



Bibliothèque nationale de France

**NISO Bibliographic Roadmap
Meeting : 15-16 avril 2013,
Baltimore, USA : rapport de
mission**

**Bibliothèque nationale
de France**

direction des Services et des réseaux
département de l'Information bibliographique et
numérique

version 1 du 15 mai 2013
émetteur : Gildas ILLIEN
affaire suivie par : Françoise BOURDON
référence : BnF-ADM-2013-049367-01



Participants¹ :

Dans l'annonce de cette manifestation on lisait que les travaux étaient ouverts à 70 personnes maximum sur inscription. En fait il y a eu :

27 participants présents

+ 26 « followers » à distance (très discrets, en fait absents du débat !)

Parmi les présents :

11 fournisseurs de services (OCLC, Proquest, Ex Libris, Book Industry Study Group, Portico, etc.),

7 bibliothèques universitaires américaines (dont Harvard, Yale, UCA),

3 bibliothèques numériques (Europeana, California Digital Library, Université de Pennsylvanie),

2 bibliothèques nationales (Finlande et France),

2 DCMI.

2 consultants (Gordon Dunsire – UK : Chair IFLA Namespaces Technical Groups & Chair-Elect JSC/RDA ; Karen Coyle - USA),

Parmi les absents :

Library of Congress (LoC), pourtant invitée.

Organisateurs de la rencontre :

Le NISO (National Information Standards Organisation, USA), association à but non lucratif accréditée par ANSI (American National Standards Institute) a organisé cette rencontre internationale. Todd Carpenter, qui en est le directeur depuis 2006, est aussi le secrétaire du ISO/TC46/SC9 « Identification et description ».

NISO est un organisme de normalisation, donc il a vocation à rechercher le consensus entre des partenaires concurrents à qui il offre un terrain neutre pour la discussion.

Objectifs de la rencontre :

Après la publication en 2008 des résultats de son étude sur « The Future of Bibliographic Control »² et après ses tests d'utilisation de RDA, la LoC conclut que MARC n'est pas suffisant pour formaliser l'information bibliographique à l'ère du Web sémantique et des Données liées. Elle propose alors BIBFRAME, mais sans une large consultation des parties prenantes (et sans prise en compte des remarques exprimées).

L'objectif du NISO est d'ouvrir le champ des investigations en donnant la parole aux bibliothèques, aux fournisseurs de SID, au monde de la recherche et aux usagers non traditionnels de l'information bibliographique. L'idée est donc de réfléchir à un nouvel écosystème de l'échange de l'information bibliographique, et de discuter des projets pilotes en cours. Quels sont les besoins non satisfaits aujourd'hui ? Quelle pourrait être la feuille de route ?

La rencontre en face-à-face devrait se prolonger par des webinaires.

Après un rappel des objectifs de la rencontre par Todd Carpenter, Gordon Dunsire (consultant) et Diane Hillmann (DCMI) ont demandé la parole pour exposer le fait qu'il n'est pas indispensable d'abandonner nos savoir-faire actuels dans la production de l'information bibliographique pour mettre nos données sur le Web³ :

The Road Forward / Gordon Dunsire, Diane Hillman

- laisser localement les données dans leur format d'origine pour répondre aux besoins des applications locales, mais les introduire dans les graphes globaux grâce à des mappings s'appuyant sur la sémantique des ontologies locales : « think local, act global »
- les développements locaux peuvent se faire à leur propre rythme, inutile de s'attendre les uns les autres ou d'attendre des consignes : « do what you need »
- les dangers :
 - o continuer à penser « notices » comme unités de gestion de l'information bibliographique, alors qu'il faut penser « éléments de données » ; la notion de « notices maîtres » disparaît : les données de niveau local sont disponibles pour toutes sortes d'agrégation au niveau global.
 - o s'inquiéter du cadre dans lequel sont gérées les métadonnées (« metadata residence » in one schema) alors que c'est surtout les « vues » que l'on a sur elles qui sont importantes.

¹ <http://www.niso.org/topics/tl/BibliographicRoadmap/attendees/>

² <http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>

³ <http://fr.slideshare.net/BaltimoreNISO/dunsire-roadmap-meeting-proposal>



- le rôle des organismes de normalisation (tel que le NISO) est
 - o de rechercher le consensus (notion de « contrat ») dans l'identification des données nécessaires aux services globaux
 - o en s'appuyant sur les points forts des bibliothèques (données de confiance, tradition de collaboration)
- importance de la provenance pour filtrer les données
 - o la confiance dans les données est difficile à établir s'il n'y a pas de mention de provenance de ces données
 - o question (non résolue aujourd'hui) : à quel niveau indiquer la provenance ? au niveau de chaque triplet ? comment rendre compte clairement des diverses modifications faites sur une donnée source ?
- quels sont nos besoins ?
 - o développer les espaces de noms, sans y introduire de hiérarchie
 - o encourager l'innovation à tous les niveaux, celle des institutions et celles des individus, et s'appuyer sur les expériences
 - o aller de l'avant, sans régression

Cette introduction a suscité quelques réactions :

- nous avons tant de belles normes qui ont nécessité beaucoup de temps d'élaboration ...
 - o Réponse de Gordon Dunsire : le Web sémantique permet de faire ce que les humains ne peuvent pas faire, par conséquent il n'est pas indispensable d'uniformiser tout ce qu'on fait au niveau international.
- Comment échanger des données au niveau global si on n'a qu'une cohérence au niveau local ?
 - o Réponse de Gordon Dunsire : il faut exposer les données selon les besoins des communautés qu'on sert.
- quel contrôle qualité sur les transformations de données qui sont un véritable business !
- pourquoi toujours parler des données des bibliothèques (« library data ») et jamais des usagers des bibliothèques (« library users ») ?
 - o réponse : parce que la mise à disposition des données crée les usages ... bien au-delà des seules bibliothèques.
 - o

En écho à cette présentation, mais un peu plus tard dans la journée, Ted Fons a présenté les travaux d'OCLC sur BibFrame.

New directions in metadata exchange : the early experimenters experience and what is happening at OCLC / Ted Fons, OCLC

- expériences à la Library of Congress, George Washington University, DNB, BL, National Library of Medicine, Princeton University, OCLC.
- calendrier de BibFrame : lancement en octobre 2011, première publication du modèle et premier atelier en octobre 2012, contrat avec Zepheira, second atelier en décembre 2012 et ouverture du site bibframe.org en janvier 2013
 - o premier atelier : discussion du modèle et de la phase d'expérimentation, démonstration par la LoC
 - o second atelier : discussion des résultats de l'expérimentation, conversion de tout WorldCat selon BIBFRAME par l'Equipe Recherche
- ce que fait OCLC :
 - o participe à l'expérimentation de BibFrame et aux discussions
 - o diffuse les données expérimentales de WorldCat en BibFrame à ceux qui font les expérimentations
 - o étudie et discute les notions « collection » et « autorités » dans BibFrame, et explore l'exposition des données de bibliothèques sur le Web
 - o aligne schema.org et BibFrame : comment prendre en compte le domaine spécifique des bibliothèques
 - o but à long terme : fournir des services pour les nouveaux flux de métadonnées et l'exposition sur le web des bibliothèques membres
- orientations
 - o exposer des ensembles de données en tant que données liées pour mettre en valeur les savoirs présents dans les données de bibliothèques
 - o agréger des données pour optimiser leur exposition sur le Web
 - o scruter les données (« data mining ») pour créer de la valeur ajoutée



- déconstruire le MARC et identifier ses composants comme des ressources (MARCR)

Vocabulary management : an essential part of the roadmap / Diane Hillmann⁴

Le Vocabulary Management Community du DCMI, créé en 2011, est un forum et un espace de collaboration pour ceux qui s'intéressent à la définition de bonnes pratiques en matière de vocabulaires, à la production de conseils pour les nouveaux praticiens et à la gestion responsable du changement. La communauté aborde également les questions de conception et de qualité des vocabulaires et des mappings dans l'environnement du Web sémantique.

Un Vocabulary Day est prévu lors de la réunion du DCMI à Lisbonne en 2013.

La présentation a abordé les thèmes suivants :

- durabilité des vocabulaires
 - Quels critères retenir pour la définir ? Quelles mesures prendre pour l'accroître ? Que faire des vocabulaires « orphelins » ? Quel rôle pourrait jouer le DCMI dans ce domaine ?
- vocabulaires et multilinguisme
 - Quelles sont les standards applicables ? Faire des exemples ... Développer des bonnes pratiques pour créer des vocabulaires utilisables au niveau global.
- faire de l'extension des vocabulaires un principe au service des communautés spécialisées
- faire des vrais mappings correspondant à des vrais besoins
 - s'appuyer sur des activités "de terrain" pour identifier les questions qui surgissent quand on passe du « parler » au « faire » comme par exemple <http://www.bbc.co.uk/arts/yourpaintings/>
 - avancer vers les bonnes pratiques en partant des besoins réels
- l'Open Metadata Registry pour promouvoir l'interopérabilité
 - http://wiki.metadataregistry.org/Open_Metadata_Registry

IFLA in RDF (+ RDA) / Gordon Dunsire⁵

Rappel des normes maintenues par l'IFLA depuis 40 ans (ISBD, UNIMARC (ISO 2709), FRBR, FRAD, FRSAD, MulDiCat), de l'apparition des principes du contrôle bibliographique universel dans les années 70, les représentations RDF depuis 2009, l'entreprise de l'Open Metadata Registry et le statut actuel des mappings.

Présentation de data.bnf.fr / Françoise Bourdon⁶

Un exemple qui démontre qu'il est possible d'exposer sur le web, en données liées, des données produites en MARC et d'expérimenter des alignements ...

3 questions m'ont été posées :

- sur l'existence d'un SPARQL endpoint : en projet, mais sans calendrier pour l'ouverture au public ;
- sur l'adresse où trouver des infos en anglais sur les techniques employés dans data : <http://data.bnf.fr/semanticweb-en> (le lien va être rendu plus visible sur les pages data.bnf.fr) ;
- sur ce qui a manqué dans les données quand nous avons fait les transformations : des notices d'autorité Oeuvres et des liens entre les notices bibliographiques et ces notices d'œuvres. Nous avons dû recourir à des algorithmes spéciaux pour rassembler les différentes éditions d'une œuvre. Certes des notices d'autorité Oeuvres existent dans notre catalogue général, mais elles étaient jusqu'à présent établies pour les points d'accès sujet. Ainsi beaucoup de notices bibliographiques n'ont pas de lien aux autorités Oeuvres dans leur zone de titre.

D'autres présentations d'expérimentations ont eu lieu, dont 2 sont aujourd'hui disponibles sur le Web :

- une application de découverte⁷ mise au point au Colorado College / présenté par Jeremy Nelson, Metadata and Systems Librarian, Colorado College
- le modèle de données d'Europeana / Antoine Isaac⁸

Le reste du temps les débats ont eu lieu dans des petits groupes de discussion

Quelques semaines avant la réunion il y a eu un large « appel à idées » sur le site du NISO, qui devait aider à formaliser le programme. 7 thèmes ont été identifiés⁹.

- Quels prérequis les données liées supposent-elles ?

⁴ http://wiki.dublincore.org/index.php/DCMI_Vocabulary_Management_Community

⁵ <http://fr.slideshare.net/BaltimoreNISO/iflardf>

⁶ <http://fr.slideshare.net/BaltimoreNISO/linked-data-efforts-at-the-bibliotheque-nationale-de-france>

⁷ <http://discovery.coloradocollege.edu/apps/discovery/>

⁸ <http://fr.slideshare.net/antoineisaac/isaac-niso>

⁹ <https://sites.google.com/site/nisobibrm/>



- Echange de données, échange de notices, compatibilité
- Servir différentes communautés (IFLA et son arsenal normatif, la communauté RDA, les éditeurs avec ONIX, la LoC avec BIBFRAME, Europeana, etc.) ; et différents utilisateurs (concepteurs de systèmes, chercheurs, grand public, etc.).
- Qui et comment ? Que devient le partage international du travail conçu aujourd'hui sur une base nationale dans le cadre du contrôle bibliographique universel ? Qui va payer ?
- Mappings vs migration de données

Ces idées ont été reprises pour organiser le travail en groupes : 8 groupes se sont réunis autour des thèmes suivants qui ont fait l'objet de rapides comptes rendus tous disponibles sur le site du NISO (voir note 9):

- Business models
- Goals
- Interoperability
- Open/Share
- Prototyping
- Rules
- Users
- Provenance/Authority

La dernière demi-journée a été consacrée à un grand brainstorming entre tous les participants, dont le but était de définir une feuille de route pour la suite. Voici un aperçu de ces échanges dont un script intégral est disponible en ligne (voir note 9) :

- C'est paralysant de réaliser qu'il se passe quelque chose qui mettra bien 20 ans à prendre forme ; la transition ne durera pas éternellement, mais elle ne se fera pas rapidement ; il faut vaincre l'inertie ...
- RDF : attention à ne pas créer un système d'encodage propre aux bibliothèques : il faut entrer dans le monde de l'échange de données global ;
- NISO pourrait jouer un rôle dans la définition de la tactique à adopter : ce ne doit pas être une compétition mais une exploration ;
- Nous avons une « vision » mais ça reste vague, nous manquons de buts concrets ;
- Il faut faire des choses concrètes, maîtrisables, et ne pas définir plus de 2 ou 3 priorités ; s'inspirer des recommandations du rapport de l'Incubator Group du W3C ; travailler quelques use cases avant d'aller plus loin ;
- L'important c'est d' « exposer la valeur », mais tout le monde n'a pas la même compréhension de la valeur, il faudrait étudier ce point.
- Dans les résultats de recherche il ne faudrait pas obtenir une liste, mais un graphe : pour cela nous devrions nous appuyer plus fortement sur nos systèmes de classification, nos notices d'autorité. Quelle tactique pour créer les autorités et pour les faire respecter ? Pourquoi attendre pour lier nos notices bibliographiques à nos notices d'autorité ?
- Il faut avancer à petits pas, faire des prototypes. C'est autant une affaire de pédagogie qu'une question technique ; il ne faut pas être trop ambitieux, on a besoin de petites réussites pour avancer ...
- Souvent on ne sait même pas ce qu'on a le droit de faire avec les données, notamment celles élaborées en partenariat : on risque des sanctions ...
- Il n'y a pas que la Library of Congress et OCLC qui font des expérimentations ! NISO devrait aider à faire connaître ce qui est expérimenté ici ou là
- Ce qu'il faut faire : « Things and strings » : rendre les identifiants explicites dans les vedettes ! OCLC le fait et tout le monde pourra le récupérer ...
- La facilité : faire confiance à BibFrame ... parce que personne n'a de Plan B ! Oui mais est-on sûr que BibFrame soit un vrai plan A ? BibFrame est en gestation ... Mais c'est le cas de beaucoup de projets en cours dans les bibliothèques : aucun ne peut se prévaloir d'être la solution finale.
- Dans les années 60 la Library of Congress a pris le leadership car elle seule avait la puissance informatique nécessaire. Aujourd'hui n'importe qui peut développer des choses sur un simple PC !
- Les organismes de normalisation ne devraient-ils pas aider à trouver une solution de remplacement à BibFrame ? Et si c'est un vrai casse-tête il n'y a qu'à l'ignorer complètement !
- Si la Bibliothèque du Congrès ne fait pas ce qui doit être fait, alors regardons ailleurs. Nous ne voulons pas construire un silo de meilleure qualité, nous voulons travailler avec le reste du monde !
- Il manque un endroit où référencer les mappings existants (tout ce qui lie 2 identifiants l'un à l'autre est un mapping), savoir qui les a faits, etc. Par exemple pour référencer ceux faits par la BL, la DNB, la BnF ... TALIS propose un tel endroit mais mieux vaudrait avoir affaire à une entité sans but lucratif



- Recenser les outils dont on dispose pour faire les mappings : il existe déjà <http://www.w3.org/2001/sw/wiki/LLDtools>, mais il faudrait identifier ce qu'il manque. On pourrait proposer des exercices, par exemple, lister les outils en lien avec les prototypes...
- Il faut trouver un moyen d'évaluer les projets existants. Par exemple prendre en compte le nombre de services rendus par rapport à l'investissement en personnel ?
- Le projet GLIMIR (Global Library Manifestation identifier¹⁰) d'OCLC attribue un identifiant aux grappes de notices représentant la même manifestation dans WorldCat. OCLC voudrait savoir comment exposer ces identifiants pour qu'ils soient réutilisés ... Comment concevoir ce service ?
- Ne pas confondre Linked Data et Linked *Open* Data ? Comment gérer le fait que certaines données ne sont pas gratuites ? L'American Library Association va plaider la cause des bibliothèques ...

Les échanges vont se poursuivre en juin-juillet, sur WebEx probablement. Un rapport final va être rédigé, qui pourrait être discuté au DCMI de Lisbonne en septembre, puis rendu public avec appel à commentaires cet automne.

¹⁰ <http://journal.code4lib.org/articles/6812> : article de Janifer Gatenby et al.