

NoDEfr-1 : Métadonnées pédagogiques et Web des données

Nicolas Delestre, Yolaine Bourda, Erwan Le Gall, Rosa-Maria Gomez de Regil, Yohan Colmant, Marie Peterlongo, Romuald Verrier

▶ **To cite this version:**

Nicolas Delestre, Yolaine Bourda, Erwan Le Gall, Rosa-Maria Gomez de Regil, Yohan Colmant, et al.. NoDEfr-1 : Métadonnées pédagogiques et Web des données. Environnements Informatiques pour l'Apprentissage Humain, Jun 2019, Paris, France. hal-02161156

HAL Id: hal-02161156
https://hal.science/hal-02161156

Submitted on 20 Jun 2019

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.

NoDEfr-1 : Métadonnées pédagogiques et Web des données

Nicolas Delestre[1], Yolaine Bourda[2],Erwan le Gall[3], Rosa-Maria Gomez de Regil[4], Yohan Colmant[5], Marie Peterlongo[6], Romuald Verrier[3]

[1] LITIS Insa Rouen Normandie, nicolas.delestre@insa-rouen.fr
[2] LRI, CentraleSupélec, Université Paris Saclay, Gif-Sur-Yvette, France. Yolaine.Bourda@lri.fr
[3] Ministère de l'Éducation nationale et de la Jeunesse, prenom.nom@education.gouv.fr
[4] CNED, Rosa-Maria.Gomez@cned.fr
[5] Université Polytechnique Hauts-de-France, yohan.colmant@univ-valenciennes.fr
[6] Unisciel, marie.peterlongo@unisciel.fr

Résumé. Les descriptions normalisées des nombreuses ressources d'apprentissage existantes reposent actuellement majoritairement sur un des profils d'application du LOM. Ces descriptions (i.e. métadonnées) sont accessibles soit via des portails soit sous forme de données ouvertes mais dans des formats difficilement traitables par des ordinateurs. D'un autre côté, la montée en puissance des données ouvertes et liées (LOD) permet de favoriser la réutilisation et le partage des données entre applications et d'imaginer de nouveaux services. La structure du LOM (arbre) rend difficile la création d'une expression RDF/OWL (graphe) faisant consensus. Enfin, l'ISO a normalisé le MLR (Metadata for Learning Resources) qui bénéficie d'une expression en RDF/OWL. Cet article commence par présenter brièvement le MLR et son profil d'application français, le NoDEfr-1. Ensuite, nous présenterons le *binding* RDF/OWL conçu ainsi qu'un outil de transformation automatique des métadonnées existantes.

Mots-clé. Description de ressources pédagogiques · Métadonnées · Profil d'application · Web des données · Données ouvertes et liées

Abstract. The descriptions of many existing educational resources are currently based mainly on LOM application profiles. These descriptive records (i.e. metadata) are accessible either via portals or as open data sets but in formats that are difficult to process by humans and machines alike. On the other hand, the rise of linked open data (LOD) makes it possible to promote the understanding and consumption of this metadata, reuse and sharing of data between applications and to imagine new services. The LOM structure (a tree) makes it difficult to reach a consensus on an RDF/OWL binding (a graph). Hence ISO has standardized the MLR (Metadata for Learning Resources) which has an RDF/OWL binding. In this paper, we begin with a brief presentation of the MLR and its French application profile : NoDEfr-1. Then we present the RDF/OWL binding and a tool for automatic transformation of existing metadata.

Keywords: Description of educational resources · Metadata · Application profile · Web of Data · Linked Open Data

1 Introduction

De très nombreuses ressources pédagogiques sont actuellement disponibles librement sur le Web. Néanmoins ces ressources ne peuvent pas toujours être indexées de façon satisfaisante sur la seule base de leur contenu, l'indexation « plein texte » étant notamment source d'ambigüités sémantiques. Par conséquent des métadonnées leur sont généralement associées. Le schéma le plus courant pour la description de celles-ci est le LOM[7]. Cette solution a néanmoins de nombreux inconvénients : les différents profils d'application peuvent être incompatibles entre eux, la sémantique de certaines métadonnées est équivoque, le multilinguisme n'est pas natif, certaines informations sont dupliquées, il n'existe aucun consensus sur une expression RDF/OWL, etc.

En France, depuis 2003, le MESRI soutient les UNT (Universités Numériques Thématiques) qui ont produit de nombreuses ressources de qualité, pour la plupart librement accessibles. Celles-ci sont décrites par des métadonnées conformes au schéma SupLOMFR - profil d'application du LOMFR, lui-même profil d'application du LOM - dédié à l'enseignement supérieur français. Dans ce contexte, les métadonnées produites par différents établissements sont publiées grâce à un réseau de portails OAI-PMH. Bien qu'utile, ce type de service est dédié uniquement à des personnes et ne permet pas l'indexation de matériel pédagogique par les moteurs de recherche généralistes. Cela limite donc la réutilisation de ces métadonnées par des logiciels et empêche leurs liaisons avec d'autres types de données.

Par ailleurs, l'explosion du Web sémantique et des données liées [2] a permis la production d'un grand nombre de données avec une sémantique univoque, augmentant la pertinence des résultats de recherche qui les exploitent et donc la visibilité de ces données. Dans notre contexte, le respect des principes des données liées peut être bénéfique à bien des égards. En effet, si des ressources pédagogiques, issues de différents entrepôts, sont liées entre elles ou avec d'autres types de ressources : (1) des services inter-entrepôts peuvent être créés ; (2) les ressources d'apprentissage peuvent référencer des ressources proches de leurs sujets pour offrir des informations utiles (i.e. DBpedia [3]) ; (3) les ressources d'apprentissage peuvent être liées à des taxonomies spécifiques ; (4) les ontologies sous-jacentes peuvent être mises en correspondance et servir de base à des inférences.

UNIT (l'UNT dédiée aux sciences et technologies) a ainsi financé deux projets : (1) le projet SemUNIT ([4]), preuve de concept de l'insertion des descriptions des ressources des UNT dans le LOD ; (2) le projet LOFAR qui propose un moteur de recherche de ressources pédagogiques à partir de sélection de cours [5]. Ces deux projets ont mis en évidence l'intérêt d'une telle approche et ont permis de sensibiliser une partie de la communauté française, peu au fait de ces technologies.

Cet article commence par présenter brièvement la norme ISO MLR. Ensuite nous présentons notre contribution : tout d'abord le profil d'application français NoDEfr-1, puis son expression en OWL pour le Web des données et enfin la transformation automatique des métadonnées respectant un des dérivés du LOMFR vers cette expression.

2 MLR

L'objectif de la norme ISO/IEC 19788 Metadata for Learning Resources (Métadonnées pour ressources d'apprentissage) dite « MLR » est de décrire des ressources pédagogiques ainsi que des ressources qui leur sont liées dans un contexte international multilingue et multiculturel. Le MLR, reposant sur des fondations explicites et précises, permet, non seulement, de pallier les insuffisances du LOM, largement connues, mais surtout il offre de nouvelles facilités dans la description des ressources pédagogiques ainsi que dans la diffusion de ces descriptions. La visibilité des ressources en est ainsi élargie. Dans le contexte du MLR, un élément de métadonnées est défini comme un élément de données utilisé pour décrire une ressource. Par la suite, nous parlerons donc d'éléments de données (et non plus d'éléments de métadonnées).

Le MLR est une norme en plusieurs parties facilement extensible, au gré des besoins. Chaque partie du MLR a un objectif très précis :

Partie 1 – *Framework* : elle sert de fondation à l'ensemble de la norme. Elle décrit comment spécifier : (1) les éléments de données et leurs caractéristiques, (2) les classes et leurs caractéristiques, (3) les règles de contenu applicables aux valeurs que peuvent prendre les éléments de données, (4) les profils d'application, (5) les vocabulaires.

Partie 2 – Éléments du Dublin Core : elle spécifie chacun des éléments de données du Dublin Core en utilisant la partie 1 du MLR.

Partie 3 – Profil d'application de base : elle définit un profil d'application fondé sur la partie 2 du MLR. Elle exprime un exemple de spécification d'un profil d'application.

Partie 4 – Éléments techniques : elle contient les spécifications des éléments de données permettant de décrire des informations de type technique.

Partie 5 – Éléments pédagogiques : elle contient les spécifications des éléments de données permettant de décrire des informations de type pédagogique.

Partie 7 – Expressions : elle propose une expression en RDF des différentes entités constituant le MLR. Elle explicite la représentation des identifiants MLR en IRI et le *mapping* des différentes entités introduites dans le MLR en RDF. Elle fournit de plus une ontologie OWL2 des classes et propriétés introduites dans les parties du MLR connues à ce jour.

Partie 8 – Enregistrements MLR : elle contient les spécifications des éléments de données permettant de décrire des enregistrements MLR.

Partie 9 – Éléments pour les personnes : elle contient les spécifications des éléments de données permettant de décrire des personnes physiques ou morales (individus ou organisations).

Partie 11 – Migration du LOM vers le MLR : elle est un guide pour la conversion d'enregistrements au format LOM en ensemble d'éléments au format MLR.

3 NoDEfr-1

La norme « Normalisation des descriptions pour l'éducation en France – Partie 1 : description des ressources » dite « NoDEfr-1 », profil d'application du MLR est en cours de finalisation. Cette norme a vocation, à terme, à remplacer le LOMFR et ses deux profils d'application que sont le ScoLOMFR (à destination des acteurs de l'enseignement primaire et secondaire) et le SupLOMFR (à destination des acteurs du supérieur).

Les travaux autour du NoDEfr-1 ont été menés dans des cadres différents (AFNOR, InterUNT...) avec des objectifs différents. D'une part, dans le cadre de l'AFNOR, il s'agit de produire un document normatif (i.e. une norme française) qui a vocation à faire foi mais qui, comme tout document normatif, peut être difficile à lire. D'autre part, dans un groupe plus informel mêlant acteurs du MENJ, du MESRI, de l'inter-UNT, l'objectif est de produire : (1) une proposition d'expression en RDF/OWL, non normative, et mise à la disposition de la communauté française voire francophone afin que celle-ci puisse l'utiliser pour ses besoins, (2) un document d'accompagnement explicitant entre autres les différences avec le LOMFR (3) une possibilité de transformation automatique de descriptions de ressources respectant le LOMFR (ou un de ses dérivés) exprimées en XML en descriptions de ressources respectant le NoDEfr-1 exprimées sous forme de triplets RDF.

3.1 La norme

La définition de la norme NoDEfr-1 a suivi les recommandations de la partie 3 du MLR. À cela s'ajoutent trois règles que s'étaient fixés les experts de l'AFNOR. Tout d'abord il fallait que l'on retrouve dans le NoDEfr-1 tous les éléments de données du LOMFR qui étaient réellement utilisées. Certains étaient présents dans le MLR, d'autres non, ces derniers ont donc été ajoutés. Ensuite lorsque des éléments de données ont été ajoutés, le « type IRI » a été privilégié, favorisant ainsi l'insertion des descriptions des ressources dans le Web des données. Enfin certains éléments de données du MLR n'ont pas d'utilité pour la communauté francophone. Ils ont donc été écartés de la norme. La figure 1 synthétise ce travail. Les classes reprises du MLR ou créées pour le NoDEfr sont reliées par des éléments de données (DES ou SED). Ces éléments de données issues du MLR ou créés *ex nihilo* sont utilisées pour décrire les classes.

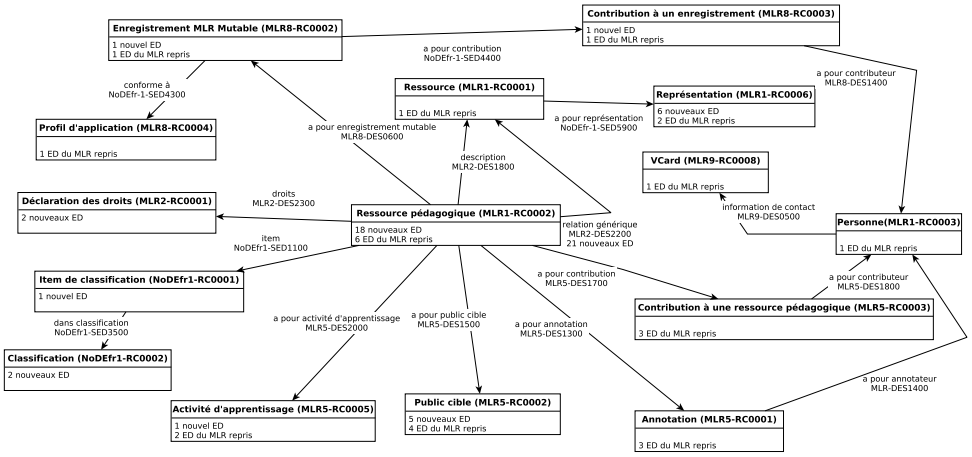


Fig. 1. Classes, nombre d'éléments de données (ED) et relations du NoDEfr-1

3.2 NoDEfr-1 et le Web des données

Représentation en OWL2 du NoDEfr-1 Le MLR, dans sa partie 7, propose une représentation du MLR en OWL2. Nous nous sommes donc fondés sur cette implantation pour définir le schéma OWL2 du NoDE-fr. Nous avons ajouté les classes et propriétés spécifiques au NoDEfr-1. Nous avons aussi redéfini certaines classes du schéma initial de façon à prendre en compte les cardinalités définies par le NoDEfr-1. Nous avons de plus privilégié l'utilisation des `owl:ObjectProperty` plutôt que les `owl:DatatypeProperty` ayant un intervalle de type `xsd:anyURI`. Enfin nous avons utilisé le schéma « Collection » pour représenter des ensembles ou des listes ordonnées [8]. Ce schéma est disponible à l'adresse suivante : `https://sourcesup.renater.fr/nodefr`

Transformation automatique Afin de favoriser l'acceptation du NoDEfr-1, il est indispensable que toutes les descriptions réalisées, et elles sont nombreuses, utilisant le LOMFR ou l'un de ses profils d'application (ScoLOMFR ou SupLOMFR) puissent être automatiquement transformées en descriptions conformes au NoDEfr-1 et à son expression définie ci-dessus. Les descriptions existantes étant conformes à un schéma XML, il s'agit donc ici de peupler une ontologie OWL, à partir de documents XML. Ce problème n'est pas nouveau et il a donné lieu à différentes approches le plus souvent basées sur XSLT dont certaines sont référencées par le W3C (cf. `https://www.w3.org/wiki/ConverterToRdf`).

La solution que nous avons développée a été réalisée en SWI-Prolog. Cette transformation a été testée sur plus de 50 000 fichiers contenant des descriptions reposant soit sur le ScoLOMFR soit sur le SupLOMFR.

4 Conclusion et travaux futurs

Après avoir présenté le MLR, nous avons décrit nos trois contributions principales : le profil d'application français du MLR : NoDEfr-1, sa représentation pour le Web des données et un outil de traduction des données XML du LOMFR vers des triplets RDF. Il faut maintenant que tout ceci soit adopté par la communauté. Outre un travail de communication, de formation, d'accompagnement et de maintenance , deux axes de travail sont envisagés. Le premier concerne la liaison des données. En effet il faut que les données publiées soient les plus formelles possible et surtout liées. Par exemple les personnes ou les établissements ne doivent plus être désignés et décrits à l'aide de chaînes de caractères mais bien identifiés par des IRI. Ce travail de désambigüisation est un problème difficile mais indispensable. Le second concerne les deux autres parties de la norme NoDEfr. Tout d'abord un groupe d'experts de l'AFNOR travaille actuellement sur le représentation des offres de formation. L'objectif est de proposer une norme en 2019. Ensuite, il faudra créer un profil d'application de la future norme ISO-IEC 22602 *Competency models expressed in MLR* qui permet de représenter les compétences. Une fois ces trois parties normalisées, les acteurs de la communauté éducative pourront publier des données ouvertes structurées qui permettront à quiconque de les réutiliser et de proposer des applications innovantes.

Références

1. Mohan, P., Brooks, C. : Learning objects on the semantic web. In IEEE International Conference on Advanced Learning Technologies (ICALT), 195–199 (2003)
2. Heath, T., Bizer, C. : Linked Data : Evolving the Web into a Global Data Space. 1st edition. Synthesis Lectures on the Semantic Web : Theory and Technology. Morgan & Claypool (2011)
3. Waitelonis, J., Sack, H. : Towards exploratory video search using linked data. Multimedia Tools and Applications, v59, N.2, 645–672 (2012)
4. Isaac, Y., Bourda, Y., Grandbastien, M. : SemUNIT - French UNT and Linked Data. In Proceedings of the 2nd International Workshop on Learning and Education with the Web of Data, LiLe@WWW 2012, Lyon, France (2012)
5. Delestre, N., Malandain, N., Ouchenne, B. : Une ontologie OWL pour le CDM-fr.In : Conference des Technologies de l'Information et de la Communication pour l'Enseignement. Beziers, France (2014)
6. Bourda, Y., Gauthier, G., Gomez de Regil, R-M., Catteau, O. : Métadonnées pour ressources d'apprentissage (MLR) - Nouvelle norme ISO de description de ressources pédagogiques. Rubrique de la revue STICEF, Volume 17, ISSN : 1764-7223, mis en ligne le 10/05/2011, `http://sticef.org` (2010)
7. IEEE, 1484.12.1-2002 - IEEE Standard for Learning Object Metadata, `https://ieeexplore.ieee.org/document/1032843`, (2002)
8. Ciccarese, P., Peroni, S. : Collections Ontology : Creating and handling collections in OWL 2 DL frameworks, Journal : Semantic Web, vol. 5, no. 6, 515–529, (2014)
9. Wielemaker, J., Schrijvers, T., Triska, M. and Lager, T. : SWI-Prolog. Theory and Practice of Logic Programming **12**(1-2), 67–96 (2012)