

Le local dans le mondial : le contrôle bibliographique universel du bas vers le haut

French translation of the original paper: “The local in the global: universal bibliographic control from the bottom up”.

Translated by: Françoise Bourdon, Bibliothèque nationale de France, Département de l'information bibliographique et numérique, Paris, France.

Le texte de ce document a été traduit en français et peut contenir des différences par rapport au texte original. Cette traduction est uniquement fournie à titre indicatif.

Gordon Dunsire

Independent Consultant, Edinburgh, Scotland.

E-mail address: gordon@gordondunsire.com

Mirna Willer

Department of Information Sciences, University of Zadar, Zadar, Croatia.

E-mail address: mwiller@unizd.hr



This is a French translation of “*The local in the global: universal bibliographic control from the bottom up*” copyright © 2014 by Françoise Bourdon. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

Résumé

Cette communication traite de la mise en œuvre du Contrôle bibliographique universel (CBU) dans l'environnement du Web sémantique et des données liées. Les tentatives visant à instituer le CBU en tant que système mondial de contrôle et d'échange de l'information bibliographique grâce à des méthodologies top-down (du haut vers le bas) n'ont que partiellement abouti à l'échelle mondiale. Parmi ces tentatives il y a eu le recours à des normes monolithiques pour des domaines spécifiques du contrôle bibliographique comme l'encodage des métadonnées en UNIMARC, la création de contenu avec l'ISBD, et l'identification des vedettes d'autorité avec GARR et FRAD. L'abandon de la dernière de ces tentatives a été en partie influencée par les technologies naissantes du Web sémantique, à savoir lier et réutiliser des données de sources hétérogènes et permettant ainsi à des façons de faire locales de se fondre dans des communautés de pratiques toujours plus vastes au niveau mondial. Cette communication décrit quelques-unes de ces nouvelles techniques qui permettent aux métadonnées locales d'être partagées avec des applications non locales, et aux métadonnées mondiales d'être utilisées dans des applications locales. La communication traite des questions soulevées par ces méthodes appliquées au contrôle bibliographique. Dans un environnement de données liées, la principale différence méthodologique vient de ce qu'on n'affine plus un ensemble mondialement défini d'éléments de description et de relations en vue de leur application locale, mais qu'on agrège des

éléments locaux préexistants pour que de la pratique locale émergent des points communs au niveau mondial. Cette communication s'appuie sur des exemples réels tirés de normes IFLA et non IFLA telles que RDA, Dublin Core, BIBFRAME et schema.org. Cette communication conclut que la conception du CBU développée durant des années par l'IFLA reste valable et n'est pas vouée à disparaître, mais qu'un changement radical de point de vue doit s'opérer si l'on veut à l'avenir capitaliser sur les investissements passés : ne plus penser à partir du monde mais du local.

Keywords: CBU, données liées, Web sémantique, localisation, interopérabilité

1. Introduction et contexte

Le CBU : contrôle bibliographique universel, « expression suprême de la maturité nouvellement découverte de l'IFLA », comme l'a écrit Herman Liebaers, alors Président de l'IFLA, dans l'introduction du document IFLA « *Universal Bibliographic Control: A long term policy – A plan for action* » publié en 1974, a permis aux services bibliographiques que nous avons et que nous utilisons aujourd'hui d'exister. Le document insiste sur la responsabilité des agences bibliographiques nationales dans la création de notices bibliographiques de référence pour les publications faites dans leurs pays respectifs et dans leur mise à disposition auprès des autres agences bibliographiques. Le processus repose uniquement sur le respect des normes internationales pour la création à la fois des notices bibliographiques et des notices d'autorité (Anderson, 1974).

Pour réaliser ce programme, l'IFLA a, depuis 1961, endossé la responsabilité de développer et de promouvoir des normes internationales, des principes directeurs, des règles et des terminologies et, depuis la fin des années 1990, des modèles conceptuels. Le principal concept du CBU était l'idée d'« uniformité » : le choix d'une vedette uniforme pour le nom ou le titre devait être basé sur « le nom le plus fréquemment utilisé ou le titre apparaissant dans les éditions des ouvrages catalogués » pour faire en sorte que la même forme du nom soit bien présente dans tous les catalogues et que ce soit à la fois économique et performant pour les bibliothèques qui échangent les notices, et pour les utilisateurs qui consultent différents catalogues (Paris Principes, 1961). Par exemple, l'ISBD, description bibliographique internationale normalisée, a été conçue dans le même esprit (IMCE, 1969).

Dorothy Anderson, première responsable du programme CBU et auteur du document cité ci-dessus, a exprimé très clairement le désaccord de certains bibliothécaires considérant le concept d'uniformité comme « impossible ». Elle a répondu que la vision du système CBU « est l'uniformité basée sur des normes internationales, ce qui ne se traduit pas toujours par des règles rigides, mais plutôt par des principes internationaux reconnus sur des points particuliers, ce qui autorise des ajustements nationaux » (Anderson, 1974, 29-30).

Il en résulte pour le CBU une structure qui combine des approches mondiales et locales s'efforçant d'équilibrer la dérivation auto-descriptive des vedettes uniformes avec le besoin de les distinguer à différents niveaux d'agrégation de collections aussi bien qu'à l'intérieur d'un périmètre d'application. S'il est évident que l'on a surtout besoin de différenciation au plus haut niveau, à savoir celui de la collecte de toutes les éditions patrimoniales, culturelles et bibliographiques du monde, de la même façon, à des niveaux plus locaux, les méthodes de construction des vedettes distinctives sont sujettes à variation. Ceci a mis en évidence le besoin de distinguer le cadre et les méthodologies qu'on applique dans ce cadre. La différenciation au niveau mondial requiert une infrastructure d'identité mondiale ; l'auto-description est un processus de transcription appliqué au niveau le plus local, celui de l'item

(article) que l'on a en main ; la collocation reposant sur les vedettes se situe quelque part au milieu.

2. Questions soulevées par une approche top-down (descendante)

La gestion des identités à l'échelle mondiale appelle un dispositif d'identifiants à l'échelle mondiale. Ce besoin est couvert par les normes ISO pour les systèmes de numérotation tels que l'ISBN, l'ISSN, l'ISRN, l'ISRC, et plus récemment l'ISNI : Code international normalisé des noms. De tels identifiants font partie intégrante de la description bibliographique puisqu'ils fournissent un lien nécessaire entre une description particulière (locale) et un système de numérotation mondial qui peut fournir une identification unique.

Cependant, les conventions de construction sous forme de points d'accès autorisés qui, avec les variantes de formes constituent des points d'accès contrôlés, exigent un effort intellectuel pour les apparier à la même entité. C'est-à-dire que le concept de point d'accès autorisé, qui a remplacé celui de vedette uniforme, donne la préférence à l'« utilisateur national » et à ses besoins, et par conséquent la forme du point d'accès préféré est devenue dépendante d'une agence bibliographique spécifique utilisant des règles de catalogage spécifiques. Il s'agit moins de « respecter la culture de l'autre » que d'énoncer « ceci est ma vision culturelle de l'autre, telle du moins que je la comprends ». Un tel changement de point de vue dans le concept du CBU s'explique par le fait qu'utiliser mondialement la même forme d'une vedette pour une entité « n'est pas pratique [...] et n'est plus nécessaire, [puisqu'] avec des capacités informatiques qui offrent plus de sophistication nous pouvons lier les notices d'autorité créées dans un pays selon un ensemble de règles de catalogage avec celles d'un autre pays pour faciliter le partage des notices d'autorité et potentiellement laisser la commutation assistée par ordinateur afficher les formes autorisées » (GARR, 2001, ix).

Le premier fichier d'autorité collectif a été le Thésaurus du CERL, fichier d'autorité de la base de données Hand Press Books (aujourd'hui The Heritage of the Printed Book in Europe, c. 1455-c 1830) développé à la fin des années 1990 par le Consortium of European Research Libraries (Consortium des bibliothèques de recherche européennes). Le Thésaurus du CERL « fusionne les formes normalisées et d'autorité des noms qui sont utilisées dans les catalogues de bibliothèques d'aujourd'hui sans que soit imposée la forme d'autorité du nom propre au CERL. Au lieu de cela, les formes normalisées et les variantes de forme sont classées alphabétiquement avec une indication claire des institutions qui utilisent une forme normalisée particulière » (CERL Thesaurus). Les différences de pratiques de catalogage dans des catalogues sur fiches rétro-convertis ou encore dans des inventaires excluent toute possibilité d'atteindre l'objectif d'une vedette/d'un point d'accès uniforme pour les entités décrites notamment les lieux, les imprimeurs, les auteurs, etc. Par exemple, « Lyon » en tant que lieu d'impression est référencé sous 41 formes de nom différentes (Permalien <http://thesaurus.cerl.org/record/cnl00011383>). Il faut cependant noter que pour apparier les noms, il a fallu un effort intellectuel, et pas simplement des algorithmes.

Ainsi, la compatibilité rétroactive qui permet de réduire les coûts de conversion des données et de développement de système, a motivé la décision de concevoir un thésaurus, plutôt qu'un fichier d'autorité reposant sur le concept de vedette uniforme. Un tel point de vue peut facilement l'emporter aujourd'hui, au vu des changements majeurs qui s'annoncent avec l'adoption et le développement de nouvelles règles de catalogage : la notion de « vedette uniforme » contraint les contenus du futur à s'adapter aux contenus hérités du passé.

VIAF, Virtual International Authority File (Fichier d'autorité international virtuel) est un fichier d'autorité collectif comparable, bien qu'offrant certaines fonctionnalités différentes (VIAF, 2014). L'intention n'est pas tout à fait la même : il s'agit de lier les différentes formes de points d'accès autorisés définis pour une même entité par différentes agences bibliographiques nationales. Les variantes de forme des noms ne sont pas prises en compte. La même requête pour Lyon (Permalien <http://viaf.org/viaf/158189703>) montre seulement le nom de lieu tel que l'utilisent huit bibliothèques nationales.

Il n'y a cependant aucune méthode clairement définie pour gérer un changement ou une rétro-conversion au sein d'une seule et même base de données bibliographiques ou d'un catalogue collectif. Par exemple, l'un des principaux problèmes imputables à l'implémentation de l'édition intégrée de l'ISBD est la nouvelle zone 0 *Zone de la forme du contenu et du type de médiation* introduite pour remplacer l'élément GMD (General Material Designation) *Indication générale du type de ressource* des ISBD spécialisés. La question se pose de que faire de l'ancien élément 1.2 qui apparaissait systématiquement dans les recommandations de l'ISBD et les notices descriptives, une fois que le concept de GMD est déclaré obsolète et que la même désignation numérique est utilisée pour l'élément *Titre parallèle* de l'ISBD dans l'édition intégrée. Le format bibliographique UNIMARC n'a pas encore déclaré obsolète cet élément 200 \$b *Indication générale du type de ressource* (UNIMARC Bibliographic Format, Update 2012) et on peut s'attendre à ce qu'il soit présent dans les notices héritées du passé. Comment devrions-nous traiter quelque chose qui a un temps été considéré important pour les utilisateurs d'un catalogue, puis a été déclaré obsolète dans la norme plus récente ? Est-ce que cela veut dire qu'un élément de la norme peut perdre son contexte « international/mondial » et devenir un élément local encore suffisamment important pour être conservé ?

3. Le local et le mondial

Le précepte « Pensez mondial, agissez local » (« Think global, act local ») repose sur le travail de Patrick Geddes dans le domaine de l'urbanisme : « chaque plan valable devrait exprimer la complète utilisation des conditions locales ou régionales dans lesquelles il s'insère et être l'expression d'une identité locale et régionale ». « Le caractère local » ... n'est atteint que si l'on saisit et traite de façon adéquate l'environnement tout entier, et qu'on exprime une sympathie active avec la vie essentielle et caractéristique de l'endroit concerné » (Geddes, 1915). Ces dernières années il a été surtout employé dans le contexte de l'environnement durable, mais on peut avec profit l'ériger en mot d'ordre des données liées du patrimoine culturel.

Dans le contexte du patrimoine bibliographique et culturel, il est nécessaire de structurer localement des données afin d'adapter un contenu local à un public local dans une culture locale. En même temps, chaque utilisateur peut avoir des objectifs qui ne peuvent être réalisés que par du contenu non local traitant de sujets étrangers ou situés au-delà du périmètre local. C'est une situation universelle que connaît chaque usager : le besoin d'une vue locale sur tout, comme une collection de bibliothèque nationale qui traite tous les sujets présentant un intérêt pour les habitants de son pays, et le besoin d'une vue mondiale sur le local, comme une collection de bibliothèque nationale qui traite les sujets d'intérêt national selon la culture, la langue, etc., où qu'ils soient trouvés dans le monde. C'est-à-dire qu'une bibliothèque nationale possède un exemplaire de chaque livre publié dans le pays, et un exemplaire de plusieurs livres sur le pays publiés à l'étranger. L'information locale est fixée sur support dans le monde entier, et l'information mondiale est localement fixée sur support.

Le « local » et le « mondial » sont entièrement à considérer l'un par rapport à l'autre, avec le « régional » au milieu. On a donc besoin d'une approche à granularité fine (plus granulaire encore que la notice bibliographique), qui puisse rendre compte des variations locales des schémas de métadonnées tout en restant cohérente au niveau mondial. L'unité de niveau notice du modèle actuel est trop large et rigide : des applications locales peuvent ne pas exiger tous les éléments de données ; des sources locales peuvent ne pas fournir de contenu pour des éléments « obligatoires ». Le problème a été bien exposé pour des contextes comparables, par exemple dans le cadre du développement de la National Science Digital Library (Hillmann, Dushay and Phipps, 2004) et d'Europeana (EDM Primer, 2013). Dans ces deux cas la solution retenue a été de transférer le niveau de granularité des métadonnées de la notice aux mentions particulières contenues dans la notice sous la forme de paires nom/étiquette-valeur pour les attributs ou les champs.

4. Données liées et RDF

RDF (Resource Description Framework) est le format des données du Web sémantique mondial et de l'environnement des données liées. Il représente les métadonnées sous la forme de déclarations simples relatives à une chose identifiée. Chaque déclaration se décline en trois parties et est appelée un triplet. On donne d'abord le sujet de la déclaration, suivi de l'aspect identifié de la chose qu'on décrit, appelé prédicat ou propriété, suivi de la valeur ou de l'identifiant de l'aspect de la chose, appelé objet : un triplet est une déclaration sujet – propriété – objet. En RDF les identifiants sont des URI (Uniform Resource Identifiers, Identifiants uniformes de ressources) qui sont uniques au niveau mondial, et donc local. Voici un exemple de triplet **ex:Resource1 isbd:P1016 "Zagreb"**. On utilise le langage « terse triple language » (ttl, syntaxe compacte pour exprimer un triplet), format où les URI sont données sous une forme compacte et où les valeurs littérales sont mises entre guillemets. L'URI compacte *isbd:P1016* se développe en URI complète <http://iflstandards.info/ns/isbd/elements/P1016> qui a pour étiquette « a pour lieu de publication, de production, de distribution » (ISBD Elements, 2014). L'abréviation « ex: » (pour « exemple ») remplace tout nom de domaine local pour les URIs des ressources. Donc le triplet donné en exemple déclare :

"La chose identifiée par l'URI compacte **ex:Resource1** a pour lieu de publication, de production, de distribution **"Zagreb"**.

La structure d'un triplet RDF implique que seule la propriété détermine la granularité de la déclaration ; la granularité de tout sujet est fixée par définition et l'objet reflète la granularité de la propriété. Par exemple une déclaration à granularité plus fine pour le même sujet est :

" La chose identifiée par l'URI compacte **ex:Resource1** a pour lieu de publication **"Zagreb"**

On peut représenter cela par le triplet **ex:Resource1 rdam:P30088 "Zagreb"** grâce à la propriété de la Manifestation de RDA ayant pour étiquette « a pour lieu de publication » (RDA Registry, 2014)

Un exemple de triplet à granularité plus grossière est **ex:Resource1 bf:provider "Zagreb : Hrvatsko knjižničarsko društvo, 2012"** grâce à la propriété de BIBFRAME ayant pour étiquette « Information sur le lieu, le nom et/ou la date relatifs à la publication, l'impression, la distribution, la sortie, la parution, ou la production de l'instance » (BIBFRAME 2014).

<i>Triplet</i>	<i>Etiquette de la propriété en anglais / en français</i>	<i>Granularité</i>
ex:Resource1 rdam:P30088 "Zagreb"	RDA: has place of publication / RDA : a pour lieu de publication	Fine
ex:Resource1 isbd:P1016 "Zagreb"	ISBD: has place of publication, production, distribution / ISBD : a pour lieu de publication, production, distribution	Moyenne
ex:Resource1 bf:provider "Zagreb : Hrvatsko knjižničarsko društvo, 2012"	BIBFRAME: Place, name, and/or date information relating to the publication, printing, distribution, issue, release, or production instance / BIBFRAME : Information sur le lieu, le nom et/ou la date relatifs à la publication, l'impression, la distribution, la sortie, la parution, ou la production de l'instance	Grossière

Tableau 1: Les différents niveaux de granularité des propriétés des triplets.

Notons que, bien que le Tableau 1 classe BIBFRAME, ISBD et RDA selon leur granularité relative plus ou moins fine pour rendre compte d'un lieu de publication, dans l'absolu chaque nom de domaine a des propriétés plus fines ou plus grossières qui lui sont propres pour rendre compte d'un lieu de publication ; par exemple RDA a la propriété plus grossière « a pour lieu de publication », ISBD a la propriété plus grossière « a pour zone de la publication, production, distribution, etc. », et BIBFRAME a la propriété plus fine « Lieu relatif à la publication, l'impression, la distribution, la sortie, la parution, ou la production de l'instance ».

Le Tableau 1 illustre aussi les deux types de granularité qui apparaissent dans la description bibliographique : sémantique, et syntaxique. La séquence « lieu de publication » < « ... lieu de publication, de production, ... » est un exemple de granularité sémantique de la plus fine à la plus grossière, du nom de lieu seul à une combinaison de plusieurs types de lieu. La séquence « ... lieu de publication ... » < « information sur le lieu, le nom ... relatifs à la publication ... instance » est un exemple de granularité syntaxique de la plus fine à la plus grossière, du nom de lieu seul au nom de lieu assorti du nom de l'éditeur. La propriété la plus grossière a aussi la signification la plus large mais la relation sémantique observe une structure combinatoire plutôt que hiérarchique. Dans RDA la catégorisation des éléments en élément, sous-type d'élément, et sous-élément, reflète ces deux types de granularité (JSC, 2009).

Le partage et la réutilisation des données définis par le CBU requièrent des méthodes automatiques qui déplacent les données d'un niveau de granularité à un autre. On peut voir ces méthodes comme un ensemble de « boîtes noires » prenant des triplets en entrée pour produire d'autres triplets en sortie, grâce à un logiciel interne qui traite les données d'une façon cohérente. Détail qui a son importance, la boîte produit de nouveaux triplets de données sans remplacer les triplets utilisés en entrée, ceux-ci restant disponibles pour alimenter d'autres processus.

Six exemples de boîtes sont proposés ici : Traducteur de schéma, Traducteur de terme, Créateur de déclaration, Fractionneur de déclaration, Créateur de notice et Fractionneur de notice.



Figure A: La boîte noire Traducteur de schéma

Sur la Figure A, la boîte noire « Traducteur de schéma » « traduit » des données créées avec le jeu d'éléments d'un nom de domaine dans un autre nom de domaine. Une déclaration de métadonnées créée dans un schéma bibliographique donné est traduite en déclaration compatible avec un schéma apparenté. Dans l'exemple, un triplet utilisant la propriété P30088 (a pour lieu de publication) de la manifestation RDA est utilisé en entrée, une option de sortie en ISBD est sélectionnée, et on obtient en sortie la propriété P1016 (a pour lieu de publication, de production, de distribution) de l'ISBD. L'exemple repose sur les alignements entre les éléments RDA et ISBD, représentés comme un graphe d'alignements RDF utilisant des relations de sous-propriété du schéma RDF (Dunsire and ISBD Review Group, 2012). La relation porte en elle une règle actionnable par machine qui crée un clone du triplet d'entrée avec un prédicat venant du jeu d'éléments utilisé en sortie.

Il faut noter que cette boîte noire ne fonctionne que dans un sens. Elle simplifie à l'excès les données pour garantir la cohérence sémantique de la déclaration de sortie dans le schéma cible. Mais on ne peut pas retricoter une déclaration à partir d'une déclaration plus spécifique. Par exemple, si un Traducteur de schéma utilise une ontologie recommandée pour les éléments « titre » de l'ISBD, de RDA et du DC (Dunsire, Hillmann, and Phipps, 2012, diagramme 6), il traduira la propriété P1004 (a pour titre propre) de l'ISBD dans la propriété titre du Dublin Core (DCMI Usage Board, 2012) :

Schema Translator (ex:Resource1 isbd:P1004 "Poem") = ex:Resource1 dc:title "Poem".

Si on se réfère à l'ontologie, on obtiendrait le même résultat à partir de données d'entrée utilisant des propriétés de la Manifestation de RDA comme P30156 (a pour titre propre), de sorte que le processus inverse devrait deviner que donner en sortie. Utiliser la propriété RDA implique que l'URI sujet identifie une *Manifestation*, alors que la propriété ISBD implique une *Ressource* ; elles ne sont pas identiques (Dunsire, 2013). Si ce qui est obtenu en sortie du Traducteur de schéma ci-dessus est remis en entrée, on ne peut obtenir en sortie qu'un schéma plus grossier comme schema.org (schema.org, 2014).

Schema Translator (ex:Resource1 dc:title "Poem") = ex:Resource1 schema:name "Poem".



Figure B: La boîte noire Traducteur de terme

La boîte noire « Traducteur de terme » de la Figure B « traduit » des valeurs de données d'un système d'organisation des connaissances (Knowledge Organization System, KOS) vers un

autre. Une déclaration de métadonnées contenant une valeur issue d'un vocabulaire contrôlé est traduite dans une déclaration avec une valeur issue d'un autre vocabulaire. Dans l'exemple, un triplet utilisant la propriété U110_1a du format UNIMARC bibliographique (Périodicité de parution dans les Données codées d'une ressource continue dans le Champ Données codées : Ressources continues) avec l'URI objet correspondant au code « h » dans le vocabulaire contrôlé *unimarc* (Ressources continues : Périodicité de parution) apparaît en entrée, et une option en RDA est sélectionnée pour la sortie : l'URI objet pour une valeur semblable dans le vocabulaire *rdaf* (RDA Périodicité).

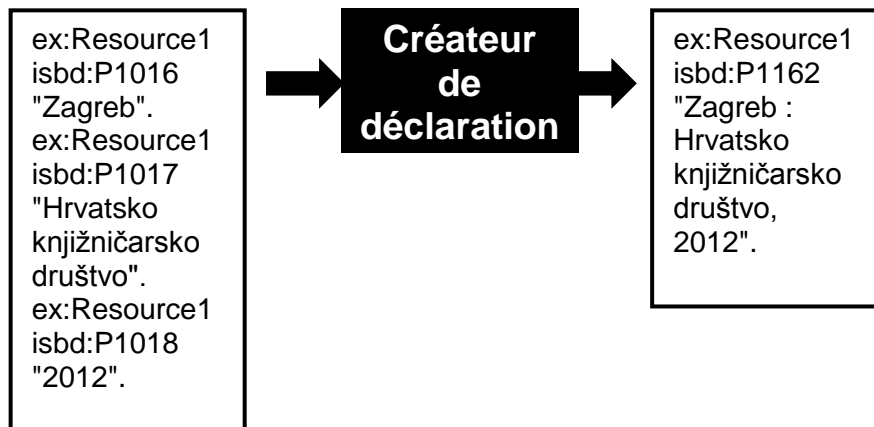


Figure C: La boîte noire Créateur de déclaration

Sur la figure C, la boîte noire « Créateur de déclaration » concatène les valeurs d'un jeu de déclarations de métadonnées à granularité fine dans une déclaration agrégée de granularité plus grossière en ajoutant des délimiteurs comme la ponctuation ou l'encodage pour la sortie ou l'affichage. Dans l'exemple, des triplets utilisant les propriétés de l'ISBD isbd:P1016 (a pour lieu de publication, de production, de distribution), isbd:P1017 (a pour nom d'éditeur, de producteur, de distributeur), et isbd:P1018 (a pour date de publication, de production, de distribution), apparaissent en entrée, et la boîte noire donne en sortie la déclaration agrégée isbd:P1162 (a pour zone de la publication, production, distribution, etc.) en utilisant les règles de ponctuation de l'ISBD (ISBD, 2011)

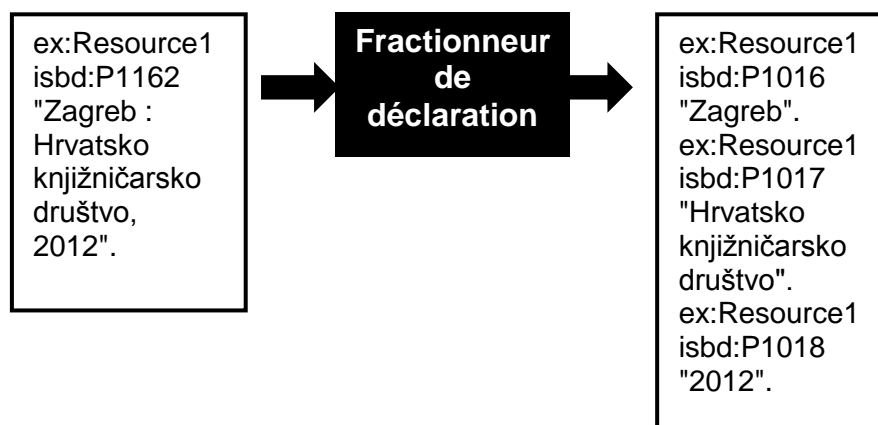


Figure D: La boîte noire Fractionneur de déclaration

La boîte noire « Fractionneur de déclaration » de la Figure D a la fonction inverse de la boîte noire « Créateur de déclaration » de la Figure C. Elle décompose la valeur d'une déclaration de métadonnées agrégée en un jeu de déclarations à granularité plus fine en reconnaissant ses délimiteurs avant de les ôter. Dans l'exemple, un triplet utilisant la propriété de déclaration agrégée isbd:P1162 (a pour zone de la publication, production, distribution, etc.) apparaît en entrée, et la boîte noire donne en sortie les triplets qui la composent en utilisant les propriétés ISBD isbd:P1016 (a pour lieu de publication, de production, de distribution), isbd:P1017 (a pour nom d'éditeur, de producteur, de distributeur) et isbd:P1018 (a pour date de publication, de production, de distribution) après avoir ôté la ponctuation ISBD.



Figure E: La boîte noire Créateur de notice

La boîte noire « Créateur de notice » de la Figure E élargit la fonctionnalité du « Créateur de déclaration » de la Figure C. Il utilise un profil d'application pour préciser les composants et la structure d'une notice de métadonnées selon un schéma donné. Un profil d'application définit des notices de métadonnées qui répondent aux besoins spécifiques d'une application en agençant ce qui est requis des données de ce qui peut en être attendu (Coyle and Baker, 2009). Dans l'exemple, un jeu de triplets utilisant des propriétés de RDA, y compris des déclarations agrégées, apparaît en entrée et la « notice » RDA est donnée en sortie sous la forme d'un jeu de triplets correspondant à un profil d'application qui précise que seules les propriétés « essentielles » doivent être utilisées, ainsi que leur ordre de présentation, et le statut obligatoire et répétable de chacune.



Figure F: La boîte noire Fractionneur de notice

La boîte noire « Fractionneur de notice » de la Figure F a la fonction inverse de celle de la boîte noire « Créateur de notice » de la Figure E. Elle décompose une notice de métadonnées en autant d'éléments qu'elle en a grâce à un profil d'application pour le schéma de notice. Dans l'exemple, une notice RDA encodée dans une syntaxe particulière apparaît en entrée et la boîte noire produit en sortie un jeu de triplets utilisant des propriétés et des valeurs du nom de domaine RDA.

On peut enchaîner différentes sortes de boîtes noires pour prendre en compte des fonctions d'interopérabilité plus complexes. Par exemple, des notices de données liées pour un service de découverte particulier peuvent être produites par le regroupement de déclarations issues de sources diverses, avec des éléments et des vocabulaires différents, que l'on verse ensuite dans

un processus en chaîne Traducteur de schéma, Traducteur de terme, Créateur de déclaration, et Créateur de notice.

Les processus internes requis par chaque type de boîte noire sont bien compris des communautés du Web sémantique, et aucun obstacle technique insurmontable ne s'oppose au développement de tels agents logiciels traitant des données liées. Ils pourraient devenir des composants implicitement embarqués dans des navigateurs du Web sémantique, comme le sont pour les navigateurs standards du web de documents les pluggins automatisés de traducteur de langue et de convertisseur de fuseau horaire. Toutefois, pour faire fonctionner les boîtes leurs développeurs ont besoin des représentations RDF des normes bibliographiques, ainsi que des graphes d'alignements établissant des relations entre elles et les profils d'application pour les jeux de déclarations de métadonnées qui se présentent comme des notices.

5. Comment penser mondialement

Il faudrait partir du principe que tout est connecté à tout au niveau mondial, sinon aujourd'hui du moins dans le futur. Les liens entre les choses seront dupliqués, soit directement soit via de multiples chemins à travers le Web sémantique. Par exemple une ressource bibliographique pourrait être liée à son lieu de publication, directement par un attribut *lieu de publication* ou indirectement par une *déclaration de publication*, ou par le *nom de l'éditeur* qui lui-même est lié au *lieu de l'éditeur*. Il suffit donc de se concentrer sur une seule chose ou sur un seul type de chose à la fois. Il est plus efficace de créer des métadonnées descriptives pour une ressource, puis de la lier à un lieu, ou de créer des métadonnées pour le lieu puis de le lier à la ressource, plutôt que de décrire à la fois la ressource et le lieu dans une seule opération locale. D'autres communautés appliqueront la même méthode ; par exemple, des services d'information gouvernementaux créeront aussi des métadonnées de lieux dans le cadre des données de recensement.

Certaines caractéristiques locales d'une entité peuvent être traitées comme des raffinements conceptuels d'une assertion mondiale. Par exemple, l'attribut spécifique *titre propre* utilisé dans une description bibliographique raffine l'attribut plus général *titre* utilisé dans le secteur du patrimoine culturel, tel que le titre d'un tableau. L'attribut *titre* lui-même raffine le concept général d'*étiquette* ; toute chose peut être étiquetée pour fournir un identifiant textuel. L'attribut *nom* d'une personne, d'une famille, ou d'une collectivité raffine également *étiquette*, et est raffiné à son tour par *pseudonyme*, une caractéristique de l'entité bibliographique persona.

Des caractéristiques particulières propres aux entités bibliographiques peuvent être ajoutées au jeu mondial d'attributs si nécessaire, quand il n'existe pas d'attribut général à raffiner. Le cas se présente rarement car les caractéristiques propres à une communauté locale ont tendance à être adaptées en vue d'un usage plus général si elles comblent une lacune au niveau mondial. Par exemple, l'attribut *édition* s'appliquait au départ aux monographies imprimées, mais est à présent utilisé pour des « éditions spéciales » de films ou même de barres chocolatées.

Il est important de séparer les choses, à savoir éviter de rassembler des identités différentes au point de vue mondial. Par exemple, même si le contenu est le même, une ressource et son substitut numérique gagneront à être traités comme des choses distinctes plutôt que de se voir attribuer la même identité (IME ICC, 2009). Une photographie numérisée n'est pas une

photographie numérique, et la différence est importante dans tout schéma qui dispose d'entités distinctes pour décrire des caractéristiques de contenu et de support, comme les entités *Work* et *Instance* de BIBFRAME.

6. Comment agir localement

Le Web sémantique repose sur le World-Wide Web, qui repose lui-même sur l'Internet. L'environnement de données liées est par conséquent disponible à quiconque ayant un accès Internet. Il y a peu d'obstacles techniques à la publication en RDF de schémas de métadonnées et de vocabulaires locaux, permettant la publication de notices bibliographiques locales sous forme de triplets de données sans perte d'information. On peut par exemple générer des triplets à partir de simples éditeurs de texte. L'infrastructure comprend des outils en open access pour publier et gérer des ensembles d'éléments et des dictionnaires de valeurs, tel que l'Open Metadata Registry (Répertoire des données ouvertes) utilisé pour les noms de domaine IFLA et RDA (Open Metadata Registry, 2014). Il existe par ailleurs une vaste littérature sur le sujet, des normes techniques aux sites de réseaux sociaux.

Si l'utilisation locale d'un schéma mondial est totalement conforme, on peut utiliser les éléments du nom de domaine du schéma, s'ils sont disponibles, pour publier des triplets de jeux de données sans simplifier les données source. Réutiliser des éléments mondiaux diminue la prolifération inutile de jeux de données et de dictionnaires de valeurs à la sémantique très proche et des Traducteurs de schéma, qui ne sont rien moins qu'un Apparieur d'identité (autre type de boîte noire nécessaire). Cependant, si la variante locale fait sens dans le contexte, alors la simplification peut être évitée si l'on présente la variante comme un raffinement du schéma mondial, ou si l'on présente le schéma local sous son propre nom de domaine. Préserver le statut local des données permet leur réutilisation sans censure préalable dans l'environnement mondial.

Les jeux d'éléments locaux et leurs variantes ont besoin d'être liés aux jeux d'éléments mondiaux par un graphe d'alignements en vue de leur réutilisation par un Traducteur de schéma d'un schéma local vers un schéma mondial. Les dictionnaires de valeurs locaux et mondiaux doivent aussi être alignés en vue de leur réutilisation par un Traducteur de terme approprié.

Si un jeu d'éléments local comprend des classes de types locaux d'entités bibliographiques, il est alors utile de publier des versions non contraintes des propriétés, sans lien à une entité quelconque, pour permettre des graphes d'alignements vers le schéma mondial qui sont incompatibles avec les entités locales. Par exemple, le nom de domaine RDA comprend des propriétés qui ne sont pas contraintes par les entités RDA/FRBR (*WEMI*), pour permettre des alignements vers un schéma non-FRBR comme l'ISBD. De même, on publiera une version du jeu d'éléments de l'ISBD sans l'associer à l'entité *Ressource* pour permettre un alignement complet de l'ISBD sur RDA.

Quand une application de données liées locale utilise des jeux de données ouvertes mondiaux, des dictionnaires de valeurs, des jeux d'éléments, des alignements, des profils d'application, etc., il est généralement préférable d'établir les liens de façon dynamique ou d'utiliser un cache local fréquemment rafraîchi plutôt que de copier un instantané statique des données externes. Les bénéfices du partage des données sont plus grands si les données sont stockées dans un réservoir commun et ne sont pas coupées du web de données, le « cloud » (nuage)

pour parler couramment. Toute nouvelle donnée mondiale est immédiatement disponible, et toute duplication locale superflue peut être réduite au minimum

7. Conclusion

Cette communication se situe dans un futur proche où le Web sémantique sera devenu aussi banal et invisible que l'est à ce jour le web de documents pour nos smartphones et téléphones, nos tablettes et ordinateurs. Ce à quoi on accède c'est du contenu numérique, né numérique ou numérisé, et qui nécessite un contrôle bibliographique « intelligent » pour faire fonctionner les services de découverte de ressources. Les principes et l'infrastructure en sont définis et modernisés ; ce sont les méthodes qui doivent changer si la communauté des bibliothèques et du patrimoine culturel entend tirer profit de l'environnement des données liées, et y contribuer.

Un CBU intelligent devrait s'efforcer de soutenir tous ceux qui souhaitent penser mondialement et agir localement, dans un savant dosage entre méthodologies « bottom-up » (ascendante) et « top-down » (descendante).

C'est le CBU du bas vers le haut, et c'est le défi que nous adressons aux gestionnaires et aux décideurs :

- Aligner les identifiants locaux sur les identifiants mondiaux, par exemple VIAF ou ISNI ;
- Publier les données locales et les schémas locaux tels qu'ils sont, directement en RDF ;
- Gérer les alignements entre les éléments locaux et les éléments mondiaux ;
- Fournir l'infrastructure pour raffiner ou élargir les éléments mondiaux afin de s'adapter aux applications locales ;
- Fournir l'infrastructure pour reverser les éléments locaux courants dans les éléments mondiaux.

Les boîtes noires sont le défi que nous adressons aux développeurs. Les auteurs de cette communication et beaucoup de collègues des communautés IFLA et RDA ont contribué à la mise au point « imminente » d'un Traducteur de schéma ISBD vers RDA, nous pensons donc qu'il s'agit d'un défi honnête. Et l'expérience nous dit que ce type d'« exercice » permettra de faire la lumière sur les manques des normes bibliographiques et des standards du patrimoine culturel tels qu'ils sont aujourd'hui.

Que l'on ne voit là rien de surprenant : il en va de notre nature humaine. Après tout, du fait de notre individualité et de notre culture, nous portons tous un regard différent sur une seule et même chose, mais nous avons en partage le même cerveau d'*Homo sapiens*. La façon dont on décrit les choses doit par conséquent être un mélange d'observations intrinsèques et extrinsèques. Dans l'environnement des données liées, la sémantique du local s'exprime dans une syntaxe mondiale ; la sémantique du mondial est inhérente au local.

Références bibliographiques

(Anderson, 1974) Dorothy Anderson, *Universal Bibliographic Control: A long term policy – A plan for action*, (Pullach/München: Verlag Dokumentation, 1974).

(BIBFRAME, 2014) BIBFRAME vocabulary, <http://bibframe.org/vocab/>

(CERL Thesaurus) HPB, CERL Thesaurus,
http://www.cerl.org/resources/cerl_thesaurus/main

(Coyle and Baker, 2009) Karen Coyle and Thomas Baker, Guidelines for Dublin Core Application Profiles, 2009-05-18, <http://dublincore.org/documents/profile-guidelines/>

(DCMI Usage Board, 2012) DCMI Usage Board, DCMI Metadata Terms, 2012-06-14, <http://dublincore.org/documents/dcmi-terms/>

(Dunsire, 2013) Gordon Dunsire, Resource and Work, Expression, Manifestation, Item, Amended 6 October 2013,
<http://www.ifla.org/files/assets/cataloguing/isbd/OtherDocumentation/resource-wemi.pdf>

(Dunsire, Hillmann, and Phipps, 2012) Gordon Dunsire, Diane Hillmann, and Jon Phipps, Reconsidering Universal Bibliographic Control in Light of the Semantic Web, *Journal of Library Metadata*, v.12, issue 2-3 (2012),
<http://ecommons.library.cornell.edu/handle/1813/36288>

(Dunsire and ISBD Review Group, 2012) Gordon Dunsire and ISBD Review Group, Alignment of the ISBD: International Standard Bibliographic Description element set with RDA: Resource Description & Access element set, Version 1.1, 24 September 2012,
http://www.ifla.org/files/assets/cataloguing/isbd/OtherDocumentation/ISBD2RDA%20Alignment%20v1_1.pdf

(EDM Primer, 2013) Europeana Data Model Primer, 14/07/2013,
<http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5>

(GARR, 2001) Guidelines for authority records and references, Revised by the Working Group on GARE Revision, 2nd ed., (München: Saur, 2001).

(Geddes, 1915) Patrick Geddes, *Cities in Evolution*, (London: Williams, 1915),
<https://archive.org/details/citiesinevolutio00gedduoft>

(Hillmann, Dushay and Phipps, 2004) Diane Hillmann, Naomi Dushay, and Jon Phipps, Improving Metadata Quality: Augmentation and Recombination, DC-2004-Shanghai Proceedings, <http://dcpapers.dublincore.org/pubs/article/view/770/766>

(IMCE, 1969) International Meeting of Cataloguing Experts, Copenhagen, 1969: Report, IFLA Annual, 1969; also published in: *Report of the International Meeting of Cataloguing Experts*, Copenhagen, 1969, *Libri*, 1, 20, 1970.

(IME ICC, 2009) IME ICC: IFLA Meeting of Experts on an International Cataloguing Code, Frankfurt, 2003-Pretoria 2007. Statement of international cataloguing principles (2009)
http://www.ifla.org/files/cataloguing/icp/icp_2009-en.pdf

(ISBD, 2011) ISBD : International Standard Bibliographic Description, Consolidated Edition, (Berlin ; München : De Gruyter Saur, 2011), ISBN 978-3-11-026379-4,
http://www.ifla.org/files/assets/cataloguing/isbd/isbd-cons_20110321.pdf

En français :

http://www.bnf.fr/fr/professionnels/normes_catalogage/a.normes_isbd_presentation.html#SHDC_Attribute_BlocArticle2BnF

(ISBD Elements, 2014). ISBD elements, <http://iflstandards.info/ns/isbd/elements/>

(JSC, 2009) Joint Steering Committee for Development of RDA, RDA Scope and Structure, 1 July 2009, <http://www.rda-jsc.org/docs/5rda-scoperev4.pdf>

(Open Metadata Registry, 2014) Open Metadata Registry, <http://metadataregistry.org/>

(Paris Principles, 1961) International Conference on Cataloguing Principles, Paris, 9th-18th October, 1961, Report, London: International Federation of Library Associations, 1963. Reprinted in: *International Conference on Cataloguing Principles*, Paris, 9th-18th October, 1961: Report, (London: Clive Bingley, 1969).

(RDA Registry, 2014) RDA Registry, <http://www.rdaregistry.info>

(schema.org, 2014) schema.org, Organization of Schemas,

<http://schema.org/docs/schemas.html>