

Pancatalogue un exemple de produit en réseau

*par Anne-Marie Motais de Narbonne
Bureau de modernisation des bibliothèques
du Ministère de l'Education Nationale*

P

ancatalogue est une base de données bibliographiques produite par le Ministère de l'Education Nationale, de la Jeunesse et des Sports, dont les finalités sont l'identification et la localisation des documents acquis et conservés dans les universités.

En tant que catalogue collectif d'ouvrages, cette réalisation s'inscrit dans la tradition et son principe de fonctionnement reprend le schéma classique de mise en commun de notices catalographiques et des localisations associées, pour favoriser les échanges entre les bibliothèques participantes. Cependant, le travail en réseau informatique introduit des différences fondamentales, principalement dûes à l'action combinée du principe de réutilisation multiple des données initiales et des enchaînements automatiques des traitements informatiques sur ces données.

La participation des bibliothèques universitaires à Pancatalogue suppose en effet une adhésion à un réseau de

catalogage partagé, et chaque action des bibliothèques sur ces bases - création, correction ou dérivation de notices - entraîne une double livraison des données correspondantes. Une livraison est destinée à la bibliothèque pour la mise à jour de son catalogue local, l'autre à Pancatalogue, cela grâce à des accords contractuels entre les bases catalographiques, les bibliothèques universitaires et la Sous-direction des bibliothèques, gestionnaire du catalogue collectif.

Produits sources et produits résultats

Le choix par les bibliothèques universitaires de leur base catalographique est ouvert entre les bases OCLC, BN-OPALE et SIBIL. Il repose sur la recherche de la meilleure adéquation entre le contenu de la base et les grandes tendances des acquisitions, de façon à trouver le meilleur ratio entre les créations et les dérivations de données. Ces bases de catalogage partagé apparaissent donc comme des

réalisations communes au service d'un gain de productivité pour chaque bibliothèque participante, et en même temps elles constituent les moyens d'une réalisation nouvelle : un catalogue collectif.

Si, par ailleurs, on situe Pancatalogue dans l'organisation générale du réseau des bibliothèques universitaires, ce catalogue collectif apparaît à son tour comme le moyen, l'outil intermédiaire, d'autres réalisations ou d'autres services. De façon très classique, ce catalogue collectif des ouvrages est le complément logique du catalogue collectif des publications en série CCN et de Téléthèses. Il s'articule aussi de façon évidente avec PEB, le système informatique du prêt entre les bibliothèques qui sera lui-même prochainement lié, grâce au projet ION, aux systèmes équivalents de la Grande-Bretagne et des Pays-Bas. Mais en outre, grâce à l'actualité de ses données et à la richesse de ses index, un catalogue collectif informatisé des ouvrages pourra être aussi l'outil nécessaire

d'une évaluation nationale et locale des collections et l'instrument indispensable d'une politique concertée en matière d'acquisition et de conservation.

Dans ce schéma d'ensemble, Pancatalogue apparaît comme un aboutissement du catalogage en réseau, comme l'instrument du prêt entre bibliothèques, mais aussi comme un des outils d'aide à la décision tant pour les universités elles-mêmes que pour une politique nationale de la documentation universitaire. Il est ainsi exemplaire du travail en réseau informatique qui conduit à des produits en cascade où les réalisations ne sont plus seulement des fins mais deviennent à leur tour autant de moyens utilisés pour d'autres objectifs.

La circulation des données

Produit alimenté par des données en provenance de bases sources, Pancatalogue doit réorganiser ces données en fonction de ses propres finalités, qu'il s'agisse des données de localisation ou des données bibliographiques.

- Les données locales

Les données de localisation sont créées par les bibliothèques sur les bases catalographiques à chaque utilisation de notice et sont transmises telles quelles à Pancatalogue. Dans le projet initial, Pancatalogue devait être aussi une source de données pour les catalogues locaux et c'est pourquoi il était prévu que Pancatalogue traite la totalité des informations relatives à l'exemplaire. L'application du principe de séparation claire entre les fonctions des bases bibliographiques, des catalogues collectifs et des bases locales a conduit à mettre sur les sys-

tèmes locaux le circuit qui organise la mise à jour des informations relatives à l'exemplaire. Pancatalogue présentera donc les informations en fonction de son rôle national : signallement de la présence (ou par défaut de l'absence) d'au moins un exemplaire dans telle ou telle bibliothèque.

La question plus spécifique de l'identifiant de la bibliothèque se pose également de façon différente selon les systèmes. En effet, en tant que produit-source du prêt entre bibliothèques, Pancatalogue doit assurer à l'usager les facilités nécessaires pour l'accès au document primaire. Cela suppose notamment que soit établie et mise à jour une correspondance entre, d'une part, les données de localisation du document telles qu'elles sont reçues des bases catalographiques et, d'autre part, l'adresse à laquelle pourront être obtenues les données locales complémentaires : cote, disponibilité, délais... C'est dire que la notion de bibliothèque ou plutôt d'unité de catalogage (bibliothèque, section, unités associées...) déclarée sur les bases catalographiques doit être compatible avec celles du catalogue collectif, celles du répertoire des bibliothèques et celles du système PEB.

- Les données bibliographiques

Les données bibliographiques sont transmises dans leur globalité par les bases sources, et les formats de présentation permettent de distinguer les informations descriptives et celles qui servent d'accès à cette description. Si l'intégration des données descriptives relève d'une opération relativement simple de conversion de format, la création d'index nécessite des traitements plus sophistiqués. Elle est toujours marquée par les caractéristiques internes du système informatique qui porte l'application

et par les finalités du produit. Ainsi, dans le système DOBIS-LIBIS, c'est-à-dire dans Pancatalogue, tous les titres sont cumulés dans un même index, titres des documents, titres des collections, titres originaux et il en est de même pour tous les noms de personnes physiques et morales auxquels une notion quelconque de responsabilité auteur est associée. De plus, pour chaque index, tous les mots non vides sont autant de points d'accès.

Passer d'un produit bibliographique à un autre, d'un système informatique à l'autre, implique donc de pouvoir faire toutes sortes de "variations" sur les données, que ce soit sur les informations bibliographiques ou sur les informations de localisation.

Méthode et phases de réalisation

Produit résultant d'une action en cascade, Pancatalogue doit intégrer des données de types divers parce que fabriquées dans trois organisations qui diffèrent sur bien des points : principes de fonctionnement, dimensions, systèmes informatiques utilisés, choix des documents traités, formats d'entrée, de traitement, de sortie...

Distinction faite entre données bibliographiques et données de localisation, l'objectif de cette intégration est, par des procédures automatiques, de regrouper toutes les localisations d'un même document et de les associer à sa description bibliographique. En théorie cela suppose que soit réalisée la reconnaissance d'un document à travers des descriptions bibliographiques différentes.

En pratique cela conduit à trois types d'actions. Il faut fusionner toutes les données d'indexation quelles que

soient les sources et quelles que soient les notices auxquelles elles sont rattachées, il faut aussi fusionner les notices se rapportant à un même document et enfin, associer toutes les localisations à la notice retenue.

La première phase de réalisation comprenait l'établissement du cahier des charges, le choix du système informatique de base, DOBIS-LIBIS, le développement de logiciels complémentaires tant pour l'application en général que pour l'intégration des données provenant de l'OCLC. Cette première phase est achevée et la base en cours de chargement avec quelques 200 000 notices OCLC.

La deuxième phase, qui commence actuellement, comprendra le changement de version de DOBIS-LIBIS avec la reprise de toutes les adaptations spécifiques et l'intégration des données provenant de BN-OPALE et de SIBIL. Cette intégration comportera la fusion des index, mais seulement la juxtaposition des notices. La phase 2 se terminera sur une étude préalable des phases ultérieures et principalement une étude de faisabilité de la fusion des notices.

La faisabilité de Pancatalogue

L'intégration en très grand nombre dans une base unique de ces données différentes présente de nombreuses difficultés et la faisabilité de l'opération repose sur trois facteurs principaux : les normes, la stabilité des données sources et les fichiers d'autorités

- Les normes

Les premiers résultats mettent en évidence la solidité et la précision de l'appareil normatif élaboré depuis des décennies par les bibliothèques.

Qu'il s'agisse, en effet, des règles de catalogage définies avec une précision telle que les différences nationales ne soient pas un obstacle insurmontable, ou des formats de codification qui, parce qu'ils sont riches et stables, permettent une identification précise de ces mêmes données, qu'il s'agisse encore de la structure des bandes d'échange bibliographique, toutes ces normes sont les piliers d'un travail en réseau.

- Les données

Cependant, les premiers résultats montrent aussi les risques d'écarts entre la pratique et la théorie, entre les usages et la règle et les conséquences de ces écarts sur un produit-résultat tel que Pancatalogue.

Le principe même de fonctionnement en réseau, la réutilisation des données, suppose que les données initiales soient parfaitement conformes à leur définition théorique.

Les traitements informatiques ultérieurs sont fondés sur ce postulat. Ainsi la création de données (le catalogage original) doit-elle être assortie de précautions et de contrôles destinés à assurer dès ce stade la correction immédiate de la plus grande partie des erreurs.

Les écarts résiduels entre la définition théorique des données et leur réalité concrète sont dus alors à la part d'interprétation personnelle ou d'inattention des catalogueurs et présentent, en conséquence, un caractère aléatoire qu'aucun traitement systématique ne peut complètement compenser.

Ce type de correction relève donc de procédures dites manuelles qui, malgré leur nom, exigent des connaissances théoriques solides, beaucoup de jugement, beaucoup de temps et donc aussi beaucoup d'argent.

Dans le schéma fonctionnel de Pancatalogue, la production courante de données homogènes et conformes à leur définition est garantie par les organisations de catalogue partagé qui forment les sources du catalogue collectif.

- Les fichiers d'autorités

Les fichiers d'autorités, parce qu'ils sont à la fois norme et contenu, constituent l'appareil idéal du travail en réseau.

En étant le système de référence commun à la création des données initiales et à tous les produits dérivés, ils garantissent non seulement la cohérence interne de chaque produit mais aussi la solidité de l'ensemble du réseau.

Le développement des fichiers d'autorités, auteurs, titres uniformes, matières, leur usage en voie de généralisation dans toutes les bases catalogographiques, l'utilisation pour la maintenance de Pancatalogue de ces mêmes fichiers d'autorités constituent un ensemble de facteurs essentiels à l'intégration des systèmes.

Conclusion

A l'augmentation et à la diversification des activités d'un réseau correspondent une sophistication et une imbrication accrues des systèmes informatiques. De par cette complexité, les interventions humaines sur les données, passé le stade de la création, sont des activités à haut risque pour l'équilibre du système et la cohérence des produits. Il convient donc de prendre toutes les dispositions organisationnelles pour en diminuer le nombre et pour encadrer leur mise en oeuvre. C'est pourquoi une première conséquence du travail en réseau pourrait bien être un renouveau du métier de catalogueur.