

À défaut d’enterrement : les défis et les promesses de l’INTERMARC nouvelle génération

French translation of the original paper: “Still Waiting for That Funeral: the Challenges and Promises of a Next-Gen INTERMARC”.

Translated by: Mélanie Roche, Bibliothèque nationale de France, Paris, France

Le texte de ce document est une traduction en français et peut présenter des différences par rapport au texte original. La présente traduction n’est fournie qu’à titre de référence.

Sébastien Peyrard

Département des métadonnées, Bibliothèque nationale de France, Paris, France.
Adresse mél : sebastien.peyrard@bnf.fr

Mélanie Roche

Département des métadonnées, Bibliothèque nationale de France, Paris, France.
Adresse mél : melanie.roche@bnf.fr

This is a French translation of “Still Waiting for That Funeral: the Promises and Challenges of a Next-Gen INTERMARC” copyright © 2018 by Mélanie Roche. This work is made available under the terms of the Creative Commons Attribution 4.0 Unported License:
<http://creativecommons.org/licenses/by/4.0>

Extrait

Depuis près de quarante ans la Bibliothèque nationale de France utilise comme format de production l’INTERMARC, une déclinaison maison des formats MARC qu’elle revisite aujourd’hui pour la rendre compatible avec les technologies du Web sémantique et les modèles bibliographiques récemment développés tels FRBR et IFLA LRM. Garder MARC comme format de production, c’est garantir la continuité professionnelle et mettre en valeur l’expertise des bibliographes ; l’exporter en Linked Data, c’est identifier et répondre aux besoins d’une communauté d’usagers principalement composée de non bibliographes. Notre objectif est de passer de notices à des données liées, réutilisables et de confiance et d’améliorer la portabilité des notices MARC grâce à un niveau de granularité plus fin — le tout en respectant le formalisme ISO 2709. Notre contexte est celui de la Transition bibliographique, qui vise un catalogage natif en LRM grâce à une déclinaison française de RDA. Les enjeux sont d’autant plus élevés qu’il s’agit d’accompagner le changement pour toutes les parties concernées.

Mots-clés : Transition bibliographique, MARC, INTERMARC, Web sémantique, IFLA LRM

INTRODUCTION

Les catalogues ont été le point d'entrée de l'informatique en bibliothèque. À la fin des années 1960, le couple formé par les catalogues de bibliothèque et les technologies de l'informatique était officialisé par l'avènement du premier format de catalogage lisible par machine (MARC, *MAchine-Readable Cataloguing*), USMARC — devenu en 1999 MARC21, la couleur de MARC la plus fréquemment utilisée dans le monde anglo-saxon, et le format d'OCLC pour Worldcat. Mais l'histoire d'amour eut tôt fait d'être menacée quand il devint évident que les nouvelles technologies évoluaient plus vite que les formats bibliographiques. Pour envenimer les choses, au tournant du XXI^e siècle un nouveau modèle conceptuel pour les données bibliographiques appelé FRBR (*Functional Requirements for Bibliographic Records*, Fonctionnalités Requises pour les Notices Bibliographiques) devait semer la discorde dans cette idylle vieille de trente ans. Dix ans plus tard, RDA (*Ressource Description and Access*, Ressource : Description et Accès), rejeton de FRBR et des idéaux du Web sémantique, consommait le divorce entre formats bibliographiques et nouvelles technologies en instaurant de nouvelles règles de catalogage agnostiques vis-à-vis des formats d'encodage.

Il est indéniable que jusqu'à présent les formats MARC ne se sont pas montrés à la hauteur des multiples défis posés par les nouveaux modèles et règles de catalogage. Pourtant il semble que, trop souvent, on parle des « formats MARC » pour le MARC21 ; or pointer du doigt les défauts de MARC21 ne devrait pas aboutir à une condamnation de l'ensemble des formats MARC. C'est en tout cas l'opinion en vigueur à la Bibliothèque nationale de France (BnF), qui a engagé une refonte radicale de son format de production maison INTERMARC pour le rendre compatible avec les nouveaux standards du Web sémantique, les nouveaux modèles bibliographiques et les nouvelles règles de catalogage.

Cet article se concentrera sur la méthodologie et les résultats : comment l'idée d'un INTERMARC « nouvelle génération » a émergé, comment on en a fait un processus véritablement participatif, et à quoi ressemblera le futur format. Sachant que le projet en est à la moitié, tout ce que l'on peut tenir pour acquis à ce jour est que la prochaine génération d'INTERMARC sera orientée entités, s'appuiera fortement sur les référentiels, et inclura des méta-métadonnées.

1. LA PRISE DE CONSCIENCE

1.1. C'est le MARC 21 qui doit mourir

Il y a plus de quinze ans, le monde des bibliothèques recevait comme un choc l'avis de décès de MARC¹. Dans une tribune du *Library Journal* qui a fait date, Roy Tennant prédisait la fin proche et certaine des formats MARC, techniquement obsolètes et généralement inadaptés aux besoins des utilisateurs et des catalogueurs. Les auteurs de cet article n'étaient pas nés lorsque les formats MARC ont vu le jour, et n'étaient pas même bibliothécaires quand l'article de Tennant a été publié : on ne saurait donc les suspecter d'un quelconque biais personnel ou professionnel en faveur d'une époque ou d'un format particulier. Pourtant, de notre point de vue de *millennials*, nous pensons que si l'on attend depuis aussi longtemps son enterrement, c'est peut-être tout simplement parce que MARC n'est pas tout à fait mort à l'heure qu'il est.

¹ <http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/#>

Les arguments de Tennant sont tous très justes... Dès lors que l'on parle de MARC21. Pour autant, il serait injuste de réserver le même sort funeste aux autres formats fondés sur le MARC. L'UNIMARC, par exemple, s'est toujours appuyé sur des identifiants locaux pour faire des liens entre deux notices (qu'elles soient d'autorité ou bibliographiques).

Tennant lui-même faisait remarquer que le terme « MARC regroupe plusieurs choses liées. Il y a la syntaxe MARC, les éléments de données MARC, et les *Anglo-American Cataloguing Rules* (AACR)² ». Cela aussi est très vrai pour MARC21 mais moins pour l'UNIMARC, comme le montrent les évolutions récentes des deux formats. Depuis la décision de la Bibliothèque du Congrès d'adopter RDA en 2012, MARC21 a évolué de façon à s'aligner sur les nouvelles règles, mais n'a pas engagé de révision structurelle en termes de modélisation. Pourtant, être compatible avec le modèle qui sous-tend un code de catalogage peut en fin de compte se révéler plus payant pour l'interopérabilité des données – d'autant plus maintenant que RDA a officiellement implémenté IFLA LRM, le modèle conceptuel ayant succédé à la famille de modèles FR-³. C'est cette voie qu'a privilégiée UNIMARC il y a quelques années : le format UNIMARC autorités (UNIMARC/A) a engagé une révision structurelle dès 2010 pour s'aligner sur FRAD⁴, le nouveau modèle pour les données d'autorité qui venait tout juste d'être publié. De même, ces quatre dernières années, le format UNIMARC bibliographique (UNIMARC/B) a évolué pour prendre en compte les concepts d'Œuvre et Expression requis pour un catalogage FRBRisé. Certes, UNIMARC est toujours loin d'avoir pleinement implémenté FRBR ou IFLA LRM, mais il a ouvert une voie dans laquelle nous avons décidé de nous engager encore plus avant.

1.2. Les enjeux d'une agence bibliographique nationale

D'après Tennant, « il n'y a que de types de personnes qui se croient capables de lire une notice MARC sans avoir à se référer à une pile de manuels : une poignée de nos catalogueurs les plus aguerris, et ceux que la drogue a sérieusement atteints⁵ ». À la Bibliothèque nationale de France, nous avons une « poignée » de quelque 300 agents manipulant du MARC au quotidien, sans compter tous nos experts en métadonnées et nos formateurs — qui, nous pouvons l'attester, ne sont pas (tous) « sérieusement atteints par la drogue ». Depuis le début du catalogage informatisé à la fin des années 1980, les catalogueurs de la BnF se sont familiarisés avec les formats MARC, et en tant que francophones, sont en réalité bien plus à l'aise avec des étiquettes à trois chiffres (aussi ésotériques puissent-elles paraître de l'extérieur) qu'avec un format reposant sur l'anglais comme l'XML que Tennant voyait comme l'avenir du catalogue.

² *Ndlr* : traduit par nos soins.

³ <http://www.rda-rsc.org/ImplementationLRMinRDA>

⁴ *Functional Requirements for Authority Data*, Fonctionnalités requises pour les données d'autorité, un modèle conceptuel rendu obsolète par IFLA LRM.

⁵ *Ndlr*: traduit par nos soins.

Cela fait une vingtaine d'années que la BnF possède son propre système de production de données en MARC, et plus de quarante ans qu'elle utilise son propre format MARC. INTERMARC est né en 1974⁶ et présente un mélange entre le MARC21 et l'UNIMARC. Les étiquettes et la structure générale sont plus proches de celles du MARC21, mais certaines fonctionnalités de base font plutôt penser à l'UNIMARC, comme l'établissement systématique de liens entre notices bibliographiques et d'autorité, ou entre notices bibliographiques ; le format est également structuré à un niveau de granularité plus fin que celui de MARC21 ou même de l'UNIMARC. INTERMARC est un format très souple et adaptable. Il y a de cela vingt ans, on a déjà dû le faire passer d'une série de formats taillés sur mesure pour des types particuliers de ressources à un format intégré pouvant convenir à la description de tout type de document. La problématique d'alors était celle de la granularité : il nous fallait inventer un format adapté à l'ensemble de nos départements de collections, tout en fournissant la richesse d'information attendue d'une agence bibliographique nationale. Aujourd'hui, la problématique n'est plus tellement la même, puisque nous pensons en termes d'interopérabilité des données ; pourtant une question fondamentale demeure : est-ce qu'on réinvente la roue, ou est-ce qu'on s'appuie sur ce que nous connaissons et maîtrisons déjà ?

La réponse aujourd'hui est la même qu'hier : pour l'établissement, il serait dangereux de tout bonnement ignorer une telle force de travail hautement qualifiée, non seulement dans une optique d'accompagnement au changement, mais aussi parce qu'en tant que productrice de la Bibliographie Nationale Française, la Bibliothèque nationale se doit de produire des notices dans un délai raisonnable après le Dépôt Légal, et ne peut par conséquent pas prendre le risque d'allonger les délais de catalogage pour cause de nouveau format.

En tant que bibliothèque nationale, nous servons également une communauté de bibliothèques publiques qui utilisent toujours le MARC — et l'utiliseront vraisemblablement toujours dans un futur proche, car UNIMARC est encore un format d'échange largement utilisé, non seulement en France mais aussi en Europe. Il nous semblait irréaliste de demander à cette communauté de patienter indéfiniment en vue d'une hypothétique révolution bibliographique. Notre responsabilité envers ces utilisateurs professionnels étant tout aussi grande qu'envers le public, nous avons donc inscrit notre réflexion sur INTERMARC dans le contexte plus large de la Transition bibliographique⁷.

1.3. Vers un INTERMARC Nouvelle Génération

À la BnF, nous ne sommes pas convaincus que les récents développements de BIBFRAME l'imposent comme remplaçant des structures fondées sur le MARC, surtout en ce qui concerne l'interprétation du modèle FRBR — à présent IFLA LRM. Si nous pensons que l'exposition en RDF est la clé pour des données trouvables et réutilisables dans le Web sémantique (c'est la raison pour laquelle nous avons développé data.bnf.fr⁸), nous établissons une distinction nette entre la façon dont on produit les données, et celle dont on les stocke et les diffuse : car stocker et diffuser des triplets RDF ne veut pas forcément dire les produire en catalogage courant.

⁶ http://www.bnf.fr/fr/professionnels/f_intermarc/s.format_intermarc_histoire.html?first_Art=non

⁷ <https://www.transition-bibliographique.fr/enjeux/bibliographic-transition-in-france/>

⁸ <http://data.bnf.fr/>.

Tennant pensait que « Si les bibliothèques s'accrochent à des normes dépassées, elles verront qu'il leur est de plus en plus difficile de servir leurs clients comme ils l'attendent et le méritent⁹ ». Pourtant, depuis 2002 nous pensons que les bibliothèques sont bel et bien devenues les « organisations souples et réactives » que Tennant appelait de ses vœux. Or cela a moins à voir avec la technologie qu'avec la façon dont les bibliothécaires s'en sont servis : la technologie, en soi, n'est ni nécessaire ni suffisante pour amener des changements significatifs, tant qu'elle ne s'accompagne pas d'un changement de paradigme dans les mentalités.

Les arguments de Roy Tennant nous ont en fin de compte convaincus qu'un tel changement d'état d'esprit importait plus que n'importe quelle révolution technologique. « Afin de créer des normes qui soient à la fois adaptées aux besoins d'aujourd'hui et assez souples pour offrir de nouvelles opportunités, il nous faudrait commencer avec les fonctionnalités de la description bibliographique (voir *Functional Requirements for Bibliographic Records*, par exemple) et concevoir un format d'encodage à la fois souple et puissant¹⁰ ». Nous ne saurions être plus d'accord (nonobstant le fait que l'on parlerait aujourd'hui d'IFLA LRM), et c'est précisément ce à quoi nous nous sommes attelés.

2. LA METHODE UTILISEE

2.1. Deux séances de préfiguration

Le concept d'un « INTERMARC nouvelle génération » devait être testé par les experts du format les plus pointus à la BnF. Deux sessions de brainstorming ont été organisées afin de discuter la pertinence de conserver un format basé sur le MARC, son adéquation avec les objectifs des données bibliographiques à long terme, et sa comptabilité avec les principes du Linked Data. Ces sessions ont confirmé que le fait de garder un format interne qui ne soit pas révolutionnaire pour les catalogueurs était compatible avec la nature même de la Transition bibliographique. Elles ont aussi fait émerger les grands domaines à explorer, et le besoin d'un contexte aussi collaboratif que possible pour que les experts du catalogage s'approprient la méthode.

2.2. Un « INTERMARC Camp »

Les sujets devant être étudiés en détail étaient les suivants :

- *Entités et relations* : comment exprimer FRBR en MARC, de façon convaincante et efficace ?
- *Méta-métadonnées* : comment fournir des données utiles sur nos métadonnées, de sorte que les réutilisateurs et les gestionnaires de métadonnées puissent en appréhender la qualité dans un contexte de production donné ? À quel(s) niveau(x) de granularité ?
- *Référentiels* : dans un monde où le Web de données est la pierre angulaire pour des données interopérables compréhensibles par machine, comment faire passer la vitesse supérieure à MARC et atteindre une compatibilité complète avec le Linked Data ?

⁹ *Ndlr* : traduit par nos soins.

¹⁰ *Ndlr* : traduit par nos soins.

- *Syntaxe et conversions* : la syntaxe actuelle (ISO 2709) est-elle toujours pertinente ? Quels mécanismes du format faut-il changer, ou simplifier ? Avec quel impact sur les conversions courantes (car les données de la BnF se doivent de rester interoperables avec le reste du monde des bibliothèques) ?

Il fallait un large groupe d'experts pour échanger autour de ces sujets, aussi représentatif que possible des différents contextes de catalogage de la BnF (chaînes de catalogage, types de documents, etc.), et réuni dans une configuration collaborative. Chaque participant devait être en mesure d'exprimer son avis sur toutes ces questions à un moment ou à un autre, de façon à contribuer activement à l'évolution du format et à rendre un livrable que le groupe entier pourrait assumer. Plutôt que de rester dans l'abstraction, les discussions devaient prendre en compte, autant que faire se peut, des considérations concrètes d'encodage des données et de faisabilité technique. Pour des raisons pratiques, l'atelier devait par ailleurs se tenir sur une période de temps courte mais pas non plus trop courte, au vu de la complexité du sujet qui nous occupait.

De toute évidence, un format de réunion traditionnelle n'était pas adapté à une telle méthode, et c'est pourquoi nous avons opté pour une configuration de « World Café »¹¹. Cet « INTERMARC camp » s'est tenu sur une période de deux jours (les 23 et 24 février 2017), afin de gagner en efficacité, de rester concentrés et d'éviter la distraction des tâches quotidiennes.

Les quatre domaines identifiés dans le cadre des sessions de brainstorming ont fait l'objet de quatre ateliers distincts, organisés en sessions parallèles. Chaque atelier se composait de trois sessions le premier jour, et deux le second jour. Une introduction générale rappelait les consignes :

- Chaque atelier avait un animateur et un rapporteur permanents tout au long des deux jours ;
- Chaque participant-e participait à un atelier le premier jour, et à un autre le deuxième, avec les missions suivantes :
 - Le premier jour, il s'agissait d'imaginer les futures fonctionnalités de l'INTERMARC Nouvelle Génération, et d'en dresser une liste priorisée¹² ;
 - Le second jour, le second groupe remettait en question la liste rédigée le jour précédent par le premier groupe.
- Afin d'éviter trop de répétition durant les ateliers, trois *liberos*¹³ ont été nommés : leur mission était de participer à un atelier différent à chaque session et de faire circuler l'information.
- La fin d'après-midi du deuxième jour était consacrée à une restitution plénière rassemblant tous les participants, et préparée en amont par les animateurs, rapporteurs et *liberos* en début d'après-midi : quatre brèves synthèses résumaient les conclusions (et les questions ouvertes) de chaque atelier, permettant ainsi à chacun-e de contribuer aux deux ateliers auxquels il ou elle n'avait pas eu l'occasion de participer dans le cadre d'une discussion ouverte.

¹¹ https://en.wikipedia.org/wiki/World_caf%C3%A9

¹² Les fonctionnalités ont été priorisées grâce à la méthode « MoSCoW » : https://en.wikipedia.org/wiki/MoSCoW_method

¹³ https://en.wikipedia.org/wiki/Defender_%28association_football%29

- Enfin, une boîte à idées a circulé dans la salle afin de trouver un nom à cet INTERMARC Nouvelle Génération dont les contours se dessinaient un peu plus nettement.

De l’avis général, cet atelier a été un franc succès : les personnes sollicitées maîtrisaient bien INTERMARC, et étaient donc des interlocuteurs de choix pour préparer l’évolution du format. Cela nous a permis de partager une vision commune ; d’avoir une idée des besoins en jeu pour les différentes chaînes de catalogage et/ou cultures de catalogage au sein de la BnF ; et enfin de poser les bases des principes directeurs qui devaient présider à l’évolution d’INTERMARC.

2.3. Des groupes de travail

L’INTERMARC Camp ayant apporté les premiers éléments, encore fallait-il les affiner et en tester la faisabilité technique à l’aune de notre contexte concret de catalogage. Ce travail a été entrepris par un sous-groupe des Campeurs à travers une série de réunions plus traditionnelles abordant les sujets suivants :

- réunions entités-relations : Œuvres/Expressions ; Manifestations ; Items ; Agents ; Indexation sujet ; Nomen et attributs. Ces réunions avaient pour but de délimiter les contours de chaque entité, de dresser la liste des attributs et des relations requis, et d’établir une analyse des lacunes par rapport à l’existant (quelles zones et sous-zones conserver, réassigner, créer) ;
- réunions orientées traitement bibliographique : recueils, spectacles ;
- réunions orientées mécanismes : méta-métadonnées, référentiels de valeurs contrôlées.

Ces réunions se sont déroulées de mai à décembre 2017. Ensuite, quatre réunions distinctes se sont tenues entre le Département informatique et le Département des métadonnées pour évaluer la faisabilité des conclusions ainsi atteintes, mettant au jour plusieurs points à creuser. Le temps était donc venu de la validation des principes de l’INTERMARC Nouvelle Génération par notre comité de pilotage des métadonnées (COMET) — ce fut chose faite le 12 février 2018. La prochaine étape pour le futur format était la rédaction d’une Charte qui en énoncerait les principes fondamentaux et fixerait un cadre aux experts chargés de mener la cure de jouvence de ce bon vieil INTERMARC.

3. LES FONCTIONNALITES DU FUTUR FORMAT

L’INTERMARC Nouvelle Génération se profile désormais à grands traits dans une Charte, et se décline en trois principes fondamentaux : implémenter le schéma entités-relations du modèle IFLA LRM ; exprimer des relations à un niveau de granularité plus fin que celui de la « notice » ; exprimer des valeurs contrôlées sous la forme d’entités à part entière.

3.1. Entités/reliations

Les catalogueurs pensaient « données liées » bien avant la naissance même du Web : cela fait longtemps qu’ils décrivent des choses dans des documents (les notices), reliés entre eux par des identifiants. L’ère FRBR a défini un modèle fondamental qui réconcilie, à travers un petit noyau dur d’entités, ce que les catalogueurs décrivent avec ce que les usagers recherchent sur le Web. Une des idées principales derrière l’INTERMARC-NG est de faire coïncider ces entités abstraites avec les notices les décrivant, ce qui aboutit aux principes suivants :

- Chaque entité décrite dans l'INTERMARC-NG aura sa propre notice. Ainsi, un Lieu n'aura qu'une seule notice, contrairement à aujourd'hui, où le même lieu peut être décrit dans deux notices distinctes : l'une en RAMEAU pour l'indexation matière, et l'autre dans le fichier d'autorités géographiques de la BnF identifiant des lieux mentionnés dans nos ressources cartographiques.
- Chaque notice INTERMARC-NG décrira une et une seule entité. Cela implique de devoir détricoter des informations qui sont aujourd'hui mélangées dans une notice unique. Ainsi, une notice MARC bibliographique contient traditionnellement des informations qui relèvent à la fois de l'Œuvre, de l'Expression ou de la Manifestation. Les éléments de niveaux Œuvre et Expression seront donc déplacés de la notice bibliographique dans des notices d'Œuvre et d'Expression, laissant une notice de Manifestation qui sera *grosso modo* une notice bibliographique actuelle allégée. Les notices d'Expression seront créées ex-nihilo, tandis que les notices d'Œuvres s'inspireront de notre format pour les notices d'autorité Titres Uniformes, enrichi de nouveaux attributs et relations.
- Les liens entre entités seront exprimés par des liens entre les notices qui décrivent ces entités. Ainsi, un lien entre une Manifestation et une Expression sera exprimé par un lien depuis la notice de Manifestation vers la notice d'Expression, au moyen de son identifiant. L'idée est d'utiliser les mêmes mécanismes partout dans le format, contrairement à aujourd'hui, où ils diffèrent selon les formats bibliographique ou d'autorité. Cela se résume à un noyau de trois sous-zones : l'identifiant de l'entité cible ; la nature de la relation ; le degré de certitude du lien, pour les cas douteux (par exemple, un créateur *prétendu* ou *présumé* d'une Œuvre).

Nous avons cependant décidé de ne pas traiter les entités LRM E9-Nomen et LRM-E11 Laps de temps comme des notices à part entière, mais comme des attributs de la notice. La relation entre une entité et un Nomen ou un Laps de temps sera directement exprimée dans les entités qu'elles décrivent, l'étiquette ou code de sous-zone qui les contient pouvant expliciter leur utilisation. Cela a l'avantage de coller d'aussi près que possible à la structure existante et d'éviter la prolifération d'entités.

Les sous-classes spécifiques de « Res » devront être créées localement pour exprimer des entités spécifiques, telles les agents fictifs, les lieux, les concepts, les événements, les événements culturels, les concepts Dewey et les valeurs contrôlées.

Le modèle de données qui en résulte est résumé dans la figure 1.

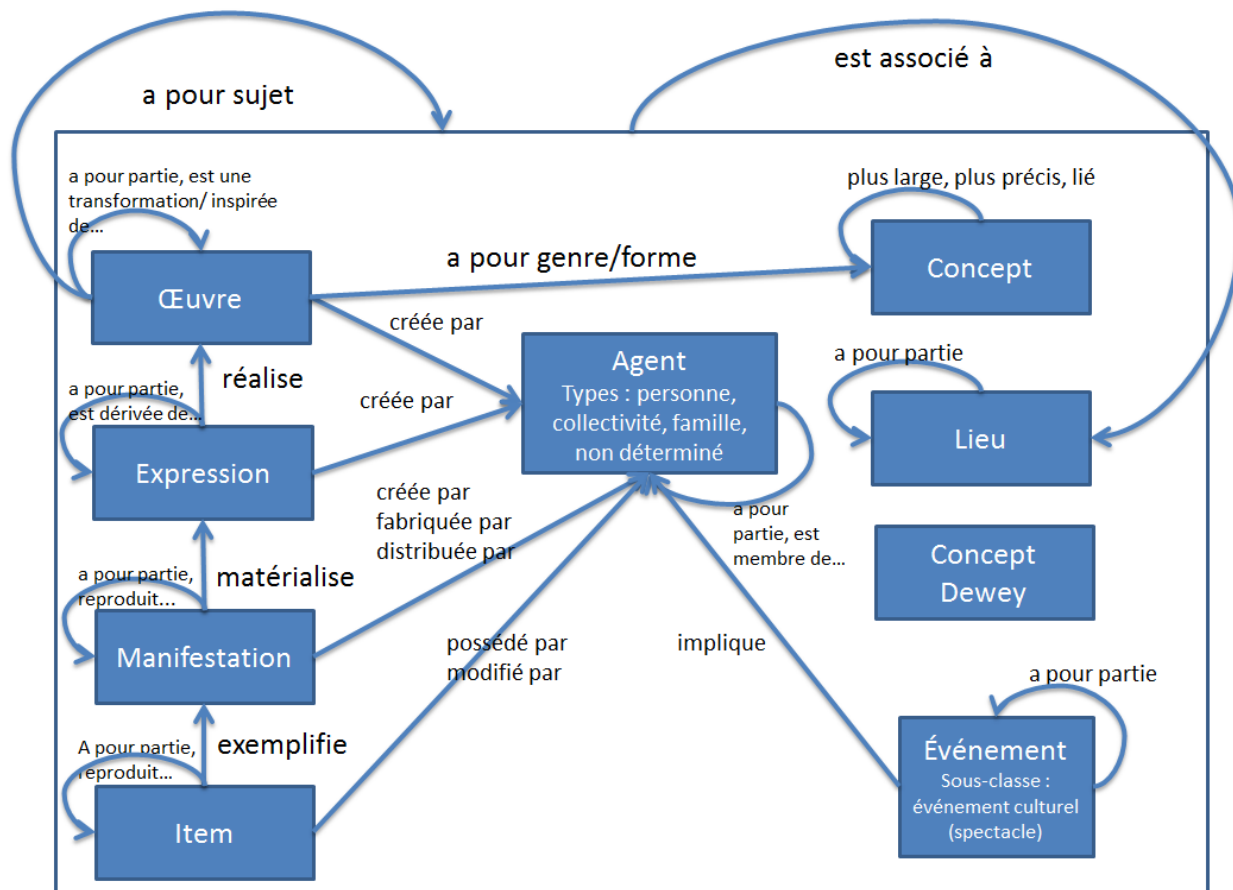


Figure 1— Modèle de données simplifié en INTERMARC-NG : une notice par entité et une zone MARC par relation.

Nous avons par ailleurs étudié la question de l'indexation matière, à un moment où RAMEAU, le langage français d'indexation matière, est en train d'être refondu. Le défi principal est d'exprimer des relations qui ne sont pertinentes que dans le contexte d'une œuvre donnée : la France n'a pas en soi de relation consubstantielle avec la cuisine (☺), ni avec le XIX^e siècle, mais cette relation (entre un concept, un lieu et un laps de temps) devient vraie dans le contexte d'une Œuvre sur la cuisine française au XIX^e siècle. Le rôle des sous-zones MARC dans la zone de l'indexation matière ne sera plus d'exprimer la structure des vedettes matière mais évoluera pour exprimer la façon dont le sujet d'une Œuvre se compose d'entités regroupées ou liées entre elles.

3.2. Méta-métadonnées

Depuis 2014, les données bibliographiques de la BnF sont ouvertes et librement réutilisables¹⁴. Les données de la BnF sont donc potentiellement intégrables à des jeux de données externes, du moment que la source (BnF) est explicitement citée. Cependant, la BnF utilise aussi ses propres sources pour vérifier des informations (dates et lieux de naissance,

¹⁴ La licence « Etalab » est une licence ouverte française équivalent à peu près au CC-BY. Voir <http://www.etalab.gouv.fr/licence-ouverte-open-licence>.

noms privilégiés et variantes, etc.). Dans l'idéal, ces informations aussi devraient être fournies afin d'évaluer la fiabilité de tel ou tel fait dans nos données ; quelqu'un pourrait alors citer explicitement cette source dans son propre jeu de données. Par exemple, dans Wikidata la langue du compositeur Georges Brassens a été importée depuis le champ d'information correspondant dans les données de la BnF. La BnF est citée comme source : or ce qu'il faudrait, c'est fournir la source utilisée réellement par la BnF (à savoir le *Who's Who, 20th Century*) afin d'établir ce fait précis.

Pour les données en entrée, cela signifie qu'une notice, mais aussi un champ d'information, devrait avoir sa propre source. Dans la structure existante d'INTERMARC, c'est déjà le cas pour la première (sources pour la notice d'autorité), mais pas pour le second. On a donc décidé que quatre éléments de méta-métadonnées seront créés au niveau de l'étiquette MARC :

- Source de la mention de manifestation ;
- Lien à une source interne (si la BnF possède la Manifestation correspondante) ;
- Lien à une source externe via son URL ;
- Pour une ressource Web externe, date à laquelle l'URL a été consultée.

Pourtant, la plupart du temps, une source particulière peut fournir des informations pour toutes les étiquettes de la notice. Afin d'éviter la répétition, il a été décidé d'avoir deux niveaux d'information pour la source en catalogage courant : une information générique au niveau de la notice, et une information spécifique au niveau de l'étiquette. En d'autres termes, une étiquette sans mention de source spécifique a pour source les mentions de source génériques de la notice. Cela fait un bon compromis entre un catalogage efficace et une information granulaire, quand il y en a besoin. À l'export, chaque champ d'information MARC sera fourni avec des sous-zones de source ; en l'absence de celles-ci, les informations de source seront déduites de celles au niveau de la notice.

Les attributs de LRM-E9 Nomen (c'est-à-dire, un point d'accès donné ou une chaîne de caractères donnée) sont également définis comme des éléments de méta-métadonnées. Il a donc été décidé qu'ils devront avoir leur propres sous-zones d'information, telles que : schéma d'encodage (pour les points d'accès autorisés), public destinataire, contexte d'utilisation, langue, écriture et système de translittération. Puisque de tels éléments de méta-métadonnées sont pratiquement applicables à n'importe quelle zone d'information MARC, chaque élément de méta-métadonnée doit avoir un code de sous-zone générique qui sera toujours le même dans quelque zone que ce soit. Dans ce but, nous avons poussé la syntaxe MARC un pas plus loin en utilisant les lettres capitales comme codes de sous-zones réservés aux éléments de méta-métadonnées.

D'autres éléments de méta-métadonnées sont nécessaires pour évaluer la qualité des métadonnées et prendre des décisions, tels que :

- Historique de la notice : corrections et enrichissements successifs ;
- Origine de la zone d'information ;
- Licence pour les métadonnées.

Ces informations n'ont pas été jugées pertinentes pour la structure MARC, car elles ne sont pas utiles dans un processus de catalogage courant, mais seulement pour le catalogage rétrospectif (l'historique de la notice d'information, l'origine de telle étiquette d'information, etc.) ou pour des réutilisateurs (informations sur la licence). De telles informations sont

toujours vitales, mais seront exprimées dans un autre format orienté événements qui reste à définir.

3.3. Référentiels

Les données INTERMARC devront être compatibles avec les objectifs du Web sémantique, c'est-à-dire un Web dans lequel « les machines deviennent capables d'analyser toutes les données du Web »¹⁵. Cela signifie que chaque élément de donnée qui n'est pas un libellé devra être façonné dans un vocabulaire contrôlé composé de valeurs individuelles toutes exprimées selon les mêmes mécanismes. Cela rencontre un besoin interne de permettre d'associer à chaque valeur contrôlée les attributs suivants, autant que de besoin :

- Formes privilégiées et variantes ;
- Liens à des valeurs plus précises, plus larges, ou en relation ;
- Notes d'historique et d'utilisation ;
- Code, dans le cas de valeurs codées (par exemple, les codes de langue ISO) ;
- Le jeu de valeurs (référentiel) auquel elle appartient.

Ces zones d'information peuvent déjà être exprimées en INTERMARC, notamment pour les vedettes matière RAMEAU. Il a été décidé d'essayer de réutiliser ces mécanismes pour les appliquer aux valeurs contrôlées.

De nombreuses sous-zones et zones de valeur fixe utilisent déjà des valeurs contrôlées dans le format INTERMARC, mais ces valeurs se contentent d'associer un code à un libellé. En INTERMARC-NG, il sera donc nécessaire de :

- considérer chaque valeur contrôlée comme une entité distincte, avec sa propre notice (cf. principe 1) et son propre identifiant ;
- réutiliser les mécanismes existants pour RAMEAU afin d'exprimer les différents champs d'information sur une valeur contrôlée donnée.

3.4. Les impacts sur la syntaxe globale du format

Les principes énoncés plus haut et le mécanisme générique entrent en conflit avec certaines spécificités de la structure MARC, qu'il a fallu abolir :

- Les méta-métadonnées ne peuvent pas être exprimées dans des zones de longueur fixe dans la structure INTERMARC, à moins d'utiliser des mécanismes internes de liens très complexes pour lesquels MARC n'a pas été conçu. Il en va de même pour les indicateurs, pour les mêmes raisons. Elles pourront toutefois fort bien être exprimées dans des zones non fixes, grâce à des sous-zones dédiées aux méta-métadonnées.
- Les valeurs contrôlées ne peuvent être exprimées qu'à travers des codes, et non des identifiants, des zones de longueur fixe et des indicateurs (les deux derniers étant par ailleurs limités dans le nombre de valeurs possibles). Elles peuvent être exprimées dans des sous-zones, en stockant l'identifiant de la valeur plutôt que le code existant dans la sous-zone.

Pour toutes ces raisons, il a été décidé **d'abandonner les indicateurs et les zones de longueur fixe**, et de les remplacer par des sous-zones standard à l'intérieur d'une étiquette MARC, qui stockera dans la sous-zone concernée l'identifiant de la notice de la valeur contrôlée. Cela évitera de devoir définir trois mécanismes différents dans le format pour gérer

¹⁵ Berners-Lee, Tim, *Weaving the Web: the Original design and Ultimate Destiny of the World Wide Web*, HarperBusiness: 2000, p. 157.

les valeurs contrôlées et les méta-métadonnées. La syntaxe du format qui en résultera s'en trouvera plus homogène.

CONCLUSION

En conclusion, ce que Tennant a fourni il y a seize ans est une analyse très convaincante d'un contexte local et d'un format MARC particulier, mais cette analyse s'avère moins pertinente quand on l'applique à d'autres formats lisibles par machine. Les auteurs de cet article ne sauraient être plus d'accord avec toutes les prémices de Tennant, mais plutôt que de répondre à son invitation prématurée à un enterrement, à la BnF on a plutôt choisi d'explorer les possibilités de ressusciter MARC d'entre les morts.

La définition d'un INTERMARC Nouvelle Génération s'est avérée un défi en termes d'organisation, mais nous avons bientôt pris conscience que de façon générale, les experts en approuvent les principes directeurs. Nous en sommes toujours aux premières étapes du développement du nouveau format, ce qui explique que jusqu'à présent nous n'ayons pu que délimiter le périmètre de notre INTERMARC-NG. Ce que donneront la rédaction par les spécialistes du format, son interprétation par notre département informatique, et son assimilation par notre « poignée » de catalogueurs, est une toute autre histoire.

Remerciements

Les auteurs de cet article souhaitent remercier Françoise Leresche, dont la connaissance approfondie de l'INTERMARC, de l'ISO 2709, et tout ce qui de près ou de loin touche aux formats bibliographiques s'est avérée inestimable. Nous souhaitons également adresser de chaleureux remerciements à tous les Campeurs INTERMARC pour leur enthousiasme, leur dévouement et leur expertise.

Références

BERNERS-LEE, Tim. *Weaving the Web: the Original design and Ultimate Destiny of the World Wide Web*, HarperBusiness, 2000.

LE PAPE, Philippe. « La vie quotidienne d'UNIMARC au temps de la Transition bibliographique », 2017. Disponible en ligne : http://www.bnf.fr/documents/jsyd2017_le-pape.pdf

TENNANT, Roy. « MARC must die », in *Library Journal*, 2002. Disponible en ligne : <https://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/>

Site Internet de la BnF :

http://www.bnf.fr/fr/professionnels/f_intermarc/s.format_intermarc_histoire.html?first_Art=non