

data.bnf.fr

Feuille de route - 2020-2024

Data.bnf.fr est un projet initié en 2011 ; après deux marchés de 4 ans, le site va bientôt entrer dans une troisième phase de développement entre 2020 et 2024.

Les principes du site sont réaffirmés : plateforme d'expérimentation de la BnF, articulation du site autour des entités œuvres et agents, agrégateur de données internes et externes à la BnF et rediffusion de ces données vers l'extérieur.

La période qui débute permettra de réfléchir au positionnement de l'interface dans l'offre de services numériques de la BnF, de perfectionner encore l'interface en repensant notamment son aspect graphique et ergonomique - en associant les utilisateurs du site - et d'associer les différents producteurs des données de la BnF pour une meilleure prise en compte dans data.bnf.fr de la variété des types de documents conservés à la BnF. Ce marché sera aussi l'occasion de continuer le travail initié sur le traitement des données et leur exposition dans le web de données. Une attention particulière sera portée à faire de data.bnf.fr un outil d'accompagnement et d'expérimentation de la Transition bibliographique.

Cette note résume les enjeux de positionnement de data.bnf.fr, précise quelques évolutions prévues et propose une démarche et des orientations de développement pour les années à venir (2020-2024) dans le cadre d'un nouveau marché.

1. Une application installée dans le Web et utilisée par les Internauts

1. Porte d'entrée de la BnF sur le Web

Aujourd'hui, data.bnf.fr n'est plus seulement un projet mais un produit de la BnF à part entière : il fait partie de ses quatre sites les plus consultés, avec, en moyenne, **19 000 visiteurs uniques par jour**. Avec plus de 80% de visiteurs en provenance des moteurs de recherche, il joue le rôle de porte d'entrée et de capteur d'audience depuis le web vers plusieurs applications majeures de la BnF et favorise la visibilité en ligne de ses ressources numériques et analogiques.

L'analyse des logs menée dans le cadre de la chaire BibliLab (BnF et Institut Mines-Telecom) montre que data.bnf.fr joue pleinement son rôle de "pivot" entre les différents portails de la BnF, et en particulier en direction de Gallica. Lorsque la navigation conduit les internautes à basculer entre data, Gallica et le catalogue général, c'est data qui constitue, de loin, le portail le plus exploité (d'Alché-Buc, Beaudouin, Bermès *et al.*, 2019 : p. 4). Le dernier marché a aussi été l'occasion d'accentuer le positionnement de data comme porte d'entrée vers les données de la BnF, notamment grâce à la mise en place de nouvelles fonctionnalités comme "Data vous conseille" et la première partie de la refonte graphique permettant de mettre en avant les liens entre les ressources présentées dans data et leurs sources d'origine (le catalogue général, Gallica, BAM, etc.).

2. Une offre de données ouvertes et réutilisables

Le site compte en 2020 environ 3.971.872 millions de pages pointant vers 13.303.994 notices bibliographiques du catalogue général, 99.558 liens vers BnF archives et manuscrits et propose près de 974.832 liens vers la bibliothèque numérique Gallica.

Comme en témoignent les alignements et réutilisations dont il fait l'objet à l'initiative de tiers¹, cette offre a atteint une masse critique qui en fait un service de confiance reconnu pour la diffusion de données culturelles de qualité sur le web sémantique. C'est désormais une alternative à d'autres protocoles de diffusion et de dérivation des métadonnées de la BnF, notamment grâce à la mise en place de jeux de données téléchargeables (*dumps*) et d'un requêteur *ad hoc* (SPARQL Endpoint). Data.bnf.fr illustre la politique de la BnF et son positionnement fort au plan national dans le *Linked*

¹ On peut prendre comme exemple le site France Archives qui aligne une partie de ces autorités et référentiels avec ceux de la BnF par le biais de data.bnf.fr, ou l'Abes, qui télécharge régulièrement les *dumps* de Rameau.

Open Data : ses données exposées en RDF sont bien « liées » à plus d'un titre tout en étant largement « ouvertes » en application de la licence ouverte de l'Etat (Etalab) choisie par la BnF en 2014 pour l'ensemble de ses métadonnées.

3. Une machine à LRMiser

Le rôle d'outil d'ingénierie de data.bnf.fr, de retraitement et de reversement de métadonnées au service du Catalogue général de la BnF, et potentiellement au service d'autres applications a aussi été confirmé. L'exploitation de ses algorithmes contribue en effet à la transformation et à l'enrichissement automatique ou semi-automatique de données et à leur transition vers le modèle FRBR. Le service data.bnf.fr s'appuie sur l'application interne de traitement de données RobotDonnées, qui utilise les technologies, les compétences et le développement logiciel de data.bnf.fr, mais qui est désormais autonome de l'interface destinée au public et peut donc être utilisée pour d'autres applications. Parallèlement à l'enjeu de diffusion et de référencement en ligne, cette dimension d'ingénierie et de support a pris beaucoup d'importance dans le développement du projet. Grâce aux algorithmes de RobotDonnées, il a notamment été possible de créer automatiquement plus d'1 million d'œuvres supplémentaires.

4. Un outil d'expérimentation au service de toute la BnF

Data.bnf.fr se positionne enfin, dans sa démarche et sa méthodologie de projet, comme une application pilote ou exploratoire, qui anticipe, teste et répond aux évolutions des usages, des standards et des pratiques du web.

Les évolutions proposées ici entendent conforter l'ensemble de ces fonctions : là où data.bnf.fr a fait ses preuves, il est indispensable de maintenir une offre de service stable, fiable et robuste, capable de mises à jour régulières de ses contenus comme de ses fonctionnalités.

Toutefois, il ne s'agirait pas d'enfermer data.bnf.fr dans ce contour fonctionnel, au risque de le voir rapidement concurrencé par d'autres produits et de lui faire perdre son caractère innovant. Pour l'équipe projet, la dimension exploratoire, innovante et transverse de data.bnf.fr doit être préservée, ce qui implique de **continuer à développer ce service dans un esprit expérimental et agile en étant particulièrement attentif à l'évolution des usages numériques.**

1. Axes de développement 2020-2024

Il s'agit maintenant de détailler les axes de travail qui seront particulièrement approfondis pendant la période 2020-2024. Ces axes sont interdépendants.

Axe 1 : Optimiser l'existant

- Poursuite de la refonte graphique et ergonomique de l'interface
 - Documentation de data.bnf.fr
 - Rétro-ingénierie du site pour l'optimiser et en nettoyer le code
 - Analyse et révision du RDF
 - Amélioration du référencement du site sur les moteurs de recherche
 - Communication et positionnement de data
-

1. Poursuite de la refonte graphique et ergonomique de l'interface

Data.bnf.fr a été mis en ligne en 2011. Après une période d'expérimentation sur l'affichage des données, l'application est entrée dans une période de perfectionnement et de stabilisation de son interface. Le site a trouvé son public (institutions publiques, acteurs privés et grand public) et ses usages. Il s'agit désormais de perfectionner l'architecture de l'information et de la rendre pleinement intelligible par le plus grand nombre, en réfléchissant notamment à la répartition des contenus entre le site institutionnel, le portail api.bnf.fr et le site data.bnf.fr. Il s'agit aussi de développer l'interface pour qu'elle puisse accueillir toutes les entités prévues par le modèle IFLA-LRM - notamment l'intégration dans l'interface des expressions - tout en maintenant la priorité sur les pages œuvres et auteurs.

Sur la fin de la période du dernier marché, un travail important a été initié pour repenser la structure des pages et améliorer la compréhension et l'appropriation du contenu par les internautes. Ce travail doit être poursuivi pour améliorer l'ergonomie générale du site, repenser certaines sections, améliorer l'accessibilité du site et sa responsivité. La navigation pourra notamment être facilitée avec l'implémentation de filtres, de facettes et de tris, particulièrement utiles pour l'affichage de listes qui se sont considérablement allongées depuis le lancement du projet.

Ces objectifs pourront être réalisés en associant des internautes et des professionnels de la BnF (voir l'axe 2 ci-dessous).

Priorité	Haute
Applications ou projets BnF en lien	api.bnf.fr
Minimum à atteindre	Réflexion sur les pratiques actuelles et le design des sites internet et chiffrage de l'implémentation dans data
Maximum à atteindre	Passer plusieurs marchés graphisme pour "challenger" les prestataires et implémenter le résultat qui convient le plus à la BnF

2. Documentation de data.bnf.fr & optimisation du code du site

Le travail qui a été effectué sur le graphisme des pages du site a mis en avant des problèmes de stratification du code du site risquant de rendre ses évolutions structurelles plus difficiles. Pour préparer l'arrivée de NOEMI et du FNE dans le paysage applicatif de la BnF, il est important de documenter le code existant et d'en nettoyer les parties obsolètes.

Il est primordial que les équipes de data.bnf.fr disposent d'une documentation fonctionnelle permettant de connaître et d'explicitier le fonctionnement du site, son contenu précis, les règles de gestion et de sélection des documents. Cette documentation pourra être utilisée pour les futurs développement du site mais aussi pour mettre au point des services à destination des utilisateurs de data. Le but est de préparer au mieux, dès ce marché, les futurs changements opérés sur les données de la BnF.

Intermarc laissera place à Intermarc-NG (Intermarc Nouvelle Génération), un format qui permettra aux catalogueurs de l'établissement de cataloguer nativement des données conformes au modèle IFLA-LRM. La BnF produira donc des œuvres, des expressions et des manifestations nativement : data.bnf.fr serait l'application la mieux placée pour bénéficier le plus rapidement de ces changements, ce qui nécessite de travailler dès maintenant sur ce volet pour les anticiper.

Priorité	Haute
Applications ou projets BnF en lien	- NOEMI - FNE - api.bnf.fr - Lotus Notes (et son successeur)
Minimum à atteindre	Avoir un dispositif de documentation à jour et actualisable par l'équipe data
Maximum à atteindre	Avoir un dispositif de documentation à jour et actualisable par l'équipe data et correction et mise à jour du code pour gommer les strates et améliorer la performance du site

3. Analyse et révision du RDF & amélioration du référencement du site sur les moteurs de recherche

La phase d'étude de l'existant de la documentation permettra aussi de réaliser un travail nécessaire de mise à niveau du RDF de data.bnf.fr. En effet, certaines ontologies RDF ont évolué et les données utilisées pour construire l'interface du site n'ont pas toujours été intégrées au RDF. Le maintien du RDF de data est un enjeu important puisqu'il est la source des données réutilisées par différentes communautés. Il s'agit donc de le faire évoluer et de trouver des moyens de le maintenir pour qu'il soit toujours à jour et puisse répondre aux attentes des utilisateurs du site.

L'un des enjeux de la construction d'un outil comme data.bnf.fr est de permettre aux données fiables et produites par des professionnels de l'information d'être visibles. Dans cette optique, data.bnf.fr est un outil qui s'illustre notamment par son référencement sur les moteurs de recherche et qui permet d'accéder à des données qui ne peuvent être produites qu'au sein des institutions culturelles en général et de la BnF en particulier.

Une étude sur le référencement sera demandée afin d'enrichir et de développer les mécanismes existants pour conforter la place de data sur le web.

Priorité	Haute
Applications ou projets BnF en lien	api.bnf.fr
Minimum à atteindre	Mettre à jour les ontologies utilisées.
Maximum à atteindre	Améliorer le référencement de data.bnf.fr

4. Définition d'une stratégie de communication

Afin de valoriser au mieux le projet, tant au sein de la BnF qu'à l'extérieur, il importe de mettre rapidement en œuvre une stratégie de communication, et de planifier à court et moyen terme les actions qui en découleront.

La communication interne permettra notamment de fluidifier le partage d'information au sein des équipes de la BnF (quels sont les utilisateurs actuels et potentiels ? quels usages ont-ils ou pourraient-ils avoir de data ? où peuvent-ils trouver l'information pertinente sur le projet ?) Actuellement, l'information est disséminée sur plusieurs sites :

- le site data.bnf.fr lui-même,
- le dossier pivotdoc de la base de production²,
- la documentation accessible depuis la page outils SIM/SRE du portail du département des métadonnées³,
- la présentation du projet sur le site institutionnel⁴, sous forme de pages web et de vidéos,
- les informations de téléchargement des données sur le portail api.bnf.fr⁵,
- la présentation du projet sur l'intranet⁶.

Sur les trois sites accessibles au public, la répartition de l'information manque de cohérence. Il convient de définir plus précisément ce qui doit apparaître ici ou là, en accord avec les responsables des différents portails et en tirant parti des échanges qui sont préfigurés dans l'axe 4 (ci-dessous).

Par ailleurs, les évolutions du site et de ses usages (mises à jour majeures, nouvelles expérimentations, actions de médiation, réutilisations particulièrement marquantes, etc.) gagneraient en visibilité si elles faisaient l'objet de publications régulières. Parmi les pistes à envisager figurent notamment :

- la publication de brèves sur Biblionauts,
- la publication d'actualités dans les lettres d'information diffusées en interne et dans la profession,
- la publication ponctuelle sur les réseaux sociaux, avec l'accord du service Internet et réseaux sociaux,
- la publication d'articles dans les revues professionnelles,
- la publication de billets, par exemple sur le nouveau portail api.bnf.fr..

La communication externe, à destination des réutilisateurs connus et/ou potentiellement intéressés par les données de data s'appuiera sur plusieurs canaux et sur plusieurs formes de médiation. Les échanges préfigurés dans l'axe 4 y contribueront, et permettront de renforcer les relations qui se sont nouées dans le passé.

Pour ce faire, un plan de communication formalisé permettra de définir les étapes du processus de communication.

² <http://intranotes/ABN/abnprodu.nsf/0/1EE705BA7CD81D93C12576780030664A>.

³

⁴ <https://www.bnf.fr/fr/databnffr>.

⁵ <http://api.bnf.fr/dumps-de-databnffr>.

⁶ <http://biblionauts.bnf.fr/data-bnf-fr>.

Priorité	Haute
Applications ou projets BnF en lien	- site institutionnel - api.bnf.fr - Biblionauts
Minimum à atteindre	Mettre en place un dispositif de communication interne et externe autour de data
Maximum à atteindre	Impliquer des relais en interne et à l'extérieur pour valoriser data.bnf.fr

Axe 2 : Traitement et accroissement de données

-
- Intégration et alignement avec de nouvelles sources de données (BnF et externes)
 - Intégration de l'ontologie RIC-O
 - Perfectionnement des traitements de données existants
 - Développement de nouveaux traitements de données
-

1. Intégration et alignement avec de nouvelles sources de données (BnF et externes)

Data.bnf.fr a atteint une masse critique de données qui en fait un outil très utilisé et complet. Cependant, il reste encore des travaux à accomplir pour faire du site le pivot documentaire de la BnF à proprement parler. Des réflexions devront être menées pour intégrer d'autres données BnF : on pense notamment au CCFr, à Mandragore, à la Bibliographie nationale française. D'autres pistes de sources BnF à intégrer à data.bnf.fr pourront émerger lors des différents travaux en interne induits par l'axe 2 de cette feuille de route. Ce nouveau marché permettra de poursuivre le travail initié sur le perfectionnement de l'alignement des deux référentiels spécialisés sur les lieux de la BnF (le référentiel Géographique du Département des Cartes et Plans et la partie de RAMEAU consacrée aux lieux). A l'heure actuelle, dans data.bnf.fr, si un alignement n'est pas déclaré entre un lieu géographique et un lieu RAMEAU, alors chaque lieu renvoie à deux pages distinctes. Cette amélioration de l'alignement entre les deux types de lieux permet de limiter les doublons d'entités, de rendre plus intelligible pour le grand public le contenu de data.bnf.fr et de regrouper sur une même page les ressources à propos d'un même lieu (que la description du lieu vienne du référentiel géographique ou de RAMEAU). Pour la BnF, cet alignement entre les deux référentiels est important : data.bnf.fr préfigure ainsi l'émergence de l'entité lieu, en cours de définition normative dans le cadre de la Transition bibliographique.

Un autre volet de cet axe de développement sera consacré à aligner les données de la BnF avec de nouvelles sources extérieures. Ces alignements vers les référentiels extérieurs sont très importants, ils permettent à data d'être présent dans le graphe RDF, d'augmenter les possibilités de réutilisation des données et de compléter les données de la BnF. Ce nouveau chantier sera encore l'occasion d'accroître le nombre d'alignements vers des données extérieures à la BnF. Les coordonnées géographiques pourront être enrichies par alignement entre les données BnF et GEONAMES laissant ainsi envisager de nouvelles perspectives d'affichage des données.

On pense également aux données de la British Library, Europeana, les différents produits d'OCLC, DOREMUS (DOing REusable MUSical data), LOTERRE (Linked open terminology resources), les autorités lieux de France Archives, le DicoTopo de l'École nationale des Chartes, le référentiel de l'UNESCO, IAML, les données de la RMN-Grand Palais ou encore celle de Paris Musées. Il s'agira aussi de développer des mécanismes permettant de maintenir plus efficacement la fraîcheur des alignements existants et de les améliorer. Dans le cas des alignements vers les données de Wikidata, ce marché sera l'occasion de réfléchir sur les moyens de reverser aussi souvent que possible les alignements que nous calculons dans Wikidata pour qu'ils puissent bénéficier largement à la communauté.

Priorité	Haute
Applications ou projets BnF en lien	Projets de recherche BnF Wikidata
Minimum à atteindre	Suivre les projets interne BnF et aller aussi loin que possible dans l'intégration de nouvelles données produite en interne
Maximum à atteindre	Suivre les projets interne BnF et aller aussi loin que possible dans l'intégration de nouvelles données produite en interne et à l'extérieur

2. Amélioration du traitement des données de BAM

Durant ce marché, plusieurs chantiers concernent des évolutions nécessaires du modèle de données et des améliorations de la modélisation de certains types de documents, notamment les archives. L'intégration des documents en EAD provenant d'Archives et manuscrits dans data.bnf.fr est imparfaite, notamment parce que le modèle de données sur lequel le site est construit est pensé pour les ressources bibliographiques (le modèle IFLA-LRM). Or, data.bnf.fr est pensé comme un pivot documentaire de la BnF : il faut donc que son modèle de données permette d'exposer toutes les données produites par l'établissement.

L'amélioration du traitement des données provenant de BAM va suivre deux axes :

- Meilleure prise en compte et révision des règles d'interprétation de l'EAD. L'objectif est d'intégrer à data plus de données provenant de BAM et ainsi de les rendre plus visible ;
- Exploration de nouvelles ontologies pour décrire au mieux les ressources de BAM, particulièrement les archives. En décembre 2019, l'International Council on Archives (ICA) a publié l'ontologie Records in Contexts⁷ (RIC-O). L'intégration de nouvelles ontologies comme RIC-o sera indispensable, entraînant donc un travail avec les spécialistes des archives à la BnF et les collègues des Archives Nationales ayant participé à la rédaction de cette ontologie. Il s'agira aussi de faire un travail équivalent pour la description des manuscrits.

Priorité	Haute
Applications ou projet BnF en lien	BAM, BnF Catalogue général, web service sur les données de BAM
Minimum à atteindre	Pointer jusqu'à la pièce dans l'affichage des pages
Maximum à atteindre	Diffusion des données dans le SPARQL

3. Nouveaux traitements de données & expérimentations

Le traitement automatique des données est un aspect très important de data.bnf.fr : il s'agit de continuer à perfectionner et développer les algorithmes de RobotDonnées pour la création d'œuvres à partir des données des notices bibliographiques. Il s'agit aussi de continuer à créer de nouveaux traitements automatiques sur les données de la BnF : on pense notamment à l'indexation automatique, mais d'autres pistes pourront être explorées. Le positionnement de data.bnf.fr en fait naturellement un hub de traitement et de test sur les données de la BnF.

Les alignements évoqués plus haut (axe 2.1) sont très importants pour rendre les données de la BnF plus visibles mais aussi pour envisager des enrichissements de données et des calculs de nouveaux liens ou de nouvelles métadonnées. La récupération du *full text* des œuvres peut aussi permettre d'envisager des calculs de métadonnées supplémentaires, notamment la création d'indexations, de résumés, ainsi que de proposer des formes alternatives de recommandation de contenus pour les usagers du site. Les alignements sont aussi un défi technique qui mêle des règles métiers fines et des compétences informatiques exigeantes. Ce marché sera l'occasion de développer et de normaliser les outils et algorithmes existant pour améliorer la qualité des alignements et pouvoir les élargir à d'autres sources de données.

Au titre des expérimentations, de nouveaux chantiers pourront être développés comme, par exemple, la recherche d'entités nommées dans les métadonnées descriptives des notices et dans les tables des matières de Gallica pour accroître le nombre de liens entre les entités.

Il s'agit aussi de poursuivre les réflexions initiées sur la mise en place de nouveaux outils de diffusion des métadonnées comme une API JSON. Cela permettrait de toucher la communauté des développeurs informatiques et de construire de nouveaux outils de connaissance et de curation des données. La mise

⁷ Pour plus d'information voir : <https://www.ica.org/standards/RiC/ontology.html>

en place d'une API JSON permettrait d'utiliser les outils basés sur Elasticsearch, notamment Kibana (outil permettant de visualiser le contenu des données).

Priorité	Moyenne
Applications ou projet BnF en lien	Gallica Département des Cartes et Plans / RAMEAU MISAOA
Minimum à atteindre	Aligner les lieux avec GEONAMES Génération d'oeuvres films et enrichissement des oeuvres films par des données extérieures
Maximum à atteindre	Recherche d'entités nommées Expérimentations sur Elasticsearch et Kibana

Axe 3 : Accompagner la Transition bibliographique

- Réfléchir à la présentation des données dans les catalogues de demain
 - Diffusion des outils de traitement des données pensés pour data.bnf.fr
 - Modification du modèle de données pour qu'il soit conforme au modèle IFLA-LRM
 - Meilleure prise en compte du code de catalogage RDA-FR
 - Construction de liens avec NOEMI et le FNE
-

Les bibliothèques françaises sont engagées dans la Transition bibliographique depuis 2015. La BnF est à la fois engagée dans cette transition en interne, ce qui se traduit par différents projets (écriture d'un nouveau format de catalogage, NOEMI, FNE, etc.), mais également vers l'extérieur, notamment par ses activités en liens avec les autres établissements culturels (formation, rédaction des normes, etc.). La Transition bibliographique est une phase de réalisation concrète d'actions visant à modifier le modèle de catalogage utilisé par les établissements. Cette modification a des conséquences lourdes à la fois en termes de formation des catalogueurs, mais aussi en termes d'infrastructure informatique, puisqu'elle peut impliquer un changement de logiciel de catalogage et de catalogue public. Dès lors, pourquoi s'engager dans de tels changements ?

1. *data.bnf.fr : un outil de LRMisation des données*

Les raisons de ces changements sont déjà largement explicitées; il s'agit pour les bibliothèques d'être plus visibles sur le web, de mieux répondre aux attentes du public, de mieux prendre en compte les spécificités des ressources conservées dans les bibliothèques.

Dans ce contexte, data.bnf.fr est un outil qui permet à la BnF de communiquer et de montrer les évolutions et les atouts d'un système utilisant le modèle IFLA-LRM pour exposer ses données. Il s'agit aussi de développer la visualisation des données conformément au modèle IFLA-LRM pour préfigurer la forme et les contours des catalogues de demain. La navigation dans des données LRMisées n'a pas encore trouvé de modalités complètement satisfaisantes. Data.bnf.fr doit être aussi un environnement d'expérimentation pour la visualisation de données structurées suivant le modèle IFLA-LRM.

Pendant la période 2020-2024, il s'agit d'accroître ce rôle et de positionner data.bnf.fr comme un outil de visualisation des données LRMisées. Il s'agit également de développer le volet de data.bnf.fr portant sur les outils, les chantiers, les algorithmes de LRMisation pour accompagner et aider les bibliothèques tout au long des changements induits par la Transition bibliographique. Data.bnf.fr pourrait également fournir de la documentation sur les chantiers et les algorithmes de LRMisation, afin de donner une visibilité sur le processus de transformation des données.

Il s'agit de penser l'interface pour deux types de public :

- le grand public : Data.bnf.fr est une interface agrégeant les données produites par la BnF et permettant de naviguer facilement à l'intérieur de cette galaxie ;
- les professionnels des bibliothèques, chercheurs, familiers des données produites par la BnF : Data.bnf.fr est une interface permettant de visualiser, de tester, d'expérimenter et de se documenter sur des données répondant aux principes du modèle IFLA-LRM.

Priorité	Moyenne
Applications ou projet BnF en lien	Programme de la Transition bibliographique (Groupe système et données, Groupe formation) FNE NOEMI
Minimum à atteindre	Création et intégration d'expressions dans l'interface de data.bnf.fr
Maximum à atteindre	Rendre certaines données (dans les pages ateliers) complètement conformes à IFLA-LRM et RDA-FR pour utiliser l'interface de data comme interface de test et de médiation pour la Transition bibliographique

2. Rapprocher le modèle de données du modèle IFLA-LRM & mieux prendre en compte du code de catalogage RDA-FR

Engagé dans la Transition bibliographique, data.bnf.fr n'est pas seulement un outil de diffusion et d'agrégation des données de la BnF, c'est aussi un outil permettant l'appropriation des changements induits par le nouveau modèle et leur apport.

Pendant la période 2020-2024, des modifications dans l'interface de data.bnf.fr devront être réalisées, notamment l'intégration des entités expressions, l'ajout et l'amélioration des liens entre les entités.

Dans la même optique, il s'agira de suivre la publication des sections publiées de RDA-FR pour mieux prendre en compte le futur code de catalogage des bibliothèques françaises.

De manière plus générale, ce marché sera aussi l'occasion de réfléchir et d'initier un travail profond sur les liens entre data et la Transition bibliographique pour que le site puisse être utilisé comme un outil de démonstration, voire de formation.

Priorité	Moyenne
Applications ou projet BnF en lien	Groupe normalisation de la Transition bibliographique
Minimum à atteindre	Création d'expression avec RobotDonnées
Maximum à atteindre	Création d'expression avec RobotDonnées Calcul de liens supplémentaires entre les données par algorithme ou sources extérieures (liens entre œuvres)

3. Construction de liens avec NOEMI et le FNE

En fonction de la possibilité de faire converger les calendriers des projets NOEMI, FNE et data.bnf.fr, on pourrait aussi imaginer des liens étroits entre ces projets et l'utilisation de data.bnf.fr pour faire tester les modifications en cours sur les données. Par ailleurs, il serait aussi envisageable que data.bnf.fr soit utilisé pendant cette période comme une réserve d'outils, d'algorithmes (pour la création des clusters et pour les alignements) et de compétences pour aider ces projets sur les aspects relatifs au traitement des données.

Priorité	Haute
Applications ou projet BnF en lien	NOEMI FNE
Minimum à atteindre	Mettre à disposition et faire de la médiation autour des algorithmes présents dans data pour les mettre à disposition d'autres projets de la BnF
Maximum à atteindre	Intégration de données dans l'interface de data en avance de phase pour rendre le travail sur ces deux projets visibles

Axe 4 : Faire émerger de nouveaux usages et répondre aux besoins des utilisateurs internes à la BnF et des internautes

- Association des départements producteurs des données pour améliorer la structuration et l’affichage des données
- Association des personnes responsable des applications dont les données sont agrégées dans data pour améliorer l’interface et les données
- Amélioration de la prise en compte de certains types de données
- Association des utilisateurs de data.bnf.fr pour optimiser l’interface et les données
- Amélioration des outils de diffusion des données de data.bnf.fr

1. Associer les départements producteurs au processus d’amélioration

Un des objectifs déjà énoncé pour la période 2020-2024 est la finalisation de la refonte graphique et ergonomique du site. Afin d’y parvenir, l’association étroite de professionnels de la BnF et des utilisateurs du site est nécessaire.

Au sein de la BnF, la période qui s’ouvre sera l’occasion de réaffirmer le caractère de pivot documentaire du site. À l’heure actuelle, l’interface du site est principalement pensée pour les ressources imprimées et la structuration de chaque type de page est uniforme. Il s’agit maintenant de raffiner la présentation des données pour mieux rendre compte des spécificités de chaque type de document et mettre ainsi en valeur la richesse des collections de la BnF. Il s’agit de modifier la structure du contenu des pages, sans revenir sur l’organisation du site autour des pages auteurs, œuvres, etc. Pour ce faire, une attention forte sera portée à la communication, des ateliers seront organisés avec les départements producteurs des données pour recueillir les besoins et les préconisations d’évolutions.

Dans data.bnf.fr, certains pans de modélisation et d’affichage dans l’interface du site sont à améliorer — on pense particulièrement aux périodiques et aux archives et manuscrits de BAM. Revoir leur modélisation permettra d’exposer les données de manière optimale et de mieux répondre aux attentes des usagers.

Priorité	Haute
Applications ou projet BnF en lien	Groupe de suivi métier Acco{ab
Minimum à atteindre	Réactiver le groupe de suivi métier Avoir des relais dans les départements producteurs des données pour faire remonter des demandes
Maximum à atteindre	Intégrer les préconisations des départements producteurs

2. Associer les responsables des applications internes

Parallèlement à l’association des producteurs des données, les responsables des applications qui sont déjà moissonnées par data.bnf.fr seront aussi sollicités pour interroger la pertinence de la structure des données et de leur affichage actuel. L’objectif est de travailler en avance de phase et avec une certaine agilité sur des développements qui pourront s’avérer utiles pour les catalogues et sur Gallica, en particulier sur ceux relatifs à la diffusion des données sur le web, à la modélisation, à la navigation et à la visualisation dans les collections. Data.bnf.fr est ainsi un produit qui aide la BnF à réfléchir à

l'évolution de ses catalogues, à celle de sa bibliothèque numérique et, plus largement, à son positionnement dans le web culturel et scientifique. Ce qu'il découvre et teste de manière expérimentale, sur le mode du « bac à sable » ou de l'atelier en ligne peut alimenter la vision et le développement d'autres produits de l'établissement en même temps qu'une vision plus globale de l'architecture d'ensemble des services que propose la BnF. Ces développements permettront de préfigurer les travaux de refonte du catalogue général de la BnF après la mise en production de NOEMI.

Priorité	Moyenne
Applications ou projet BnF en lien	Catalogue général Gallica BnF Archives et manuscrits MISAOA
Minimum à atteindre	Atelier avec tous les fournisseurs de données de data pour identifier les besoins et interconnexions entre les projets Intégration de visualisations innovantes sur la page d'accueil de la partie atelier
Maximum à atteindre	Intégrer de nouveaux services proposés et les évolutions dans data Intégrer des visualisations innovantes dans toutes les pages ateliers

3. Association des utilisateurs de data.bnf.fr pour optimiser l'interface et les données

Les utilisateurs du site (particuliers et amateurs, chercheurs, réutilisateurs institutionnels, professionnels d'autres secteurs) seront interrogés pour mieux les connaître et être capable de cerner leurs usages et leurs besoins. L'objectif est d'enrichir le site de nouvelles fonctionnalités, d'améliorer la lisibilité de l'interface existante et d'intégrer aux développements ultérieurs les préconisations de ses réutilisateurs grâce aux échanges qui seront initiés. Il s'agira par exemple de mettre en place des focus groupes et des tests utilisateurs pour créer un lien pérenne avec les publics et s'assurer que data.bnf.fr corresponde effectivement aux besoins de sa cible.

Ces différentes périodes de recueil des besoins auprès de personnes ne faisant pas partie de l'équipe data.bnf.fr impliquent que la gestion agile du projet soit maintenue et explicitée, afin de ménager une part de souplesse et de réactivité dans les développements informatiques.

Priorité	Haute
Applications ou projet BnF en lien	BnF Data Lab, Acco{ab
Minimum à atteindre	Définir un panel de réutilisateurs et construire une relation durable
Maximum à atteindre	Animer un "datathon" sur les données de data

4. Amélioration des outils de diffusion des données de data.bnf.fr

Un volet spécifique du travail concernera la diffusion des métadonnées afin de développer des outils de médiation favorisant l'appropriation réelle et la réutilisation des données de data.bnf.fr. Data.bnf.fr fournit des outils performants de récupération des données, en particulier via son [SPARQL Endpoint](#). Néanmoins, ces techniques ne sont pas encore bien maîtrisées par les utilisateurs professionnels historiques de la BnF, ni par les développeurs. Dans les deux cas, favoriser les réutilisations nécessite accompagnement, pédagogie et médiation. Le service pourrait de fait être amélioré par la refonte du SPARQL endpoint avec l'intégration d'un outil permettant de l'interroger en langage naturel (SPARNatural, ASKELYS, Spartacus, etc.). Il sera aussi indispensable de travailler sur la documentation et sur des modalités de médiation efficaces. On pense particulièrement à la mise en place de tutoriels et à une refonte des pages statiques de data.bnf.fr. Ce volet permettra également de mieux valoriser les réutilisations des données de data.bnf.fr afin de capitaliser de manière plus transverse sur ces retours d'utilisateurs.

Priorité	Moyenne
Applications ou projet BnF en lien	Spartacus
Minimum à atteindre	Implémenter un outil d'exploration des données en RDF plus accessible (ex : SparNATURAL)
Maximum à atteindre	Intégrer les expérimentations des réutilisateurs