

RAPPORT DE RECOMMANDATIONS SUR L'OUVERTURE DE JEUX DE DONNÉES DE BAnQ

**par Synapse C
pour Bibliothèque et Archives nationales du Québec**

Janvier 2023

Table des matières

Table des matières	2
Liste des tableaux et des figures	2
1. Introduction	3
1.2 Contextualisation	3
1.3 Problématisation du mandat	3
2. Méthodologie.....	4
2.1 Évaluation des données disponibles	4
2.2 Identification de jeux potentiels de données	6
2.3 Collecte d'information par des entrevues	6
2.4 Constats et recommandations	7
3. Résultats.....	8
3.1 Identification et pré-évaluation de jeux potentiels de données	8
3.2 Résultat des entrevues.....	11
4. Recommandations	16
4.1 Recommandation #1	16
4.2 Recommandation #2	17
4.3 Recommandation #3	18
4.4 Recommandation #4	19
Conclusion.....	20
Bibliographie sélective	20
Annexe – documents d'accompagnement	21

Liste des tableaux et des figures

Tableau 1. Synthèse des caractéristiques principales des jeux de données dont l'ouverture est suggérée.	8
Figure 1. Représentation graphique de l'évaluation des variables.....	5

1. Introduction

1.1 Mandat

Synapse C a été chargé d'évaluer l'utilité des jeux de données de Bibliothèque et Archives nationales du Québec (BAnQ)¹, ainsi que de produire des recommandations quant à l'ouverture de ces jeux de données, par l'analyse des besoins de quatre acteurs culturels québécois et d'un acteur québécois du secteur de la recherche.

1.2 Contextualisation

BAnQ identifie l'ouverture de données comme un moyen privilégié pour faire découvrir son contenu, l'utilisation de sa collection de documents, les créateurs et créatrices d'œuvres, ainsi que l'usage de ses services. L'ouverture de données s'inscrit également dans le plan stratégique 2021-2028 et fait suite à la politique de partage et valorisation des données (juin 2022).

Afin d'ouvrir et de partager des données pertinentes non seulement pour le public, mais aussi pour les acteurs québécois du milieu culturel et du secteur de la recherche, BAnQ souhaitait obtenir une meilleure connaissance de l'utilisation et des besoins de ces derniers en termes de données. Les entrevues qui ont été menées auprès des acteurs culturels et de l'acteur du secteur de la recherche ont été conçues à cet effet.

Plus qu'une prise de connaissance des besoins du milieu culturel québécois, ces entrevues visaient à engager une dynamique d'ouverture des données relatives à ce milieu, favorisant ainsi le potentiel de réutilisation des données ouvertes par BAnQ.

1.3 Problématisation du mandat

Afin de répondre au mandat, ce rapport apporte des réponses aux questions suivantes :

- Quels sont les besoins en données à moyen ou à long terme des acteurs québécois des milieux culturels et du secteur de la recherche, représentés par les cinq acteurs interviewés ?
- Quel est l'intérêt de la part de ces acteurs québécois pour les données de BAnQ ?
- Quel est l'intérêt de la part de ces acteurs québécois pour les données que BAnQ propose d'ouvrir ?
- Quelle priorité doit être accordée aux différents jeux de données de BAnQ qui ont été jugés pertinents à ouvrir ?

Ce rapport tente de répondre également à des questionnements auxiliaires, à savoir :

- Les acteurs québécois culturels et du secteur de la recherche connaissent-ils la plateforme *Données Québec* et les données ouvertes partagés par BAnQ sur cette plateforme ?
- Le cas échéant, ces acteurs utilisent-ils les données partagées par BAnQ sur *Données Québec* ?

¹ L'acronyme BAnQ et le terme « institution » sont employés de façon interchangeable dans le présent rapport.

2. Méthodologie

Les recommandations émises dans ce rapport sont fondées notamment sur une évaluation des données disponibles pour ouverture, effectuée par BAnQ ainsi que sur les analyses d'entrevues menées auprès des quatre acteurs culturels et de l'acteur du secteur de la recherche identifiés par l'institution.

L'examen de l'évaluation des données de BAnQ et la documentation auxiliaire fournie par l'institution a permis l'identification de quatre jeux de données potentiellement intéressants pour le milieu culturel québécois. Ces jeux de données ont été identifiés et évalués selon la qualité des données qui les composent ainsi que d'après l'intérêt qu'ils représentent pour les acteurs interviewés.

Dans le cadre de ce projet, nous avons effectué une activité de collecte d'information (activité 2.3) et trois activités distinctes d'analyse et d'évaluation de données (2.1, 2.2 et 2.4), dont les résultats sont présentés dans ce rapport :

- 1.1 Évaluation des données disponibles;
- 1.2 Identification de jeux potentiels de données;
- 1.3 Collecte d'information par des entrevues;
- 1.4 Constats et recommandations.

2.1 Évaluation des données disponibles

L'évaluation des données de BAnQ est fondée notamment sur l'évaluation préliminaire de ces données effectuée par BAnQ. L'évaluation préliminaire des données de BAnQ effectuée par cette même institution se présente sous la forme d'un dictionnaire de données comportant des statistiques relatives aux contenus de 49 variables. Ces variables sont associées à quatre catégories, à savoir :

- notice bibliographique (données descriptives relatives aux documents de la collection de BAnQ);
- notice d'autorité (données descriptives normalisées, relatives aux créateur.rice.s, contributeur.rice.s et institutions québécois.e.s);
- usager (données descriptives ayant trait aux usager.ère.s de BAnQ);
- transaction (données d'usage portant sur l'utilisation des services de BAnQ par les usager.ère.s).

Certaines catégories sont basées sur une seule table de données (notice bibliographique et notice d'autorité), alors que d'autres (usager, transaction) peuvent regrouper des variables provenant de différentes tables relationnelles.

Valeur des variables et effort de production

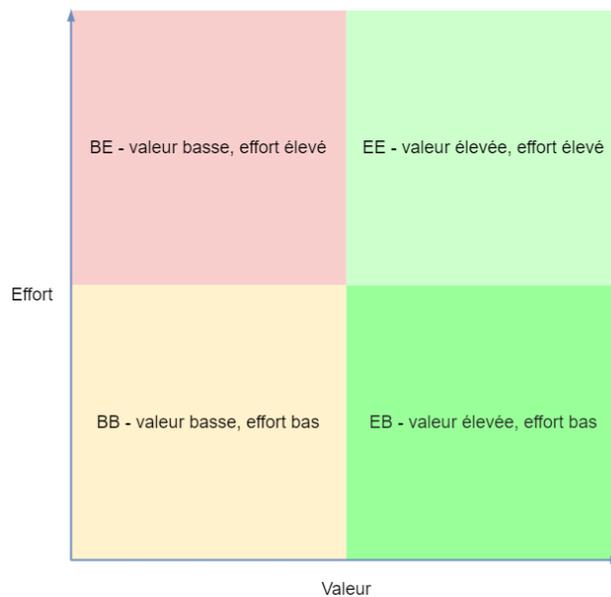
Durant cette activité, nous avons évalué en termes de valeur et d'effort chacune de 49 variables selon sa pertinence pour (ou son potentiel d'apport envers) les jeux de données potentiels. Le critère « valeur » indique l'apport direct d'une variable à un jeu de données, à savoir si le jeu de données gagnerait ou perdrait de façon significative en exhaustivité si la variable en faisait partie ou non respectivement. Le critère « effort » indique le niveau relatif d'effort qui sera nécessaire afin de rendre une variable pertinente pour un jeu de données.

La valeur de chaque variable et l'effort devant lui être consacré, le cas échéant, ont été évalués sur la base de l'évaluation préliminaire faite par BAnQ ainsi que d'après les considérations additionnelles suivantes :

- l'**interopérabilité** du contenu de la variable, à savoir si la donnée peut être publiée sous sa forme actuelle, dite « brute », ou si elle doit être formatée afin de répondre à une norme;
- l'intérêt que la variable représente, notamment en termes de **compatibilité** avec les jeux de données de bibliothèques déjà publiés ou d'un point de vue de recherche et de développement dans le milieu culturel (par ex., en lien avec les recherches portant sur la découvrabilité de contenu);
- les enjeux potentiels relatifs au contenu de la variable, notamment ceux de **confidentialité** ou de représentativité (ou équivocité) ainsi que les risques² potentiels que représenterait la publication du contenu en question;
- le **coût de production** estimé³, à savoir si une normalisation ou tout autre traitement préliminaire des données est nécessaire, ou encore, si la production d'une table de correspondance permettant la compréhension de données encodées s'impose, etc.

Sur la base de ces critères et considérations, quatre combinaisons distinctes de critères de valeur et d'effort ont été attribuées aux 49 variables individuelles, à savoir, *valeur élevée, valeur basse, effort élevé, effort bas* (voir fig. 1).

Figure 1. Représentation graphique de l'évaluation des variables.



² Par exemple, la publication de la variable pourrait-elle contribuer indirectement à l'identification d'individus, notamment par le croisement avec d'autres variables provenant d'autres sources de données. Des études ont établi que « 87 % des Américains étaient identifiables par le seul croisement de leur date de naissance, de leur sexe et de leur code postal » ([Guide de mise en oeuvre des énoncés d'orientations pour les données ouvertes \(quebec.ca\)](#), p. 5 et n. 3). Voir aussi [Avis WP-2016 du G29 sur les techniques d'anonymisation](#), p. 38.

³ Le coût de production des différentes variables est estimé notamment sur la base de l'évaluation préliminaire des données effectuée et partagée par BAnQ.

Cette évaluation des différentes variables en termes de valeur et d'effort a permis l'évaluation de la pertinence des jeux de données que ces variables permettent de former, à savoir si une combinaison de variables forme un jeu de données de valeur (en termes de qualité des données) requérant un effort de production réaliste.

La priorisation des jeux de données ainsi identifiés a été évaluée sur la base d'entrevues menées avec quatre acteurs culturels québécois et un acteur québécois du secteur de la recherche. L'identification des jeux de données pertinents et leur évaluation sont présentées dans les sections suivantes de ce rapport.

2.2 Identification de jeux potentiels de données

Sur la base de l'évaluation valeur/effort des données de BAnQ ainsi que du *Cahier de recommandations* réalisé par l'équipe de BAnQ visant à déterminer quels jeux de données de l'institution seraient les plus pertinents à ouvrir selon différents secteurs culturels, nous avons suggéré la formation de quatre jeux de données à porter à l'attention des acteurs culturels et du secteur de la recherche sélectionnés pour des entrevues (voir 3.1).

2.3 Collecte d'information par des entrevues

Étant donné que l'objectif des entrevues était la collecte de données qualitatives provenant d'un milieu commun – celui des acteurs québécois des milieux culturels et du secteur de la recherche –, mais aux contextes multiples et variés – ceux des différents acteurs interviewés, œuvrant dans différents milieux –, l'approche méthodologique adoptée a été l'entrevue semi-dirigée. En effet, cette approche a permis d'explorer différentes facettes d'un même cadre – celui de l'utilisation des données par le milieu culturel –, tout en créant des occasions de partage de points de vue et de discussions ainsi que de contextualisations des éléments d'information. Les entrevues semi-dirigées menées sous la forme de visioconférences auprès des cinq acteurs ont été d'une durée maximale d'une heure chacune.

Préalablement aux entrevues, chaque acteur a reçu un document de deux pages (Annexe) explicitant l'objectif des entrevues et présentant les thématiques ciblées (le cadre des entrevues) ainsi qu'un tableau présentant succinctement les quatre jeux de données dont l'ouverture est suggérée (tab. 1).

L'animation des entrevues était basée sur un document d'accompagnement (pense-bête) contenant les questions ouvertes couvrant les thématiques mentionnées ainsi que des questions d'orientation (Annexe). Ce document d'accompagnement permettait d'attribuer une structure aux entrevues, alignant les discussions sur les objectifs explicités dans le mandat du projet. Dans le but d'obtenir des réponses plus ou moins spontanées, le document d'accompagnement (contenant les questions de base et celles d'orientation) n'a pas été distribué aux acteurs interviewés.

BAnQ a identifié comme interlocuteurs pour chaque entrevue cinq acteurs québécois représentant quatre milieux culturels différents et le secteur de la recherche, à savoir :

- livre – De Marque;
- musique – MétaMusique;
- audiovisuel – Cinémathèque québécoise;
- patrimoine – Musée de la civilisation;

- recherche – INRS.

Outre les représentant.e.s des acteurs interviewés, deux membres de BAnQ et deux membres de Synapse C (dont l'animateur d'entrevue) étaient présents dans quatre des cinq rencontres. Les membres de BAnQ étaient au nombre de trois dans l'entrevue. Toute personne assistant aux entrevues était libre d'intervenir à tout moment avec des questions, des précisions ou des commentaires.

Outre que par leur capacité de répondre aux besoins des acteurs québécois des milieux culturels et du secteur de la recherche (*infra*), les recommandations portant sur l'ouverture des jeux de données suggérés s'appuient autant que possible sur la qualité estimée de ces jeux, évaluée selon les critères généraux des données ouvertes⁴, à savoir :

- intégrité – propriété d'une donnée de ne pas avoir été altérée ou détruite de façon erronée ou sans autorisation;
- exhaustivité – des données aussi complètes que possible reflétant l'ensemble de ce qui est colligé sur un sujet donné;
- disponibilité – propriété d'une donnée d'être accessible en temps voulu et de la manière requise par une personne autorisée;
- granularité – des données présentées au niveau de détail le plus fin possible, tout en empêchant l'identification de personnes, assurant ainsi la sécurité des données et la protection des renseignements personnels;
- interopérabilité – la possibilité pour des systèmes d'information hétérogènes d'échanger des données.

2.4 Constats et recommandations

Enfin, nous avons compilé, synthétisé et analysé l'information collectée par le biais des entrevues dans le but d'évaluer la pertinence des jeux de données dont l'ouverture a été suggérée. Les recommandations relatives à ces jeux de données sont basées tant sur cette dernière évaluation que sur l'analyse des données effectuée préalablement aux entrevues (2.1).

⁴ Voir, entre autres, DAMA UK (2013); Gouvernement ouvert (2018).

3. Résultats

3.1 Identification et pré-évaluation de jeux potentiels de données

Tableau 1. Synthèse des caractéristiques principales des jeux de données dont l'ouverture est suggérée.

No.	Type de jeu	Objet	Contenu
1	Autorité	Artistes, créateur.rice.s, chercheur.euse.s	Nom principal; ISNI; date de naissance; pays de résidence; activité (secteur); profession; genre; langue; ...
2	Palmarès	Documents québécois les plus empruntés	Notices bibliographiques des documents physiques et numériques (livre, musique, audiovisuel) ...
3	Listes bibliographiques	Collection physique et numérique	Description bibliographique des documents (notices bibliographiques) ...
4	Statistiques d'usage	Utilisation des services selon les publics, agrégation par catégories d'âge.	Nombre d'abonnements et de renouvellements d'abonnements; nombre de prêts et de réservations de documents; nombre de réservations et temps d'utilisation total des postes; ...

3.1.1 Jeu #1 - données d'autorité

Description

Le jeu de données d'autorité est entièrement basé sur la table de notices d'autorité (catégorie notice d'autorité). Il comprend huit variables composant la description normalisée d'une entité (créateur.rice d'œuvres culturelles ou institution culturelle), tels le Code international normalisé des noms (ISNI), le nom principal de l'entité, sa date de naissance, son secteur d'activités, etc. (tab. 1).

Valeur

L'intérêt potentiel de ce jeu de données, ou sa valeur, réside notamment dans sa capacité de contribuer envers la découvrabilité des créateur.rice.s de contenu et de leurs œuvres par son potentiel de croisement avec les données internes des acteurs des milieux culturels et du secteur de la recherche ou avec les données provenant d'autres institutions ou organismes.

Effort

Étant basé sur des données normalisées, ce jeu de données devrait avoir un coût de production relativement bas et devrait présenter une très bonne qualité des données. Il est à noter, cependant, que la publication de données d'autorité en général fait face à un enjeu important de confidentialité et d'auto-identité - fait qui pourrait affecter l'exhaustivité de ce jeu de données, par exemple, si l'on décidait d'en retirer des variables (le lieu de naissance d'un.e créateur.rice, ou son genre dans l'éventualité où celui-ci n'a pas été défini par la personne concernée).

3.1.2 Jeu #2 - palmarès

Description

Il s'agit d'un jeu de données basé sur une seule table, comportant 17 variables dont au moins cinq sont normalisées (MARC 21), deux autres (pays, langue) sont encodées et une est générée automatiquement par le système de gestion de bibliothèque (SIGB).

Valeur

Les variables de ce jeu contiennent des données bibliographiques décrivant les œuvres faisant partie de la collection de BAnQ (par ex., titre de l'œuvre, année de parution, langue de publication, ISBN, etc.). À la différence d'une liste bibliographique (jeu #3, voir infra), les notices incluses dans ce jeu de données sont celles des œuvres les plus populaires, notamment en termes d'usage (par ex., top 1 000 des documents les plus empruntés).

Valeur

Se présentant sous la forme de listes comprenant des œuvres à la notoriété notable, les palmarès présentent un intérêt notamment pour la découvrabilité de ces œuvres (à l'instar des listes de sélection des prix littéraires). Les palmarès contribuent également aux recherches sur la découvrabilité des œuvres d'un type/genre spécifique regroupant des œuvres selon des critères linguistiques (œuvres en langue française), géographiques (œuvres québécoises), etc.⁵

Un autre intérêt que pourrait présenter la publication de palmarès est le croisement des données de palmarès avec celles provenant d'autres jeux de données, par exemple, le jeu de données de référence *Documents publiés au Québec* partagé par BAnQ sur Données Québec. De tels croisements pourraient fournir au milieu culturel québécois et américain francophone des informations pertinentes sur la consommation d'œuvres québécoises.

Enfin, la publication de palmarès sous la forme de données ouvertes anticipe certains besoins médiatiques (publications dans la presse de listes de documents les plus populaires, etc.), contribuant ainsi, en retour, envers la notoriété des œuvres comprises dans les palmarès.

Effort

Le coût de production de ce jeu de données est estimé comme étant bas, notamment en raison du fait que les données bibliographiques qu'il contiendra pourraient être celles qui composent d'autres jeux de données, notamment celui de référence⁶ *Documents publiés au Québec*. Pour cette même raison, la qualité des données composant ce jeu est estimée comme étant bonne.

Étant basés par définition sur des données d'usage (nombre d'emprunts et de réservations), les palmarès représentent certains milieux (livre, musique, audiovisuel) mieux que d'autres (patrimoine). En d'autres termes, les systèmes de gestion de données, notamment les SIGB, ne sont habituellement pas adaptés pour effectuer un suivi de la consommation d'œuvres culturelles lorsque cette consommation n'est pas exprimée en termes d'emprunts et/ou de réservations. Un effort additionnel pourrait être

⁵ Cette contribution consiste, notamment, en l'adoption d'œuvres de palmarès comme des étalons de découvrabilité dans les évaluations semi-automatisées (analyse de statistiques compilées automatiquement) et automatisées (moissonnage) de plateformes de diffusion d'œuvres.

⁶ D'après Données Québec, les jeux de données de référence sont des « jeux liés à des caractéristiques en vue d'en favoriser la fiabilité. Ces caractéristiques concernent notamment les pratiques appliquées en matière d'identifiant unique, de qualité, d'intégrité, de disponibilité, d'interopérabilité, de métadonnées et de documentation. » Voir [Données ouvertes de référence : des données pour bâtir - Données Québec \(donneesquebec.ca\)](https://donneesquebec.ca).

requis pour la compilation de palmarès portant sur des documents ou objets patrimoniaux hors circulation dont la consommation n'est pas mesurée en transactions, mais plutôt en termes de consultations sur place ou virtuelles.

3.1.3 Jeu #3 - listes bibliographiques

Description

À l'instar du palmarès (jeu #2), ce jeu de données présente des données bibliographiques relatives à la collection de BAnQ. Les 17 variables composant ce jeu sont les mêmes que celles qui composent le palmarès et proviennent de la même table. À la différence du palmarès, la liste bibliographique présente les données (ou notices bibliographiques) de tous les documents d'un, de plusieurs, voire de tous les types de supports (livres imprimés et/ou numériques, disques compacts, etc.) représentés dans la collection de BAnQ. Ainsi, la différence principale entre ce jeu de données et le palmarès est dans le nombre de cas (ou de notices) que ces deux types de listes contiennent⁷.

Valeur

Outre d'un point de vue de compatibilité avec les types de jeux habituellement publiés par d'autres institutions possédant des collections de documents accessibles au public, les listes bibliographiques servent de sources d'information pour l'enrichissement des données descriptives compilées par d'autres acteurs des milieux culturels ou autres. En contribuant envers la bonification des données bibliographiques de ces autres acteurs québécois, la liste bibliographique contribuerait directement à l'amélioration collective de la découvrabilité des œuvres qui s'y trouvent.

Effort

Pouvant contenir des types de documents variés, les listes bibliographiques peuvent être relativement plus coûteuses à produire que les palmarès, notamment en termes de nettoyage, de complétion ou de normalisation des données. Sur la base de l'évaluation préliminaire de ses données effectuée par BAnQ et de nos observations ayant trait aux critères de « valeur » et d'« effort » décrits précédemment, nous évaluons le coût de production de ce type de jeu de données comme étant potentiellement moyen et sa qualité comme étant acceptable.⁸

3.1.4 Jeu #4 – statistiques d'usage (données transactionnelles)

Description

Le quatrième, et dernier, jeu suggéré comprend des statistiques relatives à l'utilisation de différents services offerts par BAnQ, notamment le nombre d'abonnements et de renouvellements, le nombre de prêts et de réservations de documents, le nombre de réservations et le temps total d'utilisation des postes informatiques, etc. Pour des raisons de confidentialité et d'anonymat, nous suggérons que celui-

⁷ Dans le cas du palmarès, ce nombre est fixé par les besoins spécifiques du classement et est relativement bas (100, 500, 1 000,...), alors que dans le cas de la liste bibliographique, il est fixé par le nombre d'œuvres présentes dans la collection de l'institution et peut contenir des dizaines, voire des centaines, de milliers de notices.

⁸ Il est à noter que le taux relativement grand de données a priori qualifiées de manquantes affiché par certaines variables pourrait être dû aux différentes propriétés des différents types de documents, et non pas à des données manquantes véritablement. Par exemple, si la variable attribuée au ISBN affiche un grand taux de données manquantes sur l'ensemble de la collection de documents de BAnQ, cela pourrait être dû au fait qu'un nombre relativement grand des documents de cette collection (enregistrements sonores ou audiovisuels, objets, etc.) ne comportent pas d'ISBN. Si cette induction s'avérait correcte, le taux de données manquantes devrait être révisé à la baisse et la qualité de ce jeu de données à la hausse.

ci soit ouvert sous forme d'agrégation par groupes d'âges des usager.ère.s.⁹ Sous sa forme de données agrégées, ce jeu est composé de sept variables calculées à partir de données provenant de cinq tables différentes.

Valeur

Ce jeu de données se présente comme un compromis acceptable entre la publication de données à l'échelle de l'individu.e pouvant comporter des enjeux d'anonymat et de confidentialité et l'absence complète de données ouvertes portant sur l'utilisation et l'appropriation des services offerts par BANQ.¹⁰

En plus d'être compatible avec des jeux de données similaires publiés par d'autres institutions, ce jeu présente un intérêt d'un point de vue sociodémographique par le potentiel de croisement ou de comparaison avec d'autres données, notamment par l'association des indicateurs de transactions (abonnements, réabonnements, réservations, etc.) avec différents groupes d'âges.

Effort

Découlant directement de pratiques quotidiennes (données d'utilisation) et se présentant sous forme d'agrégations par groupes d'âges dont la composition peut évoluer de façon notable en un espace de temps relativement court, ce jeu de données pourrait exiger des mises à jour plus régulières comparativement à d'autres types de jeux de données afin que sa pertinence soit maintenue.

Nous évaluons le coût de production de ce jeu comme étant moyen, notamment en raison des calculs nécessaires à la production des indicateurs qu'il enferme. Ne présentant pas de données manquantes et ne nécessitant pas de normalisation, la qualité de ce jeu de données est évaluée comme étant bonne.

3.2 Résultat des entrevues

3.2.1 Utilisation générale des données

Trois des acteurs interviewés – milieux du livre, de la musique et de l'audiovisuel – créent, utilisent et enrichissent des données notamment descriptives (bibliographiques et métadonnées). L'acteur du milieu de la recherche s'intéresse davantage aux données d'usage, tout en contextualisant celles-ci au moyen de données descriptives. L'acteur du milieu du patrimoine crée et utilise tant des données descriptives que des données d'usage.

Tous les acteurs interviewés démontrent un intérêt pour le croisement de données descriptives et d'usage afin d'acquérir une meilleure connaissance en ce qui a trait aux dynamiques de consommation de contenu culturel propres à leurs milieux spécifiques ou entre milieux culturels différents.¹¹ Toutefois,

⁹ En effet, alors qu'aucune variable dans ce jeu de données ne permet l'identification d'individu.e.s en soi, comme certaines de ces variables portent sur l'utilisation de services sur place, leur croisement entre elles et avec d'autres variables du même jeu de données pourrait fournir suffisamment d'indices pour que des individu.e.s puissent être identifiés indirectement.

¹⁰ Ce type de données est déjà collecté et publié dans le cadre de l'Enquête annuelle des bibliothèques publiques du Québec (EABP); malgré sa forme agrégée, le jeu de données suggéré dans ce document se veut plus détaillé que les résultats publiés dans l'Enquête, car ce premier présente une agrégation par groupes d'âges, alors que cette dernière agrège les données par secteur (jeunes, adultes, aînés).

¹¹ En ce qui a trait aux acteurs des milieux du livre, de la musique et, à moindre degré, de l'audiovisuel, cet intérêt pour le croisement de données est expliqué par une volonté d'amélioration des services offerts, mais aussi par le souhait de mieux comprendre les dynamiques des contenus culturels dont ils sont les gestionnaires et les diffuseurs. Quant à l'acteur du secteur de la recherche, le croisement de données descriptives avec les données d'usage qui font l'objet de ses études lui permet une meilleure contextualisation socio-culturelle et démographique de ce dernier type de données.

certaines facteurs limitent la capacité des acteurs en question de croiser des données de différents types. Ainsi, l'acteur du milieu du livre n'applique pas à ses activités les informations acquises par les données d'usage qu'il reçoit de la part des diffuseurs de livres outre que par la transmission de ces informations à sa clientèle d'éditeurs de livres. Néanmoins, cet acteur exprime son intérêt pour l'incorporation future de données d'usage dans ces activités de base, notamment dans le but d'acquérir une meilleure connaissance de la consommation d'œuvres sur différents types de supports ainsi qu'afin d'analyser le cycle de vie des différents types de support d'une œuvre. L'acteur du milieu musical reconnaît également le potentiel informationnel du croisement de données d'usage avec des données descriptives (en l'occurrence, métadonnées). Cependant, étant un nouvel acteur dans son milieu et en alignement avec sa mission de produire, de bonifier et de diffuser des données descriptives (métadonnées), il ne prévoit pas utiliser les données d'usage (des œuvres musicales) dans un court ou moyen terme. Étant dans un processus de transition de système de gestion de données, l'acteur du milieu audiovisuel ne voit pas à court terme d'utilité pour les données d'usage dans le cadre de ces activités, outre en ce qui a trait aux données de transaction (nombres de visites des événements qu'il organise) qu'il compile déjà. Enfin, l'acteur du milieu du patrimoine s'intéresse tant aux données descriptives qu'aux données d'usage ainsi qu'au croisement de données des deux types.

3.2.2 Données d'autorité (jeu #1)

Quatre des cinq acteurs interviewés démontrent un grand intérêt pour le jeu de données d'autorité. Cependant, les raisons qui poussent chacun de ces acteurs à démontrer un tel intérêt pour ce jeu de données varient.

L'intérêt pour le jeu en question exprimé par l'acteur du milieu du livre semble basé sur la grande importance que ce type de données représente presque par définition pour le milieu; cet acteur précise, cependant, qu'il obtient déjà des données d'autorité au moyen d'un API¹², dans le cadre d'un partenariat avec un autre acteur québécois du milieu du livre.

L'acteur du milieu de l'audiovisuel s'intéresse vivement à ce jeu de données et précise qu'il l'utiliserait pour enrichir ses propres données qui comportent actuellement des manques, notamment en ce qui a trait aux dates de naissance des créateur.rice.s/contributeur.rice.s.

Les données d'autorité étant le type de donnée principal utilisé par l'acteur du milieu de la musique, celui-ci démontre un intérêt particulièrement grand pour ce jeu de données, ainsi que pour un travail en collaboration (« connexion bidirectionnelle ») avec BANQ dans le but d'assurer un bon lien entre, d'un côté, les identifiants (ISNI) et, de l'autre, les créateur.rice.s.

L'acteur du milieu du patrimoine s'intéresse aux données d'autorités notamment pour enrichir sa documentation sur les créateur.rice.s dont il conserve et/ou expose les œuvres. Cet acteur ne conserve pas des données biographiques, ce qui ferait des données ouvertes d'autorité sa source principale d'information sur les créateur.rice.s. Cet acteur s'intéresse également au croisement de données d'autorité provenant de différentes sources, notamment dans un cadre de création et d'utilisation de données liées.

¹² *Application Programming Interface* : interface de programmation d'application par laquelle un logiciel (ou système) offre des services à d'autres logiciels (ou systèmes), appelé communément API dans le milieu québécois culturel.

Quant à l'acteur du secteur de la recherche, il ne trouve pas d'utilité pour ce jeu de données dans le cadre de ses activités professionnelles qui ne portent pas sur des individus ou sur des œuvres individuelles.

3.2.3 Palmarès (jeu #2)

Comparativement au premier jeu de données pour lequel l'intérêt est quasi unanime, l'intérêt pour des palmarès est moins grand. Néanmoins, tous les acteurs interviewés y voient des applications potentielles dans leurs activités.

Le plus grand intérêt pour des palmarès est démontré par les acteurs des milieux du livre et de la musique. D'après l'acteur du milieu du livre, les palmarès sont utiles pour orienter les communautés canadiennes et américaines francophones hors du Québec, notamment dans le contexte d'acquisition de documents francophones « à succès » par les bibliothèques publiques.

L'intérêt perçu dans la publication de palmarès par l'acteur du milieu de la musique se situe davantage dans la possibilité de croisement et de comparaison des données collectées par cet acteur avec celles collectées par BAnQ. Cet acteur ayant souligné le fait que son organisme est encore en processus de formation, les besoins pour des données de palmarès sont des besoins anticipés, qui n'ont pas encore émergé.

Quant à l'acteur du milieu du patrimoine, celui-ci verrait un intérêt pour ce type de données si elles portaient également sur les objets (par ex., collection d'affiches) et sur autres documents patrimoniaux.

Les acteurs du milieu de l'audiovisuel et du secteur de la recherche voient une utilité moindre dans les palmarès. L'acteur du secteur de la recherche y voit même des enjeux. Il est d'avis que les palmarès peuvent créer et renforcer une demande en portant le public général à s'intéresser surtout aux œuvres de la liste de palmarès, confirmant la notoriété déjà établie de ces œuvres aux dépens de celles qui n'y sont pas incluses. Une fois publié et rediffusé, le palmarès se retrouverait alors à s'auto-renforcer.

3.2.4 Données descriptives (jeu #3)

Les acteurs des milieux du livre et de l'audiovisuel démontrent tous deux un intérêt marqué pour ce jeu de données. Cependant, leurs conceptions respectives des contextes d'utilisation de ce type de données diffèrent notablement.

Ainsi, l'acteur du milieu du livre utilise des données descriptives « contrôlées » (produites ou compilées et traitées par des experts dans le milieu) et ne peut pas utiliser professionnellement des données ouvertes dans le but d'enrichir ses propres données descriptives. Par contre, cet acteur est très intéressé à des données descriptives (bibliographiques) croisées avec des données d'usage. Un tel croisement permettrait à cet acteur de suivre les tendances de la consommation d'œuvres culturelles (par ex., l'intérêt pour les adaptations d'œuvres sur différents supports, pour les différents genres, etc.).

L'acteur du milieu de l'audiovisuel s'intéresse aux données bibliographiques davantage afin d'enrichir ses propres données descriptives. Cet acteur est moins intéressé par le croisement des données descriptives avec des données d'usage (au sein d'un même jeu de données), car il ne possède pas pour le moment les moyens techniques qui lui permettraient d'intégrer des données d'usage au niveau de l'exemplaire à ses propres données descriptives.¹³

¹³ L'acteur du milieu de l'audiovisuel compile des données d'usage au niveau de l'événement (expositions, projections, etc.).

L'acteur du milieu de la musique travaille davantage avec des données d'autorité plutôt que bibliographiques, mais il démontre néanmoins un intérêt pour ce dernier type de données en précisant qu'elle pourrait servir notamment à bonifier le portrait de la consommation d'œuvres musicales.

Les données bibliographiques sont d'une utilité réduite pour l'acteur du milieu du patrimoine, car le système de gestion de données dont celui-ci dispose ne permet pas l'intégration automatisée de telles données à ses propres bases de données.

Enfin, un jeu de données bibliographiques serait de peu d'intérêt pour l'acteur du secteur de la recherche s'il n'était pas possible de croiser les données d'un tel jeu avec des données d'utilisation non pas en termes d'emprunts et de renouvellements, mais davantage en termes de services. En effet, dans le cadre de ses activités professionnelles, cet acteur s'intéresse davantage aux façons (en termes de pratiques) dont les contenus culturels sont consommés ainsi qu'aux contextes de leur consommation par le public.

3.2.5 Données d'utilisation des services (jeu #4)

L'acteur du secteur de la recherche et celui du milieu du patrimoine sont les seuls à démontrer un intérêt marqué pour ce jeu de données. En effet, l'acteur du secteur de la recherche s'intéresse vivement aux actions et aux pratiques qui engendrent, ou qui peuvent expliquer les données d'usage même dans l'éventualité où ces données étaient présentées sous forme agrégée (par groupes d'âges) et non pas à l'échelle de l'individu.e.

L'acteur du milieu du patrimoine porte un grand intérêt pour ce type de données notamment pour la bonification de ses descriptions de persona et de sa compréhension de ces persona.

Le manque d'intérêt de la part des autres acteurs pour ce jeu de données s'explique notamment par l'incapacité des systèmes qu'ils utilisent d'assimiler et d'intégrer de telles données. Ces acteurs voient également un moindre intérêt dans la publication de données agrégées qui ne peuvent pas être directement associées avec la consommation des types spécifiques de documents (par ex., le type de document réservé ou consommé sur un poste informatique de la bibliothèque).

3.2.6 Plateformes de données ouvertes (Données Québec)

L'acteur du milieu de l'audiovisuel déclare avoir régulièrement recours aux données ouvertes¹⁴ en général ainsi que sur Données Québec, plateforme sur laquelle il publie également des jeux de données. Cet acteur démontre un grand intérêt pour les données descriptives publiées par BANQ sur Données Québec, percevant ces données comme une source d'enrichissement de ses propres bases de données descriptives.

L'acteur du milieu musical, lui aussi grand producteur et consommateur de métadonnées, connaît la plateforme Données Québec et y voit un potentiel, mais son niveau de développement actuel, qu'il qualifie d'embryonnaire, ne lui permet pas d'utiliser cette plateforme comme source de données dans le cadre de ses activités professionnelles.

¹⁴ Il s'agit notamment de données descriptives publiées par Wikidata, mais aussi, à moindre mesure, celles rendues disponibles par BANQ ou d'autres institutions similaires notamment sur Données Québec.

L'acteur du secteur de la recherche affirme connaître Données Québec, mais il démontre un intérêt pour cette plateforme exclusivement en tant qu'objet, à savoir qu'il s'intéresse aux façons dont le contenu de la plateforme est utilisé et aux contextes et impacts socioculturels de son utilisation.

L'acteur du milieu du patrimoine dépose des jeux de données sur Données Québec (le dernier jeu déposé date de 2016) et démontre un très grand intérêt pour la plateforme ainsi que pour les jeux de données de BANQ qui y sont déposés. Cependant, les systèmes informationnels utilisés par cet acteur (qualifiés par celui-ci de « cloisonnés » tant en termes inter- qu'intra-institutionnels d'interopérabilité) ne lui permettent pas d'y intégrer de façon plus ou moins automatique des jeux de données externes.

L'acteur du milieu du livre ne connaissait pas les données de BANQ sur la plateforme Données Québec et n'utilise pas celle-ci. Cet acteur précise que l'enrichissement des métadonnées qu'il effectue doit être basé sur des données adaptées par des experts à ses besoins. Il démontre également un intérêt pour l'acquisition de données au moyen d'APIs.

3.2.7 Autres besoins

Outre l'utilisation et les besoins en données décrits dans les paragraphes précédents, l'acteur du milieu de l'audiovisuel a exprimé un besoin spécifique particulier à son milieu, à savoir, l'obtention de données de référence relatives à l'ISAN¹⁵ des œuvres audiovisuelles.

Constats

Les échanges avec les cinq acteurs interviewés portent à conclure que les données détenues par BANQ dont l'ouverture a été proposée ont le potentiel de répondre aux besoins en données des acteurs en question. En effet, de l'intérêt a été démontré pour tous les quatre jeux de données suggérés. Cependant, selon le jeu de données et selon l'acteur interviewé, le niveau de cet intérêt varie; c'est cette variabilité d'intérêt à l'échelle de l'acteur individuel qui permet d'accorder une priorité d'ouverture aux quatre jeux de données en question.

Les acteurs culturels démontrent presque unanimement un très grand intérêt pour l'ouverture de données d'autorité ainsi qu'un grand intérêt pour les données bibliographiques et les palmarès. L'intérêt a priori faible pour des données « transactionnelles » (jeu # 4) – seuls deux acteurs accordent une priorité au partage de ces données – est expliqué avant tout par l'incapacité des systèmes employés actuellement par la majorité des acteurs interviewés d'intégrer automatiquement ce type de données.¹⁶

Il convient de noter que l'intérêt des acteurs interviewés pour certains de ces jeux de données pourrait augmenter davantage s'ils étaient modifiés. Ainsi, les données bibliographiques et les palmarès seraient des jeux de données encore plus intéressants pour la majorité de ces acteurs si, en parallèle aux données descriptives, ces jeux comportaient des données d'usage (par ex., nombre d'emprunts, de renouvellements et de réservations par notice bibliographique).

En ce qui a trait à la plateforme Données Québec, un seul des acteurs interviewés a utilisé cette plateforme comme source de données dans un cadre professionnel et ce, de façon intermittente. Les enjeux mentionnés, relatifs aux méthodes et aux systèmes employés par les acteurs interviewés, indiquent que leur utilisation de données ouvertes dans le cadre de leurs activités professionnelles (incluant les données publiées sur Données Québec) est conditionnelle : 1) à l'automatisation des

¹⁵ L'*International Standard Audiovisual Number* est un identifiant unique et persistant de toute œuvre audiovisuelle ainsi que de ses différentes versions, incluant les films, les documentaires, les programmes et séries télévisées, etc. Voir www.isan.org.

¹⁶ Notamment si ce type de données se présentait sous la forme de jeu de données ouvertes.

processus d'accès aux données et de leur intégration aux systèmes utilisés par les acteurs culturels ou du secteur de la recherche, notamment par la mise à disposition d'APIs, et 2) au niveau de valorisation en amont de ces données.

4. Recommandations

Suite aux constats basés sur l'évaluation de la qualité de ses données effectuée par BANQ, sur les entrevues menées avec des acteurs québécois de quatre milieux culturels différents et du secteur de la recherche et, enfin, sur l'évaluation des jeux de données identifiés (3.1), elle-même appuyée par les activités de collecte et d'analyse des informations décrites dans ce rapport (3.2), nous sommes en mesure d'émettre les recommandations qui suivent. Chaque recommandation est accompagnée d'une discussion comportant des éléments justificatifs.

4.1 Recommandation #1

Accorder une priorité à l'ouverture du jeu de données d'autorité (jeu #1);

Recommandations d'implémentation

- que ce jeu soit le plus exhaustif¹⁷ possible, tout en respectant la confidentialité et l'auto-identification des créateur.rice.s dont les données y figurent
- que ce jeu soit mis à jour régulièrement afin d'en maintenir la pertinence, notamment pour y relier les changements qui pourraient survenir dans les données, notamment en lien avec les processus d'auto-identification (changement de nom, de genre, de pseudonyme, de lieu de résidence, etc.)
- envisager l'implémentation d'une API permettant aux acteurs québécois culturels et du secteur de la recherche d'extraire et d'intégrer les données d'autorité dans leurs systèmes de façon automatique ou semi-automatique

Avantages (+) et enjeux (-)

- + très grand intérêt de la part des acteurs québécois interviewés
- + jeu normalisé par définition
- + potentiel jeu de référence
- + contient des identifiants uniques
- pourrait nécessiter des mises à jour régulières afin de maintenir un bon niveau de pertinence (par ex., noms ou pseudonymes, lieux de résidence, genre, etc.)

¹⁷ Des données aussi complètes que possible reflétant l'ensemble de ce qui est colligé sur un sujet donné.

Discussion

L'estimation de la qualité des données d'autorité permet d'anticiper un jeu de données intègre, exhaustif¹⁸, disponible, granulaire et interopérable (puisque normalisé). Ce jeu de données étant d'un grand intérêt pour les acteurs interviewés, son éventuelle popularité sur la plateforme Données Québec pourrait contribuer à la visibilité des autres jeux de données ouverts par BAnQ. Toutefois, les acteurs du milieu culturel, notamment celui du livre, préfèrent travailler avec des données valorisées exploitables par le biais d'APIs qui leur sont offerts par d'autres organisations; l'offre d'un tel API de la part de BAnQ dans le but d'offrir un accès aux données d'autorité

4.2 Recommandation #2

Accorder une priorité d'ouverture au jeu de données bibliographiques (jeu #3).

Recommandations d'implémentation

- que ce jeu couvre le plus grand nombre de types de documents possible, de préférence, toute la collection de BAnQ
- que des données d'usage (nombre de prêts, de renouvellements et de réservations) soient ajoutées à chaque cas (chaque notice bibliographique) de ce jeu

Avantages (+) et enjeux (-)

- + rôle potentiel dans les projets d'enrichissement¹⁹ de données et de découvrabilité
- + compatibilité avec type de jeux de données publiées par d'autres institutions similaires
- + grand intérêt de la part de deux des cinq acteurs interviewés, relativement intéressant pour les autres acteurs
- + transparence et visibilité du contenu de la collection de BAnQ et de son usage
- peut requérir un temps de travail conséquent (nettoyage des données, compréhension de l'historique des données MARC, formatage)
- nécessite des opérations additionnelles (calcul des données d'usage par notice bibliographique)
- taux de données manquantes

Discussion

Le jeu de données bibliographiques comporte des variables qui affichent un taux relativement élevé de données manquantes ou qui sont encodées de façon non-normalisée. Un éventuel retrait des variables à taux élevé de valeurs manquantes de ce jeu de données compromettrait l'exhaustivité de ce dernier.

¹⁸ Certaines variables relatives à ce jeu de données et présentes dans la base de données de BAnQ pourraient ne pas être incluses dans le jeu de données pour des raisons de confidentialité, tel le lieu de naissance des créateur.rice.s; les entretiens avec les cinq acteurs culturels québécois portent à croire que de telles omissions ne compromettraient pas de façon significative l'exhaustivité du jeu de données en question et n'affecteraient pas négativement le grand intérêt qui lui est porté par le milieu culturel.

¹⁹ De façon sous-jacente, ce jeu de données permettrait d'éviter le dédoublement des efforts dans la création de métadonnées dans le milieu culturel.

Les variables à la nomenclature non harmonisée (ayant trait notamment aux dimensions des œuvres) pourraient affecter négativement l'interopérabilité du jeu de données. L'harmonisation préalable à leur publication augmenterait donc les coûts de production au moment de la première ouverture de ce jeu de données.

4.3 Recommandation #3

La publication de palmarès (jeu #2) serait intéressante

Recommandations d'implémentation

- couvrir, par plusieurs listes de palmarès, différents types de documents (livre, musique, audiovisuel, patrimoine)
- baser l'inclusion à la liste et l'ordre (tri) de la liste sur des critères variés, nuanciant l'indicateur de nombre de prêts habituellement retenu pour ce type de listes

Avantages (+) et enjeux (-)

- + peut être basé sur des jeux de données (bibliographiques) existants
- + rôle potentiel dans les projets de découvrabilité (notoriété des œuvres)
- + compatibilité avec type de données publiées par d'autres institutions similaires
- confirmation récursive : à moins d'adopter des critères plus variés que le seul indicateur de nombre de prêts, la liste pourrait avoir tendance à renforcer de façon circulaire la notoriété des œuvres qui s'y trouvent
- l'intérêt de la part des différents acteurs interviewés dépend du type spécifique de documents (livres, sonores, audiovisuels, patrimoniaux, etc.), ou combinaisons de types de documents, que ce jeu pourrait contenir

Discussion

Le coût de production de ce jeu de données pourrait être réduit notablement si celui-ci était basé sur le jeu de données bibliographiques (dans l'éventualité où des données d'usage étaient ajoutées à ce dernier). Afin qu'il soit utile aux acteurs relevant de différents milieux culturels, il devrait inclure (les notices bibliographiques ou d'exemplaire de) plusieurs types de documents (livres, enregistrements sonores, enregistrements audiovisuels, etc.), ou bien plusieurs palmarès, un par type de document, devraient être publiés.

Dans le milieu de la musique, de l'audiovisuel et du livre, les palmarès sont également utiles dans la composition de « paniers » d'œuvres servant aux recherches sur la découvrabilité des œuvres québécoises ainsi qu'aux recherches sur des indicateurs de découvrabilité.²⁰ Ces « paniers » sont utiles notamment lorsque les palmarès sur lesquels ils sont basés sont des jeux de données dits de référence.

Il est important de noter que les acteurs s'attendent, explicitement ou implicitement, à ce que leurs milieux respectifs soient relatés dans les éventuels palmarès. En d'autres termes, un seul palmarès portant sur un seul type de document satisferait un seul parmi les acteurs culturels, celui du milieu auquel correspond le type de document relaté dans le palmarès. D'un autre côté, un palmarès mixte

²⁰ Voir notamment Danvoye (2021, p. 14 et suiv).

comportant plusieurs types de documents ne satisferait aucun des acteurs (sauf, possiblement, l'acteur du secteur de la recherche). Ainsi, la seule façon dont la publication de palmarès pourrait satisfaire les besoins en données de tous les acteurs serait de publier des palmarès individuels pour chacun des principaux types de documents correspondant aux milieux de ces acteurs (livre, musique, audiovisuel, patrimoine).

4.4 Recommandation #4

Le partage de statistiques d'usage (données transactionnelles) agrégées (jeu #4) pourrait être intéressant.

Recommandations d'implémentation

- prioriser le partage de ce jeu de données avec des acteurs culturels ou du secteur de la recherche de façon restreinte, à la demande, sous forme agrégée ou désagrégée anonymisée

Avantages (+) et enjeux (-)

- + contextualisation démographique de la consommation de contenus culturels
- + source d'enrichissement des persona utilisés par les acteurs culturels ou du secteur de la recherche
- + grand intérêt de la part de deux acteurs sur cinq
- intérêt de la part de seulement deux acteurs sur cinq
- relativement complexe (données de sources et de nature hétéroclites)
- agrégation spécifique compromet interopérabilité

Discussion

Se présentant sous forme de statistiques agrégées par groupes d'âges (variable intéressante du point de vue des acteurs interviewés), ce jeu de données transgresserait le critère de granularité et, possiblement, celui d'interopérabilité (les groupes d'âges pouvant être définis différemment par les différentes institutions publiques et organismes culturels). En effet, la nature des variables composant ce jeu de données ne permet pas une bonne granularité, même dans l'éventualité où des super- ou sous-agrégations (par exemple, par type de document) étaient ajoutées au jeu; de telles agrégations additionnelles ne pourraient être appliquées à toutes les variables.

D'un autre côté, même agrégé (pour des raisons de confidentialité), un jeu de données transactionnelles mis à jour de façon régulière contribuerait à contextualiser les données sommaires publiées à l'occasion de l'Enquête annuelle sur les bibliothèques publiques ou d'autres enquêtes similaires dont les résultats sont publiés ouvertement. Comme il a été établi par le biais des entrevues, une telle contextualisation de la consommation de contenu culturel par le public québécois présente un intérêt général relatif pour l'ensemble des acteurs culturels québécois et est d'intérêt particulier pour les acteurs du secteur de la recherche. De plus, l'ouverture d'un tel jeu de données contribuerait envers l'enrichissement et la compréhension des persona que certains acteurs emploient dans le but de bonifier leurs services.

Le partage de ce jeu de données, plutôt que son ouverture, contribuerait à maintenir sa pertinence.

Conclusion

Tout processus d'ouverture de données doit répondre à des critères et à des exigences de base ou spécifiques à la plateforme hébergeant les données ouvertes. Ces critères et exigences peuvent se présenter comme un défi considérable en ce qui a trait à la collecte et à la préparation des données à ouvrir. L'effort posé en amont pour relever ce défi peut être compromis par la validité du besoin en données²¹ ou par l'absence d'un tel besoin. C'est alors que la nouveauté que représente un jeu de données récemment ouvert est rapidement usée et que le jeu de données perd sa pertinence.

À l'exception de BANQ, peu de bibliothèques publiques au Québec ont relevé les défis associés avec l'ouverture de données. Par l'ajout de nouveaux jeux de données ouvertes à ceux qu'elle publie déjà, BANQ se démarquera davantage tout en confirmant son statut d'exemple à suivre. Mais c'est surtout par son approche de la problématique de l'accès aux données que BANQ se démarque. Cette approche, fondée sur l'analyse contextualisée et sur la consultation effectuée en collaboration éthique et équitable avec différents acteurs québécois (tant dans le cadre de ce projet spécifique, que dans celui des autres projets développés par l'institution) transforme les démarches d'ouverture ou de partage de données en une dynamique de cocréation qui favorisera la réutilisation des données de BANQ en vue de leur transformation en information et en connaissances utiles à l'ensemble du milieu culturel québécois.

Bibliographie sélective

DAMA UK (2013). *The six primary dimensions for data quality assessment: Defining data quality dimensions*. Repéré à : [\[PDF\] THE SIX PRIMARY DIMENSIONS FOR DATA QUALITY ASSESSMENT - Free Download PDF \(silo.tips\)](#).

Danvoye, M. (2021). *Conception d'indicateurs de rayonnement de l'offre culturelle québécoise dans les réseaux numériques, Étude de faisabilité* [En ligne], Québec, Institut de la statistique du Québec, 56 p. [statistique.quebec.ca/fr/fichier/conception-indicateurs-rayonnement-offre-culturelle-quebecoise-reseaux-numeriques-faisabilite.pdf].

Gouvernement ouvert : Québec (Province). (2018). *Guide de mise en oeuvre des orientations pour les données ouvertes : gouvernement ouvert*. Secrétariat du Conseil du trésor. Repéré à : <http://collections.banq.qc.ca/ark:/52327/3686499>.

²¹ Par exemple, lorsqu'il s'agit de répondre à des demandes récurrentes, mais relativement ponctuelles, comme les demandes annuelles de la part des médias pour une liste « top » des documents les plus empruntés.

Annexe – documents d’accompagnement

Document envoyé aux acteurs culturels préalablement aux entrevues

BAnQ - Projet d'ouverture de données Entrevues

Introduction à l’entrevue

Bibliothèque et Archives nationales du Québec (BAnQ), c’est à la fois une bibliothèque nationale, des archives nationales, une bibliothèque publique – la plus fréquentée de l’espace francophone – et une bibliothèque virtuelle accessible dans le monde entier.

Plus grande institution culturelle du Québec par sa fréquentation et la diversité de ses missions, pilier essentiel de la société du savoir, BAnQ a pour mandat d’offrir un accès démocratique à la culture et à la connaissance. Elle rassemble, conserve et met en valeur le patrimoine documentaire québécois ou relatif au Québec et une vaste collection universelle au bénéfice des générations présentes et futures.

En pleine transformation numérique, l’institution vise une meilleure découvrabilité de ses collections pour soutenir sa mission. L’ouverture de données est un moyen privilégié pour faire découvrir son contenu, l’utilisation de sa collection de documents, les créateurs et créatrices d’œuvres, ainsi que l’usage de ses services.¹ Elle s’inscrit dans le plan stratégique 2021-2028 et fait suite à la politique de partage et valorisation des données (juin 2022).

Afin de partager des données pertinentes non seulement pour le public, mais aussi pour les acteurs du milieu culturel québécois, BAnQ souhaite obtenir une meilleure connaissance de l’utilisation et des besoins de ces derniers en termes de données. La présente entrevue est conçue à cet effet.

Plus qu’une prise de connaissance des besoins du milieu, cette entrevue vise à engager une dynamique d’ouverture des données relatives à ce milieu, favorisant ainsi le potentiel de réutilisation de ces données.

Liste thématique de l’entrevue

Les thèmes suivants sont abordés durant l’entrevue.

1. La place de données dans le cadre de vos activités et la nature des données que vous produisez.
2. Votre utilisation de données ouvertes, notamment les données de BAnQ sur la plateforme Données Québec.
3. Vos besoins en données ouvertes de BAnQ et les enjeux relatifs aux données ouvertes et à leur utilisation au sein de votre organisme/institution.
4. Votre avis sur les jeux de données que BAnQ propose d’ouvrir (voir liste).

Liste des jeux de données que BAnQ propose d'ouvrir

Type de jeu	Objet	Contenu
Autorité	Artistes, créateur-trice-s, chercheur-euse-s	Nom principal; ISNI; Date de naissance; Pays de résidence; Activité (secteur); Profession; Genre; Langue.
Palmarès	Documents québécois les plus empruntés	Notices bibliographiques des documents physiques et numériques (livre, musique, audiovisuel).
Listes bibliographiques	Collection physique et numérique	Description bibliographique des documents (notices bibliographiques).
Statistiques d'usage	Utilisation des services selon les publics, agrégation par catégories d'âge.	Nombre d'abonnements et de renouvellements d'abonnements; Nombre de prêts et de réservations de documents; Nombre de réservations et temps d'utilisation total des postes.

Document pense-bête utilisé par l'animateur des entrevues

Entrevue

1. *La place de données dans le cadre de vos activités et la nature des données que vous produisez.*

1.1 Quelle place occupent les données dans le cadre des activités de votre organisme/institution?

Quelles sont les activités qui impliquent une utilisation de données?

Quels types de données sont utilisés dans ces activités?

S'agit-il principalement de données descriptives ou de données d'usage?

1.2 Produisez-vous des jeux de données dans le cadre de vos activités?

De quel type de données s'agit-t-il? S'agit-il de données descriptives ou de données d'usage?

Partagez-vous ces données? Qui sont les destinataires principaux des données que vous partagez?

2. *Votre utilisation de données ouvertes, notamment les données de BAnQ sur la plateforme Données Québec.*

2.1 Utilisez-vous des données ouvertes? (Si pas utilisées, passer à 5.)

Quelles sont les sources des données ouvertes que vous utilisez?

Organismes culturels? BAnQ? Collection musicale sur BAnQ Numérique? Autres bibliothèques publiques?

Connaissez-vous la plateforme *Données Québec*?

Si oui. Y avez-vous recours?

Si non. Y a-t-il des raisons particulières?

Quels types de données ouvertes utilisez-vous?

Données descriptives, d'usage? (Préciser, au besoin : Et pour ce qui est des données de BAnQ?)

Quel type de données ouvertes utilisez-vous principalement? (Et pour ce qui est des données de BAnQ?)

Utilisez-vous les données ouvertes pour créer des outils? Pouvez-vous nous donner des exemples de tels outils?

3. Vos besoins en données ouvertes de BAnQ et les enjeux relatifs aux données ouvertes et à leur utilisation au sein de votre organisme/institution.

3.1 Quels sont les enjeux relatifs à la nature des données que vous utilisez, notamment les données ouvertes de BAnQ qu'on trouve sur la plateforme *Données Québec*?

Trouvez-vous que les données sont exhaustives?

(exhaustivité des données ouvertes : proportion de données sélectionnées pour publication sur le total de données disponibles)

Êtes-vous en mesure d'intégrer les données dans votre propre système de données sans devoir recourir à des transformations majeures?

Si non. Y a-t-il des éléments dans les jeux de données qui pourraient être modifiés afin de vous permettre de les utiliser plus facilement? Dictionnaires de données exhaustifs, tableaux de correspondances (lorsque les données sont encodées), une normalisation des données?

Trouvez-vous les données suffisamment granulaires, par exemple, temporellement (données à échelle mensuelle, plutôt qu'annuelle) ou socio-démographiquement (à l'échelle de l'individu, plutôt que de l'agrégat d'individus), ...?

(granularité : niveau de désagrégation)

Pourriez-vous donner un exemple de modèle de données ouvertes que vous trouvez particulièrement intéressant dans votre discipline?

3.2 Y a-t-il des enjeux au niveau de l'utilisation des données au sein de votre organisme/institution?

Par exemple, au niveau de la connaissance des données, ou des enjeux organisationnels ou de gouvernance?

Si oui. (Voir section *Freins à l'utilisation des données ouvertes - questions auxiliaires*)

4. *Votre avis sur les jeux de données que BANQ propose d'ouvrir (voir liste).*

4.1 Données pas encore ouvertes dont l'ouverture pourrait répondre à des besoins.

Y a-t-il parmi ces jeux de données des données dont l'ouverture serait prioritaire de votre point de vue?

Y a-t-il d'autres données qui pourraient vous être utiles?

5. Quelles sont les raisons pour lesquelles vous n'utilisez pas de données ouvertes?

Connaissez-vous la plateforme *Données Québec*?

Les données ouvertes actuellement disponibles (généralement et sur DQ) ne répondent pas aux besoins de votre organisme/institution?

Y a-t-il des freins à l'utilisation de données ouvertes? Lesquels?

(Voir section Freins à l'utilisation des données ouvertes - questions auxiliaires)

6. Projets en partenariat avec BANQ numérique et mettant en valeur les données

Exemples de projets utilisant des données ou métadonnées de BANQ numérique:

- [Montréal d'un autre siècle: cartographie des images des albums de rues E.-Z. Massicotte](#)
- [Cartes-index des plans d'assurance-incendie de Montréal](#) (une réalisation du Laboratoire d'histoire et de patrimoine de Montréal)
- [Carte-index des plans d'utilisation du sol de ville de Montréal](#) (une réalisation du Laboratoire d'histoire et de patrimoine de Montréal)
- [Carte-index des atlas de Montréal de Charles Goad](#) (une réalisation du Laboratoire d'histoire et de patrimoine de Montréal)

Freins à l'utilisation des données ouvertes - questions auxiliaires

Q. Y a-t-il des freins à l'utilisation de données ouvertes sur un plan relatif à la **connaissance** des données dans votre organisme?

Pensez-vous que le niveau de connaissance de votre organisme de la disponibilité de données ouvertes pourrait être amélioré?

Pensez-vous que le niveau de connaissance de votre organisme quant au potentiel d'analyse ou de création d'outils au moyen de données ouvertes pourrait être bonifié?

Pensez-vous que la visibilité des jeux de données ouvertes relatives au milieu culturel québécois est adéquate?

Q. Quels sont les freins à l'utilisation des données ouvertes sur un plan **organisationnel**?

Pensez-vous que le niveau des besoins en données de votre organisme ne justifie pas l'utilisation systématique de données ouvertes?

Pensez-vous que votre organisme ne dispose pas des effectifs nécessaires pour assimiler et employer de façon régulière les données ouvertes?

Q. Quels sont les freins à l'utilisation des données ouvertes sur le plan de la **gouvernance** de données?

Des enjeux de confidentialité et d'anonymat (anonymisation, agrégation des données) affectent-ils votre utilisation de données ouvertes (ou autres)?

Conclusion

Discussion informelle. Remerciements. Fin.