

L'avenir de l'UNIMARC : MARC est mort, vive le MARC !

En 2002 paraissait un article célèbre intitulé « MARC Must Die »¹, qui annonçait la fin des formats MARC développés dans les années 1970 pour faciliter l'échange de données bibliographiques, et presque universellement utilisés dans le monde des bibliothèques dans les décennies qui ont suivi. Centrés sur la ressource, et associés à des formats d'encodage informatiques vieillissants (ISO 2709), ils sont parfois présentés comme inadaptés à l'exposition des données de façon conforme aux besoins du web et à l'adoption des nouveaux modèles entité-relation tels qu'IFLA LRM.

20 ans plus tard, le MARC est pourtant toujours vivant. Défini par l'arrêté du 3 novembre 1993 comme format d'échange des données bibliographiques pour le réseau du ministère de la Culture, l'UNIMARC notamment continue de s'adapter aux nouveaux besoins des catalogues des bibliothèques. Seul format maintenu et diffusé par l'IFLA (*International Federation of Libraries Association*), il évolue grâce aux travaux du *Permanent UNIMARC Committee* (PUC), qui a entamé une réflexion sur sa stratégie à la suite du congrès de l'IFLA en août 2023.

Un peu d'histoire : adapter UNIMARC aux Principes internationaux de catalogage

En 2009, l'IFLA publie les Principes internationaux de catalogage (PIC), qui définissent les nouvelles recommandations pour le catalogage à l'heure du web de données et remplacent les Principes de Paris de 1961. Afin de pouvoir être mis en pratique, ces nouveaux principes doivent être traduits dans un nouveau modèle conceptuel de type entité-relation, et une adaptation des codes de catalogage et formats bibliographiques et informatiques existants utilisés par les bibliothèques.

Pour l'IFLA, qui assure la maintenance et la mise à disposition pérenne et gratuite à l'échelle internationale d'un ensemble cohérent de normes, depuis les principes (PIC), jusqu'au format (UNIMARC), en passant par un modèle conceptuel (IFLA LRM, pour *Library Reference Model*) et une norme de contenu (ISBD), un travail d'évolution et d'adaptation de cet ensemble de normes commence alors.

¹ Roy Tennant, « MARC must die [archive] », sur pbworks.com, 15 octobre 2002 (consulté le 19 mars 2024), <http://soiscompsfall2007.pbworks.com/f/marc+must+die.pdf>

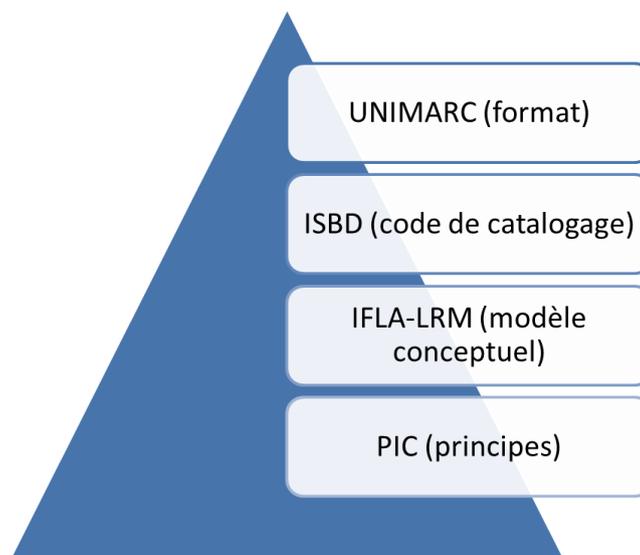


Figure 1 : La suite de normes de catalogage maintenues par l'IFLA

Les modèles FRBR (1997), puis FRAD (2009) et FRSAD (2010) sont développés puis intégrés dans IFLA LRM (2017). Ces évolutions sont intégrées dans les ISBD avec la création d'ISBD (M) pour les manifestations² à l'été 2023, et c'est dans ce sens que l'UNIMARC évolue également depuis 2012. Lors du congrès de l'IFLA d'août 2023, les différentes sections concernées (*Cataloguing, Bibliography, Subject Analysis and Access Sections*) et le Comité des normes ont réaffirmé leur volonté de développer et de maintenir cette suite de normes de façon cohérente et harmonisée.

Quelle stratégie pour l'UNIMARC ?

Un format, deux options de catalogage

La compatibilité d'un format de catalogage avec un modèle conceptuel entité-relation peut prendre plusieurs formes. Une première option consiste à permettre un catalogage par entités, c'est-à-dire la création d'une notice pour chaque entité (Œuvre, Expression, Manifestation, Item...). Une autre option consiste à intégrer directement à la structuration traditionnelle des catalogues de bibliothèque (notices bibliographiques et éventuellement d'autorité) tous les éléments d'information requis par le modèle conceptuel.

Un catalogage et une gestion directe par entités dans un SGB ont l'avantage de permettre une meilleure gestion des relations entre les entités n'ayant pas d'équivalent dans un modèle traditionnel – en particulier l'Expression. Il permet également un gain de temps pour les catalogueurs, qui peuvent mutualiser et harmoniser les informations du niveau Expression et Œuvre, alors que ces données doivent être répétées en catalogage traditionnel. Pour une exposition et une diffusion des informations du catalogue sous forme d'entités, il n'est dès lors pas nécessaire de gérer d'opérations complexes de conversion.

Le maintien d'une structuration traditionnelle (notices bibliographiques et éventuellement d'autorité) évite en revanche d'avoir à refondre ou remplacer son système de gestion de bibliothèques (SGB) et de transformer ses données existantes. Cette seconde option, qui est celle

² <https://www.ifla.org/g/isbd-rg/isbd-for-manifestation-task-force/>

qui a été retenue par le format MARC21, présente certaines contraintes dans l'expression des relations entre entités du modèle conceptuel, mais permet néanmoins une réelle compatibilité.

Pour s'adapter à IFLA LRM, l'UNIMARC a fait un choix audacieux : conserver un seul format de catalogage (UNIMARC), mais définir deux « profils » : un pour les catalogues gérant des entités (UNIMARC dit « ER » pour entités-relations) et un pour les catalogues gérant des notices bibliographiques et éventuellement d'autorité (UNIMARC dit « traditionnel »). Concrètement, cela signifie que certaines zones et sous-zones du format UNIMARC ne sont utilisables qu'en catalogage « ER », et d'autres uniquement en catalogage « traditionnel » ; d'autres encore peuvent être utilisées quel que soit le type de catalogage adopté. Chacun de ces deux modes de catalogage peut être compatible avec le modèle conceptuel IFLA LRM, sous réserve que tous les éléments d'information requis par le modèle peuvent y être exprimés.

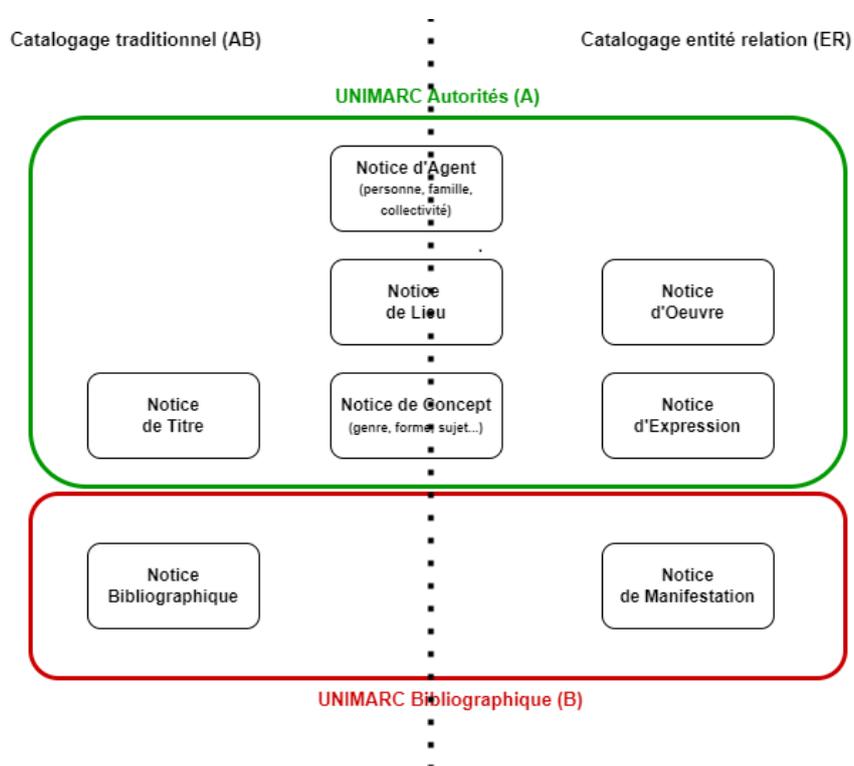


Figure 2 : Les deux profils d'UNIMARC

D'UNIMARC en ISO 2709 à UNIMARC en RDF

Le format UNIMARC (*UNiversal MACHine-Readable Cataloguing*), créé en 1977 à partir du format MARC (devenu MARC21 en 1999), segmente l'information bibliographique en champs de données. Ce découpage intellectuel de l'information bibliographique a traditionnellement été utilisé conjointement au format informatique d'échange de données ISO 2709, créé et maintenu par l'Organisation internationale de normalisation (ISO). C'est ce format d'échange qui a dicté l'association des éléments d'information composant une notice à des zones, des sous-zones et/ou des positions spécifiques. L'UNIMARC n'y est toutefois pas intrinsèquement lié. Il est par exemple tout à fait possible de diffuser de l'UNIMARC dans un encodage XML depuis 2008 grâce au format d'encodage MarcXchange (ISO 25577).

Le format intellectuel UNIMARC et le format d'encodage choisi pour le diffuser sont donc deux choses distinctes : s'il n'est pas exagéré de dire qu'ISO 2709 est un format vieillissant, qui répond de moins en moins aux besoins des utilisateurs, cela n'implique en rien une péremption du découpage intellectuel de l'information proposé par l'UNIMARC. MarcXchange a proposé une première modernisation des outils liés en permettant la mise en place de services comme les SRU (*Search and Retrieve by URL*).

Le PUC travaille depuis deux ans en collaboration avec le *Linked Data Technical Subcommittee* (LIDATEC) de l'IFLA à la création d'un espace de noms permettant d'attribuer des URI (*Uniform Resource Identifiers*) à chacun des éléments d'information définis par le format UNIMARC³. Le format B (bibliographique) est déjà disponible⁴, et le format A (autorités) le sera courant 2024. La réalisation de ce projet permettra l'exposition en RDF des données structurées en UNIMARC, ce qui est la clé de données ouvertes, trouvables et réutilisables dans le web. Un travail complémentaire a également été effectué en 2022 pour une meilleure gestion des URI au sein du format.

Où en sont les travaux du *Permanent Unimarc Committee* ?

Suite à l'aboutissement des travaux sur l'ISBD pour les manifestations, ISBD (M), le PUC entreprend en 2024 un travail d'alignement afin de compléter le format UNIMARC B. Par ailleurs, un groupe de travail spécifique dédié à une réflexion de fond sur l'avenir du format UNIMARC a été créé dans le prolongement du congrès 2023 de l'IFLA. Son objectif est de mener une enquête auprès des utilisateurs d'UNIMARC sur leurs besoins et de produire une déclaration officielle qui sera publiée par l'IFLA à l'été 2024. Ce document précisant la stratégie d'évolution d'UNIMARC et son calendrier permettra aux utilisateurs et aux éditeurs de SGB d'avoir une meilleure visibilité sur l'avenir et de prévoir leur propre stratégie en conséquence. Le groupe mettra en ligne et maintiendra également une page web au sein du site de l'IFLA où seront rassemblés la documentation, les actualités et un formulaire de contact. Une réunion des utilisateurs d'UNIMARC sera également organisée à l'automne 2024, à Maribor (Slovénie) et à distance.

BIBFRAME, une alternative ?

BIBFRAME est à la fois un modèle conceptuel (comme IFLA LRM) et un format (comme UNIMARC ou MARC21). Il a été créé en 2012 par la Bibliothèque du Congrès (*Library of Congress*) pour répondre aux mêmes besoins qu'IFLA LRM : améliorer la visibilité des données bibliographiques dans le contexte du web. Là où l'UNIMARC ER arrive à peine à maturité, le format BIBFRAME est utilisable depuis plusieurs années et a donc déjà fait l'objet d'implémentations, notamment par des éditeurs majeurs de SGB ayant une clientèle internationale. C'est pourquoi on trouve aujourd'hui des systèmes implémentant du BIBFRAME, mais pas encore de l'UNIMARC ER.

Contrairement à IFLA LRM et à l'UNIMARC, dont la gouvernance est assurée à l'IFLA sur la base de comités internationaux comme le PUC auxquels la France participe, BIBFRAME est maintenu par la Bibliothèque du Congrès, bibliothèque nationale des États-Unis. Il est donc plus difficile pour un utilisateur européen de peser dans les décisions concernant le format, et notamment dans son

³ Un URI permet d'identifier une ressource de manière permanente. Un espace de noms (*namespace* en anglais) permet de définir un contexte particulier d'attribution d'URI : au sein de ce contexte (qui fonctionne comme un préfixe commun à tout identifiant), chaque URI peut être employé pour parler d'une ressource sans que cela préjuge de son emplacement ou de la manière de la référencer.

⁴ <https://www.iflstandards.info/unimarc>

évolution. Cela peut être essentiel dans des domaines aussi importants que les données d'accessibilité, pour lesquelles les législations américaine et européenne diffèrent significativement. Les groupes d'utilisateurs européens qui se sont formés (*BIBFRAME Workshop in Europe*, utilisateurs des services de Casalini Libri...) se concentrent pour l'heure sur des questions d'implémentation technique plutôt que sur des enjeux de contenu.

En termes d'échange de données, l'UNIMARC est aujourd'hui le format prépondérant en France. La plus grande part des données mises à la disposition des bibliothèques françaises, et notamment celles de la Bibliographie nationale, sont en UNIMARC. Pour les établissements qui choisissent d'autres formats et souhaitent récupérer ces données, une conversion est nécessaire. De ce point de vue, les différences de choix de modélisation entre BIBFRAME et IFLA LRM (et donc UNIMARC ER), qui sont souvent mises en avant, ne sont pas sans importance. Un modèle conceptuel sous-tend la structuration des systèmes d'information qui vont permettre de créer, de gérer et de diffuser des informations bibliographiques, et conditionne donc dans une certaine mesure leur compatibilité. Les conversions de données entre BIBFRAME et tout format compatible avec IFLA LRM risquent d'être ainsi plus complexes d'entraîner des pertes d'informations, qu'entre formats compatibles IFLA LRM. Le modèle BIBFRAME tend cependant de plus en plus à converger avec IFLA LRM (création de l'entité « Hub », conformité à RDA...), ce qui devrait contribuer à atténuer cet effet.

BIBFRAME est donc complémentaire d'UNIMARC ER plutôt que concurrent, tout comme l'UNIMARC « traditionnel » est complémentaire de MARC21 et des autres formats MARC : c'est pourquoi les contraintes et les bénéfices liés à chacune des deux solutions sont à évaluer par les établissements selon leurs impératifs et leurs besoins propres.

UNIMARC et les enjeux français de la Transition bibliographique

Vers UNIMARC pour le catalogage par entités

Le statut de l'UNIMARC, confirmé comme format d'échange pour le réseau de la Culture lors du Comité Stratégique Bibliographique (CSB) d'octobre 2023⁵, ainsi que la volonté, matérialisée dans le projet national Transition bibliographique, d'adopter IFLA LRM à l'échelle nationale, font que la France est un contributeur majeur aux évolutions du format UNIMARC afin de permettre sa compatibilité avec le modèle IFLA LRM.

Par le biais du Comité français UNIMARC (CfU), des propositions d'évolution ont été soumises régulièrement depuis 2012 afin de rendre le format UNIMARC compatible avec IFLA LRM, ce qui rend désormais possible le catalogage des entités en UNIMARC. Afin d'en permettre les premières manipulations concrètes, le programme Transition bibliographique a produit en 2022 un jeu de données en UNIMARC ER qui a fait l'objet d'une présentation lors de la journée d'étude organisée par le groupe Systèmes & Données le 25 novembre 2022⁶. Les ajustements restant à mener concernent l'alignement de l'UNIMARC avec l'ISBD (M) publié cet été et le repérage des compléments à apporter afin de le rendre pleinement compatible avec RDA et, pour la France, avec RDA-FR. Sur ce second point, le CfU a préparé une liste de propositions portant sur les entités Agent (Personne, Famille, Collectivité) et Lieux qui seront soumises au PUC au printemps 2024.

⁵ <https://www.transition-bibliographique.fr/enjeux/comite-strategique-bibliographique/>

⁶ <https://www.transition-bibliographique.fr/systemes-et-donnees/7e-journee-professionnelle-metadonnees-en-bibliotheques-du-25-novembre-2022/>

... mais aussi un UNIMARC « traditionnel » enrichi ?

Dans un premier temps, les propositions françaises ont porté sur des évolutions propres au catalogage par entités (points d'accès, attributs et relations des Œuvres et Expressions...) ou applicables aux deux options de catalogage (nouvelle zone de l'adresse, forme du contenu, type de médiation et de support...). La question se pose à présent de savoir si l'UNIMARC souhaite également développer spécifiquement son option « traditionnelle » en y incluant tous les éléments d'information pertinents pour sa compatibilité avec LRM et RDA, sur le modèle de ce qui est fait par MARC21.

Il est en effet probable qu'un certain nombre de bibliothèques françaises ne seront pas en mesure d'adopter, à court ou à moyen terme, un format entités-relations. Le développement de l'UNIMARC « traditionnel » constitue donc une opportunité, pour les utilisateurs de l'UNIMARC dont les systèmes ne pourraient pas évoluer vers la gestion directe d'entités, de bénéficier d'une compatibilité permettant l'échange de données avec les formats entité-relation tels que l'UNIMARC ER ou BIBFRAME, et l'exposition de ces données sous forme d'entités. Ainsi pourraient-ils bénéficier eux aussi de l'enrichissement des données consécutif de la mise en œuvre du catalogage par entités par des fournisseurs de données tels que la Bibliothèque nationale de France.

Septembre 2023