

Politique de qualité des données de la BnF : recommandations

Contenu

Politique de qualité des données de la BnF : recommandations	1
1. Définir un référentiel qualité des données produites par la BnF.....	2
2. Reprendre la politique de catalogage pour y inclure les autres catalogues jusqu'ici hors périmètre.....	3
3. Mettre en place un plan d'action pour l'exhaustivité du signalement des collections de la BnF...	4
4. Repenser l'offre des services fondés sur la diffusion des métadonnées de la BnF	5

Ces recommandations sont issues des travaux du groupe de définition de la politique de qualité des données de la BnF qui s'est réuni au premier semestre 2018. Elles ont vocation à compléter le document de politique de qualité des données, en proposant des axes concrets de mise en œuvre.

L'éventuelle mise en œuvre de ces recommandations doit nécessairement être concrétisée par un calendrier programmatique sur les années à venir. Ce calendrier devrait pouvoir s'appuyer sur les grandes dates fixées par les étapes de validation du Comité des métadonnées, telles qu'elles ont été fixées lors de la session du 22 juin 2018 :

- juin 2019 : adaptation de la politique qualité aux différentes filières de production de la BnF (mise en place du référentiel qualité)
- novembre 2019 : optimisation des processus de catalogage (reprise de la politique de catalogage)

Le Contrat d'objectifs et de performance 2017-2021 de la BnF prévoit par ailleurs une action spécifique pour instruire l'enjeu de « Tendre à l'exhaustivité du signalement », préconisation qui a donné lieu à la constitution en 2018 d'un groupe de travail spécifique qui s'est appuyé sur les premières conclusions formulées en la matière par le groupe de travail Définition de la politique de qualité des données.

1. Définir un référentiel qualité des données produites par la BnF

La qualité des données produites par la BnF ne saurait se résumer à la qualification des filières de traitement *via* le binôme *Statut de la notice* et *Niveau de validation du catalogueur*, mais elle doit être exprimée au niveau de chaque entité et fondée sur les métadonnées que ces notices portent, leur nature et leur mode de création, de correction et de vérification.

Le but de ce référentiel est de constituer un outil d'évaluation non de la qualité du travail des agents, mais bien de la qualité des données produites par la BnF dans une multitude de contextes et de temporalités.

Une logique de "labellisation BnF" de données de qualité pourrait facilement en découler, en marquant les valeurs de ce référentiel correspondant au niveau de qualité qui souhaite être mis en exergue par l'établissement.

Il semble particulièrement urgent de se doter d'un tel référentiel car il permettrait de définir plus précisément et de manière factuelle les objectifs qualitatifs de chaque filière de production, courante ou rétrospective, en matière de création ou de reprise des données traitées. Il s'agit d'un enjeu d'importance eu égard aux évolutions actuelles des contextes de production : massification des données – et notamment données externes - enjeu de l'exhaustivité du signalement rétrospectif des collections, mise en œuvre des filières de documents dématérialisés (DLSD, DLLN...), accroissement du Dépôt légal...

Le groupe de travail a ébauché les contours et les possibles modalités de mise en œuvre d'un tel référentiel qui doit permettre avant tout de retranscrire le caractère évolutif de la qualité des données, afin de mieux répondre aux attentes diversifiées de nos usagers, en garantissant autant que faire se peut la « fraîcheur » des données mises à dispositions, tout en documentant, le cas échéant, les reprises postérieures de ces données.

Le référentiel qualité de la BnF doit être **fondé sur une pluralité de critères** permettant de rendre la diversité des éléments traités tout en restant synthétique, structuré et documenté. La qualité des données documentaires doit être **exprimée au niveau de chaque notice** (et non plus de chaque filière), en se fondant sur des éléments objectifs portées par les métadonnées et les méta-métadonnées de ces notices.

Pour que ce système soit opérationnel, il importe qu'il soit **automatisé sur la base d'un algorithme** qui pourrait à chaque nouvel enregistrement de la notice recalculer le niveau de qualité de la notice, afin de ne pas introduire de notions subjectives d'évaluation de la qualité de son propre travail, et dans un souci également d'économie du catalogue.

La portabilité de cette information se ferait *a priori* **dans le cadre de la nouvelle application Noemi**, dans une zone à créer de l'Intermarc NG, structurée en sous-zones correspondant chacune à un élément.

Eu égard au temps imparti au groupe de travail, il n'a pas été possible de pouvoir livrer d'ores et déjà un référentiel qualité finalisé. Le groupe a néanmoins essayé de définir une cible à partir de plusieurs scénarios possibles, qui tendrait vers une **logique de datavisualisation de type diagramme** fondé sur plusieurs axes qu'il reste à définir. Un premier séquençage de la mise en œuvre a également été proposé :

- la définition des axes, des composantes de ce référentiel ;
- la recension puis l'élaboration des critères de qualité fondés sur les métadonnées et méta-métadonnées de chaque type de documents, aboutissant à un référentiel à plat (ce référentiel devrait être articulé aux axes précédemment définis) ;
- choix du mode de visualisation de données permettant de traduire le plus lisiblement et le plus simplement possible ce référentiel.

2. Reprendre la politique de signalement pour y inclure les outils jusqu'ici hors périmètre

La politique du catalogage a été publiée dans sa dernière version en 2016. Elle vise à fixer les directives appliquées pour le catalogage courant dans BnF catalogue général. Ce faisant, elle présente plusieurs lacunes notables, qui nuisent à la cohérence de traitement de la BnF :

- *aucune prise en compte de la dimension rétrospective*, que ce soit par traitement rétrospectif de données déjà existantes ou la création *ex nihilo* de notices de documents détenus par la BnF et jamais encore signalés ;
- *une politique pensée uniquement dans le cadre du catalogue général*, sans prise en compte de l'existence des autres catalogues de la BnF, sans même parler des autres environnements de production de données documentaires.

Or, le groupe juge primordial pour la BnF de se doter d'une **politique globale de signalement**, qui doit pouvoir fédérer la politique de catalogage (pratiques de description orientées bibliothéconomie), le signalement des fonds (pratiques de description orientées archivistique), et même la logique d'enrichissement documentaire, qui se concentre sur la description de l'information et non de son support (pratiques orientées documentation). Cette politique doit **embrasser les différentes filières d'entrées et types de documents traités**, et doit pouvoir « **réconcilier** » **les différents modes de traitement proposés** : schématiquement, d'une part le catalogue général, sous format MARC et normes Afnor, avec un traitement bibliothéconomique fondé sur la description « par le bas » (à la pièce, à l'unité documentaire) ; d'autre part la base Archives et manuscrits, sous format EAD et norme ISAD(G), avec un traitement archivistique fondé sur le traitement « par le haut » (le fonds documentaire, qui peut le cas échéant être décliné, suivant la pertinence, la nécessité ou les moyens, jusqu'à la pièce). Il est nécessaire qu'un nouveau document de politique du signalement, capable d'englober les politiques de description dans chaque filière ainsi que la politique du sujet, puisse définir l'articulation et la complémentarité des différentes bases bibliographiques de la BnF, ainsi que les critères scientifiques et organisationnels qui justifient le recours à un type de description plutôt qu'à l'autre.

3. Mettre en place un plan d'action pour l'exhaustivité du signalement des collections de la BnF

Dans son contrat d'objectifs et de performance 2017-2021, sous l'axe « Garantir la continuité des collections physiques et numériques et faciliter leur accessibilité », la BnF se fixe comme résultat à atteindre *Tendre à l'exhaustivité du signalement (selon une granularité adaptée à la nature des collections)*.

Plusieurs axes d'action semblent possibles en la matière, avec une nécessaire hiérarchisation des tâches, et en partant du postulat suivant qui doit être admis par l'ensemble de l'établissement : eu égard aux masses considérables de documents détenus par la BnF qui ne font l'objet d'aucun signalement, fût-il élémentaire, **l'exhaustivité du signalement de l'ensemble des collections doit primer sur l'exhaustivité du catalogage de chaque document.**

La première étape nécessaire doit être une **action de cartographie systématique des fonds non encore signalés**, visant à en évaluer globalement la nature, la quantité, les difficultés intrinsèques de signalement. Cette action est à mener dans le cadre de l'instruction du groupe de travail Signalement de la Direction des collections.

Dès lors, une politique de reprise systématique pourra être mise en place, qui puisse mettre en regard de ces fonds des moyens, des solutions de traitement pertinentes et élaborer un calendrier pluriannuel rationnel et réaliste.

En particulier, doivent à cette étape être privilégiés tous les modes de traitement favorisant un signalement rapide, économique et au plus haut « niveau d'entrée » possible. Ainsi, la description archivistique doit dans ce contexte se focaliser sur le signalement par grands fonds ; une description plus détaillée pourra être opérée dans un second temps dans des contextes de production plus spécifiques (chantiers semi-automatisés de corrections et d'enrichissement, projets de recherche, préparation d'une exposition...). De même, il faut pouvoir tenir compte des apports de la numérisation en matière d'automatisation de traitement de documents : par le biais par exemple de l'OCR ou de la reconnaissance de formes et d'images, la numérisation offre des possibilités d'indexation automatique et, plus largement, de production automatique de métadonnées, qui semblent pertinents dans bien des cas (images, dépôt légal du Web, documents très normalisés et structurés comme les cartes nautiques...)

Il est à noter que la création d'un référentiel qualité (*cf* recommandation n°1) serait un facilitateur de la mise en œuvre de cette politique d'exhaustivité de signalement des collections, en permettant d'en fixer les objectifs en matière d'exhaustivité et de richesse des données à renseigner et de donner à voir de façon factuelle ces critères aux utilisateurs des données de la BnF.

Dans le cadre de cette réflexion, un groupe de travail en charge de l'instruction de cette question a été mis en place par la Direction des Collections dès le premier semestre 2018, afin de pouvoir rendre ses propres conclusions à la fin de l'année.

4. Documenter la qualité de l'offre des services fondés sur la diffusion des métadonnées de la BnF

L'offre des services de la BnF fondés sur la diffusion de ses métadonnées s'est considérablement diversifiée ces dernières années, et il apparaît nécessaire de la repenser à l'aune des enseignements tirés des échanges avec les usagers et de l'instruction du groupe de travail. L'enjeu principal est celui de la **transparence des choix politiques et opérationnels effectués**, élément indispensable d'une politique qualité qui doit s'adresser aussi bien aux équipes de l'établissement qu'aux usagers des services qu'il propose. Trois axes d'action se dessinent :

- **L'adoption d'un plan, d'une stratégie des métadonnées**, qui fixe une vision globale et cohérente dans un calendrier pluriannuel des grands axes de développement de l'offre de services centrée sur l'utilisation des métadonnées de la BnF. Ce plan pourrait s'inspirer par exemple de la [British Library's Collection Metadata Strategy 2015-2018¹](#), même s'il semble nécessaire *a minima* d'en adapter forme et contenu aux choix stratégiques propres à la BnF.
- **La reprise, la complétude et l'harmonisation de la documentation** déjà disponible sur le site institutionnel de la BnF qui décrit les normes, standards, modes de traitement, interfaces de consultation et de diffusion de données. Ce travail peut constituer une opportunité pour harmoniser, voire simplifier les pratiques et consignes actuelles, parfois fruit de compromis organisationnels passés, de moyens de contournement d'impossibilités techniques à réinterroger aujourd'hui, etc. Surtout, il est primordial que la documentation en la matière prenne en compte l'aspect historique de la constitution des catalogues. Le périmètre de cette réflexion devrait naturellement englober non seulement le catalogage courant, mais également la masse des conversions rétrospectives. L'enjeu d'un tel travail est double : transparence pour nos utilisateurs et surtout capacité de leur fournir les moyens de mieux gérer en aval leurs chantiers de reprise des données. Une documentation synthétique de l'ensemble serait précieuse, y compris pour un usage en interne, mais d'autres dispositifs seraient peut-être à explorer (FAQ, cartographie / état des lieux par stratifications chronologiques et chantiers menés...).
- **Un questionnaire sur l'association plus étroite des usagers à ces processus** : quels objectifs, quels profils, quelles modalités ? Des éléments plus poussés de réflexion ont été avancés dans le cadre d'une réunion du groupe, qui ont permis d'aboutir aux points de synthèse suivants : il apparaît d'une part nécessaire de constituer un panel représentatif d'usagers qui aurait vocation à suivre la mise en œuvre des mesures découlant des propositions du groupe de travail, et qui pourrait ainsi être sollicité ponctuellement pour avis, propositions et préconisations ; d'autre part, il semble prometteur d'explorer les possibilités d'association des usagers aux processus de signalement et de contrôle qualité, selon plusieurs modalités envisageables : saisie de corrections voire d'enrichissement (exemple de la plateforme [Correct²](#)), identification par exemple de fonds iconographiques non légendés (exemple des [Herbonautes³](#)), mise à disposition de corpus à indexer avec une limitation des interventions sur des champs prédéfinis et selon des référentiels contraignants (exemple de [Mémoire des Hommes⁴](#)).

¹ <https://www.bl.uk/bibliographic/pdfs/british-library-collection-metadata-strategy-2015-2018.pdf>

² http://www.bnf.fr/fr/professionnels/anx_numerisation/a.projet_correct.html

³ <http://lesherbonautes.mnhn.fr/>

⁴ <http://www.memoiredeshommes.sga.defense.gouv.fr/fr/article.php?larub=52&titre=annotation-collaborative>