

Bibframe, un nouveau modèle de données pour les bibliothèques

Initiative visant à faire évoluer les standards de description bibliographique vers un modèle de données liées, Bibframe (*Bibliographic Framework Initiative*) doit ainsi faciliter l'accès à l'information bibliographique et maximiser son utilisation par les différentes communautés d'usagers.

L'initiative Bibframe a été lancée en 2011 par la Bibliothèque du Congrès afin de fournir « une base pour l'avenir de la future description bibliographique, pour le web mais aussi plus largement pour le monde interconnecté »¹. Le projet vise tout d'abord à remplacer le format Marc 21, qui fut la pierre angulaire de l'informatisation des bibliothèques et de la gestion des métadonnées. Bien que Marc 21 soit un standard très bien structuré et maintenu – offrant un riche panel d'éléments adapté à la complexité de l'information bibliographique –, il montre toutefois aujourd'hui ses limites. Initialement format d'échange, il a été largement implémenté dans les SIGB en tant qu'interface de catalogage. Depuis, Marc est devenu véritablement la *lingua franca* des catalogueurs. Mais les SIGB n'étant malheureusement pas suffisamment adaptés aux nouveaux enjeux du catalogage, Marc est également devenu un format dépassé.

QUAND MARC 21 RENCONTRE LE LINKED LIBRARY DATA

Pour définir le successeur du format Marc, il était nécessaire de prendre en compte les approches du web de données. Aujourd'hui, le web est la plateforme où se trouvent nos utilisateurs, c'est un outil incontournable. La construction du web sémantique est un effort entrepris par un grand nombre de communautés. Cependant, pendant longtemps, les bibliothèques n'ont pas réussi à prendre part au développement de standards qui constituent pourtant les bases de sa structure². Les « silos de données » stockent des informations accessibles via des interfaces non ouvertes sur le web, donnant ainsi sur Internet l'image d'un trou façonné par les bibliothèques. Pourtant, la valeur ajoutée portée par la richesse de leurs métadonnées les encourage à faire partie intégrante du web afin de contribuer au « *Giant global graph* »³.

Avant Bibframe, quelques implémentations allant dans le sens du *Library Linked Data* existaient déjà. Afin de fournir un meilleur service aux « consommateurs » de données, des éléments issus de différents vocabulaires (tels Dublin Core, FOAF, BIBO, RDA, ISBD...) ont été choisis. Si cette sorte de « picorage » reste une approche valide, il semble toutefois nécessaire de fonder un modèle et un vocabulaire spécifiques qui puissent offrir à l'ensemble des éléments une structure cohérente.

PETIT RETOUR EN ARRIÈRE

Tenant compte des modèles existants et s'appuyant sur des analyses approfondies des potentialités des technologies du web de données, la Bibliothèque du Congrès, en partenariat avec la société Zepheira⁴, a établi un nouveau modèle ainsi qu'un vocabulaire. Quatre classes principales ont ainsi été définies : « Œuvre », « Instance », « Autorité » et « Annotation ». Un premier groupe d'expérimentateurs a ensuite été mis en place. Le modèle et le vocabulaire ont été intensément discutés, des documents de travail et d'étapes ont été rédigés afin de mettre en lumière les cas particuliers, des outils ont été développés pour convertir les données Marc en Bibframe et créer des données Bibframe *ex nihilo*. Enfin, des tests ont été effectués afin de vérifier si Bibframe pouvait convenir autant aux données existantes qu'aux futures implémentations. Au bout d'un an, le cercle des premiers expérimentateurs a été élargi et le « Registre d'implémentation Bibframe » a été mis en place. Il s'agit d'une liste ouverte des institutions utilisant les structures Bibframe dans des situations concrètes.

DES ACTEURS DIVERS ET VARIÉS

Différents acteurs sont très actifs dans l'environnement Bibframe. On n'évoquera ici que les plus importants.

Tout d'abord, la Bibliothèque du Congrès

a préparé un prototype en vue de tester l'efficacité du modèle lors de la création de métadonnées de documents et de langues variés. Travaillant en parallèle en Marc 21 et en Bibframe, les catalogueurs ont donné leurs avis aux experts du modèle et du vocabulaire. Les outils constituant les blocs structurants du modèle pour le prototype ont été mis à disposition. On note que le Programme pour le catalogage coopératif (PCC) a été largement impliqué dans les activités Bibframe de la Bibliothèque du Congrès. Avec l'initiative LibHub, Zepheira a choisi de rendre visibles les richesses des bibliothèques par la création d'un vocabulaire Bibframe modulable. Les données, collectées principalement à partir de catalogues de bibliothèques de lecture publique, ont été converties en Bibframe. Les relations entre entités ont ensuite été créées. Enfin, les données ont été exposées sur le web afin que les fournisseurs de moteurs de recherche puissent les utiliser. Certains résultats sont déjà visibles.

OCLC s'est concentré sur schema.org, vocabulaire déterminé par les quatre plus grands fournisseurs de moteurs de recherche. Quelques éléments pour les données spécifiques aux bibliothèques ont été ajoutés et une couche additionnelle a été définie. L'analyse des résultats a été publiée par les experts de la Bibliothèque du Congrès et d'OCLC dans un article intitulé : « Common Ground »⁵. Les modèles sont maintenant utilisables de façon interactive afin que les données puissent être « mappées » puis réutilisées.

Le projet LD4L (« *Linked Data for Libraries* ») et son successeur LD4P (« *Linked Data for Production* ») – financés par la Fondation Mellon⁶ et pilotés par plusieurs bibliothèques de recherche – visent quant à eux à développer une nouvelle infrastructure pour la gestion des ressources et des

métadonnées. Dès le début, Bibframe a joué un rôle significatif. Il reste maintenant à voir comment ces différentes voies vont converger à l'avenir.

L'IMPLICATION DE LA BIBLIOTHÈQUE NATIONALE ALLEMANDE (DNB)

La Bibliothèque nationale allemande a été parmi les premiers expérimentateurs de Bibframe. En 2013, un projet a été mis en œuvre pour soutenir l'initiative. En plus d'autres activités, la DNB a mis en place un service fournissant des données basées sur un simple *mapping* du format pivot Pica+ vers Bibframe : une première table de conversion a été écrite et implémentée dans le catalogue de la DNB. Concrètement, à partir de l'affichage simple de la plupart des notices, un menu propose d'obtenir une représentation Bibframe.

Le fait d'avoir entrepris ces tests sur des données réelles a permis à la DNB de partager des résultats détaillés avec les autres partenaires Bibframe. Depuis, la DNB suit l'initiative en tant qu'observateur, mais une participation plus active est prévue pour la suite du projet sur 2016-2017. À mi-parcours de son développement, Bibframe deviendra un produit solide du portefeuille de l'offre de métadonnées de la DNB.

DE NOUVEAUX DÉVELOPPEMENTS

Très récemment, le modèle et le vocabulaire ont subi de profondes modifications. Les résultats issus des analyses du groupe pilote de la Bibliothèque du Congrès et du LD4L ont indiqué qu'un processus de révision s'imposait. En avril 2016, une nouvelle version Bibframe a été publiée. Dorénavant, Bibframe 2.0 structure l'information en trois niveaux fondamentaux d'abstraction : **l'œuvre**, qui reflète l'aspect conceptuel de la ressource cataloguée ; **l'instance**, une représentation « individuelle » de l'œuvre, et **l'item**, nouveau concept pour désigner la copie physique ou numérique d'une instance. Autres concepts clés à partir desquels des relations ont été bâties entre les trois classes fondamentales : les **agents**, i.e. les personnes physiques, les organisations, etc., associés à une œuvre ou une instance à travers des rôles spécifiques, les **sujets** qui décrivent le contenu de l'œuvre, et les **événements**. Les concepts « Autorité » et « Annotation » considérés comme non pertinents ont été abandonnés. Ce nouveau

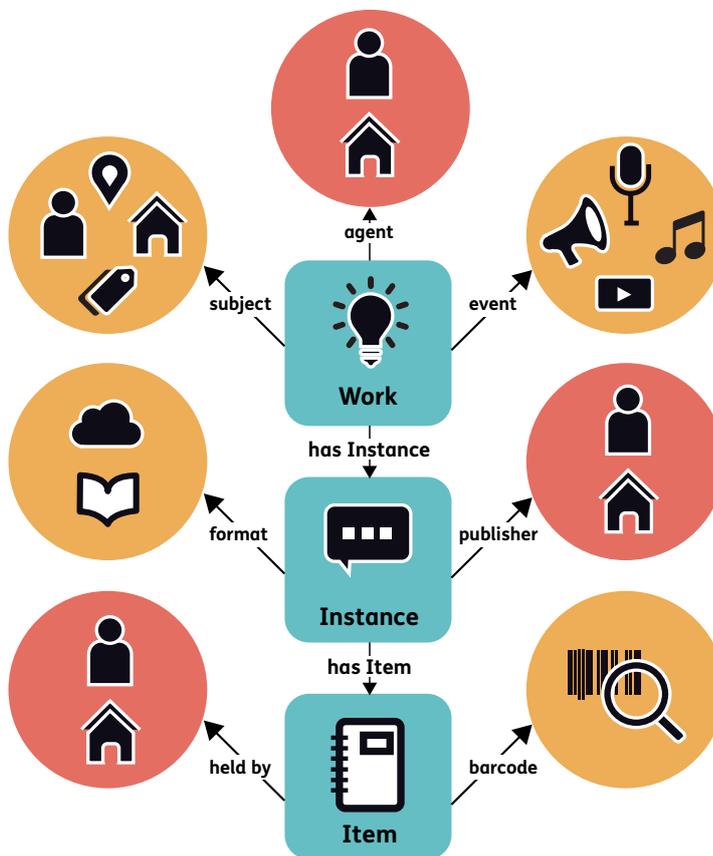


Schéma du modèle et vocabulaire Bibframe
Source : <http://www.loc.gov/bibframe/docs/bibframe2-model.html>

vocabulaire est publié en RDF et disponible sous une forme qui permet la navigation hypertextuelle et la recherche par différents critères.

QUESTIONS OUVERTES

Comme tout nouveau produit, Bibframe a parcouru le « *Hype cycle* » : il a débuté par la mise en place d'une nouvelle technologie, avant de suivre l'ascension du « Pic des espérances exagérées », puis la descente vers le « Gouffre des désillusions » et à nouveau l'ascension de la « Pente de l'illumination » pour arriver sur le « Plateau de productivité »⁷.

Quelques interrogations subsistent : Bibframe est-il encore trop fondé sur Marc 21 ? Respecte-t-il les bonnes pratiques du web de données ? Les différentes approches finiront-elles par diviser les communautés ? Comment cette initiative peut-elle devenir vraiment internationale ? Comment intégrer plus largement les secteurs culturels et patrimoniaux ? Combien de temps faudra-t-il pour atteindre un bon niveau de production de métadonnées ? Y aura-t-il un retour sur investissement ? Et dans quelle mesure la mise en œuvre de Bibframe concrétisera-t-elle la nécessité de servir les intérêts des utilisateurs finaux ?

CONCLUSION

Il est toujours fascinant de voir un nouveau standard in *statu nascendi*.

Bibframe est évolutif et n'en est qu'au début de son développement. Au fil de la transition de Marc vers Bibframe, la communauté des métadonnées se transforme également. Le succès du projet repose sur un équilibre entre les différents objectifs, rassemblant tous les acteurs dans un modèle de maintenance et de gouvernance commune. Les institutions à l'échelle mondiale, y compris les éditeurs de logiciels et de ressources, sont invitées à participer à ce projet en proposant et en « connectant » à Bibframe leurs métadonnées. Intégrer les standards existants tels que Mods, Onix et Unimarc à Bibframe est aussi une voie à explorer. Enfin, toute contribution est bien sûr la bienvenue !

REINHOLD HEUVELMANN

Bibliothèque nationale allemande
R.Heuvelmann@dnb.de
Traduction : P.Pons (Abes)

- [1] <https://www.loc.gov/bibframe/>
- [2] « Building blocks » dans le texte original
- [3] « Graphe global géant », expression employée pour la première fois par Tim Berners-Lee
- [4] <https://zephaira.com/>
- [5] <http://www.oclc.org/research/publications/2015/oclcresearch-loc-linked-data-2015.html>
- [6] <https://mellon.org/>
- [7] https://fr.wikipedia.org/wiki/Cycle_du_hype