

Journée AFNOR CN 46 / BnF du vendredi 24 juin 2016
« Bibliothèques, archives et musées à l'heure de l'*Open data*
Bonnes pratiques, normes et retours d'expérience »

Compte rendu de la table ronde interprofessionnelle « Quelle ouverture pour quelles données ? »

Modérateur :

Antoine Courtin, responsable de la cellule d'ingénierie documentaire à l'INHA

Intervenants :

Gaël Chenard, directeur des Archives départementales des Hautes-Alpes

Caroline Latour, responsable informatique/web/multimédia du Musée des Augustins de Toulouse

Benoît Deshayes (pseudo : Shonagon), Wikidata

Les trois intervenants ont présenté des projets s'inscrivant dans des contextes très différents en termes de moyens, de volumes, de contenus, ou encore de statuts. C'est pourtant l'idée d'une convergence qui se dégage de cette table ronde, fondée sur une proximité de motivations et sur une (re)connaissance mutuelle : ainsi se dessinent les contours de ce que pourrait être le périmètre effectif de l'ouverture des données, à considérer sans doute moins dans la dynamique de ceux qui les ouvrent que dans celle de ceux qui les reçoivent...

L'exploration de ce périmètre – toujours placée sous le signe du service – a abordé plusieurs questions, en particulier celle de la communauté ainsi créée. Et l'auditoire a été sensible à la façon dont ces trois cas, à leur manière, font école.

Une lecture des questions juridique, technique et politique selon l'angle du service rendu

Un projet de « service »

Conscientes que la consultation sur place du document est devenue aujourd'hui une part résiduelle de leur activité, les Archives départementales des Hautes-Alpes ont conçu un projet d'*Open data* répondant à trois objectifs : permettre la numérisation à la demande, s'adapter aux usages des lecteurs et favoriser les réutilisations. Comme l'explique Gaël Chenard, elles proposent un service de numérisation à la demande et mettent leurs données à disposition des publics sur un site conçu comme un instrument de dialogue avec l'utilisateur et dont le ton cherche à reproduire la proximité appréciée par les publics au sein de la salle de lecture physique¹. Le succès rencontré (passage de 900 demandes annuelles à 18 500) confirme les ingrédients nécessaires à toute offre se réclamant du « service » : la rapidité, la clarté, la proximité, et l'originalité.

Le bénéfice, *in fine*, de cette initiative est l'enrichissement et le progrès de la connaissance, comme le confirme Benoît Deshayes, dont l'important travail d'alignement entre les données de Wikidata² et diverses

¹ <https://www.archives05.fr/>

² <https://www.wikidata.org/>

autres bases de connaissances permet de compléter des éléments manquants – par exemple la langue d'expression – et, par effet de retour, est susceptible de profiter au catalogue de la BnF lui-même. Des langues comme le breton gagnent ainsi en visibilité : Wikidata permet d'éditer la liste des écrivains l'ayant choisie comme langue d'expression³.

Par ailleurs, Wikidata apporte des référentiels dans des domaines où ils sont inexistantes tel le champ des ordinateurs, utilise le modèle FRBR pour affiner les requêtes bibliographiques autour d'un auteur donné – ce qui permet de trouver par exemple l'ensemble des films tirés des œuvres de Jules Verne – et favorise la mise en perspective de contenus sous des formes renouvelées. Dans une preuve de concept présentée par Benoît Deshayes, plusieurs récits de voyages en ligne sont de cette façon rendus accessibles via un planisphère dynamique figurant tous les lieux qu'ils décrivent, du Mississippi à l'île Spitzberg !

Ce qui fonde la qualité du service est l'autorité dont il s'entoure : en tant que base secondaire, Wikidata n'a pas pour vocation de créer des références, mais relaie des déclarations. Il est important pour elle de pouvoir s'appuyer sur des référentiels d'identifiants validés, qui vont apporter du crédit aux données. À Toulouse, le Musée des Augustins a pu d'autant mieux répondre au projet global d'*Open data* qu'il avait terminé son inventaire et venait de procéder au récolement imposé par une obligation décennale : il disposait donc d'un jeu de données actualisé.

Le droit et la technique pour garantir l'accès

Caroline Latour comme Gaël Chenard ont précisé que la licence ouverte est une condition *sine qua non* pour mener à bien des projets de ce type. La Ville de Toulouse et le Département des Hautes-Alpes ont tous deux opté pour la licence Etalab, présentée cette même journée par Romain Talès. Wikidata, quant à elle, est une base libre.

Si l'ouverture est acquise juridiquement, elle est plus complexe techniquement : en effet, dans le projet de Gaël Chenard, il faut encore développer les outils de partage, qui, le cas échéant, rendraient possible le téléchargement de masses très importantes de données, tandis qu'à Toulouse, le duo chargé de réaliser l'*Open data* à l'échelle de la Ville cherche un moyen pour mettre les données à jour de façon automatisée, sans l'intervention d'un tiers. Les trois interlocuteurs s'accordent sur un point : le caractère essentiel de la structuration des données ainsi ouvertes.

La logique de service dans laquelle s'inscrit cette ouverture place la question technique sous le signe de l'accès : Benoît Deshayes insiste sur la dimension requêtable de Wikidata (grâce à l'endpoint SPARQL) et sur la continuité qui doit être garantie en ce qui concerne l'accès aux ressources. C'est dans ce sens qu'il faut interpréter l'une des questions préparées par Antoine Courtin : comment un utilisateur sait-il, à partir d'un site, que les données sont réutilisables ? Comment l'articulation entre le site et l'endroit où elles sont disponibles s'établit-elle ? Caroline Latour, qui est également webmestre du site du Musée des Augustins⁴, réfléchit à un moyen d'aller plus loin encore en faisant savoir aux internautes que le musée est sur Wikimedia Commons. Selon Benoît Deshayes, c'est peut-être à un échelon plus international qu'il faut envisager ces liens.

Une volonté politique à l'appui

Caroline Latour inscrit d'emblée le projet d'ouverture des données du Musée des Augustins dans un contexte politique global : dès 1998, le directeur du musée a souhaité que la base de données soit mise en ligne en même temps que le site internet, et l'*Open data* fait partie des projets forts de la Ville.

Dans les Hautes-Alpes, l'implication politique a été sous-tendue par la rencontre entre deux volontés : la Direction des services informatiques souhaitait s'impliquer dans un projet d'ouverture, et les Archives avaient envie se positionner comme une institution « qui prend de la donnée et la diffuse » plutôt que comme une institution qui thésaurise.

³ <http://tinyurl.com/EcrivainsBr>

⁴ <http://www.augustins.org/>

Créer une communauté

Cette architecture globale ainsi posée, les projets se pensent également en termes de réseaux ou de communautés, tant du point de vue de l'organisation interne qu'en ce qui concerne les publics.

En interne

Le premier défi est d'impliquer un maximum d'acteurs : tous les acteurs culturels de la Ville de Toulouse ont par exemple été conviés pour réfléchir dans un premier temps aux données qu'il était possible de libérer. Dans les établissements eux-mêmes comme à une large échelle, il semble qu'un des enjeux porte aussi sur la manière dont la saisie des données est mutualisée. Caroline Latour souhaite ainsi multiplier les liens entre le Musée des Augustins et la base Joconde.

De l'importance des réseaux sociaux

Les résultats – enviablés, reconnaissons-le – des Archives départementales des Hautes-Alpes auprès de leurs publics sont notamment permis, d'après Gaël Chenard, par les généalogistes qui disent du bien du site et du service. De même, l'initiative du musée toulousain génère de nombreuses réactions positives sur Twitter. Ses données ont en particulier été réutilisées dans un projet de data visualisation par Antoine Courtin, sur le site Crotos⁵ de Benoît Deshayes, ou encore dans le cadre de « Tous au musée ! », un exercice soumis aux étudiants du Master PluriTal.

Du microcosme au macrocosme

Caroline Latour le reconnaît : les exemples qu'elle a cités émanent surtout... des membres de cette table ronde eux-mêmes. Cette constatation la conduit à évoquer l'idée d'une communauté de « museogeeks » relativement réduite qui peine à partager son enthousiasme, ses compétences et ses projets. Cela dit, si c'est incontestablement un point de vigilance à prendre en compte dans les développements futurs, les fruits de l'ouverture se lisent également à travers le décloisonnement qu'elle permet – ce qui représente davantage que le simple enrichissement des connaissances évoqué plus haut. Les données du Musée des Augustins se retrouvent dans Europeana ou Wikimédia, et Benoît Deshayes décrit les riches promesses portées par Wikidata en matière de multilinguisme.

Perspectives de modélisation ?

À la manière du Petit Poucet semant des cailloux sur son chemin – puisque cette journée a aussi été placée sous le signe du conte (voir la présentation de Raphaëlle Lapôte et Romain Wenz) – les intervenants ont eu le souci de donner des clés à un auditoire désireux de s'inspirer de leurs réussites. Caroline Latour a énoncé quelques pré-requis (bien connaître la base de données, travailler en concertation avec les conservateurs et les chargés de recherches documentaires, publier les données même si elles sont incomplètes), et Gaël Chenard a indiqué un moyen d'anticiper une croissance subite de la demande en matière de numérisation : il est possible de deviner ce qui sera demandé, car c'est aussi ce qui est demandé dans les salles de lecture.

Deux grandes leçons se dessinent encore.

S'inscrire dans une continuité

Caroline Latour l'a rappelé : le Musée des Augustins – comme nombre d'institutions – produit des données depuis des siècles, son premier catalogue ayant été publié en 1795. Dans le présent, il s'agit avant tout d'acquérir de bonnes habitudes pour structurer les données : Caroline Latour assure dans ce but un travail pédagogique auprès de ses collègues. Le monde des archives cherche quant à lui à harmoniser l'indexation – en particulier en ce qui concerne les données très locales – pour favoriser les alignements avec des bases comme GeoNames⁶. C'est ainsi que la volonté d'ouverture peut être intégrée en amont. Enfin, ces projets, à

⁵ <http://zone47.com/crotos/>

⁶ <http://www.geonames.org/>

commencer par Wikidata, sont en développement continu,

Procéder de façon empirique

Antoine Courtin s'est montré curieux de savoir quelle ont été les compétences mises en place pour développer cette démarche d'*Open data*, en particulier si elles ont été acquises par tâtonnement. Il s'avère que le travail expérimental a des vertus dans ce domaine, à l'heure où, si nous ne sommes plus dans l'enfance de l'*Open data*, nous sommes encore dans son adolescence. À Toulouse, même si le type d'export nécessaire était maîtrisé parce qu'il avait déjà été pratiqué, l'empirisme conserve une place.