

Nouveautés et perspectives des données diffusées par la BnF

Panorama des enjeux des données exogènes proposées par la BnF

(Anila Angjeli et Pauline Moirez)

En matière de données, ce qu'on appelle la « qualité BnF » repose sur la Bibliographie nationale française et sur les notices d'autorité, dont on sait qu'elles sont structurées et pérennes. Mais en plus de celles qu'elle produit, la BnF collecte également des données exogènes : comment et dans quel but ? Quelles sont leurs provenances ? Qu'apportent-elles ?

Ne seront envisagées ici que les données du catalogue général¹, même si la dynamique d'intégration de données extérieures est globale et concerne aussi BnF Archives et Manuscrits (BAM), Gallica, data.bnf.fr...

Pourquoi ?

Selon une logique collaborative, la BnF souhaite se positionner en tant que tiers de confiance entre les producteurs de données et les utilisateurs : elle ne se contente donc pas de travailler à partir de ses propres données. Cette démarche lui permet de s'appuyer sur des expertises variées : elle mutualise des données de qualité produites par ses partenaires, ce qui enrichit le service *in fine*, par exemple grâce à l'ajout de résumés que la BnF n'a pas les moyens de rédiger elle-même.

Dans cette perspective, un cadre réglementaire et légal a été établi : un travail en amont est mené avant la diffusion des données. Il repose sur des négociations et un dialogue avec le partenaire.

Selon le principe de traçabilité, qu'elle a placé au cœur de cette politique, la BnF s'engage à mentionner la source des données et elle invite tous les réutilisateurs – professionnels comme usagers des bibliothèques – à en faire autant.

D'où viennent ces données exogènes ?

1. Données de réservoirs internationaux, fournisseurs internationaux et autres partenaires internationaux

Depuis 2010-2011, la BnF récupère des données de WorldCat² afin de faciliter la description de ses documents, notamment issus du domaine étranger (les documents du dépôt légal ne sont pas concernés) : cela permet de mutualiser le travail de signalement. Elle importe ces notices dans leur intégralité et conserve leur partie descriptive, mais retouche partiellement certaines zones (correspondant à l'indexation, aux liens avec les autorités...). Les notices sont alors intégrées dans les services et les produits proposés : selon la convention conclue entre la BnF et OCLC, la source apparaît clairement (numéro OCLC, numéro de la notice source de WorldCat, date de récupération...). Ainsi, l'utilisateur devra conserver à la fois le numéro FRBNF et les sources OCLC-WorldCat.

¹ <http://catalogue.bnf.fr/>

² <https://www.worldcat.org/>

Les notices d'autorité Personnes et Collectivités du registre des ISNI sont, quant à elles, intégrées dans le catalogue général de la BnF depuis 2013, et diffusées via les produits et services depuis 2014. L'identifiant ISNI est également présent dans les notices bibliographiques associées à chaque autorité ainsi importée³.

En ce qui concerne l'ISSN, l'essentiel vient du centre français, mais l'ISSN-L (qui permet de faire le lien entre les différentes versions – imprimée, numérique – d'une même publication) est généré par le centre international⁴.

2. Données de l'édition

Depuis 2009, la BnF permet la déclaration en ligne des livres et des périodiques. L'extranet du dépôt légal⁵ propose deux modes de saisie : à l'unité (saisie à partir de formulaires) ou en masse (par flux ONIX, depuis 2013). Les données qui parviennent ainsi comprennent des enrichissements qui sont conservés sans retouche (résumés, couvertures...).

Depuis 2013, cette dématérialisation concerne également les partitions, les cartes & plans, les estampes...

Concrètement, la pré-notice issue de l'extranet du dépôt légal apparaît dans le catalogue général de la BnF, ce qui permet à la fois la vérification du dépôt légal imprimeur et la récupération du signalement (uniquement via Z39.50). Lorsque le processus de catalogage est achevé, la notice est récupérable via les produits et services, et le site Nouveautés éditeurs⁶ renvoie vers elle.

Le projet de dépôt légal des livres numériques, actuellement en cours de développement, prévoit l'entrée directe des métadonnées au format ONIX en même temps que les fichiers numériques qu'elles accompagneront. Des informations descriptives seront injectées automatiquement dans le catalogue général. Certaines informations seront reprises et mises aux normes de la BnF, tandis que d'autres seront conservées (résumés, liens avec le livre papier, images de couverture, données techniques...). La situation sera différente en fonction de l'existence ou non d'une version imprimée pour laquelle une notice aura déjà été faite. Les modalités de redistribution de ces données ne sont pas encore fixées.

Il est à noter que des données d'Electre ont été intégrées ponctuellement en 2014 lors de l'import du catalogue du CNLJ⁷ dans le catalogue général de la BnF. Ces données sont rediffusées avec la mention de source. Comme dans le cas des couvertures provenant de l'extranet du dépôt légal, la BnF projette en 2017 de diffuser les couvertures d'Electre via ses produits et services.

3. Données des bibliothèques partenaires

L'offre « Gallica marque blanche » permet aux partenaires ne disposant pas de bibliothèque numérique

³ Remarque : il importe en effet de placer l'ISNI non seulement dans la notice d'autorité, mais aussi dans la notice bibliographique parce qu'il y a des bibliothèques qui ne gèrent pas les autorités et/ou qui gèrent mal les liens : elles ne récupèrent par exemple que les notices bibliographiques et pas les notices d'autorité.

⁴ <http://www.issn.org/understanding-the-issn/assignment-rules/the-issn-l-for-publications-on-multiple-media/>

⁵ <http://depotlegal.bnf.fr/>

⁶ <http://nouveautes-editeurs.bnf.fr/>

⁷ Centre national de la littérature pour la jeunesse

de mettre en ligne leur propre bibliothèque, à leurs couleurs, tout en profitant de l'infrastructure de Gallica. En retour, la collection nationale numérique disponible sur Gallica s'enrichit de documents locaux, complémentaires des collections de la BnF.

La convention conclue avec le partenaire stipule que la BnF est propriétaire des documents numériques ainsi diffusés. Ils sont signalés dans son catalogue général : après un travail en amont, leurs métadonnées y entrent soit via un export UNIMARC à partir du catalogue de la bibliothèque partenaire, soit à partir de bases de données spécifiques créées autour de la bibliothèque numérique en marque blanche. Là encore, la source est mentionnée dans la notice du catalogue général de la BnF.

L'exigence de qualité ne pèse pas, alors, autant que lorsqu'il s'agit des documents du dépôt légal.

Conclusion : un cercle vertueux de circulation des métadonnées

Lors de la journée professionnelle « Systèmes et données » organisée le 9 novembre dernier à la BnF dans le cadre de la transition bibliographique, David Aymonin évoquait la création d'un fichier national d'entités : cette entreprise confirme l'enjeu stratégique de la coproduction des données aujourd'hui.

Dans cette logique d'échange et de circulation, la BnF se pose en tiers de confiance et se veut entraînant, pour impulser une dynamique susceptible de permettre le positionnement et l'implication de chacun, en dépit de contextes politiques ou économiques variés. Elle a placé ses données sous la licence ouverte préconisée par la mission Etalab : elles sont donc librement réutilisables sous condition de mention de leur source, laquelle peut être multiple.

Pour en savoir plus

- ISNI : http://www.bnf.fr/fr/professionnels/isni_informer.html
- Dépôt légal : http://www.bnf.fr/fr/professionnels/depot_legal.html
- Protocole d'échange des données : http://www.bnf.fr/fr/professionnels/protocoles_echange_donnees.html
- Schéma numérique de la BnF : http://www.bnf.fr/documents/bnf_schema_numerique.pdf