

L'ouverture des données, pilier de la découvrabilité des contenus culturels de BAnQ

Guillemette Jessica, chargée de produits numériques
Russell Orphée, bibliothécaire de soutien et de développement
Niort Lucile, analyste en Intelligence d'affaires

Notre essor culturel à l'ère des algorithmes



PHOTO FRANÇOIS ROY, ARCHIVES LA PRESSE

Bibliothèque et Archives nationales du Québec (BAnQ) s'intéresse de près à la visibilité du contenu québécois lors d'une recherche en ligne.

« À l'ère des algorithmes, la découvrabilité devient la condition de tous les heureux hasards qui permettent à une culture de se construire. BAnQ s'intéresse vivement à cet enjeu, car il touche à l'essence de son action. Rendre accessibles les idées et les mots du Québec, encourager leur exploration et favoriser ces découvertes fortuites qui changent une journée ou une vie se trouve au cœur de nos missions. »

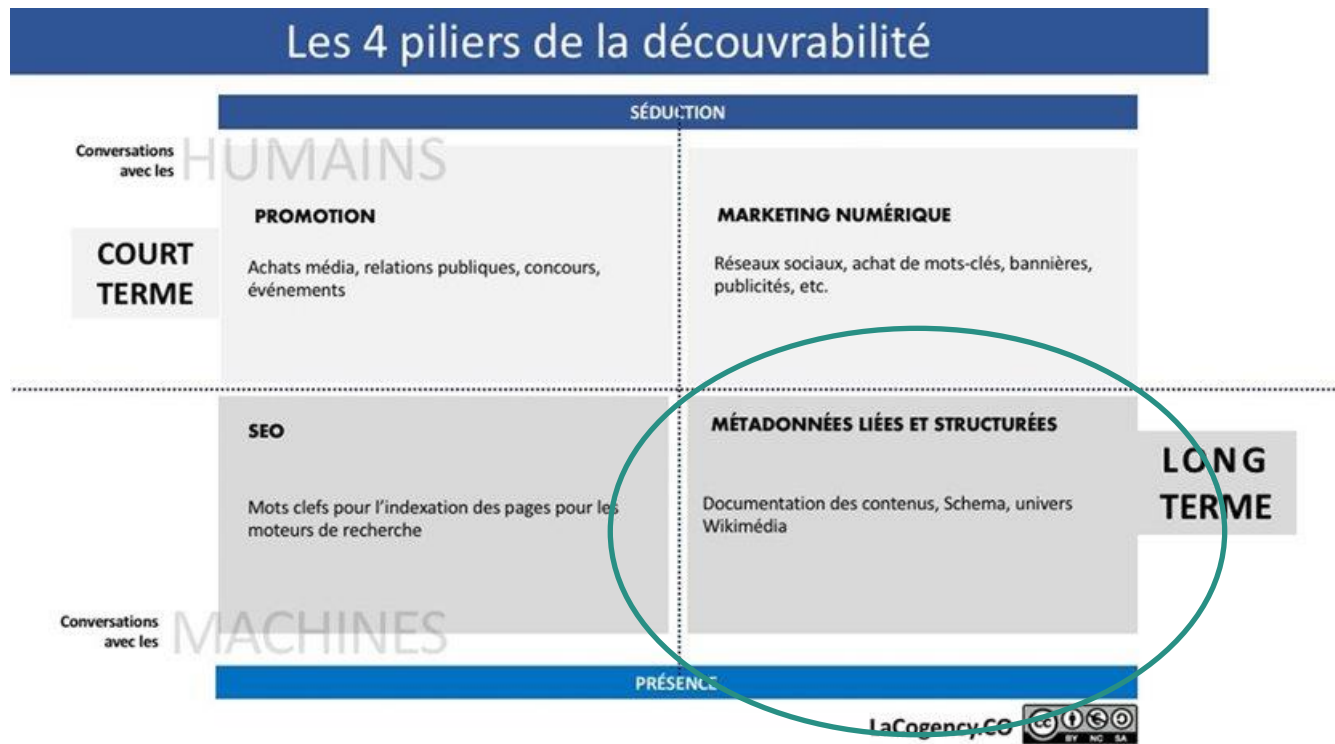
Source : La Presse (20 mars 2023)

Définition

« La découvrabilité est la capacité d'un contenu culturel de se laisser découvrir aisément par le consommateur qui le recherche et de se faire proposer au consommateur qui n'en connaissait pas l'existence. »

- Observatoire de la culture et des communications du Québec (2017)

Les quatre piliers de la découvrabilité



Métadonnées (et données) liées et structurées

Métadonnées : données servant à définir, décrire et/ou contextualiser d'autres données et leur permettant de faire sens. Ex. nom d'une organisation, date de diffusion d'un document, auteur d'un livre, etc.

Données structurées : données définies par des balises selon une structure propre à la plateforme de recherche (ex. Google). *Exemple d'action : balisage Schema de nos pages Web.*

Données liées : données de différentes plateformes qui sont liées entre elles à l'aide d'un hyperlien ou d'un identifiant unique. *Exemple d'action : utilisation d'identifiants pérennes et uniques (ISBN, ISNI, etc.)*

Les données structurées et liées peuvent être mises en relation, favorisant ainsi les partenariats et la collaboration entre différentes institutions.

BAnQ et la mutualisation de métadonnées

Depuis 2021, plusieurs demandes de projets de mutualisation de métadonnées à la DMNB (Direction des métadonnées et de la normalisation bibliographiques)

- Organismes des domaines du livre, de la musique, des arts de la scène et des arts visuels

Pour répondre aux défis de découvrabilité sur le Web

- Enjeux primordiaux
 - En 2021, seulement 7% des pistes musicales écoutées sur les plateformes de diffusion au Québec étaient québécoises – Le Devoir (2021)

Lien entre les demandes = l'identifiant ISNI (International Standard Name Identifier)

- « Identifiant unique et pérenne attribué au niveau international à toute personne, [tout] organisme ou personnage de fiction impliqué dans les chaînes de création » - ministère de la Culture et des Communications (2021)

Pourquoi l'ISNI?

ISNI = clé permettant de lier des métadonnées de différentes sources

Court/moyen terme

- Identifier les contributeurs
 - Dans le catalogue de BAnQ, 16 personnes avec le nom Michel Tremblay et 10 personnes avec le nom Pierre Lapointe – lequel est le bon?
- Minimiser le nombre de saisies et de vérifications manuelles
- Automatiser les échanges de données

Long terme

- Favoriser la création d'écosystèmes de données
- Encourager la transition vers les données liées

<input type="checkbox"/>	Lapointe, Pierre
<input type="checkbox"/>	Lapointe, Pierre, 17e siècle
<input type="checkbox"/>	Lapointe, Pierre, 1932-
<input type="checkbox"/>	Lapointe, Pierre, 1941-
<input type="checkbox"/>	Lapointe, Pierre, 1944-
<input type="checkbox"/>	Lapointe, Pierre, 1950-
<input type="checkbox"/>	Lapointe, Pierre, 1957-
<input type="checkbox"/>	Lapointe, Pierre, 1961-
<input type="checkbox"/>	Lapointe, Pierre, 1981-
<input type="checkbox"/>	Lapointe, Pierre (Producteur)

BAnQ et l'ISNI

BAnQ = agence ISNI pour le Québec

- Organisme bien placé pour remplir ce rôle grâce à l'exhaustivité de nos collections


Comment obtenir un ISNI de BAnQ

- Avoir contribué à un document dans la Collection nationale
- Remplir le formulaire Web ISNI sur le portail



Demande d'ISNI

Vous pouvez utiliser le formulaire Web pour faire une demande d'ISNI si les deux conditions ci-dessous sont remplies :

- vous avez créé une production culturelle québécoise, ou vous y avez contribué;
- cette production culturelle québécoise n'est pas assujettie au [dépôt légal](#) .

Exemple de projet de partage d'ISNI

Projet à venir avec la BTLF (Banque de titres de langue française)

- But = implanter l'ISNI dans les chaînes de travail du milieu du livre
 - Pousser les ISNI aux éditeurs
 - Automatiser les échanges de données entre les acteurs

Étapes de la collaboration

- Envoi massif d'ISNI d'auteurs à la BTLF
- Mise en place d'un flux d'échange de données régulier pour les nouveaux auteurs



Données ouvertes

« On définit les données ouvertes comme des données diffusées de manière **structurée**, en **format ouvert**, sous forme de fichiers numériques, d'origine publique ou privée, qui peuvent être **librement partagées, utilisées et mises à profit** par quiconque, **sans restriction technique, juridique ni financière**, conformément à une licence garantissant leur **libre accès et leur réutilisation**. »

- *Pour des données ouvertes en culture*, Gouvernement du Québec

Exemples de plateformes proposant des données (ou des jeux de données) ouvertes :

- Wikidata : multisectorielle, collaborative et multilingue
- Données Québec : villes, municipalités et organismes
- Données ouvertes de la Ville de Montréal : administration publique

Spectre d'ouverture des données



Publiez vos données sur le Web avec une licence ouverte (ex. Creative Commons).

Publiez vos données en tant que données structurées.

Publiez vos données dans un format ouvert et non-propriétaire (ex. CSV plutôt qu'Excel).

Utilisez des URI (identifiant uniforme de ressource) pour identifier les objets dans vos données.

Liez vos données à d'autres données pour y ajouter du contexte.

Source : <http://5stardata.info/>

Principes FAIR

DÉCOUVRABLE (FACILE À DÉCOUVRIR)

Les données sont accompagnées de métadonnées riches et d'identifiants uniques et pérennes qui en favorisent la découvrabilité.

INTEROPÉRABLE

Les données peuvent être liées à d'autres données, par exemple à l'aide d'un langage de programmation commun ou d'un vocabulaire contrôlé.

ACCESIBLE

Les données peuvent être obtenus facilement tant par les humains que par les machines et leurs conditions d'accès sont clairement établies.

RÉUTILISABLE

Les données sont richement décrites et sous une licence clairement identifiée et idéalement ouverte, ce qui en favorise la réutilisation.

Source : <https://ogsl.ca/fr/principes-fair/>

Avantages d'ouvrir ses données

Pour les administrations et les organisations	Pour les citoyens et les usagers	Pour la société
<ul style="list-style-type: none">• Offre un meilleur accès aux données d'intérêt public;• Favorise et accroît la transparence des institutions;• Améliore l'efficacité organisationnelle et assure une plus grande gouvernance;• Permettre de générer de la valeur;• Améliore la visibilité dans le monde numérique.	<ul style="list-style-type: none">• Améliore la connaissance et l'autonomie individuelle;• Permet de s'approprier les données et les utiliser de façon créative.	<ul style="list-style-type: none">• Améliore la découvrabilité des données et contenus culturels québécois;• Participe à la vitalité économique, sociale et écologique;• Favorise la démocratie• Favorise la création d'une société apprenante.

Exemple de jeu favorisant la découvrabilité : les films



BESOIN

- Liste des films empruntables à la Grande Bibliothèque
- Demande initiale de *Mediafilm* pour ouvoir.ca
- Janvier 2022 -> Diffusion en Septembre 2022



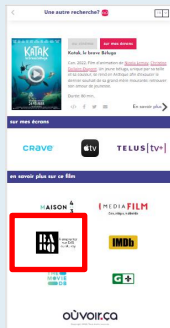
JEU DE DONNÉES

- Liste des Films sur support physique (DVD, blu-ray..)
- Environ 70k films.
- 20 métadonnées: titre, date, durée, langue, catégorie, sujets....

Exemple de jeu favorisant la découvrabilité : les films

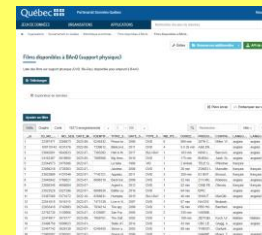
DÉCOUVRABLE

Les films sont découvrables à partir de Ouvoir.ca



ACCESSIBLE

- Dépôt sur Données Québec
- Mise à jour hebdomadaire



INTEROPÉRABLE

- Exploitable via API Données structurées et harmonisées selon les règles directrices Médiaparc



RÉUTILISABLE

- Format CSV
- Licence ouverte

✓ Format : CSV

✓ Licence :  Attribution (CC-BY 4.0)

Exemple de jeu fa

abilité : les films

DÉCOUVRABLE

Les films sont découvrables à partir de Ouvoir.ca

INTEROPÉRABLE

- Exploitable via API
- Données structurées et harmonisées selon les règles directrices Médiafil

The screenshot shows a mobile application interface for a movie. At the top, there is a search bar with the text "Une autre recherche?" and a magnifying glass icon. Below the search bar, there is a movie poster for "KATAK le brave Béluga" with a play button overlay. To the right of the poster, there are two buttons: "au cinéma" and "sur mes écrans". Below the poster, the title "Katak, le brave Béluga" is displayed, followed by the text "Can. 2022. Film d'animation de [Nicola Lemay](#), [Christine Delaine](#), [Dujont](#). Un jeune béluga, unique par sa taille et sa couleur, se rend en Arctique afin d'exaucer le dernier souhait de sa grand-mère mourante: retrouver son amour de jeunesse." and "Durée: 80 min.". There are social media icons (Facebook, Twitter, Email) and a link "En savoir plus". Below the movie information, there is a section "sur mes écrans" with logos for "CRAVE", "Apple TV", and "TELUS [tv+]". Underneath, there is a section "en savoir plus sur ce film" with logos for "MAISON 4/3", "MEDIAFILM", "IMDb", "THE MOVIE DB", and "G". At the bottom, there is a logo for "ouvoir.ca" and the text "Copyright 2020. Tous droits réservés".

Données

aire

SV
ouverte

SV

Attribution (CC-BY 4.0)

Exemple de jeu favorisant la découvrabilité : les ISBN



- Liste des documents publiés au Québec depuis 2010
- Jeu de données de référence publié sur Données Québec depuis 2021
- Mis à jour hebdomadairement
- Métadonnées suivants les principes Dublin Core

Bénéfices pour les bibliothèques

Plusieurs bibliothèques nationales mettent en place des flux d'échanges de métadonnées bidirectionnelles avec des organismes du milieu culturel

- Générer des notices en lien avec le CIP (catalogage avant publication) et/ou le dépôt légal
- Générer des ISNI et les repousser aux organismes du milieu

Projet avec la BTLF : gains potentiels pour BAnQ

- Demander aux éditeurs de fournir les ISNI dans les demandes CIP et formulaires de dépôt légal
- Générer automatiquement des liens entre les notices bibliographiques et les notices d'autorité dans le catalogue

Projet avec MétaMusique pour le dépôt légal de la musique numérique : contexte

- Immense augmentation du nombre de documents à traiter
- Même nombre de personnes pour les traiter

Réutiliser les données de MétaMusique

MétaMusique = base de données mise en place par les sociétés de gestion de droits d'auteur en musique

- Minimiser le nombre de saisies pour les acteurs

Buts de notre collaboration

- Recevoir les fichiers de musique
- Générer des pré-notices au catalogue avec les métadonnées de la base

Les données de MétaMusique

- Structure équivalente à celle des notices bibliographiques et d'autorité
- Indexation



Encadrer les initiatives : Politique de partage et de valorisation de données

Principe d'ouverture des données par défaut



« Les données institutionnelles sont ouvertes par défaut, sauf si une justification sérieuse et documentée démontre qu'il doit en être autrement »

Rôle et responsabilités



Le détenteur des données

Le gestionnaire des données

Le comité consultatif

Processus de partage



Formulaire d'autorisation de partage à compléter par le détenteur des données, et à valider par le gestionnaire des données.

L'ouverture des données, pilier de la découvrabilité des contenus culturels de BAnQ

Merci pour votre écoute !
Questions ?

Guillemette Jessica, chargée de produits numériques
Russell Orphée, bibliothécaire de soutien et de développement
Niort Lucile, analyste en Intelligence d'affaires