, Alain. (2006)Développement et validation d'une échelle de mesure des orientations de carrière individuelles. 14e congrès AIPTLF (Association Internationale de Psychologie du Travail de Langue Française). Hammamet, TUNISIE.
Mots-clés : carrière, orientation, ancre de carrière, aspirations professionnelles

Figure 2 : Détails d'une autre notice bibliographique sur le site web

Étant donné la difficulté de définir un rapport entre ces deux termes sans la prise en compte du contexte et d'avis externe auprès des chercheurs, la mise en place d'un vocabulaire contrôlé nous semble difficile. La démarche SAFARI n'explique pas comment élaborer ces vocabulaires. D'autant que, lorsque nous allons parvenir à séparer ces termes, il faudra préciser ces adjectifs qualificatifs tels qu' « attitudes de carrière protéennes et sans frontière » ou « orientation carriériste».

CERDIN, Jean-Luc. LE PARGNEUX, Marie. L'impact de la carrière sur la réussite individuelle de la mobilité internationale. 5ème journées sur les carrières : Carrières et Contextes. Lyon, FRANCE. 2008-05-23
Mots-clés : réussite individuelle mobilité internationale, ancres de carrière, attitudes de carrière protéennes et sans frontières, orientation carriériste, congruence

Figure 3 : Détails d'une autre notice bibliographique

6 Conclusion

Cette étude soulève le problème de la mise en place et de l'évolution d'une plate-forme numérique de publication scientifique. Elle point aussi le problème de l'interopérabilité des données, lorsque nous travaillons sur des métiers et des disciplines différentes. D'autres classifications (métiers ou classifications par facettes) permettant plusieurs entrées devraient être envisageables pour gérer cette multitude de termes de gestion. Et enfin cette étude souligne que le web sémantique reste encore à définir et à développer et auxquelles les chercheurs doivent prendre part eux-mêmes à cette action.

Bibliographie

[1] Ducloy J. (2004) Plaidoyer pour un réseau d'inventaire des résultats de la recherche. Disponible sur http://archivesic.ccsd.cnrs.fr/sic_00001147

[2] Ducloy, Nicolas, Le Hénaff, Foulonneau, Grivel et Ducasse (2006) *Metadata towards an e-research cyberinfrastructure,* Dublin Core Conférence

[3] Claerre T., Westeel I. (2010). *Numériser et mettre en ligne*. Presses de l'ENSSIB, la boite à outils.

[4] http://www.openarchives.org/OAI/openarchivesprotocol.html

[5] Nawrocki, F. Le protocole OAI et ses usages en bibliothèque Disponible sur http://www.culture.gouv.fr/culture/dll/OAI-PMH.htm

[6] Ron Daniel J. (2010) *Six Steps SAFARI from the Dublin Core to the Semantic Web*. Dublin Core 2010 Conference, Pittsburgh disponible sur http://www.dublincore.org/resources/training/dc-2010/Tutorial4_SAFARIDaniel.pdf

---

i