



OBSERVATOIRE
DE LA CULTURE ET DES
COMMUNICATIONS
DU QUÉBEC

Conception d'indicateurs de rayonnement de l'offre culturelle québécoise dans les réseaux numériques

Étude de faisabilité

Février 2021

Pour tout renseignement concernant l'Institut de la statistique du Québec (ISQ) et les données statistiques dont il dispose, s'adresser à :

Institut de la statistique du Québec
200, chemin Sainte-Foy
Québec (Québec) G1R 5T4

Téléphone :
418 691-2401
1 800 463-4090 (Canada et États-Unis)

Site Web : statistique.quebec.ca

Ce document est disponible seulement en version électronique.

Dépôt légal
Bibliothèque et Archives nationales du Québec
4^e trimestre 2021
ISBN 978-2-550-90850-0 (en ligne)

© Gouvernement du Québec, Institut de la statistique du Québec, 2021

Toute reproduction autre qu'à des fins de consultation personnelle est interdite sans l'autorisation du gouvernement du Québec.
statistique.quebec.ca/fr/institut/nous-joindre/droits-auteur-permission-reproduction

Décembre 2021

Publication réalisée à
l'Institut de la statistique du Québec par :

Marik Danvoye
Chargée de recherche
Observatoire de la culture
et des communications du Québec

Avec la collaboration de :

Dominique Jutras
Directeur sortant
Observatoire de la culture
et des communications du Québec

Harold Gendron
Consultant représentant le ministère
de la Culture et des Communications
(Cellule numérique)

Sous la direction de :

Christine Routhier
Directrice par intérim
Observatoire de la culture
et des communications du Québec

Révision linguistique et édition :

Direction de la diffusion
et des communications

Financé par :

Ministère de la Culture
et des Communications

Pour tout renseignement concernant
le contenu de cette publication :

Institut de la statistique du Québec
200, chemin Sainte-Foy
Québec (Québec) G1R 5T4
Téléphone :
418 691-2401
1 800 463-4090 (Canada et États-Unis)
Site Web : statistique.quebec.ca

Notice bibliographique suggérée

DANVOYE, Marik (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].

Table des matières

Introduction	7
1 Les études sectorielles	11
1.1 Approche 1 : L'assemblage de statistiques déjà colligées par différentes firmes ou par les outils d'analytique Web des plateformes de diffusion.	11
1.1.1 Constats généraux	11
1.1.2 Livre	12
1.1.3 Audiovisuel	16
1.1.4 Musique enregistrée et spectacles en arts de la scène	19
1.2 Approche 2 : La création d'un outil de moissonnage du Web qui permettrait d'analyser les différents marqueurs du rayonnement	23
1.2.1 Démarche et principaux enjeux	23
1.2.2 Livre	25
1.2.3 Audiovisuel	27
1.2.4 Musique enregistrée	30
1.2.5 Arts de la scène	33
1.2.6 Patrimoine	34
1.3 Évaluation des sources et recensement d'indicateurs potentiels	36
1.3.1 Présence	36
1.3.2 Visibilité	37
1.3.3 Recommandation	38
1.3.4 Consommation	39
2 Synthèse	41
2.1 Enjeu	41
2.2 Observations générales	41
2.3 Observations techniques	42
3 Proposition d'actions à mettre en œuvre	43
4 Recommandations	47
Description des travaux du comité et des mandataires	51

Introduction

Contexte

Le ministère de la Culture et des Communications (MCC) a soutenu, dans le cadre du Plan culturel numérique du Québec (PCNQ), diverses mesures visant non seulement à accroître la présence et la disponibilité de l'offre culturelle québécoise sur les réseaux numériques, mais aussi à stimuler la visibilité et la découvrabilité de cette offre sur ces mêmes réseaux. Le ministère souhaite maintenant s'outiller pour tenter d'évaluer différentes dimensions relatives à la présence numérique de la culture québécoise. L'obtention d'indicateurs qui pourraient témoigner du rayonnement de l'offre culturelle québécoise dans les réseaux numériques et qui seraient basés sur des données quantitatives va en ce sens.

Le MCC a mandaté l'Observatoire de la culture et des communications de l'Institut de la statistique du Québec pour réaliser une étude de faisabilité visant à analyser le cadre théorique et méthodologique ainsi que les conditions entourant l'obtention de ce type d'indicateurs. Les objectifs de l'étude de l'Observatoire sont les suivants :

- Comprendre le système référentiel de plusieurs concepts dont, entre autres, ceux de rayonnement, d'offre culturelle, de contenu québécois et de réseaux numériques ;
- Étudier les mécanismes et les canaux du rayonnement de l'offre culturelle dans les domaines du livre, de la musique enregistrée, du spectacle en arts de la scène, de l'audiovisuel et du patrimoine ;
- Déterminer les caractéristiques de l'objet à mesurer, par exemple, le caractère québécois de la production culturelle ;
- Décrire les données existantes qui pourraient servir à produire des indicateurs de rayonnement, en précisant leur disponibilité et leurs limites ;
- Révéler les options les plus prometteuses et concrètes qui s'offrent pour la production à brève échéance d'indicateurs de rayonnement, et énoncer les forces et les faiblesses de chacune d'elles ;
- Ébaucher le cheminement logique à suivre et les principales étapes à franchir pour obtenir ces éventuels indicateurs ;
- Énoncer, s'il y a lieu, les actions à entreprendre et les conditions à réunir en vue d'un processus de construction d'indicateurs de rayonnement.

En somme, la présente étude consiste à déterminer s'il serait possible de produire ou d'acquérir des indicateurs statistiques du rayonnement des produits culturels québécois dans l'univers numérique et si oui, comment procéder à cette fin.

Description des travaux et identification des parties prenantes

La conception d'un ou de plusieurs indicateurs du rayonnement de l'offre culturelle québécoise dans les réseaux numériques est un défi extrêmement complexe. Chaque domaine culturel a sa propre dynamique, caractérisée par une diversité d'intervenants, par la présence de plateformes nationales et internationales, par des pratiques de production, de diffusion et d'exploitation distinctes et par des normes et des règles de fonctionnement particulières.

La conception d'indicateurs est aussi assujettie à la disponibilité de données pertinentes, accessibles et compilables se prêtant au traitement statistique. Compte tenu de cette complexité, le MCC et l'Observatoire ont mis sur pied un comité de réflexion composé d'experts de divers horizons, de consultants externes et de représentants de différents milieux afin de mettre en commun les connaissances et les expertises des diverses parties prenantes.

Le déroulement des travaux du comité et de ceux des consultants et mandataires est présenté à l'annexe 1.

Les marqueurs de rayonnement

Les considérations théoriques et les définitions opérationnelles des concepts dont on doit tenir compte pour la conception d'indicateurs ont fait partie des réflexions des différents comités qui ont participé au projet.

Des nombreux échanges avec le comité de réflexion et le comité de concertation de l'Observatoire est ressorti un schéma des divers « marqueurs de rayonnement », qui illustre les principales dimensions du rayonnement (figure 1).

Ces travaux ont fait émerger plusieurs questionnements à propos des enjeux associés à la mesure du rayonnement de l'offre culturelle québécoise dans les réseaux numériques. Les rapports des consultants en font largement état. Ces enjeux ainsi que les réflexions qu'ils ont suscitées sont résumés dans le présent rapport.

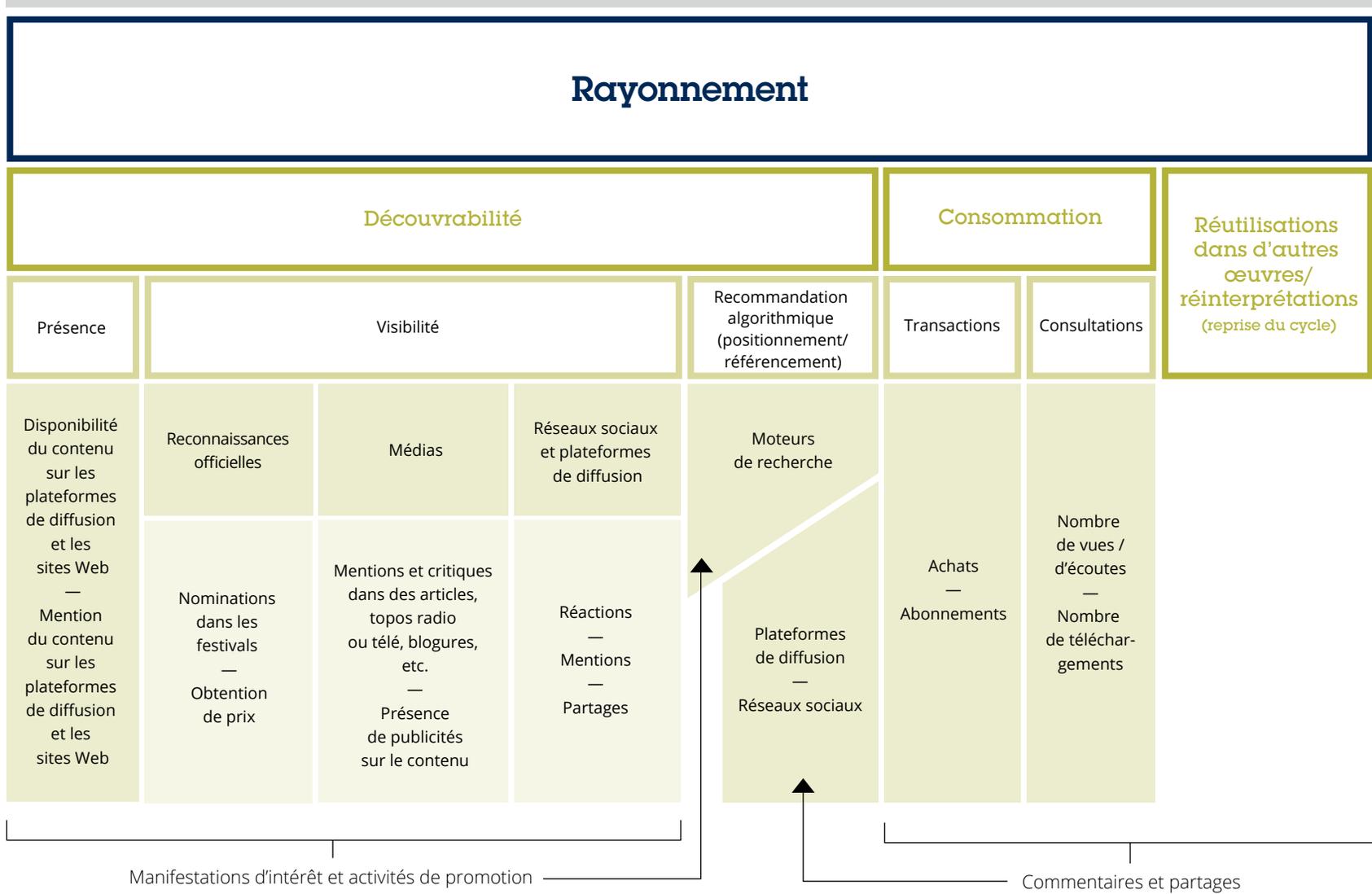
Des constats relatifs à la difficulté d'établir des balises claires pour témoigner du rayonnement sont ressortis d'emblée de ces échanges :

- L'objet dont le rayonnement est à mesurer – d'une dimension du rayonnement à l'autre, l'objet mesuré peut être l'œuvre elle-même ou ses contributeurs (créateurs, producteurs, interprètes, diffuseurs, etc.), ce qui complexifie le recensement de l'offre culturelle à mesurer. Par exemple, on peut dénombrer les adeptes d'un artiste, mais pas ceux d'une chanson.
- La « québécoïté de l'offre culturelle » – le caractère québécois d'un contenu est peu indiqué dans les réseaux numériques et les critères qui permettent de le déterminer objectivement ne sont pas nécessairement disponibles. Il est donc difficile de procéder autrement qu'à partir d'un inventaire prédéfini, en espérant qu'il soit le plus exhaustif possible.
- La portée du rayonnement des contenus québécois – le rayonnement des contenus culturels québécois peut être mesuré par marché, qu'il soit québécois, canadien, nord-américain, francophone ou étranger. Toutefois, aucune source ne permet de mesurer de façon exhaustive la portée mondiale de l'ensemble des contenus d'un domaine culturel donné et il peut s'avérer particulièrement difficile de croiser différentes sources en ce sens, s'il est même possible de déterminer la provenance des auditoires.

- La nouveauté de l'offre culturelle dont le rayonnement est à mesurer et la question du rayonnement social (visibilité auprès du public, conservation des contenus dans des institutions clés, etc.) ou économique (volume de contenus consommés/vendus) – les sources de données à privilégier et la fréquence de la mesure ne seront pas les mêmes si l'on s'intéresse plus particulièrement aux nouveautés ou au catalogue.

Enfin, il s'est avéré difficile de déceler des pistes d'actions convergentes et concrètes à partir de considérations théoriques lors des travaux des comités et dans les rapports reçus des mandataires, bien que tous ces collaborateurs aient des expériences très diversifiées. Néanmoins, à la suite d'une analyse approfondie de ces constats, des actions à mettre en œuvre ont été proposées ; celles-ci figurent au chapitre 3.

Figure 1
Les marqueurs de rayonnement



1

Les études sectorielles

1.1 Approche 1 : L'assemblage de statistiques déjà colligées par différentes firmes ou par les outils d'analytique Web des plateformes de diffusion.

Rappel des objectifs : décrire les mécanismes et les canaux du rayonnement dans un domaine culturel donné ; dresser un inventaire des marqueurs du rayonnement de l'offre québécoise pouvant être mesurés grâce aux sources de données existantes ; évaluer les complémentarités, les différences, les dédoublements et les lacunes des sources de données recensées et recommander la stratégie qui semble la plus appropriée pour produire des indicateurs de rayonnement dans leurs domaines respectifs.

1.1.1 Constats généraux

- Les trois rapports obtenus des consultants reposent principalement sur une approche théorique de la recherche-action : quelques indicateurs ou éléments de mesure sont proposés (ou nommés), et les mécanismes du rayonnement décrits demeurent plutôt traditionnels.
- Les trois rapports accordent plus d'attention aux déterminants (ou facteurs) du rayonnement qu'aux marqueurs du rayonnement.
 - Les trois consultants ont mentionné l'importance d'accroître la découvrabilité par l'implantation de bonnes pratiques en matière de métadonnées. Ils estiment que cette question est primordiale pour amplifier la commercialisation des contenus culturels sur le Web.
 - D'ailleurs, chacun d'eux souligne que la production d'indicateurs de rayonnement serait grandement facilitée si l'ensemble des contenus culturels québécois disposait de métadonnées de qualité.
- Plusieurs caractéristiques des sources de données potentielles décrites par les consultants demeurent par ailleurs floues (« données d'usage », « positionnement dans les réseaux sociaux », « données privées »). Aussi, sur les unités de mesure disponibles dans les sources analysées, peu ont été décrites comme des indicateurs valables de rayonnement.
 - Les mandataires ont décrit quelques sources primaires d'information ainsi que des bases de référence qui permettraient de dresser des listes de productions québécoises (et parfois étrangères) par domaine, celles-ci pouvant être utilisées lors d'opérations de moissonnage.
 - Cependant, il semble que recueillir des renseignements détaillés sur les sites, plateformes et applications est très difficile.
 - Un abonnement payant serait minimalement requis, mais ne garantirait pas pour autant un accès satisfaisant aux données recueillies par ces sites, entre autres parce que le volume et l'ampleur des requêtes sont limités.
 - Compte tenu de l'échéancier et de la diversité des objectifs imposés aux mandataires de l'approche 1, ceux-ci ont effectué l'inventaire des sources potentielles d'indicateurs, mais la description et l'évaluation des indicateurs disponibles dans ces sources demeurent partielles.
 - Dans ce contexte, les recommandations énoncées dans les rapports doivent être considérées avec réserve. Des travaux exploratoires devraient être réalisés pour évaluer la portée des résultats qui en émergeraient.

1.1.2 Livre

Synthèse et analyse du rapport réalisé par la Société de gestion de la banque de titres de langue française (BTLF) : *Projet de recherche d'indicateurs de rayonnement de l'offre culturelle québécoise dans les réseaux numériques dans le domaine du livre*

Le rapport de BTLF fournit un bon nombre de sources potentielles de mesures. Dans leur description des mécanismes et des canaux du rayonnement, on porte une attention particulière à l'identification de marqueurs favorisant le rayonnement, par exemple la présence de la couverture d'un livre ou de métadonnées structurées (titre, auteur, éditeur, etc.). Il y a moins de renseignements sur les marqueurs du rayonnement proprement dits (abonnements, achats, consultations, téléchargements, etc.).

Les mesures décrites sont priorisées à partir des besoins en informatique décisionnelle de l'industrie du livre : par exemple, on porte de l'intérêt à la « note moyenne » accordée à un livre (une évaluation subjective) tout en reconnaissant que le fait qu'un livre a été repéré par un nombre de lecteurs potentiels sur les réseaux numériques serait un marqueur de rayonnement plus probant.

D'après les pistes proposées par BTLF, quelques indicateurs pourraient être calculés pour analyser différentes dimensions du rayonnement des livres québécois. Le tableau 1 présente une synthèse de leur recherche.

- Ces indicateurs ne sont peut-être pas tous pertinents, mais ils témoignent des limites de ce qui peut être mesuré à partir des sources de données existantes dans le domaine du livre.
- Dans le tableau, la colonne « Statut » distingue les indicateurs « disponibles » des indicateurs « théoriques », c'est-à-dire dont la faisabilité ou l'accès aux données pour les calculer restent à démontrer.
- Dans la colonne « Portée » figurent les sources dont les données ne couvrent que le territoire québécois.

Tableau 1

Indicateurs potentiels formulés à partir des analyses présentées dans le *Projet de recherche d'indicateurs de rayonnement de l'offre culturelle québécoise dans les réseaux numériques dans le domaine du livre*

Indicateurs potentiels	Sources potentielles	Statut	Portée
Dimension – Présence			
<ul style="list-style-type: none"> Pourcentage de titres québécois disponibles parmi l'ensemble des titres échantillonnés sur un groupe défini de plateformes transactionnelles¹ (taux de présence des titres chez les détaillants) 	Hercules	Disponible	
<ul style="list-style-type: none"> Score SEO moyen d'une liste de sites d'éditeurs québécois 	Outils d'analytique Web/Hercules (à dériver des requêtes)	Théorique	
<ul style="list-style-type: none"> Part des acquisitions de titres numériques québécois par rapport à l'ensemble des acquisitions de titres numériques des bibliothèques publiques québécoises 	Bibliopresto	Théorique	Québec
	OCCQ ²	Disponible	Québec
<ul style="list-style-type: none"> Part des titres québécois dans le catalogue d'un groupe défini de bibliothèques étrangères 	Worldcat	Théorique	
<ul style="list-style-type: none"> Part des titres québécois dans les bibliothèques québécoises 	Worldcat	Théorique	Québec
	OCCQ ³	Disponible	Québec
Dimension – Visibilité			
Plateformes de diffusion <ul style="list-style-type: none"> Nombre moyen d'évaluations Nombre de produits avec évaluation Nombre moyen d'évaluations pour les produits qui ont au moins une évaluation Note moyenne de l'évaluation 	Hercules	Disponible	
<ul style="list-style-type: none"> Analyse des résultats de moteurs de recherche par catégorie de sites (médias, réseaux sociaux, etc.) 	Hercules (mesure synthèse à élaborer)	Théorique	
Réseaux sociaux <ul style="list-style-type: none"> Sentiments (« J'aime » et expression d'émotions) Engagement (nombre d'interactions/commentaires) Nombre d'évaluations Volume de mentions de titres québécois sur les réseaux sociaux Portée : nombre d'adeptes atteints par la publication d'un titre québécois (d'un artiste québécois) via le fil d'actualités 	Outils de réputation numérique ⁴ qui moissonnent les principaux réseaux sociaux ⁵ (données non accessibles)	Théorique	
<ul style="list-style-type: none"> Pourcentage de titres québécois en nomination parmi les titres en nomination aux prix Goncourt, Femina, Médicis, Renaudot et Interallié 	(non précisée)	Théorique	
<ul style="list-style-type: none"> Pourcentage de titres québécois ayant remporté un prix Goncourt, Femina, Médicis, Renaudot ou Interallié 			
Dimension – Recommandations			
<ul style="list-style-type: none"> Nombre moyen de sujets associés aux titres québécois sur les plateformes de diffusion 	Hercules	Disponible	

Suite à la page 14

Tableau 1 (suite)

Indicateurs potentiels formulés à partir des analyses présentées dans le *Projet de recherche d'indicateurs de rayonnement de l'offre culturelle québécoise dans les réseaux numériques dans le domaine du livre*

Indicateurs potentiels	Sources potentielles	Statut	Portée
Dimension – Consultation			
Plateformes de diffusion – sites éditeurs-auteurs de l'UNEQ/ANEL <ul style="list-style-type: none"> • Nombre de pages visitées • Nombre de visites sur la page • Durée des visites sur la page • Taux de rebond (usagers qui ont quitté dès la première page vue) 	Systèmes d'analytique Web	Théorique	
Dimension – Consommation			
• Pourcentage de titres québécois empruntés à BAnQ	BAnQ	Disponible	Québec
• Pourcentage de titres numériques québécois empruntés sur les principaux sites de prêt numérique québécois	Bibliopresto	Disponible	Québec
• Pourcentage de livres numériques québécois achetés dans les principales librairies en ligne (y compris le livre audio)	Demarque	Disponible	Québec
	OCCQ ⁶	Non diffusées	Québec

1. Hercules en compterait environ 50, dont Amazon Canada, Amazon France, Apple Canada, Apple France, Archambault, Fnac France, Kobo Canada, Les Libraires, Renaud Bray, iTunes Canada (English), iTunes Canada (Français) et Decitre.
2. Ces données sont aussi recueillies annuellement dans le cadre de l'*Enquête annuelle sur les bibliothèques publiques (EABP)* et sont diffusées par l'Observatoire : INSTITUT DE LA STATISTIQUE DU QUÉBEC (2021, mis à jour le 8 novembre), *Statistiques sur les bibliothèques publiques du Québec, Québec*, [En ligne], [\[statistique.quebec.ca/fr/document/statistiques-sur-les-bibliotheques-publiques-du-quebec\]](http://statistique.quebec.ca/fr/document/statistiques-sur-les-bibliotheques-publiques-du-quebec).
3. *Idem*.
4. Mention, Webmii, Keyhole, Notify, Social Searcher, Talkwalker.
5. Facebook, Twitter, Youtube, Instagram, Reddit, Dailymotion, Tumblr, Vimeo, Vkontakte, Flickr. Les outils de réputation numérique mentionnés par BTLF ne semblent pas intégrer Goodreads, Babelio et Quialu, qui pourraient aussi être considérées comme des sources de réactions propres au domaine du livre.
6. Information partiellement disponible dans l'*Enquête sur les ventes de livres numériques au Québec* (données non diffusées). Demarque est un agrégateur parmi d'autres.

De cet exercice ressortent les constats suivants :

- L'accent mis par le mandataire sur la description des possibilités d'Hercules par rapport aux autres sources de données potentielles est intéressant, mais ne permet pas de juger objectivement de la valeur de cette source.
 - À ce jour, Hercules comporte des limites importantes, dont l'impossibilité de remonter dans le temps et l'injection d'un maximum de 1 500 titres par jour¹, dans un nombre limité de sites.
 - Il faut alors constituer un échantillon, ou « panier représentatif » des éléments à mesurer. De plus, s'il est nécessaire de contextualiser le rayonnement de l'offre culturelle québécoise en la comparant à un panier représentatif de l'offre culturelle non québécoise, il faut aussi constituer un panier d'éléments non québécois.

1. Pour référence, à l'échelle du Québec, un peu plus de 8 000 monographies sont déposées au dépôt légal par année : INSTITUT DE LA STATISTIQUE DU QUÉBEC (2021, mis à jour le 29 mars), *Monographies imprimées et numériques publiées au Québec et reçues en dépôt légal, selon la catégorie de sujet, Québec*, [En ligne], [\[statistique.quebec.ca/fr/document/publications-imprimees-et-numeriques-statistiques-a-partir-de-2009\]](http://statistique.quebec.ca/fr/document/publications-imprimees-et-numeriques-statistiques-a-partir-de-2009).

- La constitution et la mise à jour de tels « paniers » à chaque fois que l'on veut mettre à jour les indicateurs à extraire comportent elles-mêmes certaines contraintes, des ajouts ou remplacements aux paniers initialement sélectionnés pouvant introduire artificiellement une variation dans la valeur des indicateurs. Rappelons que l'une des caractéristiques des industries culturelles est « que le gagnant rafle tout », ce qui signifie que la présence ou l'absence d'un titre à succès dans un panier entraînera une grande variation des résultats.
- On doit aussi porter une attention particulière aux sites sélectionnés pour mesurer le rayonnement des œuvres, le nombre et la diversité de sites retenus ayant une incidence sur le nombre de titres pouvant être inclus dans une requête. Ainsi, le maximum de 1 500 titres par jour énoncé plus haut est calculé à partir des sites déjà moissonnés par Hercules.
- Des renseignements sur les prêts de livres numériques et sur le rayonnement des œuvres littéraires en bibliothèques sont brièvement présentés. Des données comparables sont par ailleurs déjà disponibles sur le site de l'Observatoire. La mesure des prêts de livres numériques à l'intérieur des frontières du Québec repose sur Bibliopresto, De Marque et BAnQ.
 - On ne fait pas mention des données disponibles sur les fonds et acquisitions de titres québécois recueillies dans le cadre de l'*Enquête annuelle sur les bibliothèques publiques* (EABP)².
 - Les auteurs ne mentionnent pas non plus les statistiques sur les ventes de livres numériques produites par l'Observatoire à partir d'une enquête auprès d'agrégateurs (*Enquête sur les ventes de livres numériques au Québec*³). Toutefois, aucune donnée sur le nombre de titres de livres numériques québécois vendus n'est diffusée à ce jour, puisque seuls quelques répondants – les entrepôts numériques québécois (agrégateurs ou distributeurs) et les librairies numériques – sont en mesure de distinguer les ventes de livres numériques québécois des ventes de livres numériques non québécois.
 - La proposition d'étudier la part de livres prêtés d'origine québécoise par BAnQ concerne plus ou moins 12 % des prêts réalisés dans l'ensemble des bibliothèques du Québec.

Certains outils présentés par le mandataire ont une portée transversale et pourraient être adaptés pour étudier le rayonnement des contenus culturels québécois dans d'autres domaines (Hercules, outils d'analytique Web et de réputation numérique).

2. INSTITUT DE LA STATISTIQUE DU QUÉBEC (2021, mis à jour le 8 novembre), *Statistiques sur les bibliothèques publiques du Québec*, [En ligne], [statistique.quebec.ca/fr/document/statistiques-sur-les-bibliotheques-publiques-du-quebec].

3. INSTITUT DE LA STATISTIQUE DU QUÉBEC (2021, mis à jour le 2 novembre), *Ventes de livres numériques, Québec*, [En ligne], [statistique.quebec.ca/fr/document/ventes-de-livres-numeriques-quebec].

1.1.3 Audiovisuel

**Synthèse et analyse du rapport réalisé par LaCogency et La Fabrique de sens:
*Projet d'indicateurs de rayonnement de l'offre culturelle québécoise
dans les réseaux numériques – Étude des sources de statistiques
existantes quant à l'audiovisuel***

Dans le domaine de l'audiovisuel, la dynamique sectorielle d'exclusivité territoriale et de fenêtres de diffusion complique la production de statistiques, et ce, tant en ce qui concerne les longs métrages et les séries télé que les séries Web. Malgré la profusion de plateformes d'écoute en continu, l'audiovisuel demeure une industrie structurée autour des réseaux traditionnels.

Peu de sources de données tertiaires⁴ ou secondaires⁵ sont disponibles à l'heure actuelle. Aussi, plusieurs des sources de données potentielles retenues par le mandataire sont encore au stade de l'élaboration ou ont été conçues pour observer principalement le marché canadien. Le tableau 2 synthétise les mesures qui pourraient être extraites de ces sources de données potentielles.

Tableau 2

Indicateurs potentiels formulés à partir des analyses présentées dans le *Projet d'indicateurs de rayonnement de l'offre culturelle québécoise dans les réseaux numériques – Étude des sources de statistiques existantes quant à l'audiovisuel*

Indicateurs potentiels	Sources potentielles	Statut	Portée
Dimension – Présence			
• Présence des œuvres audiovisuelles québécoises dans les bases de connaissances collaboratives	IMDB/Rotten Tomatoes	Disponible	
• Nombre de recherches sur Wikipédia	Parrot Analytics (composante de la <i>Demand Expression</i>)	Disponibilité du détail à confirmer	
• Indice de découvrabilité du LATICCE (P)	Hercules (H264)	Disponibilité du (P) à confirmer	
• Présence des productions suivies sur les plateformes de diffusion	Viewer Centric	Théorique	Canada

Suite à la page 17

4. Nous entendons par « source tertiaire » toute source qui produit des statistiques sur l'utilisation qui est faite d'un contenu en ligne ou sur les mentions le concernant sur des sites Web et des plateformes de diffusion. Il s'agit d'outils d'analytique Web utilisant plusieurs sources primaires. Dans le domaine de la musique enregistrée, Chartmetric est un bon exemple de source tertiaire.
5. Nous entendons par « source secondaire » toute source qui produit des statistiques grâce aux outils d'analytique Web des plateformes de diffusion elles-mêmes. Dans le domaine de l'audiovisuel et de la musique enregistrée, Youtube Analytics un bon exemple de source secondaire.

Tableau 2 (suite)

Indicateurs potentiels formulés à partir des analyses présentées dans le *Projet d'indicateurs de rayonnement de l'offre culturelle québécoise dans les réseaux numériques – Étude des sources de statistiques existantes quant à l'audiovisuel*

Indicateurs potentiels	Sources potentielles	Statut	Portée
Dimension – Visibilité			
<ul style="list-style-type: none"> Nombre moyen d'évaluations des œuvres audiovisuelles québécoises dans les bases de connaissances collaboratives 	IMDB/Rotten Tomatoes	Disponible	
<ul style="list-style-type: none"> Score moyen de ces évaluations 			
<ul style="list-style-type: none"> Nombre d'interactions des utilisateurs sur les médias sociaux 	Parrot Analytics (composante de la <i>Demand Expression</i>)	Disponibilité du détail à confirmer	
<ul style="list-style-type: none"> Indice de découvrabilité du LATICCE (V) 	Hercules (H264)	Disponibilité du (V) à confirmer	
<ul style="list-style-type: none"> Mesures d'engagement de l'auditoire sur Youtube 	Tubular Labs	Théorique	
<ul style="list-style-type: none"> Nombre de mentions d'œuvres audiovisuelles québécoises dans les médias 	Eureka	Théorique	
<ul style="list-style-type: none"> Popularité des productions suivies sur les réseaux sociaux et les plateformes de diffusion 	Viewer Centric	Théorique	Canada
Dimension – Recommandations			
<ul style="list-style-type: none"> Indice de découvrabilité du LATICCE (R) 	Hercules (H264)	Disponibilité du (R) à confirmer	
Dimension – Consommation			
<ul style="list-style-type: none"> Gestion des redevances 	Hercules (H264)	Disponible	
<ul style="list-style-type: none"> Données d'écoute multiplateforme dans un échantillon de foyers canadiens sur les plateformes de diffusion et les réseaux sociaux 	Numéris – VAM	Disponible	Québec/Canada
<ul style="list-style-type: none"> Nombre de vues sur YouTube 	Tubular Labs	Disponible	Canada
<ul style="list-style-type: none"> Nombre de visionnements sur un groupe défini de plateformes vidéo Nombre de partages pair à pair de fichiers 	Parrot Analytics (composante de la <i>Demand Expression</i>)	Disponibilité du détail à confirmer	
<ul style="list-style-type: none"> Popularité des productions suivies sur les réseaux sociaux et les plateformes de diffusion 	Viewer Centric	Théorique	Canada

- Parmi les pistes intéressantes décrites par le mandataire, il y a les travaux réalisés par le Fonds des médias du Canada (FMC) dans le cadre de leur infrastructure de mesure des médias numériques (IMMN). Cette infrastructure regroupe des données de trois sources distinctes, auxquelles s'ajoute un projet pilote.
 - Service VAM (Video Audience Measurement) de Numéris : Service basé sur les résultats d'un panel composé de foyers canadiens ayant installé le Kantar Focal Meter sur leur routeur, qui permettra à terme (normalement à l'automne 2021) de colliger des données d'écoute multiplateforme de format longue durée déjà diffusées en linéaire, mais aussi des données d'écoute sur les médias sociaux (TikTok, Facebook et Google). La portée du service demeure cependant essentiellement canadienne.
 - *Demand Expression* (expressions de la demande) de Parrot Analytics : Indice composé de quatre variables : les interactions des utilisateurs sur les médias sociaux, les partages pair à pair de fichiers, les recherches sur Wikipédia et les visionnements sur les plateformes vidéo⁶. L'outil permet de mesurer le degré d'engagement des auditoires envers une émission grâce à un algorithme propriétaire qui suit et agrège les interactions des utilisateurs sur les médias sociaux, les flux vidéo, le partage de photos, les blogues ou microblogues, les évaluations des adeptes et des critiques, les sites de partage de fichiers, et même les téléchargements piratés et le trafic sur Wikipédia. Cet outil inclut les programmes non diffusés sur les services linéaires traditionnels, pays par pays.
 - Tubular Labs : Outil qui mesure les visionnements d'un panier de titres sur YouTube et qui fournit des données de base sur l'engagement de l'auditoire (l'entreprise espère éventuellement fournir aussi la durée du visionnement). Cet outil peut également capter des vidéos qui auraient été produites par des adeptes autour du contenu initial.
 - Projet-pilote du FMC avec Viewer Centric : Projet en cours d'élaboration, Viewer Centric est un tableau de bord conçu pour les bailleurs de fonds, les diffuseurs et les créateurs, et qui leur permet de suivre des indicateurs tirés des outils analytiques de différentes plateformes⁷ : visites de sites Web, profils sociodémographiques des visiteurs, comparatif des productions en fonction des critères de popularité et du budget investi dans la production. À l'heure actuelle, on y retrouve cependant peu de contenus québécois : en 2020, la plateforme recensait 75 productions, dont 20 québécoises. Elle ne vise aussi que le marché canadien.
- Les coûts associés à l'IMMN du FMC sont évalués à plus ou moins 500 000 \$ par année, auxquels s'ajoutent les salaires de quatre employés du service d'analytique des médias. Pour alléger le fardeau financier, le FMC partage les licences de Numéris et de Parrot Analytics avec le CRTC et le ministère du Patrimoine canadien.
- Par ailleurs, le mandataire décrit l'élaboration pour l'audiovisuel d'une version bêta d'Hercules pour l'agrégateur-distributeur H264.
 - Contrairement aux indicateurs d'Hercules décrits par l'équipe de BTLF, les indicateurs portent ici plus précisément sur la mesure de la découvrabilité et la gestion des redevances. On peut toutefois s'attendre à ce que des mesures semblables à celles élaborées pour le livre puissent aussi être opérationnalisées pour les productions audiovisuelles.

6. Dans le tableau 2 *La Demand Expression* de Parrot Analytics a été assimilée à la source Parrot Analytics, puisqu'elle est constituée de quatre mesures relevant de dimensions différentes du rayonnement (présence [1], visibilité [1] et consommation [2]). Puisque les données détaillées ne seraient disponibles que pour les productions ayant un score de *Demand Expression* élevé, la disponibilité de ces mesures spécifiques demeure à confirmer. Plus de précisions sur la contribution relative de chacune de ces mesures à la *Demand Expression* globale seraient nécessaires pour valider la pertinence de cet indicateur composite.

7. Google, Facebook, Twitter et Instagram Analytics, YouTube et Vimeo.

- Rappelons cependant que Hercules a des limites, notamment le nombre d'occurrences et l'obligation de procéder par échantillon.
- Dans le contexte de l'audiovisuel, la constitution d'un panier équilibré composé de nouveautés et de produits du catalogue peut être complexifiée par le retour en nouveauté de productions de catalogue, par exemple, les plateformes de diffusion comme Netflix ayant mis en place des stratégies de diffusion d'œuvres du catalogue s'apparentant à celles associées à des nouveautés.
- En dehors de ces sources, quelques bases de données institutionnelles et des bases de connaissances collaboratives ont été retenues par le mandataire.
 - Pour recenser les contenus audiovisuels québécois, on suggère de s'appuyer sur le répertoire audiovisuel du Québec tenu par la Cinémathèque québécoise, où on retrouve des longs métrages, des séries télé et des séries Web.
 - Des données de présence et d'évaluation par le public pourraient aussi être recueillies par moissonnage sur IMDB et Rotten Tomatoes.
 - On fait brièvement mention d'Eureka, un outil qui traite de la présence dans les médias et qui permet de faire des recherches dans des articles de journaux, de magazines et d'agences de presse, dans des transcriptions d'émissions de radio et de télévision, et dans des blogues. On y trouve des sources du Québec, du Canada et de partout dans le monde rédigées dans plusieurs langues. Cette piste, à peine abordée, mériterait peut-être d'être examinée.

1.1.4 Musique enregistrée et spectacles en arts de la scène

Synthèse et analyse du rapport réalisé par Espaces Temps :
***Projet d'indicateurs de rayonnement de l'offre culturelle québécoise
dans les réseaux numériques pour la musique enregistrée
et les spectacles en arts de la scène***

► **Musique enregistrée**

Dans le domaine de la musique enregistrée, la consommation consiste principalement en de l'écoute en continu, le téléchargement de pistes audio étant en déclin. La piste (ou le titre) circule largement et indépendamment de l'album sur lequel elle figure, le cas échéant.

Les plateformes d'écoute en continu proposent proportionnellement plus de pistes issues des fonds de catalogues que de nouveautés, ce qui pose le défi de l'indexation « à rebours » des œuvres du catalogue par des métadonnées, un processus qui favoriserait la découvrabilité des données, mais qui permettrait aussi de suivre leur rayonnement (circulation et consommation).

Ce contexte soulève de nombreux défis de mesure : volumétrie importante des titres, multiplicité des plateformes de diffusion, définition de la nouveauté des titres qu'on veut suivre, etc. Le mandataire Espaces Temps a décrit trois types de sources : les solutions « industrielles », créées pour les artistes et les distributeurs, et appelées les solutions « clé en main », les sources pour lesquelles les producteurs doivent avoir accès à leur compte et les sources nécessitant un développement applicatif.

Le tableau 3 synthétise les mesures qui pourraient être extraites de ces sources de données potentielles.

Tableau 3

Indicateurs potentiels formulés pour la musique enregistrée à partir des analyses présentées dans le *Projet d'indicateurs de rayonnement de l'offre culturelle québécoise dans les réseaux numériques pour la musique enregistrée et les spectacles en arts de la scène*

Indicateurs potentiels	Sources potentielles	Statut	Portée
Dimension – Présence			
• Indice de découvrabilité du LATICCE (P)	LATICCE	Disponible	
• Nombre de pistes d'artistes québécois pour lesquelles on trouve des données dans Last.fm	MusicBrainz/Last.fm	Disponible	Last.fm
Dimension – Visibilité			
• Indice de découvrabilité du LATICCE (V)	LATICCE	Disponible	
• Nombre d'adeptes et d'abonnés sur les plateformes de diffusion	Soundcharts/Chartmetrics	Possibilité de déterminer l'ID québécoise à confirmer	
• Nombre d'adeptes et d'abonnés sur les réseaux sociaux	Soundcharts/Chartmetrics	Possibilité de déterminer l'ID québécoise à confirmer	
• Exposition à la publicité dans les réseaux sociaux	Google Data Studio – Collaboration spéciale	Théorique	
Dimension – Recommandations			
• Indice de découvrabilité du LATICCE (R)	LATICCE	Disponible	
Dimension – Consommation			
• Nombre d'écoutes de pistes d'artistes québécois au Canada sur les principales plateformes d'écoute en continu	Nielsen/MRC Data/OCCQ	En développement	Canada
• Nombre d'écoutes de pistes d'artistes québécois sur les principales plateformes d'écoute en continu parmi les utilisateurs de Last.fm	MusicBrainz/Last.fm	Disponible	Last.fm
• Nombre de membres de Last.fm qui ont écouté des pistes d'artistes québécois sur les principales plateformes d'écoute en continu			
• Nombre d'écoutes de pistes d'artistes québécois sur les principales plateformes d'écoute en continu	Soundcharts/Chartmetrics	Possibilité de déterminer l'ID québécoise à confirmer	
• Nombre d'écoutes de pistes d'artistes québécois sur QUB	QUB Musique	Théorique	QUB
• Gestion des redevances	Exactals – RAI/Quansic	Théorique	Clients

- Solutions « clé en main »
 - Nielsen/MRC Data : Données qui permettent de recenser les volumes d'écoute sur les services interactifs de musique (Spotify, Apple Music, etc.) au Canada et en Amérique du Nord. Un projet pilote présentement en cours à l'Observatoire analyse les pistes ayant 100 écoutes ou plus au Canada durant une semaine de référence. Ces pistes sont appariées à une liste d'*International Standard Recording Code* (ISRC) de producteurs et à une liste validée d'interprètes québécois⁸. Cela permet de calculer la part des producteurs et des interprètes québécois dans l'ensemble des pistes écoutées hebdomadairement au Canada. À ce jour, il n'est cependant pas possible d'isoler l'écoute pour le territoire québécois ou encore d'identifier la langue des pistes écoutées. Cette dernière pourrait cependant être inférée à partir de bases de référence externes.
 - QUB musique : Nouveau joueur dans la diffusion de musique en continu, ce service a pour l'instant une faible part du marché, mais a une approche éditoriale de forte présence de contenu québécois et une portée internationale. Il s'agit ici d'une plateforme de diffusion parmi tant d'autres. Elle pourrait ultimement être intégrée à la collecte de Nielsen/MRC Data.
- Sources pour lesquelles les producteurs doivent avoir accès à leur compte
 - Soudcharts et Chartmetrics : Sources qui colligent par une interface applicative de programmation les données d'écoute par territoire, mais aussi les manifestations d'intérêt dans les principaux réseaux sociaux pour des artistes/groupes musicaux : nombre d'écoute de titres par territoire, adeptes et abonnés sur les réseaux sociaux (Instagram, Facebook, Twitter, etc.) et les plateformes de diffusion (Deezer, Soundcloud, Spotify, etc.), profil sociodémographique des adeptes et des abonnés sur certaines plateformes, etc.
 - Conçus pour dresser le portrait de la notoriété d'un groupe ou d'un interprète en particulier, ces services ne donnent pas accès à des données sommatives.
 - Il pourrait s'agir d'une piste intéressante pour analyser l'écoute de titres québécois à l'extérieur du Québec et le nombre moyen d'adeptes et d'abonnés d'une liste québécoise d'interprètes sur différentes plateformes.
 - Google Data Studio – Collaboration spéciale : Application créée pour une agence de publicité (Collaboration spéciale) qui permet d'agréger des statistiques qui découlent des investissements publicitaires des producteurs.
 - Cette application est notamment utilisée pour mesurer l'exposition à leurs publicités sur Facebook, YouTube, Google Ads, Spotify, Instagram, etc.
 - Exactals – RAI (Californie) et Quansic (Suisse) : Applications de gestion des redevances créées pour les éditeurs et les maisons de disques. Elles pourraient procéder à des agrégations de données anonymisées, mais de telles données ne sont pas disponibles pour l'instant. Il n'y a aucune précision quant au nombre d'éditeurs ou de maisons de disques qui ont recours à ces applications. La couverture des artistes québécois et des plateformes reste à évaluer.

8. L'identification des interprètes québécois est ici réalisée à partir des bases de données de l'ADISQ, de Spotify, de Wikipédia, de Bandcamp, etc. alors que les données sur les producteurs proviennent de la SOPROQ.

- Sources nécessitant un développement applicatif :
 - Indice de découvrabilité de LATICCE : Indice composite mesurant des scores de présence, de visibilité et de recommandation d'un panier de titres québécois sur Spotify, Apple Music et Google Play Music à partir des comportements d'écoute de « persona » contrôlées. Il mesure le potentiel de découvrabilité d'un panier de contenus culturels québécois définis pour des profils d'auditeurs particuliers, mais ne mesure pas le comportement réel des auditoires et leur exposition aux pistes québécoises.
 - MusicBrainz/Last.FM : Last.FM est une application de mesure de l'auditoire et du nombre d'écoutes de pistes audio sur différentes plateformes. L'identification des titres québécois se fait par MusicBrainz.
 - Cette approche est intéressante, surtout en ce qui concerne l'utilisation de MusicBrainz pour cibler les artistes québécois. La représentativité statistique des données colligées par Last.fm demeure toutefois à démontrer. En effet, en comparant les données de Last.fm à celles de l'application Music Connect de MRC Data, on remarque des écarts importants.

Tableau 4

Comparaison entre certaines données extraites de Last.fm et de MRC Data (Music Connect)

Territoire d'écoute	Last.fm	MRC Data (Music Connect)
	Tous	Canada seulement
Période	Données cumulatives dans Last.fm	Données cumulatives depuis le 26 mai 2014
Simple Plan	46 M	105 M
Céline Dion	30 M	468 M
Léonard Cohen	52 M	112 M

► **Spectacles en arts de la scène**

Dans le domaine des spectacles en arts de la scène, le principal enjeu relevé par Espaces Temps est de concilier les données sur l'offre et celles sur l'assistance.

Les principales sources décrites dans le rapport gravitent autour de la mesure de l'offre. Les données d'assistance en salle colligées par l'Observatoire n'ont pas de lien direct avec les réseaux numériques. Aucune piste n'a par ailleurs été recensée pour évaluer l'assistance à des performances en ligne, qu'il s'agisse de diffusions en direct ou en différé.

Espaces Temps propose d'utiliser Scène Pro – RIDEAU pour recenser l'offre québécoise. La collaboration de La culture crée au WikiProject Performing Arts avec CAPACOA fait aussi partie des sources de données structurées sur la description des œuvres et leur programmation (spectacle, artiste, date, lieu) qui pourraient minimalement permettre d'évaluer la présence des événements dans les réseaux numériques, mais sans proposer d'indicateurs spécifiques.

Des sites de billetterie tels que Tuxedo-Turbine pourraient par ailleurs être exploités pour recenser les ventes de billets de même que les annulations et les reports de spectacles. Cette firme travaille principalement avec des salles québécoises. Ticketmaster for Developer pourrait peut-être fournir des données semblables pour les représentations de spectacles québécois hors Québec, mais cette avenue n'a pas été explorée par le mandataire.

1.2 Approche 2: La création d'un outil de moissonnage du Web qui permettrait d'analyser les différents marqueurs du rayonnement

Synthèse et analyse du rapport réalisé par La culture crée

1.2.1 Démarche et principaux enjeux

Le mandataire de l'approche 2 devait initialement tenter de reproduire (ou de compléter) les indicateurs présentés par les mandataires de l'approche 1, mais le peu de données quantitatives récoltées par ces derniers nous a amenés à modifier légèrement les objectifs de l'approche 2 pour tenter de relever minimalement ce qui pourrait, à l'heure actuelle, être recensé par moissonnage pour témoigner du rayonnement dans les différents domaines à l'étude.

En l'absence de valeurs précises à reproduire et à consigner, une attention particulière a été portée aux données plus aisément accessibles et aux limites des différentes méthodes – moissonnage traditionnel et requêtes sur des données liées (SPARQL) – afin d'établir des bases pour la création de futurs indicateurs.

Les enjeux liés à la mesure d'indicateurs par moissonnage sur le Web sont les suivants.

- **L'opérationnalisation de la « québécoïté » :** Cette question s'est posée d'entrée de jeu, la nature québécoïse d'un contenu ou d'un artiste n'étant pas nécessairement évidente dans les sources de données par ailleurs pertinentes pour évaluer le rayonnement dans un domaine donné.
 - Les différents critères retenus pour juger de la « québécoïté » des contenus culturels québécoïses dans les statistiques de consommation traditionnelle sont difficiles à opérationnaliser sur les réseaux numériques, notamment lorsqu'ils reposent sur la conjonction de plus d'une condition. En effet, toutes les sources de données ne disposent pas de détails sur l'origine des différents contributeurs d'une œuvre et l'utilisation de la condition « ET » rejette de facto les sources pour lesquelles un des éléments de la condition est manquant. L'utilisation de conditions utilisant le « OU » est alors plus inclusive et devrait être privilégiée.
 - En règle générale, les bases de connaissances⁹ qui permettent d'évaluer en amont l'existence de contenus culturels québécoïses et qui sont préconisées par les mandataires de l'approche 1 ont un indice de québécoïté (IQ) élevé. Ces bases pourraient être utilisées telles quelles pour contourner l'enjeu de la sélection de critères de « québécoïté ». Par ailleurs, les sources de données permettant d'analyser la présence ou la visibilité sur les réseaux sociaux et la consommation sur les plateformes de diffusion ont pour leur part des IQ généralement plus faibles. Certaines peuvent être croisées avec les bases de connaissances pour pallier ces lacunes. Un tel croisement est évidemment facilité lorsque les deux sources ont des identifiants uniques communs.
 - Aux fins d'un exercice de moissonnage, toutes les variantes de noms de domaines des services internationaux (par exemple Amazon.fr, Amazon.ca et Amazon.com) ne permettent pas d'identifier aussi précisément les contenus québécoïses. Il est par exemple plus facile d'identifier des contenus culturels québécoïses sur Amazon.ca que sur Amazon.com.

9. Catalogue du dépôt légal de BAnQ, MusicBrainz, Répertoire audiovisuel du Québec, Scène-Pro, etc.

- Pour une même source, la qualité de la « québécoïté » peut varier d'un marqueur à l'autre. Par exemple, sur Amazon.ca, la présence d'œuvres québécoises s'établit à partir de sources externes de contenus déjà reconnus comme québécois. Par ailleurs, le site d'Amazon.ca diffuse le palmarès des 100 meilleures ventes de littérature québécoise.
- **L'objet dont le rayonnement est à mesurer :** La question soulevée lors des rencontres du comité de réflexion quant à l'intérêt de mesurer le rayonnement de l'œuvre ou de la personne est liée aux types de marqueurs qu'on veut analyser : l'œuvre est à la base de la présence dans les plateformes de diffusion, mais la personne est suivie sur les réseaux sociaux. Les œuvres ont rarement des pages qui leur sont propres. Dans les opérations de moissonnage, la conjonction du nom de la personne et de son œuvre est souvent nécessaire pour « prouver » qu'on recense les bons renseignements.
- Compte tenu du caractère exploratoire de l'exercice de moissonnage, les données recueillies peuvent comprendre certains faux positifs associés à certains noms plus communs ou passés dans la toponymie, de même que les titres d'œuvres comprenant des mots ou des expressions plus génériques
 - Ces contraintes associées à la recherche textuelle devront être prises en compte dans toutes solutions associées à un développement applicatif ne reposant pas sur les données liées, de même que dans l'élaboration des « paniers », si une telle solution est retenue. En effet, le nom de l'œuvre (ou même de ses créateurs) peut avoir en soi une incidence sur son repérage potentiel.
- **Le caractère « éphémère » de la mesure :** À moins de négocier des droits d'accès avec les plateformes et les sites dont on veut recueillir les données, le moissonnage du Web permet seulement de recueillir de façon plus ou moins automatisée les renseignements qui sont visibles pour les usagers de ces sites. Il s'agit alors de photos en temps réel d'une situation donnée.
- Les différences observées entre un moment de mesure et un autre peuvent être causées entre autres par la modification des algorithmes des services moissonnés.
 - Les mentions dans les résultats de recherche d'un moteur de recherche peuvent varier d'un utilisateur à l'autre, car elles tiennent compte des préférences de l'utilisateur afin de personnaliser son expérience.
 - Comme le souligne l'Institut national de la statistique et des économies de France, « Les séries disponibles sur Google Trends ne correspondent pas à un comptage exhaustif des termes retenus, mais à un échantillonnage avec des retraitements (par exemple des redressements pour corriger de la diffusion croissante d'Internet et de Google) qui ne sont pas documentés et peuvent introduire de l'instabilité. Après tout, Google n'a pas pour objectif la cohérence temporelle, mais une amélioration constante de son moteur de recherche, ce qui conduit forcément à une instabilité de la source. De plus, il n'est pas possible de connaître le contexte des requêtes effectuées, qui peuvent avoir des motivations très diverses et sans rapport avec l'activité économique que l'on cherche à mesurer¹⁰. »

10. Benoît OURLIAC (2000), « Google en sait-il plus que l'Insee que les Français ? », *Le blog de l'Insee*, [En ligne], [blog.insee.fr/google-en-sait-il-plus-que-linsee-sur-les-francais/].

- Les stratégies des plateformes de diffusion pour retenir le plus longtemps possible l'intérêt des usagers sont aussi régulièrement modifiées, ce qui peut induire des changements dans les données mesurées par une même requête.
 - Par exemple, l'algorithme de YouTube associé au « *Gangnam style effect*¹¹ » tendait à suggérer des vidéos de plus en plus populaires aux usagers. Ultimement, tout usager qui ne quittait pas la plateforme aboutissait à la vidéo de *Gangnam style*. Après avoir réalisé que les vidéos « hors du commun » maintenaient les usagers en ligne plus longtemps que les vidéos plus populaires, YouTube a modifié son algorithme pour faire des propositions plus hors-norme. Les différences observées entre un nombre de vues d'un même contenu à deux moments distincts peuvent ainsi être autant liées à une modification de l'algorithme de la plateforme qu'à un réel gain de popularité auprès du public.

1.2.2 Livre

- Dans le domaine du livre, la base de référence pour recenser les contenus québécois qui semble la plus pertinente pour le mandataire de l'approche 2 est le répertoire du dépôt légal de BANQ.
 - En plus de servir de base d'identification des titres québécois, le site de BANQ pourrait fournir des renseignements sur les lauréats des prix littéraires.
 - L'interface client de BANQ permet aussi à ses membres de choisir des favoris et de partager certaines publications, des renseignements qui pourraient être exploités si BANQ les rendait disponibles (ils ne peuvent être moissonnés).
- Dans ses conclusions, le mandataire indique aussi que Worldcat permet d'évaluer la disponibilité des titres québécois dans les bibliothèques internationales à partir de requêtes sur des données liées, à l'instar des travaux réalisés dans le cadre de l'étude *Feuille d'érable : Découvrir le Canada par ses publications*¹².
- Les données suivantes pourraient aussi être extraites d'Amazon par moissonnage direct :
 - Classement du produit dans les ventes d'une catégorie de produits (nombre total inconnu)
 - Rang dans les ventes (nombre total inconnu)
 - Nombre d'artistes avec ventes sur Amazon
 - Nombre moyen d'œuvres/artistes
 - Nombre d'évaluations par les usagers
 - Évaluation moyenne des produits
 - Cote de l'artiste selon la moyenne des classements de ses produits

11. Jeremy CARSON (2019), « Why YouTube's Algorithm Constantly Recommended Watching Gangnam Style », *Jeremy Carson*, [En ligne], [\[medium.com/@thejeremycarson/why-youtubes-algorithm-constantly-recommended-watching-gangnam-style-4a3e8d462ab4\]](https://medium.com/@thejeremycarson/why-youtubes-algorithm-constantly-recommended-watching-gangnam-style-4a3e8d462ab4).

12. Brian LAVOIE (2019). *Feuilles d'érable : Découvrir le Canada par ses publications*, [En ligne], Dublin, OH : OCLC Research, 33 p. doi : doi.org/10.25333/ek4v-ag09.

Pour réaliser l'exercice de moissonnage, compte tenu des délais, on a utilisé les palmarès des meilleurs vendeurs sur Amazon et Goodreads pour constituer un « panier test ».

- Les 86 titres retenus obtiennent en moyenne 24 635 résultats de recherche dans Bing et sont tous présents dans le catalogue de BAnQ et d'Amazon.
- Dans Amazon, 55 de ces 86 titres ont été évalués, en moyenne 33 fois, pour une note moyenne de 4,4/5.
- Sur Facebook, 5 auteurs n'ont suscité aucune réaction ou commentaire. Parmi ceux qui en ont reçu, on recense en moyenne près de 3 500 mentions « J'aime » et plus ou moins 50 commentaires¹³.
- Sur les 86 titres retenus, 55 ont un classement de leurs ventes dans la catégorie « Livres » sur Amazon. Ces titres se retrouvent en moyenne à la 6 801 801^e position dans les ventes de l'ensemble des livres d'Amazon.
- Si on s'intéresse seulement aux 100 meilleurs vendeurs en « littérature canadienne-française » sur Amazon (sans valider la « québécoisité » des auteurs), ces 100 livres se trouvent en moyenne à la 460 963^e position dans le total des ventes de livres sur Amazon.

13. Compte tenu du caractère exploratoire de la démarche, ces résultats sont approximatifs et peuvent comprendre des faux positifs associés à des homonymes.

Tableau 5

Résultats de l'exercice de moissonnage dans le domaine du livre réalisé par La culture crée

Indicateur	Source	Unités recherchées		Unités avec résultats		Somme des résultats	Résultat moyen
		Description	n ^{bre}	n ^{bre}	%		
Résultats obtenus par moissonnage du Web							
Dimension – Présence							
Mentions (#)	Bing	Titres	86	86	100	2 118 620	24 635
Disponibilité (%)	BAnQ	Titres	86	86	100	s.o.	s.o.
Disponibilité (%)	Amazon	Titres	86	86	100	s.o.	s.o.
Dimension – Visibilité							
Évaluations (#)	Amazon	Titres	86	55	64	1 828	33
Évaluation moyenne (#)	Amazon	Titres	86	55	64	243	4,4
Réactions (#) (« J'aime »)	Facebook	Auteur(e)	77	72	94	281 849	3 915
Commentaires (#)	Facebook	Auteur(e)	77	72	94	3 800	53
Dimension – Consommation							
Classement moyen des ventes des titres du panier dans la catégorie « Livres »	Amazon	Titres	86	55	64	374 099 051	6 801 801
Classement moyen des ventes des 100 meilleurs vendeurs de la catégorie « Littérature canadienne-française » par rapport à l'ensemble des ventes de livres ¹	Amazon	Titres	100	100	100	46 096 341	460 963

1. Ce classement est diffusé en temps réel dans Amazon et la liste des ouvrages de la catégorie fluctue. Ces résultats ne sont donc pas associés au panier défini par La culture crée.

1.2.3 Audiovisuel

Dans le domaine de l'audiovisuel, l'enjeu particulier est de trouver une source qui comprend toutes les productions audiovisuelles québécoises. Il existe plusieurs listes pouvant potentiellement être exploitées, mais elles ne reposent pas sur les mêmes critères (provenance du financement, part de la participation québécoise à une coproduction, films exploités en salle, etc.).

- Comme pour le livre, les données pouvant être recueillies sur Amazon par moissonnage révèlent pour la plupart le positionnement des produits dans les ventes sur Amazon ou leur évaluation par les usagers :
 - Classement du produit dans les ventes d'une catégorie de produits (nombre total inconnu)
 - Rang dans les ventes (nombre total inconnu)
 - Nombre d'artistes avec ventes sur Amazon
 - Nombre moyen d'œuvres/artistes
 - Nombre d'évaluations par les usagers
 - Évaluation moyenne des produits
 - Cote de l'artiste selon la moyenne des classements de ses produits

Il est par ailleurs impossible de recenser le nombre de visionnements sur Amazon Prime.

- Sur YouTube, les renseignements qui peuvent être moissonnés portent sur les aspects suivants :
 - Le nombre de visionnements
 - Le nombre de réactions
 - La cote moyenne qui peut être calculée à partir de ces réactions
- Sur IMDb, le « star meter » – c'est-à-dire le top 100 des productions les plus populaires – pourrait être exploité, en plus de la présence et des évaluations disponibles sur la page d'une production.

Aux fins de l'exercice de moissonnage, le palmarès des meilleurs vendeurs sur Amazon a été retenu pour constituer un « panier test ».

- Les œuvres de ce panier, qui sont les 47 meilleurs vendeurs québécois, sont mentionnées ou disponibles sur l'ensemble des sites et plateformes moissonnés : Bing (moyenne de 71 843 résultats de recherche/œuvre), IMDb, Répertoire audiovisuel Québec, Amazon et YouTube.
 - Il a toutefois été impossible, sur YouTube, de distinguer s'il s'agissait de bandes-annonces, de contenus créés par le public autour de l'œuvre ou de l'œuvre elle-même.
- Toutes ont été évaluées au moins une fois sur Amazon et YouTube, obtenant une évaluation moyenne de 3,8/5 sur les deux plateformes.
- Le classement des ventes dans la catégorie « Films et séries télévisées » était disponible pour 41 des 47 œuvres du panier, pour un classement moyen estimé de ces 41 œuvres à la 48 137^e position (sur un total inconnu).
- Les contenus associés aux 47 œuvres (bande-annonce, film, série télé, critiques) ont récolté un peu plus de 4 millions de vues sur YouTube, pour une moyenne de 86 013 vues par œuvre.
- Les résultats des requêtes sur des données liées révèlent par ailleurs que 42 de ces œuvres sur 47 ont un identifiant unique dans Wikidata permettant de les suivre. De ces 42, 6 ont dans Wikidata des renseignements à propos de nominations, alors que 3 auraient été lauréates d'au moins un prix.
 - Ces chiffres semblent particulièrement faibles compte tenu de la popularité des œuvres du panier, ce qui soulève la question du niveau d'avancement des entreprises du domaine quant à la structuration de leurs données en données ouvertes et liées.

Tableau 6

Résultats de l'exercice de moissonnage dans le domaine de l'audiovisuel réalisé par La culture crée

Indicateur	Source	Unités recherchées		Unités avec résultats		Somme des résultats	Résultat moyen
		Description	n ^{bre}	n ^{bre}	%		
Résultats obtenus par moissonnage du Web							
Dimension – Présence							
Mentions (#)	Bing	Titres	47	47	100	3 376 642	71 843
Mentions	IMDB	Titres	47	47	100	s.o.	s.o.
Mentions	Répertoire audiovisuel Québec	Titres	47	47	100	s.o.	s.o.
Disponibilité (%)	Amazon	Titres	47	47	100	s.o.	s.o.
Disponibilité (%)	YouTube	Titres	47	47	100	s.o.	s.o.
Dimension – Visibilité							
Évaluations (#)	Amazon	Titres	47	47	100	4 844	103
Évaluation moyenne (#)	Amazon	Titres	47	47	100	177	3,8
Évaluations (#)	YouTube	Titres	47	47	100	3 686	78
Évaluation moyenne (#)	YouTube	Titres	47	47	100	179,2	3,8
Dimension – Consommation							
Classement moyen des ventes dans la catégorie « Films et séries télévisées »	Amazon	Titres	47	41	87	1 973 635	48 137
Nombre de vues	YouTube	Titres	47	47	100	4 042 631	86 013
Résultats obtenus par requête sur des données liées (SPARQL)							
Dimension – Présence							
Taux de correspondance entre le panier et identificateurs globaux uniques Wikidata	Wikidata	Titres	47	42	89	s.o.	s.o.
Dimension – Visibilité							
Prix (#)	Wikidata	Titres	42	3	7	9	3,0
Nomination (#)	Wikidata	Titres	42	6	14	15	2,5

1.2.4 Musique enregistrée

Dans le domaine de la musique enregistrée, différentes sources potentielles ont été décrites par le mandataire, même si toutes n'ont pas été retenues pour l'exercice de moissonnage.

- Rappelons que la requête sur des données liées croisant les données de la base de connaissances de MusicBrainz à celles de l'application Last.fm présentée à la section 1.1.4 a été réalisée par La culture crée.
- Ici encore, Amazon s'avère une plateforme de diffusion d'importance, tant pour sa plateforme d'écoute en continu que pour les achats d'albums sur support physique ou numérique. Il est cependant difficile d'obtenir un classement des ventes ou de dénombrer les évaluations parce que le mode de consommation par piste décuple le nombre d'unités suivies.
- Sur YouTube, les données disponibles concernent les écoutes de pistes, de vidéoclips ou d'albums :
 - directement sur YouTube, que ce soit sur le service gratuit (avec publicité) ou payant ;
 - sur les chaînes YouTube de particuliers qui relayent les contenus ;
 - sur les chaînes YouTube des interprètes eux-mêmes.
 - Le nombre d'abonnés à ces chaînes pourrait alors être considéré comme un marqueur du rayonnement d'un artiste.
- Des données similaires à celles retrouvées sur les chaînes YouTube d'artistes sont disponibles sur les pages TikTok des interprètes.
- Il serait par ailleurs théoriquement possible dans les moteurs de recherche comme Bing ou Google de capter le nombre de vues, de commentaires ou de réactions sur les réseaux sociaux, mais ces dénombrements ne seraient pas nécessairement exhaustifs, puisqu'assujettis aux algorithmes en vigueur au moment de la recherche.
- Il serait aussi possible d'exploiter les renseignements fournis aux usagers sur les plateformes d'écoute en continu, mais toutes n'affichent pas les mêmes renseignements :
 - Soundcloud – nombre de titres, nombre d'écoutes et nombre d'adeptes ;
 - Spotify – nombre d'auditeurs par mois, nombre d'adeptes, classement des 10 chansons les plus populaires, « delta du mois » (c'est-à-dire la variation du nombre d'auditeurs par mois), nombre d'auditeurs selon la ville.

Selon l'indicateur documenté, les paniers utilisés pour l'exercice de moissonnage sur la musique enregistrée reposent soit sur le panier de chansons proposé par Espaces Temps (n=50), soit sur une liste d'interprètes (n=94¹⁴). On remarque ainsi ce qui suit.

- Moteurs de recherche
 - Tous les interprètes ont au moins un lien officiel dans un moteur de recherche (Bing), pour un nombre moyen de six liens officiels par interprète.
 - Dans le panier d'Espaces Temps, 49 des 50 chansons du panier ont des résultats de recherche dans Bing, pour un nombre moyen de 13 431 résultats de recherche par paire « titre de chanson + interprète ».

14. Créée à partir des interprètes de musique québécoise dans Amazon.

- Amazon
 - Un peu plus de la moitié (26/50) de ces chansons sont disponibles dans Amazon.
 - Celles qui le sont ont toutes été évaluées au moins une fois (nombre moyen d'évaluations de 8,3), pour une note moyenne de 4,5/5.
 - Sur 50 chansons, 24 se trouvent dans le classement des ventes d'Amazon, pour un classement moyen à la 128 088^e position des ventes de la catégorie « Musique » dans Amazon.
 - Si on s'intéresse seulement aux 100 meilleurs vendeurs en musique québécoise sur Amazon (sans valider la « québécity » des auteurs), ceux-ci se classent en moyenne à la 3 640^e position dans le total des ventes de la catégorie « Musique québécoise » sur la plateforme.
- Facebook
 - Sur 94 interprètes, 80 ont reçu au moins une mention « J'aime », pour une moyenne de 122 216 réactions par interprète.
 - Sur ces 94, 9 ont par ailleurs des adeptes, pour une moyenne de 476 017 adeptes par interprète.
- Les 36 interprètes différents du panier d'Espaces Temps ont en moyenne 58 509 auditeurs et 1 178 423 écoutes sur Last.fm.
 - En s'éloignant du panier, il serait possible de produire des résultats semblables pour 5 223 interprètes québécois reconnus par MusicBrainz.
- Par requête sur les données liées, on réussit à retrouver 72 des 94 interprètes recherchés.
 - Sur les 94 interprètes, 23 sont lauréats d'au moins un prix. En moyenne, ces 23 lauréats ont reçu 3,8 reconnaissances officielles.

Tableau 7

Résultats de l'exercice de moissonnage dans le domaine de la musique enregistrée réalisé par La culture crée

Indicateur	Source	Unités recherchées		Unités avec résultats		Somme des résultats	Résultat moyen
		Description	n ^{bre}	n ^{bre}	%		
Résultats obtenus par moissonnage du Web							
Dimension – Présence							
Liens officiels de l'artiste (#)	Bing	Artiste	94	94	100	560	6,0
Mentions	Bing	Artiste – Chanson	50	49	98	658 141	13 431
Disponibilité (%)	Amazon	Artiste – Chanson	50	26	52	s.o.	s.o.
Dimension – Visibilité							
Évaluations (#)	Amazon	Artiste – Chanson	50	26	52	219	8,3
Évaluation moyenne (#)	Amazon	Artiste – Chanson	50	26	52	119	4,5
Réactions (#) (« J'aime »)	Facebook	Artiste	94	80	85	9 777 312	122 216
Adeptes (#) (« abonnés »)	Facebook	Artiste	94	9	10	4 284 153	476 017
Dimension – Consommation							
Classement moyen des ventes	Amazon	Artiste – Chanson	50	24	49	3 074 112	128 088
Nombre d'auditeurs	Last.fm	Artistes	36	36	100	2 106 336	58 509
Nombre d'écoutes	Last.fm	Artistes	36	36	100	42 423 219	1 178 423
Classement moyen des ventes dans la catégorie « Musique » des 100 titres de la liste « Meilleures ventes en musique québécoise »	Amazon	Chanson	100	100	100	364 032	3 640
Résultats obtenus par requête sur des données liées (SPARQL)							
Dimension – Présence							
Musiciens et groupes de musique québécois	Music-Brainz	Artistes	5 223	5 223	100	s.o.	s.o.
Musiciens et groupes de musique québécois (correspondance avec MusicBrainz)	Last.fm	Artistes	5 223	5 223	100	s.o.	s.o.
Taux de correspondance entre le panier et les identificateurs globaux uniques de Wikidata	Wikidata	Artiste	94	72	77	s.o.	s.o.
Dimension – Visibilité							
Prix (#)	Wikidata	Artiste	72	23	32	87	3,8

1.2.5 Arts de la scène

Les données qui peuvent être moissonnées ou explorées par requêtes sur des données liées à propos des spectacles en arts de la scène ont trait à la présence et à la visibilité, les caractéristiques de l'offre étant les principales données circulant sur les réseaux numériques.

- Dans sa définition des contenus culturels à l'étude, le mandataire nomme certaines plateformes de diffusion (en direct ou en différé) de spectacles en arts de la scène qui n'ont pas été abordées dans le rapport d'Espaces Temps. Il s'agit de sites principalement destinés à un auditoire québécois, mais qui pourraient aussi être consultés de l'étranger : La Fabrique culturelle, Yoop, Livetoune, Latrame et Lepointdevente.com (un site de vente de billets pour les spectacles en ligne et en salle). Leur interface ne précise toutefois pas combien de fois les œuvres qu'on y retrouve ont été vues. Ces sites n'ont pas été moissonnés, mais ils peuvent avoir été dénombrés parmi les mentions dans les moteurs de recherche.
 - Les données moissonnées portent pour leur part sur un panier de 38 spectacles défini par Espaces Temps.
 - Ces 38 spectacles ont obtenu, en moyenne, 29 247 résultats de recherche dans Bing, 8 742 réactions sur Facebook et 124 commentaires.
 - En cherchant plutôt à partir des données liées, on dénombre 1 399 artistes des arts de la scène nés au Québec.
 - Des 38 artistes (ou compagnies) associés aux spectacles du panier, 29 font partie des 1 399 artistes des arts de la scène nés au Québec recensés dans Wikidata. Selon Wikidata, 14 (48 %) de ces 29 artistes sont lauréats d'au moins une reconnaissance officielle, le nombre moyen de prix inscrits dans Wikidata étant de 4,1.
 - Parmi les 1 399 artistes des arts de la scène, la proportion de lauréats est plus faible (12 %), mais le nombre moyen de prix est semblable à celui du panier.
- Les artistes et les spectacles du panier semblent ainsi déjà avoir une plus forte reconnaissance que l'ensemble des artistes en arts de la scène nés au Québec répertoriés dans Wikidata. Cet écart soulève particulièrement la question de la représentativité d'un panier type.

Tableau 8

Résultats de l'exercice de moissonnage dans le domaine des arts de la scène réalisé par La culture crée

Indicateur	Source	Unités recherchées		Unités avec résultats		Somme des résultats	Résultat moyen
		Description	n ^{bre}	n ^{bre}	%		
Résultats obtenus par moissonnage du Web							
Dimension – Présence							
Nombre de mentions (résultats) dans les moteurs de recherche	Bing	Artistes – Spectacles	38	38	100	1 115 198	29 247
Dimension – Visibilité							
Réactions (#) (« J'aime »)	Facebook	Artistes	38	38	100	332 196	8 742
Commentaires (#)	Facebook	Artistes	38	38	100	4 712	124
Résultats obtenus par requête sur des données liées (SPARQL)							
Dimension – Présence							
Taux de correspondance entre le panier (38) et la requête générale Wikidata (artistes de la scène nés au Québec)	Wikidata	Artistes	38	29	76	s.o.	s.o.
Dimension – Visibilité							
Prix obtenus inscrits dans la page – base panier	Wikidata	Artistes	29	14	48	57	4,1
Prix obtenus inscrits dans la page – artistes de la scène nés au Québec	Wikidata	Artistes	1 399	171	12	744	4,4

1.2.6 Patrimoine

Bien que les contenus patrimoniaux se distinguent des autres contenus culturels québécois étudiés par leur caractère non commercial, leur suivi sur les réseaux numériques présente une certaine complexité. Selon le mandataire, celle-ci peut s'apparenter à la complexité qui caractérise le suivi de la musique enregistrée. Selon cette analogie :

- les institutions/organisations s'apparentent aux producteurs/groupes ou interprètes ;
- les expositions et collections peuvent se comparer à des albums ;
- les objets des expositions et collections se rapprochent des pistes.

Cela montre bien à quel point il est complexe de définir le rayonnement dans le domaine patrimonial. S'intéresse-t-on plus particulièrement à la notoriété des institutions québécoises et des contenus qu'elles mettent en valeur (quelle que soit leur origine) ou voulons-nous mettre l'accent sur les contenus patrimoniaux québécois, quel que soit leur lieu de diffusion ? Et encore, l'analogie ne tient pas compte de l'origine des artistes ayant réalisé les œuvres, lorsqu'ils sont connus.

Quinze « entités » ont été retenues à partir de la liste transmise par le MCC pour constituer le panier destiné à l'exercice de moissonnage en tenant compte de la difficulté d'appliquer à la lettre la stratégie envisagée dans les autres domaines, de la complexité des mécanismes du rayonnement dans le domaine du patrimoine et de l'absence de travaux préalables dans l'approche 1.

- Les sites de certaines institutions muséales et ressources Web en archéologie et quelques sites consacrés à l'histoire et à la généalogie ont été privilégiés en raison de la richesse de leurs contenus.
- Il s'agit par ailleurs principalement d'un banc d'essai pour vérifier si, à ce jour, les requêtes sur des données liées permettraient de dégager certaines mesures qui pourraient être utilisées pour quantifier le rayonnement dans le domaine du patrimoine.

En somme, peu de résultats peuvent être recueillis de cette façon, notamment pour les raisons suivantes :

- Des 15 entités du panier, 10 ont un identifiant dans Wikidata ;
- Le nombre de pages vues dans Wikipédia est disponible pour 8 organisations, qui ont été consultées, en moyenne, 4 347 fois au cours des 12 derniers mois ;
- Aucune mention de reconnaissance officielle n'est recensée, et ce, même si on élargit le panier aux 33 musées d'histoire du Québec mentionnés sur la page en.wikipedia.org/wiki/Category:History_museums_in_Quebec.
- Sur les 15 entités, 7 peuvent être repérées à l'aide d'un *International Standard Name Identifier* (ISNI), et 6 peuvent l'être à l'aide d'un identifiant Worldcat.

Tableau 9

Résultats de l'exercice de moissonnage dans le domaine du patrimoine réalisé par La culture crée

Indicateur	Source	Unités recherchées		Unités avec résultats		Somme des résultats	Résultat moyen
		Description	n ^{bre}	n ^{bre}	%		
Résultats obtenus par moissonnage du Web							
Dimension – Présence							
Mentions	Bing	Organisation	15	15	100	564 380	37 625
Résultats obtenus par requête sur des données liées (SPARQL)							
Dimension – Présence							
Taux de correspondance entre le panier et les identificateurs globaux uniques Wikidata	Wikidata	Organisation	15	10	67	s.o.	s.o.
Taux de correspondance entre le panier et les identificateurs globaux uniques ISNI	Wikidata	Organisation	15	7	47	s.o.	s.o.
Taux de correspondance entre le panier et les identificateurs globaux uniques WorldCat	Wikidata	Organisation	15	6	40	s.o.	s.o.
Dimension – Visibilité							
Nombre moyen de pages vues sur Wikipédia au cours des 12 derniers mois	Wikipédia	Organisation	15	8	53	34 775	4 347 (ou 362/ mois)
Nombre de prix (#) sur une liste de musées de DBpedia	DBpedia	Organisation	33	0	0	0	0
Prix (#)	Wikidata	Organisation	15	0	0	0	0

1.3 Évaluation des sources et recensement d'indicateurs potentiels

À la lumière du travail effectué, on peut affirmer qu'aucune des sources recensées ne permet de mesurer les diverses dimensions du rayonnement des contenus culturels québécois sur les réseaux numériques dans l'un ou l'autre des domaines culturels à l'étude.

Certaines sources permettent d'analyser partiellement l'une ou l'autre des dimensions du rayonnement auprès de publics internationaux pour certains contenus. Quelques-unes de ces sources portent sur Québec, mais principalement sur le territoire canadien. Les renseignements recueillis jusqu'à maintenant par moissonnage direct du Web ne permettent pas d'accéder aux données indiquant la région de résidence de l'auditoire.

Certains mandataires ont exploré des sources plus vastes, mais dont l'exhaustivité demeure à démontrer, alors que d'autres ont tablé sur des approches qui leur paraissaient plus prometteuses, même si elles ne portaient que sur l'utilisation d'un service en particulier.

Tous domaines confondus, les différentes études permettent tout de même d'entrevoir des éléments qui pourraient être mesurés. En voici un bref inventaire, dressé selon la portée des mesures du rayonnement.

1.3.1 Présence

En matière de présence, les mesures suggérées portent d'abord sur la disponibilité des contenus dans les plateformes de diffusion, de leur mention dans des bases de connaissances et de leur mention dans les moteurs de recherche.

L'indicateur de disponibilité permet de mesurer l'offre, alors que les mentions dans des bases de connaissances témoignent d'une certaine notoriété, surtout lorsque la fréquence de consultation des pages est disponible, par exemple dans Wikipédia. Les bases de connaissances sont aussi plus souvent structurées en données ouvertes et liées ; c'est le cas de MusicBrainz, de Worldcat et de Wikipédia.

Les données de disponibilité dans les plateformes de diffusion sont pour leur part rarement regroupées¹⁵. Il faudra donc déterminer quelles plateformes de diffusion sont à considérer. Les données d'analytique Web sont généralement structurées de manière à suivre la production d'un créateur ou d'une œuvre en particulier. La majorité des essais effectués reposent sur la constitution d'un panier de contenus à observer, aucune plateforme de diffusion n'étant structurée pour les requêtes par données liées, exception faite de Last.fm dans le domaine de la musique enregistrée.

Souvent proposé par les mandataires, le nombre de mentions dans les moteurs de recherche est un indicateur difficile à interpréter, puisqu'il combine la mesure de l'offre et celle de la notoriété, en plus de compiler toutes les autres mentions du contenu ou de ses créateurs dans le Web. En somme, il s'agit d'un indice de présence Web qui demeure plutôt obscur. Cet indicateur est collecté par moissonnage et est donc assujéti aux algorithmes du moteur de recherche en vigueur au moment où le contenu est recherché. De plus, plusieurs moteurs de recherche utilisent les préférences personnelles des utilisateurs, ce qui oriente les résultats obtenus. Dans ce type de recherche, les mots utilisés pour nommer l'œuvre ou le nom de l'auteur peuvent induire des biais de mesure¹⁶.

15. Sauf dans un agrégateur de données comme Hercules, dont la liste de sites à moissonner est déterminée par le client.

16. Par exemple, une recherche sur le groupe de musique « WD-40 » peut comprendre plus de résultats hors sujet que si l'on recherche « Elisapie Isaac ».

Il semble que cet indicateur pourrait livrer de l'information lorsqu'il cible les pages officielles des œuvres, des artistes, des interprètes, des auteurs ou des institutions. Sa portée est donc limitée, et puisque le résultat serait difficile à interpréter d'un point de vue de rayonnement, il ne semble pas porteur de sens.

En bref :

- Indicateur de l'offre – disponibilité des contenus sur les plateformes de diffusion
- Indicateur de notoriété – mentions des contenus dans des bases de connaissances
- Indice de référencement – mentions des contenus dans les moteurs de recherche en général, soit un indicateur relatif de « l'existence sur le Web »

Ces indicateurs présentent un niveau élevé de difficulté pour ce qui est de la collecte et de l'interprétation en matière de rayonnement.

À ces mesures effectives de présence s'ajoutent les mesures de potentiel de présence suggérées par les mandataires, soit la composante « P » de l'indicateur de découvrabilité du LATICCE et le score SEO de certaines pages Web proposant des contenus. Ces mesures ne reflètent cependant pas les manifestations du rayonnement à proprement parler, puisqu'elles évaluent des facteurs favorisant le rayonnement.

1.3.2 Visibilité

Les sources de données recensées ne couvrent pas toutes les trois catégories de marqueurs de visibilité définis à la figure 1, soit les mentions de reconnaissance officielles, la présence médiatique et les manifestations d'intérêt sur les plateformes de diffusion et les réseaux sociaux.

- Mentions de reconnaissance officielles
 - Parmi les sources de données existantes, seule BANQ indiquerait dans son répertoire les lauréats de prix littéraires.
 - Des données structurées sur les nominations, les lauréats et les reconnaissances officielles sont toutefois disponibles dans Wikipédia. Il est donc possible de recueillir ces renseignements par requêtes sur des données liées, pour autant que les concepteurs de la page les aient inscrits. Il s'agit alors d'un dénombrement minimum, mais qui a l'avantage de pouvoir être réalisé pour un ensemble de contenus notés dans Wikidata, et non seulement pour une liste définie (panier) d'œuvres ou de créateurs.
- Présence médiatique
 - Mentions dans les médias : L'outil de recherche Eureka, qui permet de rechercher et de consulter de l'information diffusée dans les médias, a été suggéré, mais peu détaillé. Il reste à déterminer s'il serait possible – si encore pertinent – de distinguer les mentions dans les médias traditionnels de celles diffusées sur les réseaux numériques.
 - Mesure de la présence publicitaire dans les médias numériques : La seule source recensée à ce sujet est l'application créée par l'agence de publicité Collaboration spéciale pour mesurer les retombées des investissements publicitaires dans les réseaux sociaux de ses clients.

- Réactions, mentions, évaluations et partages dans les réseaux sociaux, les plateformes de diffusion et les bases de connaissances
 - Les réactions (mentions « J'aime »), les évaluations et les commentaires peuvent être dénombrés de façon cumulative dans les réseaux sociaux, les plateformes et même dans certaines bases de connaissances collaboratives (IMDb, Rotten Tomato, etc.).
 - Se pose alors la question du sens que l'on peut donner à un résultat moyen ou cumulatif pour ce type de renseignement. Quelle interprétation peut-on faire d'un nombre moyen de réactions ou d'évaluations sur un ensemble de contenus? L'évaluation moyenne d'un produit ou la cote d'un artiste pourrait témoigner d'un certain succès d'estime, mais l'effort associé au calcul de la moyenne des évaluations d'un panier de contenus dans un domaine donné – ou d'une cote moyenne des artistes québécois – semble considérable par rapport aux résultats qui seraient ainsi obtenus.
 - Le nombre moyen d'abonnés ou d'adeptes d'artistes dans les réseaux sociaux, qui peut se rapprocher d'un palmarès, semble, par comparaison, plus évocateur.
 - Des services tels que ceux de Parrot Analytics, de Chartmetrics ou de Soundcharts agrègent déjà certaines données en tenant compte, entre autres, des interactions sur les réseaux sociaux. Leur méthode demeure cependant opaque et leurs données surtout utiles aux fins d'informatique décisionnelle, pour suivre le cheminement d'une œuvre ou d'un interprète.

1.3.3 Recommandation

Très peu de renseignements peuvent être recueillis sur les recommandations algorithmiques. Ces dernières sont l'application informatique des stratégies d'affaires des moteurs de recherches, des réseaux sociaux et des plateformes de diffusion. Ainsi, les seules mesures suggérées à ce sujet portent sur l'évaluation du potentiel de recommandation.

- La composante « R » de l'indice de découvrabilité du LATICCE témoigne du potentiel de recommandation de certains contenus culturels québécois selon des comportements d'écoute modélisés par des persona.
- Une autre façon d'évaluer le potentiel de recommandation par les moteurs de recherche serait de modéliser une mesure de la richesse des métadonnées accolées aux contenus culturels québécois. Il s'agirait alors d'un autre sujet d'étude, portant sur la mesure des facteurs favorisant le rayonnement.
- Les pratiques des individus en matière de découverte et de consommation culturelle dans les réseaux numériques sont par ailleurs encore peu documentées. L'importance relative que l'on accorde présentement à l'impact des algorithmes de recommandation sur la découverte de contenus demeure donc spéculative.

1.3.4 Consommation

En matière de mesure de la consommation, quelques pistes se dégagent, mais aucune ne permet à ce jour d'accéder aisément aux données de consommation de l'ensemble des contenus culturels québécois disponibles sur les réseaux numériques.

- Des données d'écoute (nombre de vues/d'écoutes et nombre d'auditeurs) sur les principales plateformes sont communiquées aux clients de services tels que Parrot Analytics, Chartmetrics ou Soundcharts. Elles sont parfois classées selon la région de résidence de l'auditoire. Ces services conçus pour évaluer le positionnement stratégique d'un contenu audiovisuel (Parrot Analytics) ou d'un interprète (Chartmetrics ou Soundcharts) ont l'avantage de couvrir la consommation mondiale et de regrouper plusieurs sources de données (principales plateformes de diffusion et réseaux sociaux¹⁷).
 - Leurs méthodologies ne sont toutefois pas explicites. De plus, leurs données n'ont pas été développées dans le but de suivre un ensemble d'œuvres ou d'interprètes. Par contre, il paraît plus facile de comprendre la portée d'un nombre total d'écoutes par continent pour une liste de titres que le nombre moyen (ou total) de réactions sur les réseaux sociaux.
 - Les travaux entrepris par le FMC avec Parrot Analytics témoignent de la faisabilité d'utiliser ce type de service pour dresser un portrait plus global de la demande.
 - Dans une perspective de production d'indicateurs économiques, des entretiens avec des représentants de ces sources mériteraient d'être envisagés pour évaluer leur véritable potentiel.
- Quelques sources de données sur la consommation portent sur le territoire québécois ou canadien.
 - C'est le cas pour la majorité des sources mentionnées par les mandataires de l'approche ¹⁸. À titre d'exemple, dans le domaine de la musique enregistrée, le travail entrepris par l'Observatoire visant à exploiter les données de Nielsen/MRC Data permettra à terme de dresser un portrait de l'écoute de musique en continu de titres et d'artistes québécois sur le territoire canadien.
- L'exercice de moissonnage direct révèle pour sa part que de tenter de calculer la valeur moyenne d'un ensemble de contenus et d'établir, par exemple, le rang moyen des ventes sur une plateforme comme Amazon sont des opérations peu pertinentes, dont les résultats sont difficiles à interpréter.

17. Ces sources fournissent aussi des renseignements sur la présence dans les réseaux sociaux, ce qui permettrait de lier ces données de consommation à des indicateurs de présence et de visibilité.

18. Bibliopresto, BAnQ, Demarque, Numéris VAM, Viewer Centric, Tubular Labs, Nielsen/MRC Data, données de l'Observatoire, etc.

2

Synthèse

2.1 Enjeu

Le défi à l'origine de nos travaux était de concevoir des indicateurs de rayonnement de l'offre culturelle québécoise sur les réseaux numériques. Dans le cadre de notre travail exploratoire, nous avons mis en place un processus de réflexion et d'expérimentation pour cerner et définir des marqueurs de rayonnement. Au terme de ce processus, nous avons dégagé des lignes directrices qui malheureusement ne permettent pas de conclure que la création de ce type d'indicateurs est possible à court terme. Toutefois, nous avons pu relever des pistes à explorer et proposer des actions qui pourront éventuellement nous aider à produire des indicateurs répondant à certains besoins, que ce soit dans une perspective d'informatique décisionnelle ou d'information statistique.

2.2 Observations générales

- Les données ouvertes disponibles sur le Web ont peu à offrir pour évaluer l'apport québécois à la diversité des expressions culturelles. En général, les contenus québécois n'ont pas de caractère distinctif. Puisque tout contenu canadien francophone n'est pas nécessairement québécois et que tout contenu canadien non francophone peut aussi être québécois, l'usage du critère de provenance « Canada » est insuffisant. Pour opérer cette distinction, il faut superposer des critères et faire appel à des sources de données externes afin de croiser les renseignements avec, par exemple, une liste de créateurs ou d'œuvres reconnus comme québécois.
- À partir des outils disponibles actuellement, il est techniquement impossible de recenser tous les contenus culturels québécois offerts ou décrits sur les réseaux numériques. En conséquence, plusieurs mandataires ont indiqué que la production d'indicateurs de rayonnement des contenus québécois devrait s'appuyer sur un procédé d'échantillonnage (ou paniers) de contenus culturels représentatifs.
 - Dans ce cas, la constitution de certains indicateurs implique la constitution de paniers de produits comportant une part de produits récents (nouveau) et une part de produits déjà diffusés (catalogue), ce qui nécessite de mettre en place un processus de vigie pour signaler les nouveautés. L'établissement d'un panier de produits suppose l'instauration de critères objectifs, notamment pour le choix des produits, des titres et des auteurs, sachant toutefois qu'une des caractéristiques des industries culturelles est l'imprévisibilité du succès. Ainsi, l'absence dans le panier d'un produit à succès rendra l'indicateur non pertinent et obsolète.
- Une autre avenue serait de limiter la mesure du rayonnement au Québec ou au Canada, sans tenir compte du rayonnement international.
- Certains contributeurs aux travaux, plus familiers avec les industries de la culture (musique enregistrée, spectacle, audiovisuel) semblent privilégier une mesure du rayonnement à l'intérieur du marché québécois, notamment parce que c'est surtout là que le rayonnement peut se traduire en revenus directs.

2.3 Observations techniques

- Les mandataires ont tous soulevé le manque de maturité numérique des domaines culturels étudiés, qui limite l'utilisation à leur plein potentiel des outils exploitant les données liées (dont le langage SPARQL¹⁹) pour refléter convenablement la situation de l'ensemble des contenus québécois à court ou à moyen terme.
 - Amener les différents secteurs culturels à maturité numérique est l'un des objectifs du PCNQ. À cet effet, le gouvernement du Québec a récemment diffusé un document d'accompagnement pour les organismes culturels qui souhaiteraient ajouter des jeux de données ouvertes sur le portail Données Québec²⁰.
 - Une avenue à envisager pourrait être la création d'indicateurs prototypes à partir de requêtes sur les données liées, dont la représentativité augmenterait au fur et à mesure de l'adoption de bonnes pratiques en matière de métadonnées et de la diffusion de ces données en format ouvert et lié.
- Mentionnons que les travaux réalisés jusqu'à maintenant l'ont été sans abonnement à des services d'analytique Web. Aucune entente n'a été prise avec les plateformes dont on voudrait moissonner le contenu à l'aide d'outils d'analytique Web.
 - Des projets pilotes sectoriels qui étudieraient les résultats obtenus à partir d'abonnements aux services d'analytique Web les plus prometteurs pourraient être envisagés.
- Une autre avenue pourrait être d'orienter l'objet d'étude vers la mesure de l'adoption de pratiques favorisant le rayonnement.
 - Certains indicateurs proposés par les mandataires mesurent les facteurs pouvant influencer le rayonnement, dont le score SEO, le nombre de sujets associés à un contenu, la présence ou l'absence de certaines métadonnées clés, etc. Ces indicateurs sont intéressants pour les éditeurs, les producteurs et les diffuseurs qui cherchent à accroître la découvrabilité de leurs contenus. Ils pourraient aussi être utilisés pour étudier certaines variations observées dans le rayonnement effectif des contenus.
 - Néanmoins, ce type d'indicateurs demeure secondaire par rapport à l'objet de la présente étude, qui vise à mesurer les manifestations du rayonnement de l'offre culturelle québécoise sur les réseaux numériques. En effet, rappelons que l'objectif initial était ici de déterminer une méthodologie pour effectuer une mesure et évaluer la hausse de la visibilité et de la consultation de l'offre québécoise. Compte tenu de l'état de faible maturité des activités d'indexation des contenus culturels québécois, il pourrait être plus approprié de rediriger les objectifs vers une évaluation de l'implantation de bonnes pratiques en matière de métadonnées. En effet, l'accent mis sur les facteurs du rayonnement par les mandataires ouvre la porte à un recentrage de l'objet d'étude en ce sens.

19. Langage de requête et protocole qui permet de rechercher, d'ajouter, de modifier ou de supprimer des données RDF disponibles sur Internet. Concrètement, une requête SPARQL permet, par exemple, de chercher dans Wikidata des artistes nés au Québec, plutôt que de programmer un par un les noms d'artistes québécois dans une requête de moissonnage traditionnelle.

20. culturenumerique.mcc.gouv.qc.ca/2020-12-14-pour-des-donnees-ouvertes-en-culture/

3

Proposition d'actions à mettre en œuvre

À la lumière des résultats de ces travaux, un premier constat s'impose : il est impossible dans les conditions actuelles de produire des indicateurs de rayonnement de l'offre culturelle québécoise sur les réseaux numériques. Les discussions en comité et les rapports des consultants ont permis d'évoquer des pistes potentielles, mais toutes présentent à divers degrés des difficultés liées soit à l'accès aux données et à leur collecte, soit à leur interprétation. Cela est vrai pour tous les domaines culturels, peu importe le type de sources accessibles sur le Web. Qu'il s'agisse de moteurs de recherche, de plateformes de diffusion ou de bases de connaissances, toutes comportent des limites ou des barrières qui bloquent l'accès aux renseignements les plus pertinents.

Un second constat se dégage : compte tenu des nombreux paramètres existants sur les réseaux numériques, il est nécessaire de faire des choix, d'établir des cadres et de déterminer une problématique afin de mieux cibler et orienter les travaux visant la production d'indicateurs de rayonnement. Dans l'état des choses, il est important de rester modeste et de bien choisir les pistes qui seront explorées afin d'assurer la réussite de la collecte et surtout de livrer des résultats significatifs, c'est-à-dire qui sont pertinents et dont l'interprétation est accessible à tous.

À la lumière de ces constats, on propose de profiter de l'élan suscité par ces travaux pour entreprendre des projets-pilotes qui nous permettront d'accroître les connaissances et d'expérimenter des outils et des partenariats, afin de progresser sur ce terrain encore peu connu ; cette étape d'expérimentation est un passage obligé dans le processus de conception et de production des indicateurs qui nous intéressent.

Pour pousser cette réflexion plus loin, il faudra s'appuyer sur un cadre de travail rigoureux et adopter des règles de conduite. Voici trois règles qui permettraient d'encadrer la problématique et de consacrer les ressources aux dimensions les plus importantes.

La première règle consiste à limiter le champ d'études aux œuvres québécoises répertoriées dans des bases de connaissance reconnues, plutôt que d'adopter des critères de « québécityde », tout en reconnaissant que cela restreint l'exhaustivité de l'univers de création retenu.

La seconde règle concerne la portée géographique du rayonnement ; tous les territoires représentatifs sont à explorer, même si ceux retenus ne sont pas comparables d'un indicateur à l'autre, sauf lorsqu'il s'agit d'analyser le rayonnement sur le marché québécois.

La troisième règle vise à canaliser les travaux vers les indicateurs à portée économique (abonnement, visite, écoute, visionnement, achat, etc.). Ce choix a l'avantage de permettre d'exploiter des données se rapportant au rayonnement social des contenus culturels, notamment lorsqu'il s'agit d'évaluer le profil des consommateurs (provenance, démographie, revenu, scolarité, etc.). Ces renseignements sont utiles à l'ensemble des intervenants et nécessaires aux orientations gouvernementales, car ils sont bien souvent les mieux compris et les plus utilisés dans la gestion des politiques publiques.

Pour chacune des actions fondées sur les rapports des mandataires et les discussions avec les membres du comité de réflexion, le MCC devra mandater un expert pour prévoir des modalités de mise en œuvre.

Action 1 – Approcher le Fonds des médias du Canada

Il est très difficile de produire des indicateurs sur le rayonnement des contenus culturels. La preuve en est qu'aucun indicateur pertinent, fiable et récurrent n'a été produit à ce jour ni ici ni à l'étranger. À titre d'exemple, le Fonds des médias du Canada s'active depuis longtemps sur diverses hypothèses et stratégies pour se doter d'indicateurs sur la portée des contenus audiovisuels canadiens sur le Web. Le travail réalisé par l'organisme pourrait être très utile et contribuer à l'établissement d'un indicateur de rayonnement.

On propose donc d'approcher le Fonds des médias du Canada pour lui proposer de collaborer à son projet d'infrastructures de données sur les médias numériques, l'objectif étant de voir si ce projet pourrait fournir des données utiles à la création d'un indicateur de rayonnement des contenus audiovisuels québécois sur tous les marchés.

Action 2 – Commanditer une recherche sur les agrégateurs de données

Une autre manière d'avancer dans la production d'indicateurs serait de commanditer une recherche visant à recenser les données qui sont disponibles et exploitables avec un abonnement à une base de données agrégées comme Parrot Analytics pour l'audiovisuel et SoundCharts ou ChartMetric pour la musique enregistrée.

La Demand expression de Parrot Analytics, qui est l'une des sources examinées par le modèle du Fonds des médias du Canada, est potentiellement le seul indice de rayonnement global – qui mesure à la fois la présence, la visibilité et la consommation – recensé au cours de nos travaux.

Par ailleurs, rappelons que l'Observatoire, avec l'appui du MCC, a déjà établi des ententes de partenariats d'affaires pour produire des indicateurs, notamment avec MRC Data (Music Connect), afin de concevoir un indicateur pour l'écoute en continu des œuvres musicales québécoises au Canada.

Action 3 – Explorer l'univers Wiki

La présence des contenus culturels sur le Web et la qualité des métadonnées qui les accompagnent sont des éléments clés pour la production de certains indicateurs de rayonnement.

Par conséquent, on propose de recenser les œuvres québécoises disposant d'identifiants dans Wikidata et d'extraire les données sur la fréquentation des pages sur ces œuvres dans Wikipédia. L'exploration de l'univers Wiki devrait permettre de construire un indicateur de présence et de consultation des renseignements à propos des contenus culturels québécois selon une périodicité à déterminer (semestrielle, annuelle, etc.). L'exploration des métadonnées sur ces contenus devrait aussi permettre d'évaluer leur qualité afin de les utiliser comme base de référence pour repérer les œuvres sur les plateformes de diffusion et en extraire d'autres indicateurs, dont des indicateurs de consommation.

Action 4 – Analyser le rayonnement de la littérature québécoise à partir de Worldcat

Quelques actions ponctuelles ont aussi le potentiel de fournir des indicateurs uniques sur des données ouvertes et liées, tout en permettant d'améliorer les compétences générales et d'évaluer les possibilités d'utilisation de nouveaux outils technologiques. Par exemple, dans l'univers du livre, les bibliothèques sont dotées de données bien structurées. On propose donc d'exploiter la base de données Worldcat afin d'analyser la disponibilité des œuvres littéraires québécoises dans les bibliothèques du monde.

Action 5 – Analyser la portée des données de MusicBrainz et de Last.fm

Un autre projet expérimental d'exploitation des données ouvertes et liées dans le domaine de la musique pourrait s'avérer prometteur : il consiste à croiser les données de la base de référence MusicBrainz avec les données de Last.fm sur le nombre d'écoutes et d'auditeurs.

Il faudrait préalablement évaluer si la base de référence MusicBrainz est suffisamment exhaustive pour cibler les enregistrements québécois et la représentativité sociodémographique des auditeurs ayant téléchargé l'application Last.fm.

Action 6 – Soutenir l'utilisation des données du Web dans la production de statistiques

Les travaux en cours qui sont en mesure de livrer des indicateurs sur la consommation de contenus québécois sur le Web devraient être appuyés. Deux initiatives de l'Observatoire vont en ce sens et ont besoin d'un soutien institutionnel.

D'abord, le déploiement du plein potentiel de *l'Enquête sur les ventes de livres numériques* pourrait permettre de fournir quelques indicateurs statistiques supplémentaires, entre autres sur la nationalité, la langue ou le genre des œuvres littéraires vendues au Québec et à l'étranger par les librairies numériques et les agrégateurs québécois. Un important travail est à faire en partenariat avec les librairies numériques et les agrégateurs québécois en collaboration avec Bibliopresto et la BTLF pour que l'Observatoire obtienne les données détaillées nécessaires à la production de statistiques globales et fiables.

Puis, on devrait appuyer l'Observatoire pour développer sa stratégie visant à faire évoluer son *Enquête sur la fréquentation des spectacles au Québec* afin d'élargir et d'améliorer les statistiques sur la diffusion et la fréquentation des spectacles au Québec, et de les obtenir plus rapidement. Il y a là un fort potentiel d'indicateurs sur la dynamique des plateformes dans la promotion des spectacles vivants. Une collaboration doit être établie avec les diverses parties prenantes dans ce secteur, notamment avec les entreprises de ventes de billets (Ticket pro, La vitrine, Le Point de vente, Ticketmaster, etc.) et les diffuseurs en vue de réunir tous les renseignements sur la diffusion des œuvres sur le territoire québécois. Cette piste devrait être attentivement explorée pour la production de données sur le profil des consommateurs.

Action 7 – Évaluer les effets sur le rayonnement dans le Web de l'enrichissement par des métadonnées d'un panier de contenus culturels numériques

L'une des méthodes suggérées par les mandataires pour produire des indicateurs de rayonnement consiste à utiliser un panier représentatif du contenu culturel québécois pour le domaine visé. Cette méthode est employée en raison des limites techniques des applications actuellement disponibles.

La proposition consiste à tester la performance relative d'un panier (sous-groupe) de contenus culturels dont la qualité des métadonnées devrait, en principe, favoriser le rayonnement et de le comparer à un panier de contenus culturels similaires, mais pour lequel il y a eu peu d'efforts fournis en matière d'enrichissement des métadonnées. Une autre option serait de suivre un même panier de contenus dans le temps, soit avant et après avoir entrepris un travail d'indexation par des métadonnées de qualité.

Un mandat pourrait être confié en ce sens à une firme comme Hercules. Ce serait l'occasion de vérifier s'il existe une corrélation entre certains facteurs de rayonnement (indice de découvrabilité du LATICCE, score SEO, nombre de sujets associés) et la disponibilité des contenus visés sur les plateformes de diffusion, le nombre de contenus évalués par l'auditoire ou le nombre moyen d'évaluations reçues.

À défaut de mesurer le rayonnement des contenus culturels québécois, cet exercice permettrait de jauger si les efforts consentis entre autres par le Plan culturel numérique en matière de métadonnées portent fruit.

4

Recommandations

Parmi les réflexions stratégiques qui se dégagent de ces travaux, il apparaît clairement que l'un des paradoxes de l'ère numérique réside dans la complexification de la mesure du succès et plus encore de celle du rayonnement, alors que les outils de mesures sont de plus en plus nombreux et que les succès sont de plus en plus souvent planétaires.

Les recherches menées par l'Observatoire et ses mandataires révèlent que les sources de données exploitables à court terme pour alimenter l'un ou l'autre des marqueurs du rayonnement mentionnés à la figure 1 offrent des accès limités, des données incomplètes et des résultats difficiles à interpréter. Par ailleurs, les quelques sources de données ou approches considérées comme prometteuses par les mandataires comportent toutes d'importantes limites. Les résultats obtenus dépendent, entre autres, des paramètres qui ont été sommairement décrits dans la première phase du projet. Les résultats étant demeurés approximatifs, l'exploration des sources, trop nombreuses, s'est avérée laborieuse. Néanmoins, quelques avenues sont intéressantes et certains indicateurs mériteraient d'être examinés davantage, dans la mesure où des ressources suffisantes peuvent y être consacrées.

À cet effet, les quelques actions proposées au chapitre précédent, si elles sont réalisées, permettront de porter un éclairage sur certains indicateurs de rayonnement.

Les autorités devront toutefois se pencher sur certains facteurs conjoncturels avant d'aller plus loin dans ces démarches complexes. Vous trouverez ci-dessous des recommandations à cet effet.

Le point de départ pour sélectionner des indicateurs pertinents est de procéder par élimination (processus du tamis), et cette sélection devra être précédée par l'établissement d'un système référentiel d'indicateurs de rayonnement. Ce système devra se fonder sur les conclusions les plus pertinentes du présent rapport et préciser ce qui devrait être mesuré en priorisant un nombre limité d'indicateurs qui ont le réel potentiel de fournir des renseignements significatifs et une interprétation sans ambiguïté.

Recommandation 1

Élaborer, en collaboration avec des partenaires nationaux et internationaux, un système référentiel d'indicateurs de rayonnement.

L'un des points forts du Québec, dans l'atteinte de cet objectif, est sa capacité à travailler en concertation avec les milieux concernés, tant gouvernementaux, qu'institutionnels (recherche et enseignement universitaire), associatifs (regroupements des milieux culturels) et citoyens. Un autre de ses points forts est sa capacité à réunir des intervenants de plusieurs domaines culturels, ce qui lui permet, entre autres, de diffuser les connaissances des uns et des autres de même que d'adopter et d'adapter, s'il y a lieu, les meilleures pratiques. Ainsi, les idées, les concepts et les expérimentations sont partagés plus largement. Il en résulte une meilleure cohérence entre les actions visant un même objectif.

Recommandation 2

Poursuivre les travaux entrepris par l'entremise d'une tribune collaborative favorisant les échanges et l'acquisition de connaissances, de façon à faciliter le suivi des travaux en cours ici, au Canada et à l'étranger ; c'est en misant sur l'expérimentation et l'essai-erreur que les institutions et les entreprises nationales pourront opérationnaliser le concept de découvrabilité.

Les simulations de moissonnage visant à générer des données permettant d'alimenter des indicateurs portent à croire que les méthodes s'appuyant sur les données liées sont prometteuses. Toutefois, les résultats provenant de ces requêtes, qui sont basées sur des données ouvertes et liées, sont insuffisants pour générer les renseignements utiles à la production d'indicateurs valables.

Recommandation 3

Poursuivre les efforts de la mesure 111 du Plan culturel numérique du Québec, qui vise à accélérer la création de normes communes de description des contenus culturels, notamment en vue d'enrichir les métadonnées, en misant sur le potentiel des données liées.

Ces travaux ont également été l'occasion de remarquer l'émergence d'entreprises québécoises en technologie numérique spécialisées dans l'analyse des données du Web.

Recommandation 4

Accompagner les entreprises québécoises qui développent leurs compétences dans l'analyse des données des réseaux numériques.

Les travaux réalisés dans le cadre du projet sont riches et inspirants, mais aucun ne débouche sur des propositions concrètes comportant une évaluation des ressources humaines et financières nécessaires à leur poursuite. Lorsqu'un certain nombre d'indicateurs seront sélectionnés, il faudra les examiner afin d'établir un devis de production.

Recommandation 5

Estimer le coût de production de chacun des indicateurs retenus en faisant appel aux organismes et aux entreprises disposant de compétences en la matière.

Par ailleurs, quelques aspects importants ont été très peu examinés dans ces travaux. D'une part, il y a les enjeux concernant la représentativité et la validité des statistiques produites. D'autre part, il y a ceux concernant la propriété et la confidentialité des données. En effet, la production d'indicateurs ou de statistiques publiques est bousculée par la « révolution numérique », qui a amené un volume et une variété considérables de données potentielles qui sont perpétuellement renouvelées. Ces problématiques non résolues sont déterminantes des difficultés rencontrées dans la production d'indicateurs sur le rayonnement des contenus culturels numériques.

Aussi, les données sur le Web sont détenues par des entreprises privées, généralement des multinationales étrangères ayant une succursale accessoire dans le pays, ce qui pose un certain nombre de problèmes liés à l'accès aux données, à leur confidentialité et à leur qualité. À défaut de pouvoir obliger ces entreprises à partager l'information, on doit mettre en place une bonne collaboration avec elles pour nous assurer d'avoir accès aux données, et pour que celles-ci soient de bonne qualité.

L'Observatoire de la culture et des communications de l'Institut de la statistique du Québec examine ces questions depuis de nombreuses années. Quelques expériences sont menées notamment dans le cadre de la production de statistiques sur la consommation de musique en ligne et sur la vente de livres numériques. D'autres avenues sont explorées quant à la façon de recueillir des données sur l'offre et la fréquentation des spectacles au Québec.

Recommandation 6

Miser sur l'expertise, les connaissances et la neutralité de l'Observatoire et sur sa capacité à travailler en partenariat avec les intervenants du secteur de la culture pour conceptualiser et produire des indicateurs de qualité.

Enfin, les enjeux de propriété et la confidentialité des données expliquent peut-être le fait que l'accès aux bases de données des sociétés de gestion des droits d'auteur n'a pas été abordé durant les travaux. Rappelons que les entreprises du FAANG (l'indice boursier regroupant les cinq grandes entreprises dominantes de l'industrie des technologies de l'information aux États-Unis : Facebook, Amazon, Apple, Netflix et Google-Alphabet) profitent largement des vides juridiques et des imprécisions du droit national et international en ce qui a trait à leurs obligations de divulgation. La protection juridique apportée par le secret des affaires ne cesse de progresser et est aujourd'hui très importante. Il existe cependant quelques exceptions juridiques en Europe, mais elles restent peu exploitées pour le moment, vu leur complexité. En Amérique du Nord, peu de choses ont été entreprises à cet effet. Des approches de nature contractuelles sont sans doute possibles avec certaines entreprises, mais les résultats ne couvriraient qu'une partie du marché et ne répondraient pas aux objectifs des statistiques publiques.

Les sociétés de gestion de droits d'auteur, dans une moindre mesure, utilisent cette absence de balises claires pour retenir les données qu'elles obtiennent des plateformes de diffusion. Elles peuvent plaider que ces données découlent d'ententes contractuelles entre deux parties légitimes, qu'elles visent à octroyer à des créateurs des redevances négociées en vertu des lois sur les droits d'auteur, et qu'elles sont – c'est généralement le cas – couvertes par le secret commercial. Il faudra une véritable volonté politique pour régler ce problème, qui, pour l'instant, ne semble pas à l'ordre du jour.

Recommandation 7

Construire un argumentaire afin qu'une impulsion émerge des milieux culturels et politiques à l'égard des données privées. Pour ce faire, mandater des juristes afin de constituer un dossier sur la propriété et la confidentialité des données et sur la protection juridique utilisée notamment par les FAANG et les sociétés de gestion des droits d'auteur. Ce dossier pourrait aussi comporter une réflexion sur la formation d'un collectif chargé de travailler sur la question des données culturelles et du rôle que devraient avoir les instituts de statistique. Ces derniers sont probablement les seuls organismes à avoir la crédibilité, la notoriété, la capacité technologique et le statut juridique nécessaires pour accéder à ces données et les exploiter convenablement, pourvu qu'on leur en donne la capacité financière.

Description des travaux du comité et des mandataires

Tableau 1 : calendrier des activités réalisées

La composition et les contributions respectives des différents groupes mentionnés au tableau 1 sont les suivantes.

Comité de suivi

Un comité de suivi a été constitué en mai 2019 afin de veiller au bon déroulement des travaux. Celui-ci est composé des membres suivants.

- Pour l'Observatoire : Dominique Jutras (jusqu'en février 2020), Christine Routhier et Marik Danvoye ;
- Pour le MCC : Mathieu Rocheleau, Sylvain Provost (remplacé par Alexandre Nault en février 2020) et Harold Gendron.

Le comité, qui a participé à chacune des étapes de la démarche, s'est réuni à six reprises pour effectuer les tâches suivantes.

- Dresser la liste des personnes invitées à participer au comité de réflexion ;
- Remettre à l'Observatoire un document de réflexion pour orienter la rédaction du document de référence présentant les résultats de la recherche documentaire réalisée ;
- Préparer les deux rencontres du comité de réflexion et la rencontre spéciale du comité de concertation de l'Observatoire et revenir sur leurs conclusions ;
- Mandater des sous-traitants chargés de cerner les données ou les statistiques pouvant être à la base d'indicateurs de rayonnement, orienter la production des invitations à soumissionner et évaluer les soumissions reçues.

Comité de réflexion

Composé d'experts de divers horizons, le comité de réflexion s'est réuni à deux reprises afin de contribuer aux réflexions sur :

- la définition du concept de rayonnement, y compris ses ramifications et sa géométrie variable dans un contexte numérique ;
- les facteurs qui influent sur le rayonnement ;
- les conditions préalables et les moyens requis pour la création et la mise en œuvre d'un indicateur de rayonnement ;
- les points qu'un indicateur de rayonnement permettrait d'éclairer ;
- les solutions concrètes qui pourraient être mises en place rapidement et qui pourraient être améliorées au fil du temps (globalement ou par secteur, selon un type de produit ou un territoire donné, etc.) ;
- ce qui pourrait être retiré d'initiatives en cours ou existantes.

Ses membres sont :

- Jean-Robert Bisaillon, codirecteur, secteur de l'enregistrement sonore, Laboratoire de recherche sur la découvrabilité et les transformations des industries culturelles à l'ère du commerce électronique (LATICCE), Centre d'étude sur l'intégration et la mondialisation (CEIM), Université du Québec à Montréal ;
- Catalina Briceno, professeure invitée, École des médias, Université du Québec à Montréal ;
- Viêt Cao, gestionnaire et analyste principal, Synapse C ;
- Nathalie Guay, coordonnatrice générale, Coalition pour la diversité des expressions culturelles ;
- Marie-Hélène Raymond, coordonnatrice de la stratégie numérique, Musée national des beaux-arts du Québec ;
- Christian Roy, consultant et fondateur, Brix Labs (Hercules.report) ;
- Gregory Saumier-Finch, cofondateur et chef technologie, Culture Creates –La culture crée.

Comité de concertation de l'Observatoire de la culture et des communications du Québec

À la suite des deux rencontres du comité de réflexion, l'Observatoire a organisé le 14 janvier 2020 une réunion spéciale de son comité de concertation (l'un de ses comités consultatifs permanents) pour discuter des enjeux et des besoins des milieux culturels quant à la mesure du rayonnement numérique de l'offre culturelle québécoise. Chacun des membres du comité de concertation représente un secteur d'activité. Les membres du comité de concertation qui étaient présents sont les suivants :

- Mylène Cyr, secteur de l'audiovisuel
- Solange Drouin, secteurs de la musique enregistrée et des arts de la scène
- Jacqueline Labelle, secteur du livre
- Gilles Lessard, secteur des arts visuels et des métiers d'art
- Christian Poirier, secteur de la recherche universitaire sur la culture

Activités réalisées	Date
1 ^{re} rencontre du comité de suivi – Organisation du projet et constitution de comité de réflexion	22 mai 2019
2 ^e rencontre du comité de suivi – Document de réflexion	19 juin 2019
Transmission du document de réflexion final produit par le représentant du MCC (document déclencheur)	7 juillet 2019
3 ^e rencontre du comité de suivi – Document de référence	11 septembre 2019
Transmission du cahier de séance de la 1 ^{re} rencontre aux membres du comité de réflexion	17 octobre 2019
1 ^{re} rencontre du comité de réflexion – Cadre théorique	29 octobre 2019
Transmission du cahier de séance de la 2 ^e rencontre aux membres du comité de réflexion	17 décembre 2019
2 ^e rencontre du comité de réflexion – Sources de données	8 janvier 2020
Transmission du cahier de séance aux membres du comité de concertation de l'Observatoire	8 janvier 2020
Rencontre spéciale du comité de concertation de l'Observatoire	14 janvier 2020
4 ^e rencontre du comité de suivi – État d'avancement des travaux	21 janvier 2020
5 ^e rencontre du comité de suivi – Mise au point sur la finalité et les orientations de l'étude de faisabilité	4 février 2020
6 ^e rencontre du comité de suivi – Invitations à soumissionner	8 avril 2020
Envoi des invitations à soumissionner	11 mai 2020
Sélection des mandataires	19 juin 2020
Rencontre de démarrage des travaux avec les mandataires	30 juin 2020
1 ^{re} rencontre de suivi avec les mandataires	13 juillet 200
2 ^e rencontre de suivi avec les mandataires	2 septembre 2020
Dépôt des rapports préliminaires – mandataires de l'approche 1	6 octobre 2020
Dépôt des rapports finaux – mandataires de l'approche 1	2 novembre 2020
Dépôt des rapports préliminaires – mandataire de l'approche 2	23 novembre 2020
Dépôt du rapport final – mandataires de l'approche 2	7 décembre 2020
Remise du rapport synthèse préliminaire de l'Observatoire au comité de réflexion	11 janvier 2021
Remise du rapport synthèse au comité de réflexion et aux mandataires	27 janvier 2021
Rencontre du comité de réflexion et des mandataires	3 février 2021
Dépôt du rapport final	10 février 2021

Travaux des mandataires

Compte tenu de la complexité de l'objet d'étude, la consultation du comité de réflexion ad hoc et du comité consultatif permanent de l'Observatoire ne pouvait à elle seule permettre d'examiner suffisamment les dynamiques sectorielles et les sources de données potentielles dans les cinq domaines culturels à l'étude, ni d'établir ce qui pourrait être tiré d'éventuels outils de moissonnage du Web. On a donc convenu avec le comité de suivi de confier quatre mandats à des experts des domaines concernés.

Les trois premiers mandats, regroupés sous le vocable « approche 1²¹ », visaient, pour un domaine culturel donné, à relever et à commenter les données ou les statistiques déjà colligées par des firmes ou par des outils d'analytique Web de plateformes de diffusion et pouvant être à la base d'indicateurs de rayonnement à court terme. Ces trois mandats demandaient aussi de faire le portrait des mécanismes et des canaux du rayonnement dans le domaine en question. Le quatrième mandat, nommé « approche 2 », visait pour sa part à vérifier la possibilité de créer un outil de moissonnage du Web qui permettrait de recenser certains marqueurs du rayonnement dans les domaines de la musique enregistrée, du spectacle en arts de la scène, de l'audiovisuel, du livre et des contenus patrimoniaux.

Un appel de propositions pour ces mandats de recherche a été diffusé par l'Observatoire au printemps 2020. Le comité de suivi s'est ensuite réuni pour sélectionner des mandataires parmi les propositions reçues. Les quatre mandataires ayant exécuté les travaux sont les suivants.

Approche 1

- Espaces Temps : Étude des sources de statistiques dans les domaines de la musique enregistrée et des spectacles en arts de la scène
 - Chargé de projet : Jean-Robert Bisaillon
- LaCogency : Étude des sources de statistiques existantes dans le domaine de l'audiovisuel
 - Chargée de projet : Danielle Desjardins
- Société de gestion de la banque de titres de langue française (BTLF) : Étude des sources de statistiques existantes dans le domaine du livre
 - Chargé de projet : Patrick Joly

Approche 2

- La culture crée : Banc d'essai d'un outil de moissonnage du Web dans cinq domaines culturels
 - Chargé de projet : Gregory Saumier-Finch

Les quatre mandataires ont réalisé leurs travaux au cours de l'été et de l'automne 2020. Ils ont livré leur rapport final à l'Observatoire en décembre 2020. Les constats qu'on peut en tirer sont synthétisés au chapitre 4 du présent rapport.

21. Une invitation à soumissionner a aussi été lancée dans le domaine du patrimoine. Toutefois, le caractère non commercial du domaine, qui amène en la quasi-absence de sources de données existantes (outre les bases de connaissances), et l'absence d'une démonstration claire de la compréhension de ces particularités dans les soumissions reçues ont fait en sorte qu'aucun mandataire n'a été retenu.

« La statistique au
service de la société :
la référence au Québec »

statistique.quebec.ca