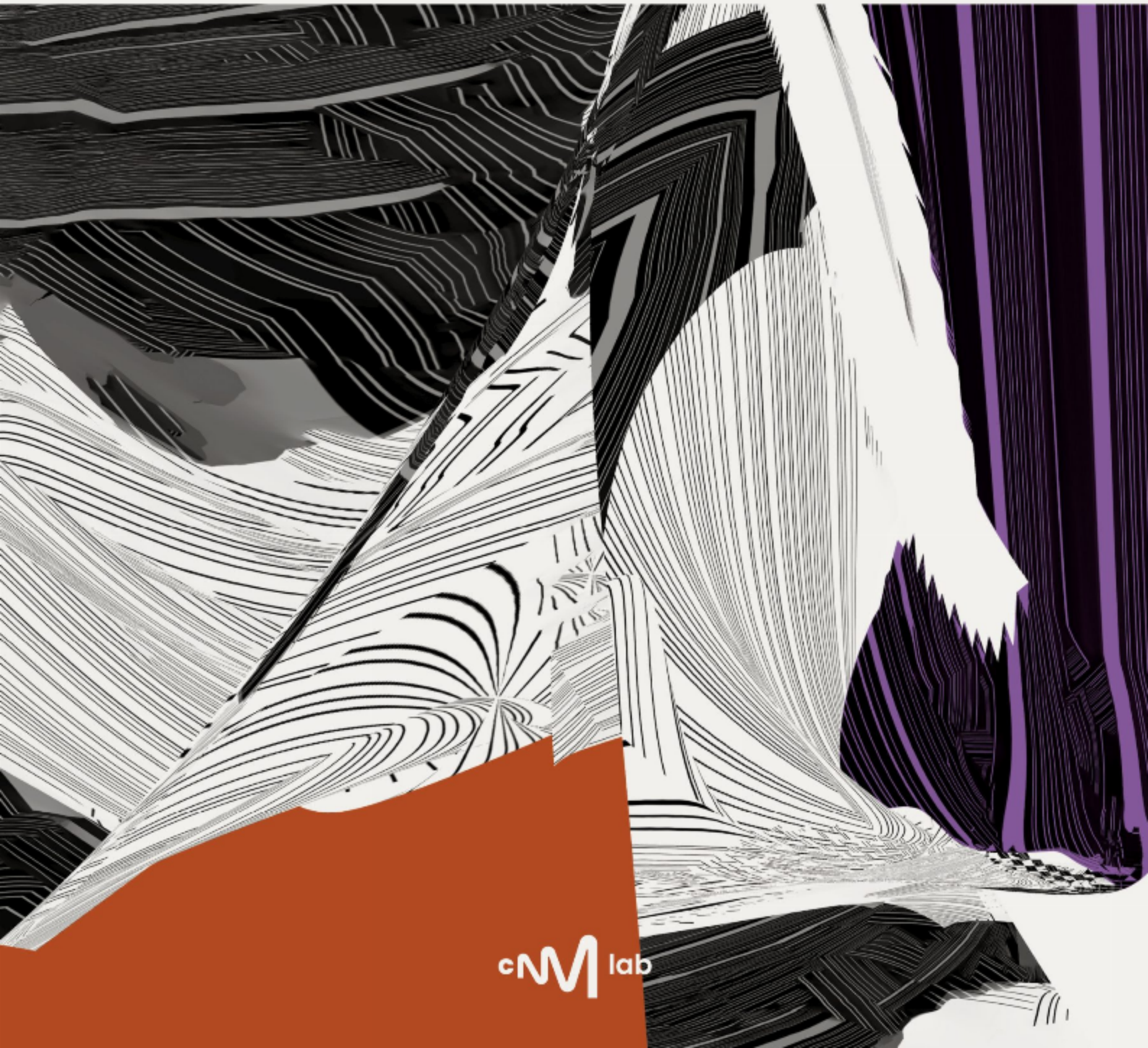
Indice de découvrabilité, streaming et diversité des contenus

Par Jean-Robert Bisaillon



Bisaillon Jean-Robert

Jean-Robert Bisaillon est codirecteur du Laboratoire de recherche sur la découvrabilité et les transformations des industries culturelles à l'ère du commerce électronique (LATICCE), à l'université du Québec à Montréal (UQAM). Musicien avec French B dans les années 1990 et pionnier du sampling au Québec, il fonde en 2006 MetaD, une entreprise de recherche et de formation sur les enjeux de la culture numérique.

Introduction

Dans une industrie phonographique marquée par une offre pléthorique, des inquiétudes naissent, non seulement autour de la difficulté pour le consommateur à faire des choix, mais aussi autour de la concentration de ces choix sur un nombre réduit de titres, deux phénomènes qui s'expliqueraient par les nouveaux modes d'éditorialisation employés par les plateformes d'écoute en ligne. Dès lors, pour maintenir la diversité culturelle de l'offre en streaming et contrebalancer certains effets d'une recommandation automatique consciemment ou inconsciemment biaisée, un levier consiste à mesurer et favoriser la propension des contenus à être découverts par les utilisateurs : ce que l'on appelle la « découvrabilité ». Néanmoins, cette mesure est conditionnée à un travail de documentation précis et d'indexation adéquate des contenus concernés [1] . Pour le mener à bien, l'adoption de bonnes pratiques industrielles, de normes et de règles communes favorisant le repérage et la traçabilité, une gouvernance des données permettant à la fois d'exploiter et de préserver les renseignements personnels des consommateurs et des artistes, selon des modalités transparentes et respectant la réglementation, sont des impératifs. Ce texte entend mettre en lumière les enjeux de la découvrabilité et le fonctionnement des algorithmes. Il aborde différents types d'approches institutionnelles et propose des pistes d'outils et de solutions : nous appelons à une normalisation des procédés industriels et invitons à une régulation du streaming, mode de diffusion encore en développement.

1. Les enjeux de la découvrabilité dans un contexte effervescent

La découvrabilité

La compréhension de la découvrabilité est en constante évolution. Une définition tirée du « Rapport de mission franco-québécoise sur la découvrabilité en ligne des contenus culturels francophones (RMFQD) » est désormais répandue :

La découvrabilité d'un contenu dans l'environnement numérique [2] se réfère à sa disponibilité en ligne et à sa capacité à être repéré parmi un vaste ensemble d'autres contenus, notamment par une personne qui n'en faisait pas précisément la recherche [3] .

Les énergies convergent désormais vers les meilleures pratiques à déployer pour activer la découvrabilité et la mesurer [4] .

Parmi les initiatives qui témoignent d'un intérêt accru pour la question, un colloque sur la diversité des contenus culturels d'expression autre qu'anglaise sur les plateformes numériques s'est tenu en France le 3 octobre 2022 au Sénat [5] . Cet événement a été l'occasion de lancer un appel à ce que les institutions responsables des politiques publiques s'impliquent dans la mise en œuvre de bases de données ouvertes recensant les artefacts culturels faisant l'objet d'une commercialisation en ligne.

Au moins deux initiatives canadiennes sont en phase avec cette proposition : le projet de loi C-11 sur la diffusion continue en ligne [6] et le projet de capture des métadonnées MétaMusique [7] . Précisons l'importance de mobiliser les métadonnées dans le travail de documentation des contenus transmis aux plateformes. Les métadonnées sont employées par les algorithmes des services en ligne pour permettre la mise en contexte et le croisement des œuvres avec les critères employés par les outils de recommandation.

Le LATICCE a quant à lui récemment démarré un chantier de recherche baptisé « Echo Chamber Research Project : métadonnées enrichies et chambres d'écho », avec le concours de la plateforme Mitacs et du label phonographique indépendant québécois InTempo Musique, qui s'emploie à définir avec plus de précision les contours de la documentation des contenus par le recours aux métadonnées. Le projet se fixe pour objectif de vérifier les gains de visibilité découlant de l'utilisation des métadonnées dites enrichies.

Toutes ces initiatives convergent et marquent l'actualité au moment où les spéculations vont bon train quant au rôle que joueront bientôt la recommandation, l'aide automatisée à la prise de décision, les algorithmes et les intelligences artificielles de toutes natures, dans les choix qui s'offrent à nous en matière de consommation, de fréquentation de biens culturels, ainsi que de création des œuvres du futur.

Pourtant, notre compréhension de la découvrabilité demeure approximative. Une fois que nous affirmons qu'il s'agit d'un défi de repérage dans une

« hyperoffre », l'enjeu soulève des questions techniques, des questions de secrets d'affaires, des questions de gouvernance des données… Une revue de littérature sur le sujet a été réalisée pour le LATICCE [8] afin d'en cerner les principaux paramètres, dont certains sont repris ici. L'objectif est de préciser notre compréhension de la découvrabilité pour le secteur de l'enregistrement sonore, de faire un bilan des connaissances, mais surtout d'identifier les axes de travail et les questions de recherche susceptibles d'orienter les efforts vers les pistes les plus porteuses.

Indice de découvrabilité

Le LATICCE a produit un prototype d'indice de mesure de la découvrabilité qui repose sur l'équation où P indique la présence, V indique la visibilité et R la recommandation. Cette dernière variable se décompose en (c) concordance, (p) pertinence et (n) nouveauté. La concordance correspond exactement au panier de référence de la personne étudiée, c'est-à-dire en correspondance exacte avec certaines variables de l'historique d'écoute, notamment la liste des artistes-interprètes écoutés. La pertinence indique une correspondance avec

divers artistes similaires déterminés par filtrage collaboratif. La nouveauté concerne la présence de propositions dont la parution est intervenue moins de 30 jours auparavant, elle est corroborée par les listes de nouveautés publiées par l'Association québécoise de l'industrie du disque, du spectacle et de la vidéo (ADISQ) [9].

Ces travaux tentent de mesurer, non pas la consommation effective des flux numériques comme nous le faisons pour établir un palmarès commercial, mais de mesurer les actions prises pour exposer certains contenus à certains publics. Les statistiques agrégées de consommation effective de type palmarès (*tops*) ou en lien avec les pratiques frauduleuses (*fake streams* [10]) nous en apprennent peu sur les approches algorithmiques de mise à disposition des contenus ou sur les barrières à la découvrabilité. L'indice de découvrabilité constitue un paradigme de mesure différent.

L'indice est une mesure fixée dans le temps qui n'indique pas le niveau de découvrabilité de façon absolue, mais plutôt l'évolution longitudinale (par la prise de mesures périodiques, régulières et échelonnées dans le temps) des variables observées qui permet d'estimer les progrès et les reculs des outils de recommandation proposés par les plateformes. C'est un outil technologiquement neutre qui jauge le succès de la recommandation sans suggérer de seuils à respecter ou indiquer de tendances de la consommation.

2. Approches institutionnelles et perspectives

Inventaire national et présence dans l'offre

La variable de la présence des œuvres et répertoires dans les catalogues en ligne est essentielle pour pouvoir, à terme, évaluer le respect d'éventuelles règles favorisant la diversité. Sans présence, la diversité et la découvrabilité sont évidemment compromises. Cette présence doit se mesurer à l'aune d'inventaires nationaux, pays par pays, qui, tel que nous l'entendons, s'incarnent dans des bases de données idéalement ouvertes et liées, contenant des métadonnées administratives et documentaires enrichies, recensant les artefacts culturels faisant l'objet d'exploitations en ligne. Ainsi, la constitution d'un inventaire national des œuvres et de leurs données (métadonnées) d'intérêt général (DIG) est un prérequis pour assurer la vérification de la présence et permettre d'y apporter des ajustements [11].

Devenant un enjeu de politique publique, ce type d'initiative relève de chantiers nationaux. Il faut songer au rôle que doit jouer le dépôt légal des œuvres dès lors qu'il est bonifié pour inclure des métadonnées névralgiques d'intérêt patrimonial dans un contexte d'économie numérique. Soulignons, à ce titre, l'initiative portée par la National Library of Finland qui s'est donné pour objectif de réunir les efforts de l'ensemble de gestionnaires collectifs de droits de son territoire autour de l'attribution de l'identifiant passerelle ISNI aux créateurs

et organisations artistiques [12] .

Le Canada et le Québec représentent l'une de trois parties (avec l'Allemagne et la Corée du Sud) qui se sont engagées à ce jour pour soutenir l'application de la Convention pour la promotion et la protection de la diversité des expressions culturelles en situation numérique [13] . Alors qu'il semblerait que la France doive bientôt faire de même, la question des mesures techniques et des bonnes pratiques à mettre en place pour activer et évaluer la découvrabilité des contenus en ligne sera un enjeu de politiques publiques des prochaines années.

Le Conseil de la radiodiffusion et des télécommunications canadiennes (CRTC), organisme public indépendant chargé de réglementer et de superviser la radiodiffusion et les télécommunications canadiennes (en France, c'est l'Arcom qui remplit ces missions), est cité dans le rapport « L'avenir des communications au Canada : le temps d'agir » connu sous le nom de rapport Yale [14] et est responsable de veiller sur l'offre de contenus en ligne au Canada.

Ainsi, en décembre 2022, le CRTC a complété un processus de révision de sa politique de radiodiffusion, et certains constats nous mènent à croire que cet organisme de réglementation a saisi l'importance de son rôle en matière de constitution d'un inventaire national. On y lit la volonté d'élaborer une base de données et de métadonnées ouverte permettant l'identification des œuvres.

Le Conseil élabore actuellement un système de surveillance numérique et une base de données ouverte pour simplifier et automatiser le processus d'identification des pièces musicales. Ce système repose sur des données probantes […] comme l'international standard recording code (ISRC), l'international standard name identifier (ISNI) et l'international standard musical work code (ISWC), des codes qui peuvent être utilisés pour confirmer l'exactitude des renseignements pour toute pièce musicale. […] Une fois rendue publique, cette base de données facilitera l'identification des pièces musicales canadiennes et atténuera les risques de non-conformité aux exigences réglementaires [15] .

Dès lors que l'on aborde la construction et le recours à une base de données d'œuvres, une question demeure toutefois entière : celle de sa qualité, de sa fiabilité, de son exhaustivité.

Communs, identifiants uniques et interopérabilité

La prescription officielle, l'exhaustivité et le niveau de fiabilité d'une base de données d'artefacts culturels et patrimoniaux du secteur musical se pose depuis de nombreuses années.

Plusieurs initiatives privées et quelques initiatives des communs numériques, produites collectivement (*crowdsourcing*), se sont intéressées à la question. Depuis 2010, je recense les bases de métadonnées de la musique qui se chiffrent à plus d'une centaine [16] . Aucune base de données n'est parvenue jusqu'ici à s'imposer comme source d'autorité, commune et fiable ; de surcroît, peu de ponts ont été créés entre les diverses initiatives.

Or, le commun numérique Wikidata donne la possibilité de saisir et d'interroger 7 709 types d'identifiants uniques normés [17] – des codes lisibles par les machines et permettant de dédupliquer, d'effectuer des croisements entre les œuvres et/ou les artistes. Ce travail de croisement aide à discerner les œuvres et à les attribuer aux bons artistes et autres ayants droit. Il existe de tels identifiants pour la vaste majorité des œuvres de propriété intellectuelle, tel l'ISBN pour le livre et l'ISRC pour la musique enregistrée [18].

L'identifiant Wikidata d'une œuvre [19], d'une personne physique ou morale, devient ainsi, par extension, un identifiant passerelle facilitant l'interopérabilité entre les répertoires et les contributeurs artistiques, l'appariement univoque des œuvres, des créateurs et des créatrices. Dans le cas de la Finlande, l'ISNI est un identifiant passerelle rendant possible de tels appariements. Un identifiant passerelle est un identifiant unique normé qui permet de croiser des références afin de lever l'ambiguïté qui pourrait exister sur elles, de faire un travail de discernement et possiblement de déduplication.

En vertu des mêmes logiques, la base de données MusicBrainz, produite de manière participative, est une source extrêmement riche de métadonnées musicales qu'il est nécessaire d'intégrer dans les dynamiques de travail, de quête d'autorité et d'exhaustivité. Parce qu'elle est créée et entretenue par des amateurs – comme c'est le cas pour Wikipédia et Wikidata –, l'industrie a eu tendance à en négliger la portée. Or, pour constituer des données de qualité et univoques, aucune source ne devrait être négligée. Il s'agit d'établir ultérieurement des protocoles appropriés de croisement, de vérification et de fusion des informations.

Ouverture des données

Outre les identifiants uniques, la base de données ouverte et liée Wikidata qualifie une multitude de propriétés associées à un élément qui y est déclaré. Par exemple, la langue d'expression ou l'origine géographique d'une œuvre, d'une personne. Une requête informatique en langage SPARQL donne la possibilité de recenser les déclarations répondant à ces critères. La propriété P407 indique, par exemple, la langue d'expression d'une œuvre.

Prenons une requête pour laquelle une composition musicale, une chanson ou un morceau de musique avec chant, est le fait d'une compositrice ou d'un compositeur ou interprète né au Québec ou en France et dont la langue est le français. Cette requête génère 3 373 résultats en 52 millisecondes [20]. Le nombre de titres des catalogues ouverts et publics de chansons adéquatement qualifiés d'origine québécoise/canadienne et française d'expression française est nettement en deçà du compte attendu [21].

Comme nous venons de le vérifier avec la requête SPARQL transmise à Wikidata, une base de données ouverte doit, par essence, être « moissonnable » automatiquement pour en permettre la réutilisation informatique par les acteurs industriels ou les institutions publiques chargées de vérifier le respect des engagements de contenus des diffuseurs. L'ouverture des données offre, à terme, l'occasion aux ayants droit de s'assurer de la qualité

de celles-ci et d'en assurer la validation, la correction et de disposer de recours à cet effet [22].

Indexation à la source

À l'instar du projet de la National Library of Finland, le projet MétaMusique repose sur une initiative collective de tous les syndicats et organisations de gestion de droits du secteur musical québécois, sous forme de consortium sans but lucratif [23].

La pratique impose désormais à la théorie des règles de préconditionnement afin de minimiser les barrières à la découvrabilité. C'est précisément ce que pourrait permettre MétaMusique : encourager de nouvelles bonnes pratiques de préconditionnement pour s'assurer que les artefacts transmis dans la chaîne de valeur et aux plateformes soient documentés adéquatement. Sans cela, la recommandation algorithmique pourrait négliger ces contenus.

Un certain nombre de choses ont été écrites concernant les algorithmes de recommandation des plateformes numériques d'écoute en ligne (PNEL), notamment par Brian Whitman, cofondateur de The Echo Nest, projet acquis par Spotify en mars 2014 et portant sur les leviers de la recommandation [24]. Whitman aborde notamment les limites du filtrage collaboratif qui a tendance à générer des bulles de filtrage, des recommandations peu stimulantes.

L'entreprise de conseil Music Tomorrow décrit les pratiques industrielles visant une meilleure compréhension des dynamiques de recommandation par l'expression *recommender system optimization* (RSO), en référence à la SEO (*search engine optimization*), qui a pour but de favoriser la remontée de résultats organiques de recherche de pages web. La RSO, notamment la *technical* RSO, s'intéresse aux métadonnées administratives et documentaires, elle doit « optimiser la chaîne de distribution technique afin de s'assurer que les catalogues et leurs métadonnées soient complets et conformes [25] ».

Or, nous ne savons pas encore avec précision quelles sont les données essentielles favorisant une découvrabilité optimale et, à terme, une reddition de comptes adéquate. Même si elles ont l'apparence de données d'intérêt général (DIG), les métadonnées descriptives semblent représenter un terrain où continue de se jouer la concurrence.

3. Les outils et solutions développés par l'équipe du LATICCE

Les bulles de filtrage et les métadonnées enrichies

C'est ici qu'entrent en jeu les travaux actuels du LATICCE sur les métadonnées enrichies. Il existe des normes d'échange informatisées de données commerciales, ou *electronic data interchange* (EDI), pour alimenter la chaîne de valeur de la musique. Le consortium Digital Data Exchange (DDEX) a été créé pour développer les protocoles informatiques nécessaires à la transmission des informations entre chacun des segments de la chaîne de valeur pour les enregistrements sonores et les podcasts. Les pratiques professionnelles des milieux culturels d'expression autre qu'anglaise doivent impérativement reposer sur ce type de normes internationales.

La chorégraphie DDEX-MEAD autorise désormais la transmission de données de type *media enrichment and description* (MEAD) entre les producteurs, les distributeurs numériques et les plateformes. Le format MEAD statue sur 30 classes d'informations enrichies et un jeu de valeurs autorisées qu'il est désormais possible d'échanger pour chacune d'entre elles (*allowed value set*, AVS).

Notre projet est développé avec le concours du label québécois InTempo Musique et divers partenaires de recherche français dont le LabEx-ICCA, la fédération FÉLIN (labels indépendants) et Musicovery. Il doit rendre possible l'analyse de la relation entre l'existence de métadonnées d'enrichissement, la qualité des recommandations et la possibilité de percer les chambres d'écho ou bulles de filtrage.

Au moment de son lancement, le projet établissait les métadonnées enrichies de la façon suivante :

- Les paroles de chansons et les mots-clés qui permettent d'activer la recherche de contenus par analyse du langage naturel par concepts, lieux, ambiances, sentiments, etc.
- Les expressions employées pour décrire un style, une période, un courant, un succès ponctuel et qui servent à traiter les requêtes vocales.
- Les filigranes numériques et les empreintes audio spectrales qui contribuent à la recherche de titres par l'écoute d'un signal via un téléphone [26].
- La liste exhaustive des contributeurs et des studios qui facilite les correspondances entre divers enregistrements et projets.
- L'inscription dans les bases de données ouvertes et liées qui permet la réutilisation de données biographiques ou phonographiques, époques et courants musicaux. Cette variable rend aussi possible l'alimentation de certains services d'appariement entre artistes similaires et genres.

- Les métadonnées techniques, juridiques et administratives utiles pour les plateformes qui soumettent les nouvelles propositions à des normes de qualité des catalogues.
- Les photographies des interprètes principaux appelées à être changées au cours de l'évolution de la carrière de l'artiste [27].

Quatre axes de travail sont retenus :

1. La typologie et la caractérisation des services, des acteurs, des marchés, des flux économiques ;
2. La mesure de la découvrabilité des contenus avant et après le conditionnement documentaire des produits à l'aide de métadonnées enrichies ;
3. La définition des meilleures pratiques d'affaires sectorielles en réponse à la nouvelle réglementation canadienne des plateformes (documentation, recommandation, découvrabilité) [28] ;
4. L'étude des impacts économiques sur le secteur, en tenant compte des différents types d'acteurs et des divers marchés étrangers.

Au nombre des informations désormais transmissibles par le protocole DDEX-MEAD, citons les suivantes : tonalité, signature rythmique, tempo, thématique, présence d'échantillons, ambiance (*mood*), genre, utilisations publicitaires ou synchronisations sur des séries audiovisuelles ou des films, prix remportés. Il est aussi à noter qu'il est possible de déclarer ces métadonnées dans l'environnement ouvert et lié Wikidata pour en donner l'accès automatique. La concurrence pourrait dès lors se jouer entre les producteurs qui parviennent à colliger, transmettre et partager des métadonnées enrichies et ceux qui négligeraient de le faire. La circulation d'objets culturels devient conditionnée par ces déterminants techniques.

L'introduction de ce nouveau standard MEAD vient confirmer les intuitions et les bonnes pratiques promues par une firme comme Music Tomorrow, le projet MétaMusique et les objectifs du projet de recherche du LATICCE, Echo Chamber Research Project, qui a pour but de percer la chambre d'écho de la recommandation pour stimuler la découvrabilité et, à terme, l'exportation de la musique.

Le phénomène de chambre d'écho, de bulle de filtres ou de bulle de filtrage est défini de la façon suivante : les systèmes décisionnels automatisés [29] génèrent, selon leur niveau de maturité, des effets de chambre d'écho ou de bulles de filtres qui enferment les auditeurs dans des univers sonores souvent très peu diversifiés, parce que conditionnés par des biais algorithmiques, volontaires ou non. Les chambres d'écho, en fonction de leur taille ou de leur hermétisme, auront tendance à freiner l'accès à des marchés extérieurs et à des publics polymorphes, aux nouveaux publics domestiques, en se cantonnant à des profils sonores conformes aux moyennes statistiques. C'est ce que l'on nomme le phénomène du *more of the same* ; autrement dit, davantage de titres, mais aux caractéristiques toujours semblables.

Ce phénomène de chambres d'écho est largement attribuable au filtrage collaboratif, technologie fondatrice des premiers algorithmes de recommandation et encore largement employée aujourd'hui. Or, si cette technologie pouvait s'appuyer sur davantage de don-

nées, les résultats qu'elle produirait seraient probablement plus raffinés.

Tests en cours

À cet effet, notre projet nous a amenés à conduire des prétests, à l'été 2022, en utilisant l'API *Get Recommendations* de Spotify [30], et a montré que les recommandations faites pour un artiste et un titre de départ (*seed artist* et *seed track*) étaient analogues, qu'importe l'historique d'écoute ou la localisation de l'abonné (et du jeton d'identification, ou *token*, de celui-ci). Nous en concluons sommairement que c'est toujours le filtrage collaboratif qui définit essentielle- ment la distance entre deux titres. C'est-à-dire, combien de fois le morceau Y a été écouté par les abonnés ayant écouté le morceau X, et ce, sans véritable considération pour les profils de goût des abonnés.

Nous en déduisons aussi qu'il est impératif de prendre en considération les modes d'accès aux flux de plateformes pour pouvoir, à terme, produire des analyses signifiantes de la recommandation. Les plateformes emploient des algorithmes différents, plus ou moins sophistiqués, selon ces modes d'accès.

Typologie des modes d'accès

Nos travaux préalables de définition de la typologie des modes d'accès permettent par exemple d'établir plus de 30 routines d'accès disponibles dans Spotify Mobile, qui sont :

- Spotify Mobile
- Playlist (Widget vertical)
- Vos mixs préférés
- Dernière sortie de [...]
- Album tendance pour vous
- Vos amis écoutent
- Pour les fans de [...] (lien vers Artist Playlist programmed)
- Dernière sortie populaire
- Album populaire
- Plus du genre de [...] (lien vers Artiste Playlist programmed)
- Plus du genre de [...] (lien vers Genre Playlist programmed)
- Spotify Wrapped (Le bilan de votre année)
- (Genre) Music picked just for you
- Écoutés récemment
- Réécoutez vos anciens favoris
- Recommandation du jour
- Daily Mix 1 (Conçu pour [vous])
- Daily Mix 2 (Conçu pour [vous])
- Daily Mix 3 (Conçu pour [vous])
- Daily Mix 4 (Conçu pour [vous])
- Daily Mix 5 (Conçu pour [vous])
- Daily Mix 6 (Conçu pour [vous])

- Bien être au quotidien
- Découvertes de la semaine
- Radar des sorties
- #SpotifyWrapped
- Radios recommandées
- Artistes recommandés
- Vos playlists
- Albums sélectionnés
- Dernières sorties rien que pour vous
- Radios populaires
- Rechercher artiste
- Rechercher titre

Typologie des flux et des requêtes

Notre équipe travaille à la définition d'une typologie des flux et des requêtes des abonnés. Cette nomenclature est devenue nécessaire pour qualifier les modes d'écoute pour lesquels d'éventuelles mesures de découvrabilité pourraient être effectuées. Elle reprend le concept de l'écoute *on demand*, employé par l'entreprise états-unienne de données LUMINATE, qui produit des statistiques pour l'audio et la vidéo à la demande [31]. Le lexique du LATICCE se décline actuellement ainsi :

1. Audio à la demande : l'abonné effectue une requête par le nom de l'interprète ou le titre d'une piste ou d'un album
    1. Vidéo à la demande : voir audio
    1. Programmation audio éditoriale : la liste d'écoute est produite par des professionnels, par exemple les éditorialistes des plateformes
    1. Programmation audio par les pairs : la liste d'écoute est produite par d'autres internautes
    1. Programmation audio algorithmique : la liste d'écoute est produite automatiquement
    1. Programmation vidéo éditoriale : voir audio
    1. Programmation vidéo par les pairs : voir audio
    1. Programmation vidéo algorithmique : voir audio
    1. Flux extrapolés : une liste ou une requête est suivie par un flux automatique
    1. Programmation hybride : deux méthodes de requête ou de programmation, ou plus, sont jumelées.

Typologie des métadonnées enrichies d'intérêt général

Enfin, nous nous sommes penchés sur une typologie des métadonnées enrichies d'intérêt général qui pourraient, à terme, être retenues pour améliorer la documentation d'artistes ou de titres. Nous avons pris en compte le jeu de données prescrit par le format DDEX-MEAD. À titre d'exemple, nous avons identifié plus de 70 actions d'enrichissement à entreprendre et champs à renseigner pour la seule base de données MusicBrainz et une vingtaine sont à

considérer comme prioritaires [32]. En voici la liste :

- ISNI à ajouter
  - Ajouter les alias
  - Lier Wikidata
  - Lier Apple ID
  - Lier Spotify ID
  - Sortie d'album
  - Lier BandCamp
  - Lier BandsIntown
  - Lier SoundCloud
  - Lier Songkick
  - Lier Last FM
  - Lier YouTube
  - Lier VIAF
  - Lier Amazon Music
  - Release Group – Entrer Titre
  - Album (Release) – Entrer Titre
  - Album (Release) – Entrer Works
  - Album (Release) – Entrer Crédits musiciens
  - Album (Release) – Entrer Tracks

Un outil d'exploration des artistes liés (*related artists search*) du label indépendant InTempo permet de suivre l'évolution des bulles de filtrage jusqu'au troisième degré de séparation, de façon longitudinale.

À titre d'exemple, nos travaux montrent que la seule voie de sortie vers l'Amérique du Nord, hors des bulles de filtrage de la chanson, de la pop et des variétés françaises pour l'artiste Clara Luciani, est l'artiste québécoise Safia Nolin qui se situe au centre d'une bulle à l'extrême gauche d'un graphe de relations produit avec le soutien de nos outils de mesure reposant sur les API de récupération des artistes similaires de la plateforme Spotify.

Le LATICCE a inauguré en avril 2023 un cycle de recherche qui a pour objectif de vérifier l'incidence d'actions d'enrichissement sur les bulles, et ce pour un vaste recensement d'artistes indépendants français et québécois.

Small data, *big data*, profil sonore et intelligence artificielle

Chaque jour, plus de 100 000 morceaux sont ajoutés sur les serveurs de Spotify [33] et 8,5 milliards de recherches Google sont effectuées [34]. La taille du web oscille aujourd'hui autour de 35 milliards de pages [35]. Les termes hyperoffre et infobésité sont depuis longtemps entrés dans le vocabulaire.

Dans son essai sur les données de Spotify, le journaliste Philippe Astor souligne le caractère massif des données soutenant un tel service :

Lors de la migration de ses infrastructures de données vers les services de cloud de Google, amorcée en 2016, c'est plus de 100 pétaoctets de données que Spotify a dû transférer depuis ses data centers, a indiqué Ramon Van Alteren, le directeur de programme qui a supervisé l'opération en interne. À l'époque, le système dorsal de Spotify était capable de transmettre à lui seul « plus de 700 000 événements par seconde à travers le monde », un « événement » désignant toute action effectuée par un utilisateur dans l'interface de Spotify, comme l'ajout d'une chanson à une playlist [36].

Les internautes parcourent ces immenses « catalogues » de contenus en faisant appel au langage naturel de tous les jours, mais ils sont accompagnés pour le faire, à leur insu ou non, par des robots, des assistants à la recherche et à la prise de décision. Pour cela, il faut organiser des masses de renseignements et de contenus, effectuer un travail éditorial qui, à terme, sera plus ou moins assisté par les algorithmes et leur offrira assistance en retour (ces informations alimentent l'apprentissage machine, *machine learning* ou *deep learning* en anglais). L'offre commerciale en ligne est aujourd'hui construite sur ces immenses gisements numériques plus ou moins structurés. On parlera de lacs de données (*data lake*) et de marécages de données (*data swamp*) pour caractériser l'état d'organisation de ces gisements [37].

Si l'on en croit une devise communément attribuée au médecin Claude Bernard, « l'expérimentateur qui ne sait pas ce qu'il cherche ne comprend pas ce qu'il trouve [38] ». Cela résume bien les dilemmes associés à l'exploitation des données massives, et l'un des défis du recours à l'intelligence artificielle : la rédaction des *prompts* ou questions en langage naturel que nous posons aux espaces de clavardages assistés par IA (*chatbot*) et assistants vocaux.

Selon une logique similaire, qu'importe la recommandation de contenus musicaux si celle-ci est inadéquate et entraîne le décrochage ou la lassitude de l'auditeur ?

De façon sommaire, la recommandation se fonde sur la playlist de l'usager pour proposer un flux radio dit personnalisé : un auditeur défini par ses traces d'activité et pris en charge par des machines. Les chercheurs Jean-Samuel Beuscart, Samuel Coavoux et Sisley Maillard ont tenté de circonscrire la place qu'occupe la recommandation dans les nouvelles habitudes d'écoute, entre autonomie et choix imposés (hétéronomie), et ont découvert que les propositions algorithmiques demeurent marginales et essentiellement fondées sur les bibliothèques des abonnés [39].

Quant au LATICCE, le bilan de nos travaux avait plutôt souligné la faiblesse actuelle de la recommandation pour un personnage aux contours définis avec précision :

Malgré l'accent mis par les services de musique sur la qualité de l'expérience musicale personnalisée offerte à leurs abonnés, aucune des 21 semaines d'écoute n'a offert ce que notre personnage attendait. Les services proposés ont réagi de manière très variée aux problèmes rencontrés dans une situation de cold start (possibilité de recommander du contenu pertinent avec très peu d'historique des données) [40].

Pour tirer de véritables bénéfices des *big data* (données massives), il faut des *small data*, à savoir des données claires, granulaires et enrichies, dont la qualité est vérifiée, afin d'obtenir un usage optimal des procédés de recommandation. Ces données sont ouvertes et partagées afin de permettre leur utilisation, leur réutilisation, leur mise à jour et leur validation en continu. L'expression : *garbage in = garbage out* (déchets à l'entrée = déchets à la sortie) est souvent utilisée pour qualifier les données de qualité impropre à faire reposer une juste reddition de comptes ou des processus de prise de décision automatisés. Ainsi, nous croyons qu'il ne faut pas négliger les *small data* ni les enfermer dans une boîte noire numérique protégée par le secret d'affaires.

Étude des impacts socioéconomiques

Le chercheur doctorant Guy-Philippe Wells, du LATICCE, poursuit en parallèle un projet de recherche qui vise à mesurer l'impact économique des plateformes d'écoute en ligne sur les revenus des auteurs et des compositeurs québécois. Il est nécessaire d'étudier les impacts de la transformation numérique d'un point de vue local pour vérifier si les dynamiques mondiales s'y reproduisent, ou si l'on y observe plutôt des dynamiques contradictoires ou divergentes. L'impact des transformations numériques sur l'industrie de la musique ne saurait être mesuré par une simple agrégation mondiale. Il doit également se mesurer sur les réseaux industriels locaux qui favorisent la création d'une musique originale qui se distingue de celle proposée par les trois grands conglomérats mondiaux (le « Big Three » : Universal, Sony, Warner) assurant une représentation artistique de la diversité des cultures mondiales. Nous avons terminé au cours des derniers mois une première étape de la recherche qui consistait en une enquête en ligne auprès des artistes musicaux québécois autoproduits ou essentiellement représentés par des labels indépendants. Cette enquête présente dix-huit questions qui visent à décrire l'impact des transformations numériques sur les revenus de ces artistes. Plus de 150 artistes ont jusqu'à présent participé à cette enquête toujours en cours, mais les résultats préliminaires nous orientent vers la nécessité d'entrevues en profondeur pour mieux comprendre l'état de la situation.

4. Gouvernance des données et politiques publiques

La qualité de la documentation des pistes musicales offertes en streaming, la gouvernance des métadonnées et des données qui les décrivent ou sont produites par cette activité, doivent gagner en maturité, afin d'établir la confiance nécessaire permettant de pérenniser durablement ce mode d'écoute. À terme, il s'agira aussi d'assurer la rémunération des artistes et par conséquent le renouvellement des contenus et de la diversité culturelle.

Nous souhaitons aborder les enjeux d'ouverture des données d'intérêt général du secteur, de leur gouvernance, de la surveillance des flux et du respect des données des abonnés. Ces paramètres font partie de la nouvelle équation régissant l'écoute de la musique et

questionnent la nécessité de régulations faisant appel aux efforts de l'ensemble du secteur.

Naturellement, ouvrir des données, les réutiliser, offrir des outils pour leur actualisation et leur validation, soulève des enjeux de gouvernance. La meilleure gouvernance des données est certainement celle qui assure au sujet auquel les données correspondent d'en conserver le contrôle et d'en autoriser l'usage selon ses besoins [41].

Nathalie Casemajor et Guillaume Sirois, chercheurs à l'INRS à Montréal, proposent une définition claire de ce que sont les données personnelles sensibles, et nous intiment à la prudence à leur égard :

L'enjeu de la confidentialité des données personnelles est particulièrement grand en ce qui a trait aux données sensibles, un type de données personnelles qui, lorsque révélées, peuvent contraindre l'exercice des libertés fondamentales de l'individu identifié ou le placer dans une situation indésirable. La CNIL explique que les données sensibles peuvent avoir trait, entre autres, à la santé, à l'origine ethnique, aux convictions religieuses ou philosophiques ou à l'orientation sexuelle d'un individu [42].

Si les données personnelles sensibles incluent l'origine et les convictions, les données concernant les goûts et les habitudes sont-elles des données sensibles ? De fait, toutes ces données contribuent à alimenter les « modèles discriminatoires » et la « reconnaissance de modèles », abordés notamment par Clemens Apprich dans l'ouvrage collectif *Pattern Discrimination* [43].

Droit à l'autodétermination informationnelle

La capacité pour un individu de décider pour lui-même, selon les principes de l'autodétermination informationnelle, quand et dans quelles limites les informations relatives à sa vie privée peuvent être communiquées à autrui, a été pour la première fois évoquée dans un arrêté constitutionnel allemand adopté pour le recensement de 1983. Cette capacité repose invariablement sur une connaissance que le sujet a de la transmission d'informations relatives à sa vie privée, et sur sa capacité à exercer son autonomie et son pouvoir. Cela fait écho à la théorie du panoptique de Bentham (1748-1832), reprise par Michel Foucault, selon laquelle une surveillance efficiente d'un sujet peut passer par son incapacité à savoir s'il est ou non surveillé [44].

Pour exercer son droit à l'autodétermination informationnelle, un principe de transparence des bases de données et d'un accès personnel protégé aux informations relatives à sa vie privée doit préexister. L'exercice d'un tel droit repose très certainement sur la gouvernementalité des données, soit un rôle assumé de gouvernance de l'État sur les enjeux de respect et protection des données, ainsi que sur leurs modes de gouvernance effective, à savoir qui détient ces informations et comment en définissent-il l'usage et le traitement.

Sur les enjeux éthiques en matière de gouvernance des données, Stefaan G. Verhulst, cofondateur et officier en chef de la recherche et du développement du GovLab de l'université de New York, établit les balises suivantes :

Face à la datafication des sociétés, il importe de réfléchir au design de la réutilisation et de penser une nouvelle licence ou contrat social encadrant la donnée. Notamment l'utilisation des données non prévue au moment de la collecte. Les éléments à prendre en compte sont :

- Le motif : pourquoi avons-nous besoin d'une donnée au départ ?
- La nature : de quelles données parlons-nous ?
- L'imputabilité : qui a accès aux données ?
- Le mode d'accès : les données sont-elles ou non anonymisées ?
- La durée : à quel moment les données sont-elles utilisées ou détruites ?
- Le lieu : à quel endroit et sous quelle juridiction les don- nées sont-elles conservées et mises à disposition ?
- Le contrat social : existe-t-il un consensus sur les usages prévus et un commissaire responsable a-t-il été désigné [45] ?

Ce dernier élément souligné par Verhulst est particulièrement névralgique. Pour réaliser certaines opérations, l'usager en ligne est tributaire de données qui sont parfois utilisées sans qu'il en soit avisé : la question de la gouvernance devient alors cruciale.

Le modèle Solid est actuellement promu par Tim Berners Lee, principal inventeur du « *world wide web* » au tournant des années 1990, dans le cadre de travaux au CERN. Avec ses PODS (*personal online data stores*), il propose de raffiner le contrôle sur les données personnelles, l'instauration de logiques de portabilité des données, d'un droit à l'autodétermination informationnelle de l'internaute (dès lors que l'on crée son profil de contributeur, d'abonné). Cette approche pourrait être explorée pour les ayants droit et leurs artefacts.

Le capitalisme de surveillance repose sur le fait que vos données sont par défaut collectées par autrui. Au contraire, avec Solid, leur contrôle ne vous est pas retiré [46].

Au Royaume-Uni, la BBC explore la technologie Solid dans le but de permettre à la fois la recommandation de contenus fondée sur le profilage et le respect éthique des données personnelles : « La technologie de gestion d'entrepôts de données personnelles pourrait transformer nos façons de faire et soutenir notre désir d'offrir des contenus personnalisés [47]. »

Surveillance

Au moment où nous sommes alertés sur les dérives potentielles et déjà en cours du capitalisme de surveillance [48] se pose la question de la régulation des entreprises multinationales, des plateformes en ligne et de l'encadrement de leur recours aux algorithmes.

Shoshana Zuboff, professeure émérite à la Harvard Business School, cite Thomas Paine sur les pouvoirs de la monarchie et, par analogie, les transpose aux grandes entreprises du numérique : « un groupe d'hommes ne s'estimant pas imputables à quiconque ne mérite la confiance de personne [49] ».

La mathématicienne Cathie O'Neil mise sur le fait que les internautes sont de plus en plus conscients des biais algorithmiques et qu'ils exigeront à terme une transparence que les plateformes ne pourront pas entièrement leur refuser [50].

Dans le même esprit, le rapport Yale, commandé en amont du processus de révision de la loi sur la radiodiffusion canadienne en cours, avançait l'idée de soumettre les acteurs du numérique à des règles de soutien à la découvrabilité (recommandations 59, 61, 63, 65, 73) et à des mesures d'audit des algorithmes (recommandation 63).

Afin que les Canadiens et les Canadiennes puissent faire des choix éclairés, que le contenu canadien ait une visibilité adéquate et que celui-ci soit facilement repérable dans les services utilisés au pays, nous recommandons que le CRTC impose des obligations de découvrabilité à toutes les entreprises de contenu audio ou audiovisuel de divertissement, comme il l'estime approprié, y compris :

- des obligations de catalogue ou de présentation ;
- des obligations de mise en valeur ;
- l'obligation d'offrir des choix de contenu médiatique canadien ;
- des obligations de transparence, notamment envers le CRTC quant au fonctionnement des algorithmes, y compris des exigences d'audit [51].

Peut-on évoquer un équilibre dans la surveillance ? Le CRTC travaille sur une base de données ouverte des enregistrements sonores canadiens, les projets MétaMusique et Echo Chamber Research Project font de même à leur échelle. Ces projets explorent ce que l'économiste Joëlle Toledano, dans son ouvrage *GAFA. Reprenons le pouvoir*, affirme être une voie à privilégier :

Il faut que des équipes pluridisciplinaires (informaticiens et spécialistes du traitement des données, mais aussi économistes, juristes…) soient constituées pour développer des méthodes et des outils d'analyse. […] Les pouvoirs publics doivent pouvoir en disposer, mais des équipes académiques ou des ONG devraient aussi avoir les moyens de mener à bien des recherches. Apprendre à tester la transparence et la loyauté des algorithmes [52].

L'ingénieur en réseaux informatiques Stéphane Bortzmeyer, dans son ouvrage *Cyberstructure*, offre quant à lui des pistes pour comprendre certains déterminismes techniques introduits par Internet et qui éclairent notre problématique :

La particularité de la « gouvernance de l'Internet » est que l'Internet n'a pas la structure bien définie d'un État-nation, d'une entreprise, ou d'une association. En fait, il n'a même pas de structure claire. […] Mais cette absence de centre a aussi des avantages : elle évite l'abus de pouvoir par une autorité. Internet est donc, à bien des égards, un cas particulier dans le monde des sciences politiques. C'est un bien commun, au sens où il s'agit d'une infrastructure partagée, et qui ne fonctionne pas toute seule. Comment se fait-il que, bien que les acteurs soient concurrents, voire franchement ennemis, il puisse fonctionner ? C'est probablement que tout le monde a intérêt à ce que l'Internet marche [53].

Conclusion

L'idée de bien commun utile à tous et évoquée plus haut par Bortzmeyer doit impérativement être validée. La mise à disposition de données culturelles d'intérêt général en mode ouvert et lié constituant, à ce titre, un terrain d'exploration tout particulièrement désigné. Ce terrain mobilise des enjeux de respect des données personnelles des internautes, de droit des créateurs et créatrices à l'autodétermination informationnelle, d'accès aux œuvres culturelles et de cadrage de l'innovation, nommément de l'intelligence artificielle et des algorithmes, de développer et d'assurer la confiance populaire en celle-ci. À défaut de fonder cette confiance, les promesses du numérique pourraient se lézarder irrémédiablement.

Pour éviter que la confiance ne repose que sur des critères subjectifs, il sera prudent de la faire reposer sur un socle normatif mobilisé à son tour dans les politiques publiques. Pour les secteurs industriels de la culture, il faut forcer l'adoption de bonnes pratiques de préconditionnement des contenus par les producteurs et les artistes et y conditionner l'accès aux subventions à la production. Nous employons le terme « forcer » pour désigner la nécessité de réglementer l'usage des métadonnées et des métadonnées enrichies, afin de nous donner davantage d'assurance que les publics soient à même de découvrir ce qui les intéresse. C'est le terrain que nous explorons actuellement.

Si, comme c'est le cas avec le projet de loi C-11 au Canada, nous imposons des barèmes en matière de découvrabilité aux plateformes et exigeons que celles-ci participent au financement des nouvelles productions, il est normal que l'industrie assume sa part de responsabilité et livre des fichiers musicaux respectant des normes de bonnes pratiques conséquentes afin de toucher ces aides. C'est aussi l'esprit des directives opérationnelles et de la feuille de route du numérique Canada-Québec de l'Unesco.

5. Remerciements

Antoine Beaubien, Denis Bouchard, Philippe Bouquillion, Vincent Castaignet, Michelle Chanonat, Véronique Desjardins, Pierre B. Gourde, Jean-Baptiste Le Friant, Céline Lepage, Mirjana Milovanovic, Jacinthe Plamondon, Michèle Rioux, Alain Saulnier, Keani Schuller, Brice-Armel Simeu-Tagno et Guy-Philippe Wells.

1. Travail également appelé « préconditionnement », qui est au fondement des procédés décisionnels de la recommandation algorithmique.
2. « L'environnement numérique », formule employée dans la définition donnée par la mission franco-québécoise, inclut, selon nous, toutes les formes de mise à disposition publique en ligne de contenus numérisés.
3. « Rapport de la mission franco-québécoise sur la découvrabilité en ligne des contenus culturels francophones », Québec, Ministère de la Culture et des Communications du Québec/Ministère de la Culture de France, 2020, p. 8, en ligne : cdn-contenu.quebec.ca/cdn-contenu/adm/min/culture-communications/publications-adm/rapport/Decouvrabilite-Rapport.pdf.
4. L'année 2023 est celle de l'innovation franco-québécoise, et elle donne lieu à une série d'initiatives de recherches bilatérales sur la découvrabilité, faisant suite à la publication du RMFQD. Un appel à projets est en cours dans le but de soutenir des activités de recherche et d'amélioration des procédés industriels : www.culture.gouv.fr/Demarches-en-ligne/Par-type-de-demarche/Appels-a-projets-candidatures/Soutien-a-la-decouvrabilite-en-ligne-des-contenus-culturels-francophones. De

surcroît, un MOOC franco-québécois sur la découvrabilité, préparé à l'initiative des ministères de la Culture, sera lancé au printemps 2023.

5. Pour revoir la communication au Sénat : videos.senat.fr/video.3007738_633abd7fd0f89. decouvrabilite-des-uvres-dexpression-francophone-sur-plateformes-numeriques-apres-midi.
6. Voir le projet de loi C-11 441 sur la diffusion continue en ligne : www.parl.ca/DocumentViewer/fr/44-1/projet-loi/C-11/premiere-lecture.
7. Voir le site officiel du projet MétaMusique : metamusique.ca/a-propos.
8. Bisaillon J.-R., « Être ou ne pas être découvrable ? Une revue de littérature », Montréal, LATICCE, 2022, en ligne : https://doi.org/10.5281/zenodo.6973451.
9. Rioux M. (dir.), « Être ou ne pas être découvrable ? Présence, visibilité et recommandation des propositions culturelles en ligne. La musique et l'audiovisuel », rapport scientifique public, Montréal, UQAM-CEIM-LATICCE, 2021, en ligne : ieim.uqam.ca/spip.php?page=article-ceim&id_article=13145.
10. « Manipulation des écoutes en ligne », Étude du Centre national de la musique, Paris, CNM, 2023, en ligne : cnm.fr/faux-streams-vrai-phenomene-le-cnm-avec-les-professionnels-pour-lutter-contre-la-fraude.
11. Voir Farchy J. et Denis J., La culture des données. Intelligence artificielle et algorithmes dans les industries culturelles, Paris, Presses des Mines, 2020 ; ainsi que Cytermann L., Morel M., Duchesne C., Aureau T. et Vachey L., « Rapport relatif aux données d'intérêt général », Paris, CGE/IGF, 2015, en ligne : www.economie.gouv.fr/files/files/PDF/DIG-Rapport-final2015-09.pdf.
12. « ISNI project launched in Copyright Management Organizations », National Library of Finland, Helsinki, 2022, en ligne : www.kansalliskirjasto.fi/en/news/isni-project-launched-copyright-management-organizations.
13. Voir les feuilles de route ou l'ensemble des communications adressées au Sénat le 3 octobre 2022, ainsi que la communication de Véronique Guèvremont prononcée dans le cadre de la Conférence N.4 : « Vers un nouvel instrument juridique de l'Unesco sur la diversité linguistique des contenus culturels en ligne. La formation d'une coalition internationale », en ligne : https://praxis.encommun.io/n/Wb5v6Zf1xBuniCjMGY27oINmrhQ/ ; et Guèvremont V. et al., « Les mesures de découvrabilité des contenus culturels francophones dans l'environnement numérique : compte rendu des tendances et recommandations », Chaire UNESCO sur la diversité des expressions culturelles, Québec, 2019, en ligne : www.unescodec.chaire.ulaval.ca/sites/unescodec.chaire.ulaval.ca/files/rapport-decouvrabilite-10_decembre_2019_-_final.pdf.
14. « L'avenir des communications au Canada : le temps d'agir [rapport Yale] », Ottawa, Innovation Canada, 2020, en ligne : www.ic.gc.ca/eic/site/110.nsf/fra/00012.html.
15. « Politique révisée sur la radio commerciale », CRTC, Gouvernement du Canada, 2022, en ligne : crtc.gc.ca/fra/archive/2022/2022-332.htm.
16. Liste des 100 bases de données de la musique et des identifiants uniques de ce secteur : bit.ly/musicalmetadata.
17. Voir la liste des identifiants qui alimentent les propriétés « Identifiants » d'une déclaration Wikidata : wikidata.org/w/index.php?title=Special:ListProperties/external-id&limit=7000 (consulté le 2 janvier 2023).
18. Les identifiants uniques sont des codes lisibles par les machines et qui répondent à des normes établies par des organisations dédiées telles l'Organisation internationale de normalisation (ISO), voir l'exemple de l'ISRC pour les enregistrements sonores : www.iso.org/standard/9515.html.
19. Voir par exemple le WKID Q925657 : www.wikidata.org/wiki/Q925657.
20. Selon la requête suivante, effectuée le 2 janvier 2023 : w.wiki/5Dg8.
21. Le LATICCE envisage actuellement d'initier ou de promouvoir un datathon autour de l'enrichissement de la propriété P407 pour étudier le potentiel d'engagement collectif autour des enjeux de contribution à un espace de données ouvertes de la musique enregistrée francophone. Le CRTC affirme aussi travailler à l'élaboration d'une telle base de données ouverte permettant de qualifier l'origine canadienne et la langue d'expression d'une pièce musicale.
22. À titre d'exemple, voir la page d'envoi de demandes de correctifs de la base ISNI : isni.oclc.org/xslt/DB=1.2/SET=1/TTL=1/WEBCAT?CI_FORMINDEX_COMMAND=update&cmd=update&PPN=466242395&formcode=IS COMMENT&CLT=ISNI.
23. Il s'inspire en partie de l'essai rédigé en 2013 soulignant la nécessité pour les artistes et ayants droit de participer directement à la construction d'une base fiable de données ouvertes : BisaillonJ.-R., « Métadonnées et politique numérique du répertoire musical québécois. Un essai de mobilisation des connaissances dans le nouvel environnement numérique », essai de recherche réalisé dans le cadre de l'obtention d'un diplôme de maîtrise en arts, Montréal, université du Québec, 2013.
24. Whitman B., « Comment fonctionne la recommandation musicale ? », Medium saignant, 2019, en ligne : mediumsaignant.media/comment-fonctionne-la-recommandation-musicale.
25. Pastukhov D., « Towards recommender system optimization. How can artists influence the Spotify algorithm ? », Paris, Music Tomorrow Blog, 2022, en ligne : www.music-tomorrow.com/blog/towards-recommender-system-optimization-how-can-artists-influence-recommendation-algorithms.
26. Ce sont les technologies de watermarking et de fingerprinting. Pour la première, il s'agit de placer des points de repères inaudibles dans le fichier audio, pour la seconde, de constituer une base de données audio de référence.
27. Bisaillon J.-R. « 10 nouvelles raisons d'État d'indexer nos oeuvres avec des métadonnées », Montréal, LATICCE - Wiki UQAM, 2019, en ligne : wiki.uqam.ca/pages/viewpage.action?pageId=54433018.
28. « Loi sur la diffusion continue en ligne C-11 441 », Ottawa, 2022, en ligne : www.parl.ca/DocumentViewer/fr/44-1/projet-loi/C-11/premiere-lecture.
29. Terme employé par le gouvernement du Canada. Voir « Projet de loi C-11 432. Loi édictant la loi sur la protection de la vie privée des consommateurs et la loi sur le tribunal de la protection des renseignements personnels et des données et apportant des modifications corrélatives et connexes à d'autres lois », 2020, en ligne : parl.ca/DocumentViewer/fr/43-2/projet-loi/C-11/premiere-lecture.

30. developer.spotify.com/documentation/web-api/reference/#/operations/get-recommendations.
31. developer.spotify.com/documentation/web-api/reference/#/operations/get-recommendations.
32. Protocole en cours d'élaboration par le label InTempo Musique et le LATICCE, avec les soutiens du programme MITACS et de la Mission franco-québécoise sur la découvrabilité en ligne des contenus culturels francophones.
33. Ingham T., « It's happened. 100,000 tracks are now being uploaded to streaming services like Spotify each day », Music Business Worldwide, 6 octobre 2022, en ligne : www.musicbusinessworldwide.com/its-happened-100000-tracks-are-now-being-uploaded.
34. www.worldometers.info.
35. www.worldwidewebsize.com.
36. Astor Ph., « Les big datas musicales, une question de souveraineté culturelle qui n'est pas posée », @music_zone, 30 novembre 2022, en ligne : musiczone.substack.com/p/les-big-datas-musicales-une-question.
37. Bisaillon J.-R., « Être ou ne pas être découvrable ? », art. cité, p. 2.
38. Devise attribuée au médecin et citée sur le site du Sénat français : www.senat.fr/evenement/archives/D28/Cbernard.html.
39. Beuscart J.-S., Coavoux S. et Maillard S., « Les algorithmes de recommandation musicale et l'autonomie de l'auditeur. Analyse des écoutes d'un panel d'utilisateurs de streaming », Réseaux, vol. 1, no 213, 2019, p. 17-47.
40. Rioux M. (dir.), « Être ou ne pas être découvrable ? », art. cité, p. 23.
41. Bisaillon J.-R., « Être ou ne pas être découvrable ? », art. cité, p. 2.
42. Casemajor N. et Sirois G., « La gouvernance des données d'usage. Enjeux éthiques et perceptions des publics dans les bibliothèques et archives », INRS, 2021, p. 4.
43. Apprich C., Chun W. H. K., Cramer F. et Steyerl H., Pattern Discrimination, Minneapolis, University of Minnesota Press, 2018.
44. Foucault M., Surveiller et punir. Naissance de la prison, Paris, Gallimard, 1975.
45. Centre for International Governance Innovation, « Digital technologies. Building global trust », 15 juin 2021, YouTube, en ligne : www.youtube.com/watch?v=IOc7Lo4ACEs.Traduction de l'auteur.
46. Harris J., « Tim Berners-Lee : "We need social networks where bad things happen less." » The Guardian, Londres, 15 mars 2021, en ligne : www.theguardian.com/lifeandstyle/2021/mar/15/tim-berners-lee-we-need-social-networks-where-bad-things-happen-less.
47. Sharp E., « Personal data stores : building and trialling trusted data services », BBC, 29 septembre 2021, en ligne : www.bbc.co.uk/rd/blog/2021-09-personal-data-store-research.
48. Voir Apprich C. et al., Pattern Discrimination, op. cit. ; Saulnier A., Les barbares numériques. Résister à l'invasion des GAFAM, Montréal, Écosociété, 2022 ; Harris J., « Tim Berners-Lee : "We need social networks where bad things happen less" », art. cité.
49. Zuboff S., The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power, New York, PublicAffairs, 2019, p. 513. Traduction de l'auteur.
50. O'Neil C., Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy, New York, Broadway Books, 2017, p. 210.
51. Innovation Canada, « L'avenir des communications au Canada : le temps d'agir (rapport Yale) », Ottawa, 2020, p. 36, en ligne : www.ic.gc.ca/eic/site/110.nsf/fra/00012.html.
52. Toledano J., GAFA. Reprenons le pouvoir !, Paris, Odile Jacob, 2020, p. 130.
53. Bortzmeyer S., Cyberstructure. L'Internet, un espace politique, Caen, C & F Éditions, 2018, p. 90-92.