

Format d'échange des données locales

Où en sommes-nous ?

par Danièle Duclos-Faure¹

Bibliothèque nationale de France

Chef du service de la coordination bibliographique

En 1994, deux articles signés Dominique Lahary et Françoise Leresche dans ces mêmes colonnes faisaient le point sur la normalisation en matière de format d'échange des données locales au niveau national. Tous deux faisaient état de l'avancement d'une réflexion théorique commune à partir des besoins des bibliothèques de lecture publique d'une part, de la BNF et des bibliothèques universitaires d'autre part. Ils mettaient en relief des orientations divergentes : les données locales incluses dans la notice bibliographique d'une part, les données locales contenues dans une notice distincte de la notice bibliographique appelée notice d'exemplaire, d'autre part.

Depuis, l'ABF, l'ADBDP et la FULBI ont mis au point ensemble la *Recommandation 995* sur la fourniture de données locales dans les échanges de notices bibliographiques accompagnant le prêt ou le dépôt d'exemplaire entre bibliothèques, tandis que débutait la mise en chantier des deux projets de catalogues collectifs : le CCF et le catalogue collectif des bibliothèques universitaires (SU), réactivant la réflexion sur le format d'échange des données locales. Simultanément, des nouveautés technologiques, comme la norme Z39.50 qui permet de consulter des bases de données hétéro-

gènes dans une architecture client-serveur, incitaient, en leur état, à réexaminer les bases du format précédemment défini par la BNF et la DISTNB, notamment le choix d'une notice d'exemplaire distincte de la notice bibliographique.

De quoi brouiller singulièrement la vision de tous ceux qui attendent un texte stabilisé sans participer de près aux travaux. Rappelons cependant qu'un format d'échange, quel qu'il soit, peut ne pas être identique aux formats internes des différents systèmes. Voir dans l'encadré, l'exposé de la maîtrise d'ouvrage du système d'information de la BNF (SI) ; les données locales, gérées séparément des notices bibliographiques sur le SI, seront exportées dans les champs UNIMARC de la notice bibliographique.

Aujourd'hui, un groupe de travail associant la BNF, le CCF (Catalogue collectif de France), la DLL (Direction du livre et de la lecture), la DISTNB (Direction de l'information scientifique, des technologies nouvelles et des bibliothèques du ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche) et l'ABES (Agence bibliographique de l'enseignement supérieur) mis en place en septembre 1996, a pour objectif d'établir un format d'échange de données locales qui prendrait en compte toutes les données locales nécessaires aux fonctionnalités du projet de chaque partenaire, les impératifs de la normalisation internationale et ceux des nouveaux

outils technologiques. Il s'appuie sur les données recensées dans le format d'échange mis au point de 1990 à 1994 par la BN et la DPDU et prend comme base la structure du format d'échange international UNIMARC. Les travaux ne sont pas achevés à ce jour on peut raisonnablement espérer qu'une définition commune d'un format d'échange des données locales incluant celles-ci dans la notice bibliographique sera établie en 1997.

On ne trouvera donc pas ci-après de document définitif, mais plutôt un état des lieux de ce dossier complexe, un arrêt sur image des travaux en cours.

1. Avec la participation de Lise Bessette, BNF-Mission pour le CCF, Marie-Claude Maddaloni, DLL, Marie-Thérèse Martini, ABES et Françoise Pellé, DISTNB.

Le format de données locales du Système d'information de la BNF

par Laurent Guillo

Bibliothèque nationale de France

Maîtrise d'ouvrage du système d'information

existantes auprès de diverses sources bibliographiques, parmi lesquelles la Bibliothèque nationale de France. Par ailleurs, le SU est l'un des trois grands catalogues, avec celui de la BNF et la base de données issue de la conversion rétrospective des bibliothèques municipales, auxquels le Catalogue collectif de France donnera accès, tout au moins dans un premier temps. On le voit, ceci entraîne plusieurs types d'échanges de données : entre le SU et les bibliothèques participantes, entre le SU et les « sources », entre le SU et le CCF. Quelles données faudra-t-il échanger ? Des notices bibliographiques cela va de soi, et il existe pour cela un format bibliographique, mais aussi des données locales. En effet, les bibliothèques participant au SU devront recevoir non seulement les données bibliographiques correspondant à leur activité de catalogage, mais aussi les données correspondant à leurs exemplaires (cote, états de collections des périodiques par exemple). Inversement, le SU, s'il veut remplir pleinement sa fonction de catalogue collectif de localisation et de prêt entre bibliothèques, doit mettre à jour, à partir d'informations fournies par les systèmes locaux, les données concernant les exemplaires possédés par les bibliothèques. Par ailleurs, le CCF doit également être en mesure de fournir ce type d'information à ses utilisateurs. Il doit donc les recevoir des catalogues auxquels il donne accès, et les traiter avant de pouvoir les afficher. Dans la mesure où ces informations locales devront être échangées, on le voit, entre de multiples acteurs, il va de soi que l'intérêt de tous consiste à en définir une structuration commune, qui n'existe aujourd'hui qu'à l'état potentiel dans le format UNIMARC. »

• Lise Bessette, Mission pour le CCF, nous rappelle les objectifs généraux du CCF : « Réaliser un catalogue collectif informatisé français en facilitant l'accès aux outils existants par la mise en place d'un guichet unique ouvert aux usagers français et étrangers, quel que soit leur statut ou leur localisation géographique. Le CCF est donc un catalogue virtuel donnant accès au catalogue de la Bibliothèque nationale de France (SI-BNF), au catalogue collectif des bibliothèques universitaires (SU) et aux catalogues (fonds anciens, locaux et particuliers) des bibliothèques municipales ayant bénéficié d'une conversion rétrospective réunis au sein d'une même base gérée par le CCF (BMR). Cet accès par une

La Bibliothèque nationale de France a finalisé le format de données locales qu'elle va utiliser dans son système d'information. Elle avait déjà finalisé son format INTER-MARC intégré pour la saisie des notices bibliographiques et des notices d'autorité, et devait dès lors s'atteler à la définition des exemplaires et des « unités de conservation ». Le format est basé sur la conception de « l'épine dorsale » du système élaborée avec le maître d'œuvre de la réalisation du système, Cap Gémini. Il a été enrichi et précisé par un groupe de travail composé de bibliothécaires ou d'ingénieurs représentatifs des divers départements et des divers systèmes actuellement exploités par l'établissement.

Ce format a les principales caractéristiques suivantes : les champs sont définis hors de tout format MARC (les données locales ne seront pas stockées dans le système dans des zones MARC mais dans des champs classiques d'un système de gestion). Ce point doit être signalé puisque, parmi les divers progiciels disponibles sur le marché, certains gèrent les données d'exemplaire dans des zones et des sous-zones MARC. Le format gère bien sûr la notion d'exemplaire, qui regroupe des aspects plutôt intellectuels comme l'appartenance, la provenance, le mode d'entrée, les cotes et les droits d'usage ; mais il gère aussi la notion d'unité de conservation (les UC), plus nouvelle dans les bibliothèques : cette notion correspond aux unités documentaires qui sont à la fois identifiées, stockées et manipulées comme une seule entité dans la bibliothèque (les exemples les plus classiques sont un volume imprimé, un coffret de CD, une boîte de diapositives). Les UC regroupent les aspects essentiellement physiques : code à barres, dimension et poids, mode de rangement, mode de communication, présence en magasin, récolement, reliure, etc. Elles sont essentiellement manipulées par les systèmes de conservation et de magasinage, alors que les notices d'autorité, bibliographiques et les exemplaires sont essentiellement manipulés par le catalogue et le système de catalogage.

Le format gère aussi une relation relativement complexe entre les exemplaires et les UC, puisqu'à un exemplaire peut correspondre plusieurs UC (cas des périodiques reliés par exemple) et plusieurs exemplaires peuvent être rassemblés dans la même UC (cas de monographies reliées ensemble, par exemple), enfin le format s'est adapté aux besoins du fonds de documents numérisés : une « UC numérique » a été définie, qui correspond en fait à une plage sur un serveur informatique. Elle contient un grand nombre de champs communs avec l'UC. Plus bas que l'UC numérique, le système permet de définir des pages numériques, qui correspondent à ce qu'un lecteur verra sur le PLS (poste de lecture simple) et le PLAO (poste de lecture assistée par ordinateur), les systèmes de consultation du fonds numérisé. Ce format tient compte des divers systèmes de gestion mis en place dans le système d'information de la BNF (catalogage, communication, conservation, magasinage et récolement, etc.) et reprend aussi certaines des particularités qui avaient été introduites dans les systèmes de gestion actuellement en exploitation. Comme il résulte de la première tentative de la BNF de « connecter » les exemplaires du catalogue aux UC gérées par l'actuel système Sycomore (communication et récolement), il doit permettre de gérer des situations problématiques qui risquent d'arriver après la migration des données des bases actuelles de la BNF vers le futur catalogue, comme l'existence d'un exemplaire sans UC ou l'existence d'une UC sans exemplaire. Il prévoit également de pouvoir rechercher une UC par sa cote même si le catalogue n'est pas disponible.

De fait, ce format est relativement lourd : il contient plus d'une centaine de champs significatifs, certains, comme les cotes, étant eux-mêmes divisés en plusieurs sections, et d'autres étant redondés entre l'exemplaire et l'UC. Compte tenu de son ambition et des impacts de l'architecture du système, il restera probablement spécifique à l'établissement, sans prétendre devenir un standard des bibliothèques. Les données locales ainsi gérées pourront être exportées en même temps que les notices bibliographiques auxquelles elles sont attachées ; il est en effet prévu que les données qui seront définies par le groupe de travail sur les données locales soient converties dans des zones du bloc 9XX des notices UNIMARC diffusées par la BNF.

interface unique permet de s'affranchir des différences d'architecture technique, de structure et de format des données. De plus, dans un souci de favoriser pour ses utilisateurs la continuité entre la localisation et la consultation des documents, le CCF met en œuvre un réseau de prêt entre bibliothèques (PIB) pour les bibliothèques de lecture publique. Il met également en place une interface entre ce réseau et le réseau de prêt entre bibliothèques universitaires (PEB). »

• Marie-Claude Maddaloni, DLL, souligne le besoin d'une normalisation au niveau national : « Par une série de recommandations : les *Recommandations pour le catalogage des livres anciens au format UNIMARC* (1993), la *Recommandation 995* (1995), les associations représentant les bibliothèques territoriales et la Direction du livre et de la lecture avaient anticipé sur l'objectif, aujourd'hui clairement défini par tous les acteurs gestionnaires de données bibliographiques, de normaliser un format de données locales. En effet, le rôle des bibliothèques départementales reposant sur le dépôt d'exemplaires, accompagnés de notices bibliographiques, dans les petites bibliothèques, ajouté au nombre croissant de bibliothèques municipales informatisées exigeaient des solutions provisoires mais rapides aux problèmes liés aux échanges. Il est aujourd'hui préférable que ces recommandations prennent le chemin de la normalisation reconnue par toutes les institutions. »

Pour l'échange de données, le CCF a fait le choix d'inclure les données locales dans les notices bibliographiques.

Quels éléments ont déterminé la position du CCF ? Pour Lise Bessette : « Les choix techniques retenus par le CCF pour permettre un accès quasi universel et une interrogation facile et conviviale sont de deux types : un accès par Internet et les outils qui y sont associés (interfaces du World Wide Web, protocole HTTP), et une interface unique pour l'interrogation des différents catalogues par l'utilisation de la norme Z39.50-1995. Ce type d'architecture est dénommé « client Z39.50 Gateway » (c'est-à-dire une passerelle permettant le passage du protocole HTTP vers le protocole Z39.50). La norme Z39.50 permet d'interroger des bases de données distribuées et hétérogènes dans une architecture client-serveur qui permet à l'utilisateur sur son poste client de s'affranchir de la diversité des syntaxes et des

structures propres à chaque base de données ou catalogue, ce que l'on nomme également l'interopérabilité de systèmes ouverts. Cette norme, particulièrement riche, nécessite cependant pour être pleinement opératoire une implémentation rigoureuse, ce qui impose un certain nombre de contraintes notamment en matière d'échange des informations du serveur vers le client. Or, dans l'état actuel de mise en œuvre de la norme, que ce soit au niveau international ou national (le SI-BNF, le SU et le CCF seront les pionniers en France dans l'application de la norme version 1995), les syntaxes d'enregistrement ne sont définies que pour l'échange de données bibliographiques. Pour pouvoir échanger séparément des données locales, dans ce que l'on appellerait une notice de données locales, un jeu d'attributs propre à ce type d'information devrait d'abord être défini et approuvé au niveau international. Dans les délais impartis à la mise en place de cette norme dans le déroulement du projet CCF, il est apparu difficile d'attendre que ce travail de réflexion soit mené. Le CCF privilégie donc l'échange des données locales au sein de la notice bibliographique dans des zones spécifiques qu'il appartient aux différents partenaires de définir ensemble. »

Le format d'échange des données locales précédemment défini en 1994 ne convenait donc pas à ces nouveaux impératifs techniques. Il fallait examiner au niveau français l'organisation des données locales dans les champs UNIMARC de la notice bibliographique.

Le groupe de travail a d'abord procédé au recensement des données locales nécessaires aux échanges entre les différents partenaires, à leur fonctionnalité, puis il a travaillé à l'organisation des données dans le format bibliographique UNIMARC.

Ces données locales doivent permettre de répondre aux fonctions suivantes : sélectionner des documents en fonction de différents critères de localisation ou de particularités d'exemplaires, localiser un document d'une manière univoque, connaître les conditions de sa communication, établir une demande de prêt et gérer celle-ci, garantir les échanges entre catalogue collectif et bibliothèques participantes.

Les données locales nécessaires à ces objectifs sont les suivantes : localisation du

document (identifiant de la bibliothèque, fonds, cote salle, dépôt permanent, etc.), état de collection (pour les documents en plusieurs unités matérielles, notamment les publications en série), informations bibliographiques propres à l'exemplaire (ex-libris et autres mentions de provenance, annotations, reliure, lacunes, etc.) et points d'accès liés à celles-ci (nom de l'annotateur, du relieur, etc.), le cas échéant, informations bibliographiques sur le document de substitution (microfilmage, numérisation, etc.), informations sur les conditions de communication ou de reproduction, informations de gestion internes à la bibliothèque (numéro d'exemplaire, numéro d'inventaire, code à barres, etc.). Bien évidemment, selon les cas, ces données sont obligatoires ou facultatives et le groupe travaille aussi à affiner ces notions.

L'état d'avancement des travaux

Le groupe de travail a convenu, à ce jour, d'un certain nombre de principes pour la gestion des données locales échangées.

Les données concernant les informations de *localisation*, de *gestion* des exemplaires seront décrites au sein de zones 9XX du format UNIMARC bibliographique. En revanche, les *données descriptives* spécifiques aux exemplaires (zones de notes et points d'accès notamment) seront gérées dans les zones *ad hoc* de la notice bibliographique (3XX, 7XX), il convient donc de pouvoir identifier clairement dans ces zones, à quel exemplaire précis la donnée appartient.

Pour ce faire, il est indispensable de définir une sous-zone qui permette d'identifier cet exemplaire de façon non ambiguë et d'apparier les données qui appartiennent au même exemplaire.

Le \$5 pourrait être adopté pour contenir l'identifiant de l'exemplaire. Le PUC, Permanent UNIMARC Committee (IFLA), a déjà validé l'utilisation de cette sous-zone pour des besoins assez comparables dans la description des particularités d'exemplaire pour les livres anciens².

2. Voir « Older Monographic Publications (Antiquarian) », *Guidelines*, 1996, n° 3.

Le \$5 sera composé de deux éléments : l'identifiant de l'établissement et l'identifiant de l'exemplaire. Le RBCCN qui identifie un établissement de façon non ambiguë a été retenu. Quant à l'identifiant de l'exemplaire, sa structure, son libellé ne seront pas définis par le groupe de travail et chaque partenaire aura donc la liberté d'en définir la composition.

Le \$5 figurera au début de chaque zone qui comporte des informations concernant un exemplaire précis et, dans une zone donnée, il ne sera pas répétable. Le \$5 s'appliquera donc à plusieurs zones du format UNIMARC bibliographique : les zones 9XX qui décrivent les données de localisation, de gestion des exemplaires, les zones 70X : 700, 701, 702 pour les points d'accès concernant un exemplaire

en particulier, les zones 3XX pour les notes qui contiennent des informations spécifiques à un exemplaire et la zone 012 contenant l'empreinte. Les zones 9XX ou 3XX ou 7XX contenant le même \$5 constitueront des groupes de zones qu'il sera facile d'afficher lors de l'interrogation des bases bibliographiques.

Il a été aussi convenu que, dans le cas particulier d'un point d'accès identique pour des exemplaires différents, la zone 7XX contenant un \$5 identifiant l'exemplaire concerné soit répétée, plutôt que de fournir une seule zone 7XX contenant plusieurs \$5.

Par ailleurs, des principes ont été retenus pour la définition des zones 9XX décrivant des données de localisation, de gestion : à même catégorie de données

(données de localisation ou données de gestion), même zone ou groupe de zones utilisé ; à même type de document concerné (monographies ou publications en série), même zone ou groupe de zones 9XX utilisé.

Bien des points sont encore en cours de discussion, comme le traitement de référence de l'exemplaire, il pourrait être celui de l'unité physique : c'est-à-dire qu'une zone 9XX donnée ne concerne qu'une seule unité physique mais il suppose que tous les partenaires puissent gérer très finement les données.

Au demeurant, on peut estimer que les bases d'un format, au sens strict, d'échange ont d'ores et déjà été établies par le groupe de travail.