

Données d'usage et usage des données à l'ère des plateformes

*De la nécessité d'un encadrement réglementaire pour
une meilleure affirmation de notre souveraineté*



Décembre 2018

Avec la collaboration de :

AVANT-PROPOS

La section québécoise de l'Internet Society, ISOC Québec, livre ici une étude exploratoire et inédite, commanditée par la Coalition pour la culture et les médias (CCM) et portant sur les données d'usage et les usages qui en sont faits par les plateformes numériques et les nouveaux systèmes intelligents.

Cette étude vise principalement, mais non exclusivement, à documenter un mémoire que la CCM doit remettre au Groupe d'experts responsable de l'examen du cadre législatif et réglementaire de la radiodiffusion et des télécommunications canadiennes, encore connu sous le nom de « Comité Yale » (du nom de la présidente de ce comité, madame Janet Yale).

L'étude s'appuie sur une revue critique de littérature et une synthèse des connaissances sur les enjeux et défis liés à l'accès, à l'utilisation et à la gouvernance des données. Sur la base d'autres études récentes, les analyses et réflexions produites dans ce document permettent de dresser un inventaire des types de données utilisées par les plateformes (canadiennes et étrangères) et par d'autres systèmes tels que les moteurs de recherche et les assistants vocaux, tout en mettant en exergue les modèles d'affaires et les différentes formes d'avantages (monétaires ou non) liés à l'exploitation de ces données.

Ce document explore également quelques pistes de réflexion sur les mesures que pourraient prendre le gouvernement canadien pour actionner les leviers réglementaires et législatifs qui sont nécessaires pour l'accès et l'utilisation des données collectées par les plateformes de diffusion et de distribution culturelles dans le but de favoriser une meilleure découvrabilité, l'accessibilité et le rayonnement du contenu canadien, et ce, dans le respect et la protection de la vie privée et des renseignements personnels des citoyens canadiens.

Nous souhaitons que cette publication serve de point de repère à toute personne intéressée à comprendre les enjeux autour de l'importance que prennent les données d'usage à l'ère du numérique et qu'elle permette une prise de conscience ou une sensibilisation aux défis et problématiques relatifs à l'encadrement de l'accès et de l'exploitation de ces données; défis auxquels il faut impérativement répondre à court ou à moyen terme afin d'affirmer la souveraineté numérique des

citoyens, des entreprises nationales et de l'État, en particulier en ce qui a trait à l'accès et à la consultation de données d'intérêt public.

À PROPOS DE LA PUBLICATION

Cette publication a été rédigée par deux experts :

Destiny Tchéhouali, président du conseil d'administration d'ISOC Québec, professeur de communication internationale au Département de communication sociale et publique de l'Université du Québec à Montréal (UQAM) et directeur de l'Observatoire des réseaux et interconnexions de la société numérique (ORISON)

Josée Plamondon, administratrice d'ISOC Québec, consultante en exploitation de contenus numériques et spécialiste du Web des données, membre du Conseil du numérique du gouvernement du Québec

Notice suggérée pour mentionner cette publication dans une bibliographie ou en reproduire un extrait :

Tchéhouali, Destiny ; Plamondon, Josée. (2018). *Données d'usage et usage des données à l'ère des plateformes : De la nécessité d'un encadrement réglementaire pour une meilleure affirmation de notre souveraineté numérique*, Montréal, ISOC Québec pour la Coalition pour la culture et les médias (CCM), 32 p.

À PROPOS D'ISOC QUÉBEC

Depuis sa fondation en 2000 comme section québécoise de l'Internet Society, ISOC Québec endosse la mission de maintien d'un Internet ouvert, sécurisé, neutre et transparent. ISOC Québec collabore activement à promouvoir le développement de l'accès, de l'utilisation et de l'appropriation d'Internet, tant par les Québécois que par l'ensemble des utilisateurs à travers le monde, contribuant ainsi à réduire la fracture numérique. Regroupant près de 750 membres à travers et hors de la province, ISOC Québec œuvre spécifiquement en faveur de : la diversité des expressions culturelles québécoises sur Internet; l'amélioration des mécanismes de gouvernance multiacteurs de l'Internet; la mise en œuvre des politiques et stratégies numériques nationales cohérentes et soutenant le développement de contenus, d'applications, de services et d'usages innovants et diversifiés de l'Internet au Québec. Notre organisation contribue également à assurer l'outillage et le renforcement des compétences numériques des acteurs francophones de la société de l'information afin de leur permettre de tirer pleinement profit de l'économie numérique ainsi que des nombreuses opportunités offertes par les

récentes innovations et les technologies émergentes de l'Internet (Internet des objets, blockchain, intelligence artificielle, Big Data, etc.), tout en les sensibilisant sur leur dimension éthique et sur leurs impacts sociétaux et environnementaux.
<https://isoc.quebec/>

SOMMAIRE

En guise d'introduction	6
1 Inventaire typologique des données	8
1.1 Données descriptives	9
1.2 Données d'usage	9
1.3 Croisement des données descriptives et d'usage	10
2 La donnée au cœur de l'économie numérique culturelle	12
3 Monétisation des données, modèles d'affaires et dynamiques concurrentielles ...	15
4 Pistes de recommandations pour un encadrement de l'utilisation des données ...	27
4.1 Souveraineté numérique et responsabilité en matière d'accès et de collecte des données d'intérêt public	28
4.2 Concurrence, innovation et accès aux services.....	29
4.3 Neutralité d'Internet	30
4.4 Découvrabilité du contenu canadien et promotion de la diversité des expressions culturelles	31

En guise d'introduction

Depuis l'avènement de l'Internet et grâce aux progrès technologiques constants dans les domaines convergents de l'informatique, des télécommunications et de la radiodiffusion, nous assistons à une augmentation exponentielle de la vitesse de transmission des flux de données produites, collectées et analysées quotidiennement. Le site « planetoscope.com » estime par exemple que chaque seconde, 29 000 Gigaoctets (Go) d'informations sont publiés dans le monde (soit 2,5 exaoctets par jour, ou 912,5 exaoctets par an¹). En termes de comparaison, cela correspondrait approximativement à cinq saisons de la série « Breaking Bad » téléchargées plus de 1000 fois. D'autres statistiques² vertigineuses, fournies par le même site, indiquent l'énorme volume de données créées chaque minute au cours de l'année 2017 : 3 600 000 recherches effectuées sur Google (soit 60 000 recherches Google par seconde); 1 115 000 vidéos regardées sur YouTube (soit 18 580 par seconde); 456 000 messages rédigés sur Twitter (soit 7600 tweets expédiés chaque seconde) ; 46 740 photos postées sur Instagram (soit 779 photos postées sur Instagram chaque seconde); 600 nouvelles pages ajoutées sur Wikipédia (soit 10 nouvelles pages créées sur Wikipédia chaque seconde) et 13 nouvelles chansons ajoutées chaque minute sur la plateforme de streaming Spotify.

Cette spirale prolifique de données massives (ou *Big Data*) transitant quotidiennement sur le Web, via des applications mobiles, des dispositifs ou systèmes intelligents ou encore via des appareils, objets et équipements connectés³

¹ <https://www.planetoscope.com/Internet-/1523-informations-publiees-dans-le-monde-sur-le-net-en-gigaoctets-.html>

² Faisant écho à ces chiffres, le rapport de 2016 sur les modalités de régulation des algorithmes de traitement de contenus souligne aussi que : « le développement rapide de l'Internet et l'utilisation généralisée des moyens informatiques ont conduit à générer une masse énorme de données hétérogènes : chaque jour Google reçoit 3,3 milliards de requêtes, qu'il traite sur quelque 30 000 milliards de pages indexées, Facebook distribue 4,5 milliards de "likes" et reçoit 350 millions de photos, les 3 milliards d'internautes échangent 144 milliards de courriers électroniques. La quantité de données augmentera encore d'au moins un facteur dix avec l'arrivée des objets communicants (Internet des objets), qui feront bientôt partie de notre environnement. Ces données peuvent représenter un potentiel économique considérable, à condition de savoir les exploiter convenablement, ce qui demande le développement permanent d'algorithmes appropriés. Pavel, I; Serris, J. (2016). Modalités de régulation des algorithmes de traitement de contenus, Rapport du Conseil général de l'économie, de l'industrie, de l'énergie et des technologies, remis à la Secrétaire d'État chargée du numérique en France en 2016.

[https://www.economie.gouv.fr/files/files/directions_services/cge/Rapports/2016_05_13_Rapport_Algorithmes\(1\).pdf](https://www.economie.gouv.fr/files/files/directions_services/cge/Rapports/2016_05_13_Rapport_Algorithmes(1).pdf)

³ Le géant américain Intel, leader mondial des microprocesseurs, estime qu'il y aura 200 milliards d'objets connectés à Internet d'ici 2030. JDN, « *L'Internet des objets comme catalyseur de l'innovation* », 5 mai 2017, <https://www.journaldunet.com/ebusiness/expert/66892/l-internet-des-objets-comme-catalyseur-de-l-innovation.shtml>

de dernière génération, constitue un phénomène d'envergure qui a contribué à l'émergence d'un véritable marché florissant de la donnée à l'échelle mondiale. Selon l'International Data Corporation (IDC), le *Big Data* et les outils analytiques ont permis de générer en 2016 près de 130,1 milliards de dollars et ce nombre pourrait augmenter à 203 milliards de dollars en 2020⁴.

L'essor de ce marché lucratif de la collecte et de l'analyse de données bouleverse profondément de nombreuses industries. Le secteur des industries culturelles et créatives n'échappe pas à ce constat. Au contraire, ces industries subissent actuellement les effets de la captation et du détournement de la valeur liée aux données d'usage des consommateurs culturels. Il va sans dire que les entreprises qui tirent actuellement le plus profit de ces données sont celles qui ont réussi à se positionner comme de nouveaux intermédiaires numériques en proposant des services de diffusion et de distribution de produits et de contenus culturels à travers des plateformes de désintermédiation ou de mise en relation directe entre l'offre et la demande culturelles. C'est par exemple le cas de Netflix, d'Amazon, de Google/YouTube, de Facebook et d'Apple. La plupart de ces géants du Web, qu'on regroupe sous l'acronyme des GAFAs, font du profilage des données des utilisateurs de leurs services grâce à leurs plateformes (vidéo et musique à la demande, moteur de recherche, réseau social, site de commerce électronique ou plateforme d'achat et de vente), de façon à créer de la valeur à partir d'une meilleure connaissance, prédiction et prescription des goûts et habitudes de consommation de ces utilisateurs ou abonnés. En effet, plus les utilisateurs passent du temps sur ces plateformes ou sur des applications de leurs smartphones, plus ils y laissent des traces et empreintes de leur présence numérique (données sur leur identité, leur personnalité, leurs préférences, sensibilités ou centres d'intérêt), susceptibles d'être algorithmiquement combinées⁵ et transformées en valeur pour alimenter des stratégies de recommandation et de personnalisation de services ou encore des publicités ciblées.

Dans un contexte où le *Rapport de surveillance des communications de 2017* du Conseil de radiodiffusion et des télécommunications canadiennes (CRTC) souligne que les plateformes et les appareils connectés à Internet constituent les

⁴ Le Big Data, « 15 faits impressionnants sur le Big Data », 23 janvier 2018. <https://www.lebigdata.fr/chiffres-big-data>

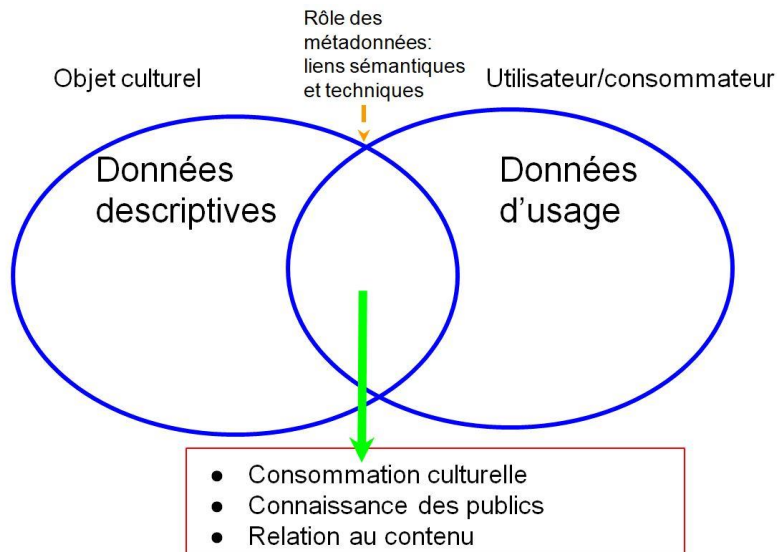
⁵ Les algorithmes de traitement effectuent un travail consistant à croiser les données pour définir des profils ou catégories d'utilisateurs. Ces derniers sont orientés par des annonces publicitaires ou par des recommandations personnalisées vers des produits (par exemple musique ou films) considérés par l'algorithme comme étant susceptibles de correspondre aux préférences, habitudes, ou attentes des utilisateurs ainsi ciblés.

nouveaux moyens ou canaux très populaires utilisés par les Canadiens pour accéder aux contenus vidéo et audio qui les intéressent⁶, il devient nécessaire de questionner les implications liées à cette tendance sur les données d'usage relatives aux contenus culturels au Canada, tant sous l'aspect de la protection de la vie privée et des données personnelles des Canadiens que sous l'aspect de la capacité des autorités publiques et gouvernementales à contraindre les plateformes à communiquer ces données, tout au moins celles d'intérêt public, dans un souci de transparence. Cependant, avant de nous pencher sur ces questions, il importe dans un premier temps de dresser un inventaire typologique des données.

1 Inventaire typologique des données

Il y a plusieurs approches de catégorisation des données. Celle que nous proposons ici distingue trois grandes familles de données : les données descriptives, les données d'usage et les données relevant du croisement des données d'usage et des données descriptives.

Figure 1 : Typologie des données



⁶ Ce rapport montre par exemple que : 23 % des Canadiens âgés entre 18 à 34 ans regardent la télévision exclusivement en ligne; 22 % d'entre eux ont écouté des stations de radio AM/FM en continu sur Internet et plus de 55 % d'entre eux ont regardé des vidéos de musique en continu sur Internet. CRTC, « *Les plateformes numériques de plus en plus populaires* », Communiqué de presse, Ottawa-Gatineau, 8 novembre 2017.

https://www.canada.ca/fr/radiodiffusion-telecommunications/nouvelles/2017/11/les_plateformes_numeriquesdeplusenpluspopulaires.html

1.1 Données descriptives

Les données descriptives sont des données de catalogage fournissant de l'information sur un **objet culturel** (chanson, vidéo, récit, illustration, sculpture, etc.). Cette information peut être de natures diverses :

- descriptive (nom, titre, auteur, dimensions, caractéristiques, langue, etc.);
- juridique (droits d'utilisation, ayants droit);
- géographique (coordonnées de géolocalisation);
- administrative (délai de conservation);
- technique (support, format, date/heure de captation).

1.2 Données d'usage

Ces données, qui constituent l'objet central de la présente étude, sont des données qui fournissent de l'information sur l'utilisateur ou le consommateur du contenu culturel (tout en l'identifiant, même si le traitement de ces données peut être anonymisé⁷). Elles sont issues de l'activité de l'utilisateur et, plus spécifiquement, de son interaction avec des contenus⁸.

À titre d'exemple de données d'usage, on peut citer :

- Les données nominatives, personnelles ou de profil (nom, adresse, identifiant, etc.);
- Les données de transaction (mode de paiement, date, montant, institution financière);
- les données sur les centres d'intérêt (contenus préférés, abonnements);
- les données de comportement (sélection, recherche, consultation, achat, partage, etc.);
- les données collectées par les téléphones mobiles, ordinateurs et objets connectés (appels, activation, localisation);
- les données de navigation.

⁷ Les enjeux liés à l'anonymisation des données se posent dès lors qu'on se place du point de vue de l'entreprise qui souhaite utiliser les données recueillies. Voir aussi à ce sujet : de Montjoye, Y.-A. *et al.* (2018). « On the privacy-conscious use of mobile phone data », in *Sci. Data*. <https://www.nature.com/articles/sdata2018286>

⁸ Rappelons que la plupart des plateformes permettent à leurs utilisateurs de télécharger les données qui sont collectées sur eux. Ces données sont généralement accessibles à partir des paramètres de confidentialité.

L'exploitation de ce type de données d'usage favorise :

- le développement d'offres plus variées et adaptées aux attentes des consommateurs grâce à la mesure des utilisations (exemple : tarification et gestion dynamique des prix pour des services identiques);
- la production de sens et d'indicateurs de mesure adaptés aux nouveaux contextes d'utilisation;
- les interactions génératrices d'information sur les publics, les usages, les marchés.

1.3 Croisement des données descriptives et d'usage

Les données descriptives et d'usage ont beaucoup plus de valeur lorsqu'elles peuvent être croisées, car les données descriptives fournissent le **contexte de l'interaction** avec le contenu et les données d'usage caractérisent **l'utilisation** même qui est faite du contenu. Plus les données descriptives sont détaillées, plus il est possible d'expliquer des choix et d'identifier des comportements récurrents.

« After we've processed and annotated every frame in a given video, the next step is to surface "the best" image candidates from those frames through an automated artwork pipeline. That way, when our creative teams are ready to begin work for a piece of content, they are automatically provided with a high quality image set to choose from. Below, we outline some of the key elements we use to surface the best images for a given title⁹ ».

C'est grâce à des métadonnées qu'il est possible de faire des rapprochements entre des données provenant de sources diverses. Par exemple, la métadonnée « ville » permet de relier des données comme Montréal, YUL et MTL. Celles-ci sont différentes, mais désignent la même chose. Dans un autre exemple, la métadonnée « ISNI » indique que le numéro qui suit est un identifiant unique et normalisé qui a été attribué à un auteur. Cette métadonnée, qui est internationalement reconnue, permet de relier des œuvres et des productions à des auteurs et contributeurs. Les métadonnées sont également utiles pour établir des liens ou pour faire des associations entre la qualification d'une œuvre ou d'un contenu et les déterminants contextuels de la consultation de cette œuvre ou de ce contenu par un utilisateur.

⁹ AVA, «The Art and Science of Image Discovery at Netflix», <https://medium.com/netflix-techblog/ava-the-art-and-science-of-image-discovery-at-netflix-a442f163af6>

Le croisement de données descriptives et d'usage permet d'analyser la consommation culturelle, d'améliorer la connaissance des publics et d'interpréter la relation au contenu.

« Les données personnelles collectées et traitées dans le contexte des contenus culturels et créatifs ont des particularités liées au caractère très intime de la relation entre une personne et les œuvres : les pratiques culturelles contribuent à la construction de la personnalité et de l'identité... Dans ce domaine, les data-scientists sont en mesure d'utiliser des données d'apparence très anodine, mais très particulières : profils, descriptions des contenus, popularité, enrichissement, goûts et contexte.¹⁰ »

*"There's a mountain of data that we have at our disposal". Netflix has a well-earned reputation for using the information it gleans about its customers to drive everything from the look of the service to the shows in which it invests. "That mountain is composed of two things. Garbage is 99 percent of that mountain. Gold is one percent. Geography, age, and gender? We put that in the garbage heap. Where you live is not that important." Stereotypes about what one region might like versus another are largely useless to Netflix¹¹. **Todd Yellin**, Netflix's VP of product innovation*

"The Database of Intentions is simply this : The aggregate results of every search ever entered, every result list ever tendered, and every path taken as a result. It lives in many places, but three or four places in particular hold a massive amount of this data (ie MSN, Google, and Yahoo). This information represents, in aggregate form, a place holder for the intentions of humankind – a massive database of desires, needs, wants, and likes that can be discovered, subpoenaed, archived, tracked, and exploited to all sorts of ends. Such a beast has never before existed in the history of culture, but is almost guaranteed to grow exponentially from this day forward. This artifact can tell us extraordinary things about who we are and what we want as a culture. And it has the potential to be abused in equally extraordinary fashion."¹²

¹⁰ CNIL, « Les données : muses et frontières de la création : Lire, écouter, regarder et jouer à l'heure de la personnalisation », Dans Innovation & Prospective (IP), no 3, p. 22.

https://www.cnil.fr/sites/default/files/typo/document/CNIL_CAHIERS_IP3.pdf

¹¹ Wired, « Netflix's Grand, Daring, Maybe Crazy Plan to Conquer the World », 27 mars 2016.

<http://www.wired.com/2016/03/netflixs-grand-maybe-crazy-plan-conquer-world/>

¹² John Battelle's Search Blog, « The Database of Intentions », 2003.

https://battellemedia.com/archives/2003/11/the_database_of_intentions

2 La donnée au cœur de l'économie numérique culturelle

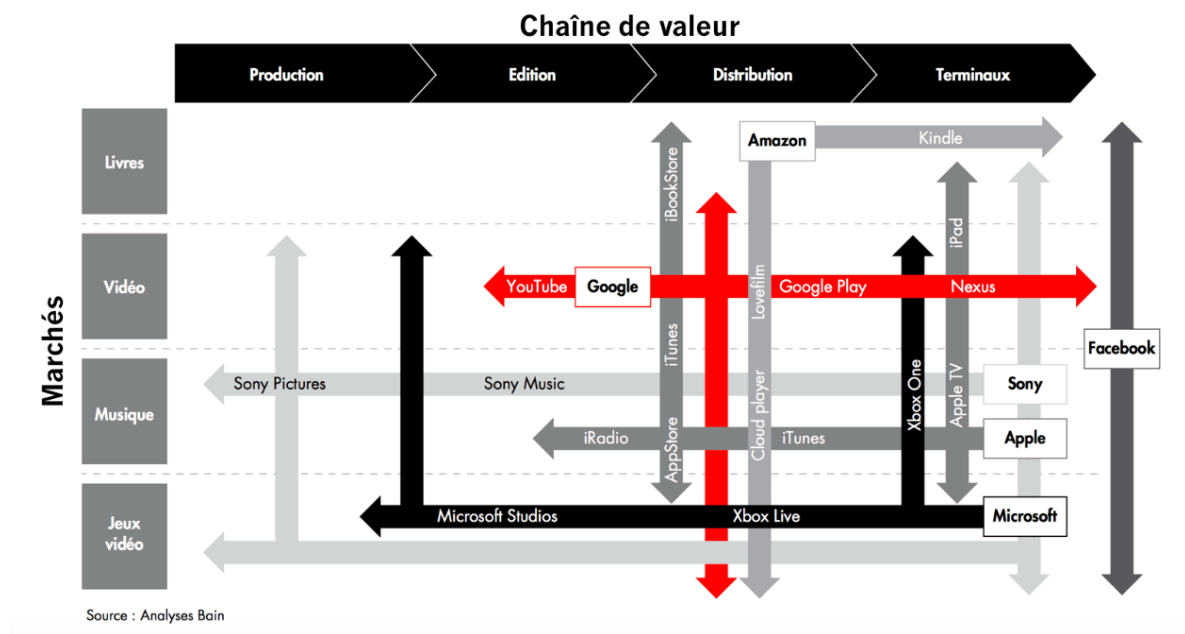
L'intelligence d'affaires des plateformes numériques et des nouveaux systèmes et objets connectés (tels que les assistants vocaux¹³) réside dans leur capacité à comprendre, à s'adapter, à satisfaire et à orienter ou influencer les choix des consommateurs et utilisateurs de leurs services. En effet, les utilisateurs sont une partie constituante de la création de valeur des plateformes puisque leurs données sont mobilisées comme ressources (matières premières) et valorisées à travers différents effets de croisement lors du traitement, de l'utilisation, de l'échange ou de la commercialisation de l'accès à cette ressource.

La dématérialisation croissante des produits et services culturels a facilité l'irruption sur le marché culturel d'acteurs spécialisés dans les contenus numériques. Si Google, Apple, Facebook, Amazon, Netflix, YouTube sont devenus à l'échelle mondiale les nouveaux fournisseurs (à la fois diffuseurs, distributeurs, mais désormais aussi producteurs et éditeurs) de contenus culturels, c'est parce que ces entreprises ont compris très tôt l'intérêt de se positionner dans la ruée vers la nouvelle mine d'or que constituent les données des consommateurs culturels afin de pouvoir générer des recettes importantes en développant des innovations de produits ou de services basés sur l'agrégation et l'analyse de ces données.

Les données, en particulier les données personnelles culturelles qui reflètent nos goûts et une partie de notre identité et de notre personnalité, sont le cœur de cible de la nouvelle chaîne de valeur résultant de la transformation numérique et de la « plateformeisation » des industries culturelles. Le phénomène de plateformeisation se caractérise par l'intégration à la fois verticale et horizontale des grandes entreprises technologiques.

¹³ Comme le mentionne le Rapport du Fonds des Médias du Canada (FMC) sur les tendances 2018 : « Les appareils dont les entrées et les sorties passent principalement par des commandes vocales gagnent en accessibilité et en popularité. [...] De nombreux robots vocaux prennent la forme d'assistants personnels intelligents, comme Alexa d'Amazon, Siri d'Apple, Google Assistant et Cortana de Microsoft, mais il y a de nombreux autres appareils qui proposent la recherche vocale tels que les lecteurs multimédias (Apple TV, Chromecast, Fire TV Stick, Roku), les téléviseurs intelligents et les consoles de jeux. » Le rapport précise toutefois que : « Malgré un intérêt évident pour les consommateurs, il reste encore des obstacles à surmonter avant que la recherche vocale ne soit adoptée à grande échelle. La personnalisation (connaître qui pose la question), l'intention (comprendre le contexte) et le respect de la vie privée (dissiper les craintes des consommateurs concernant le stockage et l'utilisation de leurs données personnelles) doivent notamment être améliorés.» FMC (2018). Rapport sur les tendances 2018 : *Le Choc du présent*, p.19 https://trends.cmf-fmc.ca/wp-content/uploads/Rapport_sur_les_tendances_2018_-_Le_choc_du_present_-_Final_.pdf

Figure 2 : L'expansion des plateformes au sein des industries culturelles¹⁴



L'économie des plateformes s'impose progressivement comme un modèle économique et technique dominant, résultant des transformations combinées des technologies, des marchés et des habitudes de consommation. Différents types d'acteurs gravitent autour de cette nouvelle économie numérique et de son écosystème très concurrentiel, notamment : les constructeurs d'infrastructures (réseaux et centres de données), les équipementiers ou fabricants de matériel et d'équipements informatiques et technologiques, les opérateurs de télécommunication fournissant l'accès aux services, les producteurs et éditeurs de contenus (films, musique, livres, information, émissions, jeux vidéo), les développeurs d'applications, de logiciels et de systèmes d'exploitation et enfin les nouveaux entrants que sont les plateformes jouant le rôle de nouveaux intermédiaires et fournisseurs de contenus.

Dès lors, la dimension écosystémique des plateformes numériques repose à la fois sur l'infrastructure matérielle (liée aux opérations d'hébergement/stockage de données sur des serveurs et à leur transmission via des réseaux de télécommunication) et sur l'infostructure ou la couche logicielle (liée aux standards, aux bases de données, aux contenus et applications, notamment les

¹⁴ Bain & Company (2013). Les nouvelles prescriptions : de l'abondance à la découverte, p. 6 http://www.forum-avignon.org/sites/default/files/editeur/2013ForumAvignon_FR_BD-23-10-2013OneonOne.pdf

Application Programming Interfaces ou API¹⁵) qui permet le bon fonctionnement et le déploiement des services de la plateforme.

Les plateformes numériques possèdent ainsi tout le potentiel pour se positionner stratégiquement sur l'intégralité de la chaîne de valeur de l'économie numérique, en captant non seulement l'attention des utilisateurs/consommateurs, mais aussi en constituant des rentes monopolistiques grâce aux avantages (monétaires ou non) qu'elles peuvent tirer de la captation et de l'exploitation des flux de données massives transitant via leur écosystème et générées par la masse critique d'utilisateurs de leurs services.

Analysant les tendances numériques dans le secteur culturel, Octavio Kulesz, éditeur numérique argentin et consultant pour l'UNESCO, illustre avec le cas d'Amazon le processus de concentration du marché culturel global et la captation de l'intégralité de la chaîne de valeur culturelle par les grandes plateformes :

« [...] une plateforme commence par occuper un espace (matériel, logiciel, communication, moteur de recherche, e-commerce, etc.) qui, grâce à la convergence numérique, est commun à toutes les industries culturelles, et finit invariablement par s'étendre et à prendre la place d'autres maillons de la chaîne. Ainsi Amazon n'est plus seulement une librairie en ligne mais également un éditeur, un vendeur d'appareils comme le Kindle, un réseau social de recommandation de livre et un imprimeur ainsi qu'un prestataire de service d'informatique en nuage. Ces nouveaux géants ne sont donc plus des acteurs de la chaîne de valeur, mais ils tendent à devenir à eux seuls l'intégralité de la chaîne de valeur. Par conséquent, ils deviennent des écosystèmes complètement clos dont les données et les rouages internes ne sont pas connus du monde extérieur¹⁶. »

Selon O. Kulesz, en plus d'entraîner des risques d'érosion des modèles économiques des acteurs traditionnels, cette situation de concentration et de

¹⁵ Les interfaces de programmation ou interfaces pour l'accès programmé aux applications permettent de faire le lien entre un utilisateur et une application. Ils sont les garants de l'interopérabilité des protocoles de communication et d'échange de données entre différents systèmes d'information, applications, programmes et plateformes. L'API permet ainsi d'accéder à des services Web ou des ressources logicielles et de contrôler l'exposition et l'utilisation des données numériques produites par ces services ou ressources. On cite souvent en exemples les «boutons Facebook». Facebook propose d'ailleurs des «API partenaires» partagées avec des entreprises tierces (avec lesquelles le réseau social conclut des ententes) dans le but d'utiliser ces API pour faciliter l'échange et l'agrégation de données d'usages et d'informations sur les profils ou comptes d'utilisateurs de Facebook, et ce parfois à l'insu des utilisateurs du réseau social.

¹⁶ Kulesz, O. (2014). *Convention de l'UNESCO pour la protection et la promotion de la diversité des expressions culturelles : Analyse transversale des rapports périodiques des Parties sur les enjeux et les tendances numériques*, CE/14/8.IGC/INF.5, novembre 2014, p. 9-10.

https://en.unesco.org/creativity/sites/creativity/files/sessions/5cp_inf9_digital_fr.pdf

contrôle de l'intégralité de la chaîne des valeurs soulève des problématiques pour les utilisateurs qui sont *de facto* confrontés à une surveillance permanente de leurs données de consommation, avec des atteintes en matière de respect de leur vie privée et de protection de leurs renseignements personnels.

*Amazon has search.
Amazon has deep understanding of consumer's needs and wants.
Amazon knows who those consumers are connected to.
Amazon knows which web services these consumers use.
Amazon knows much more than most of us will ever know¹⁷.*

3 Monétisation des données, modèles d'affaires et dynamiques concurrentielles

L'accès aux données d'usage culturelles s'impose comme une valeur d'échange puisque la valeur liée aux échanges de ces données dans l'économie de la culture est croissante et a maintenant plus d'importance que la valeur même d'un contenu¹⁹.

Contrairement aux contenus et aux données personnelles dont la valeur marchande peut s'éroder rapidement, les données d'usage suscitent davantage d'intérêt pour les services de recommandation et de publicité ciblée.

Nous identifions ici quatre formes ou possibilités de monétisation des données :

- 1- La vente de données brutes à des agrégateurs ou des opérateurs d'analyses ou directement à des clients finaux (firmes de publicité par exemple) ;
- 2- l'utilisation de données d'usage historiques fines sur l'activité d'un groupe

¹⁷ Mitch, J. The Amazon That You Don't See (That Google And Facebook Are Watching), 18 avril 2017. <https://medium.com/startup-grind/the-amazon-that-you-dont-see-that-google-and-facebook-are-watching-ff6060b60368>

¹⁹ D'après une étude (*The Internet Value Chain*) publiée en 2016 par le GSMA (l'Association des opérateurs mobiles), les contenus en ligne ont permis de générer 64 milliards de dollars de valeur en 2015, ce qui ne représente seulement que 2 % de la valeur totale créée sur le Web. Cela en fait d'ailleurs l'activité qui contribue le moins à la création de valeur en ligne. Or, au même moment, l'étude rapporte que toute la chaîne de valeur d'Internet, c'est-à-dire l'ensemble des activités qui relient un contenu, un service, ou un produit à l'internaute, a triplé en l'intervalle de seulement huit années. http://www.gsma.com/publicpolicy/wp-content/uploads/2016/05/GSMA_The-internet-Value-Chain_WEB.pdf

- d'utilisateurs ou de clients pour mieux comprendre leurs comportements de consommation et adapter son produit ou service en fonction;
- 3- la collecte et l'analyse de données massives via un service, un équipement, un produit ou objet et l'utilisation de ces données pour l'amélioration de la performance du service et la construction d'une audience;
 - 4- l'utilisation de données en temps réel pour fluidifier la rencontre entre l'offre et la demande.

*Aujourd'hui, le principal vecteur de valeur pour la donnée réside dans sa capacité à influencer sur les décisions ou à donner une information précieuse. C'est dans cette analyse approfondie de la data, le croisement des renseignements qu'elle fournit, et dans les perspectives stratégiques qu'elle laisse entrevoir que réside la nouvelle valeur de la donnée. On observe alors une mutation substantielle du secteur du Big data vers le Smart Data. Quand le premier se contentait le plus souvent de récolter massivement des données utilisateurs pour produire des analyses quantitatives de grande échelle, le second lui se concentre sur l'analyse des seules données utiles au client²⁰.
Théo Roux, Journaliste, Atelier BNP Paribas*

Il est à noter que la vente des données issues de la collecte directe ou indirecte sur les plateformes ou via tout dispositif, système ou équipement technologique connecté est devenue l'un des principaux modèles de financement des services en ligne. Ce modèle économique repose en grande partie sur la gratuité, le produit vendu ici étant l'utilisateur final. Ceci fait écho à la formule lancée par Andrew Lewis en 2010 et devenue célèbre pour exprimer le *modus operandi* des géants du Web : « If you don't pay for it, you're not the customer ; you're the product being sold »²¹ (formule qu'on pourrait traduire précisément par : « Si c'est gratuit, c'est que vous n'êtes pas la clientèle visée ; c'est vous le produit »).

« [...] En témoigne le succès de Facebook qui offre des services innovants et gratuits à ses utilisateurs, en contrepartie d'une utilisation commerciale de leurs données personnelles. La plate-forme analyse les profils de ses membres afin d'accroître la personnalisation de la relation commerciale. Cette forme de publicité contextuelle permet aux annonceurs de connaître avec exactitude le profil d'un acheteur potentiel – un gain de temps par rapport aux campagnes fondées sur la seule probabilité statistique de leur succès. Les annonceurs ne connaissent certes pas l'identité du prospect, mais ils savent qui il est. [...] »

²⁰ Roux, T. « Ces data-brokers qui font commerce de nos données personnelles », mars 2017. <https://atelier.bnpparibas/smart-city/article/data-brokers-commerce-donnees-personnelles>

²¹ Andrew Lewis, Twitter, 13 septembre 2010. <https://twitter.com/andlewis/status/24380177712>

Jusqu'à l'arrivée d'Internet, le métier qui consiste à « vendre des consommateurs à des annonceurs » relevait d'un art approximatif, fondé sur une vague idée des habitudes de consommation des différents groupes. Aujourd'hui, il s'est professionnalisé avec l'augmentation de la pertinence des cibles. De nouvelles techniques permettent de traquer les activités en ligne d'un internaute (grâce aux tracking cookies) et d'ajuster les publicités qu'on lui délivre en fonction de son comportement. Avec la publicité comportementale, les annonceurs affirment désormais savoir ce que le prospect désire.²³»

Plusieurs études tendent à démontrer que les utilisateurs considèrent souvent que leurs données (personnelles) ont moins de valeur que le service qu'ils reçoivent en retour. Ils donnent donc très aisément leur accord aux conditions d'utilisation des plateformes en cliquant sur OK, et ce, généralement sans jamais lire jusqu'au bout ces fameuses conditions générales d'utilisation (CGU) et sans prendre connaissance des implications qui y sont liées à court et à long termes. Les GAFAs nous ont ainsi habitués à délaissier de manière consentante, voire sous une fatale résignation, le sort de toutes les données liées aux traces que nous laissons sur le Web, bien que ceci soit en train de changer depuis les derniers scandales liés à Facebook (Cambridge Analytica) et aussi aux données de géolocalisation de Google²⁴. Exception faite des services de placement publicitaire, les plateformes demeurent extrêmement discrètes sur la monétisation de leurs activités relatives à l'exploitation des données. Pourtant, elles mettent en œuvre un ensemble de stratégies pour accroître sans cesse la masse de données à leur disposition. C'est ce qu'a souhaité démontrer un journaliste du Wall Street Journal en commandant une pizza lors d'une soirée entre amis. En tentant d'identifier les données qui pouvaient être mises en jeu à la suite d'une aussi banale transaction, il a soulevé les nombreuses fuites dont notre vie privée est victime par la seule interconnexion des réseaux sociaux et applications que nous utilisons²⁵.

²³ Douplitzky, K. (2009). « Le commerce du moi, modèle économique du profilage », *Hermès, La Revue*, 2009/1 (n° 53), p. 113-117 <https://www.cairn.info/revue-hermes-la-revue-2009-1-page-113.htm>

²⁴ Le Monde, « Google a collecté des données de localisation sur les utilisateurs d'Android à leur insu », http://www.lemonde.fr/pixels/article/2017/11/22/google-a-collecte-des-donnees-de-localisation-sur-les-utilisateurs-d-android-a-leur-insu_5218680_4408996.htm, 22 novembre 2017

²⁵ « Depuis le début de la soirée s'ajoutent ainsi les unes aux autres des informations « pratiques » qui ouvrent sur autant d'informations annexes récupérées par les différents programmes numériques utilisés, au prétexte de « favoriser l'expérience utilisateur ». En tout, 53 informations différentes ont été enregistrées: 15 données par Sally et Kristen (28 %), 38 collectées par les compagnies (72 %), relativement à leur politique sur la protection de la vie privée et les conditions qui l'entourent. » Polverini, L. (2018), « Une soirée pizza peut vous coûter plus cher en données privées qu'en argent », *Slate*, 16 avril 2018. <http://www.slate.fr/story/160486/soiree-pizza-donnees-privées>

Cette collecte peut aussi bien s'effectuer à travers les conditions générales d'utilisation (CGU) qui assurent une captation maximale des données produites, mais aussi des données de trafic (*analytics*), le rachat de réservoirs de contenus (YouTube par exemple, pour Google), ou encore le lancement de services dans les nuages (*cloud computing*), qu'ils soient payants ou non, tels que Gmail, Google Drive, des logiciels commerciaux de bureautique, etc.

Mais que valent au juste nos données personnelles sur le marché des données? Le tableau suivant nous fournit des éléments de réponse.

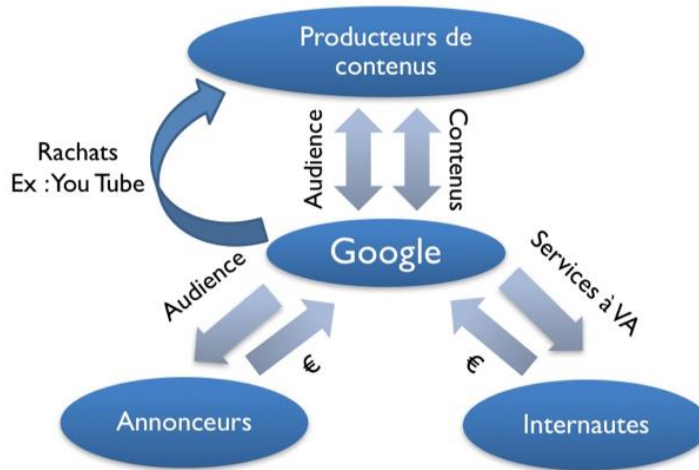
Tableau 1 : Exemples de prix pour différents types de données sur le marché²⁶

Entreprises	Types de données	Valorisation
Axiom	Adresses courriel (accompagnées éventuellement d'un profil)	2 à 5 cents par contact
Plateformes RTB et AdExchange	Profils de navigation (sans identification)	0 à 1 dollar pour 1000 affichages en Europe
Facebook	Valorisation boursière d'un profil	102,3 dollars
Federico Zannier	Données de navigation, localisation du pointeur de souris, GPS, webcam, fichiers log	2 dollars pour un jour, 5 pour une semaine
Datasift	Tweets et analyses (tarif fonction de la charge ^{SEP} de calcul; prix du Data Processing Unit : 20 cents par heure)	10 cents pour 1000 tweets plus un coût de traitement exprimé en DPU
Datacoup	Données de profil agrégées (navigation, réseaux sociaux, transactions de carte bancaire) type panel	8 dollars par mois offerts à chaque membre du panel
Appannie.com Intelligence	Données de téléchargement d'applications mobiles et ebooks par plateformes et par pays	Gratuit pour l'éditeur d'une application ; 15 000 dollars par an pour un accès à l'ensemble des données
Datamarket.com	Données mises en vente par les entreprises qui les détiennent ou les créent	Prix fixé par le vendeur

Comme l'indique le tableau ci-dessus, la valorisation boursière d'un seul profil Facebook (constitué de nombreuses données et renseignements personnels) représente en revenus la vente de 2000 à 5000 contacts d'adresses courriel ou la vente de 100 profils de navigation (anonymisés) sans identification.

²⁶ Willart, S.; Crié D. (2016). « Création de valeur par les données massives », Dans *Statistique et société*, vol. 4, n° 3, décembre 2016, p. 22.

Figure 3 : Modèle de monétisation des données basé sur l'économie des plateformes et des médias



Jean Tirole, prix Nobel d'économie, décrit l'économie des plateformes comme des systèmes bifaces de valorisation de données :

Une autre différence de taille tient dans la capacité de ces plateformes, du moins dans leurs avatars les plus aboutis, comme Google, à ouvrir une troisième face de leur marché, par la création de services à valeur ajoutée exploitant l'énorme quantité de données collectées sur la première face du marché : données de géolocalisation dans le cas de la « Google, car », données de toute sorte dans le cas des « Google glasses²⁸ ».

*What should concern you is that Amazon has expressed some degree of openness to granting **third-party app developers** access to transcripts of audio recordings saved by Alexa-powered devices—though the company says this would only be with users' consent, according to the Verge. While this may seem shocking, Amazon is considering it because its biggest competitor in the VUI market, Google Home, is already giving developers access to this data. In the same ways that we've seen Facebook and Google slice and dice data unknowingly given up by consumers and sell it to third parties with varying results, the practice should always give users slight pause. No free app or service is ever actually free²⁹.*

²⁸ Perales, C. (2016). « Future is now: « ubérisation » de la science et économie de plateforme », <http://blog.educpros.fr/christophe-perales/2016/10/02/future-is-now-uberisation-de-la-science-et-economie-de-plateforme/>

²⁹ Powers B. « “OK Google,” What Are the Problems with Speech Recognition Technology? : How Amazon and Google's big bets on the smart speaker market affect their customers' privacy », *Medium*, janvier 2018. <https://medium.com/@bnpowers8/ok-google-what-are-the-problems-with-speech-recognition-technology-531d23aa224e>

Tout récemment, une commission parlementaire britannique dévoilait des documents suggérant que Facebook aurait accordé à des compagnies amies un accès privilégié à des données personnelles d'utilisateurs en même temps que la plateforme bloquait l'accès à ses concurrents potentiels.

« Selon la BBC, des compagnies comme Airbnb, Netflix, Lyft et Badoo auraient ainsi profité des faveurs de Facebook ce qui leur a permis de développer des applications leur permettant, par exemple, de savoir ce que leurs amis avaient cherché comme destinations ou regardé comme émissions ces derniers jours. La Banque Royale du Canada et l'application Tinder auraient aussi demandé de profiter du même droit, selon le Washington Post³⁰ ».

Selon une autre enquête³¹ du *New York Times*, le réseau social Facebook aurait fourni à plusieurs géants du web (Amazon, Apple, Microsoft, Netflix, Spotify) un accès aux données personnelles de ses utilisateurs, et ce, sans jamais véritablement avertir ces derniers. Au total, plus de 150 entreprises auraient ainsi bénéficié d'« exceptions de confidentialité » de 2010 à 2017, utilisant les données des utilisateurs du réseau social à différentes fins : accès aux conversations privées sur Messenger pour Spotify et Netflix ; élargissement de la base de données de clientèle d'Amazon et d'Apple, avec des informations personnelles telles que le courriel et le numéro de téléphone et accès au fil d'actualité des utilisateurs et à l'agenda Facebook de leurs clients respectifs.

Contrairement à Facebook, Netflix se défend de l'utilisation des données personnelles à des fins de publicité parce que le modèle d'affaires de la plateforme repose avant tout sur les abonnements des utilisateurs et que les données d'usage générés à travers les interactions de ceux-ci avec les contenus de la plateforme constituent des données stratégiques et sensibles ne pouvant être partagés avec des partenaires ou des tiers. Ces données permettent plutôt à Netflix de se démarquer de la concurrence et d'anticiper les tendances des contenus à succès afin d'investir et de miser sans trop de risques dans l'acquisition et la production de nouveaux contenus qui lui permettent de fidéliser ses abonnés.

³⁰ <https://www.ledevoir.com/economie/542971/facebook-accuse-de-divulguer-des-donnees-personnelles-au-gre-de-ses-interets-commerciaux>

³¹ Dance, G.J. X; LaForgia M.; Confessore, N. (2018), *Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants*, *New York Times*, 18 décembre 2018 <https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html>

« La protection de vos données personnelles sur Internet lorsque vous allez sur Netflix est secrète, fiable et sûre. On n'est pas une boîte de publicité, on ne vend notre data à personne, et on ne le fera jamais." Et de poursuivre que le "business de la data" n'est "clairement pas un bon endroit" : "Autant faire des claquettes dans un champ de mines ! Et en plus ça ne vaut même pas le coup économiquement, ça ne rapporte pas d'argent.³²» **Todd Yellin**, vice-président en charge du produit chez Netflix

Rappelons qu'en amont au lancement de leur offre de services, les plateformes (mais aussi les fabricants de terminaux intelligents et d'objets connectés) s'assurent déjà de mettre en place l'écosystème et l'infrastructure nécessaires pour recueillir assez tôt une masse critique d'informations et de données d'usage (préférences et centres d'intérêt des consommateurs, données de comportement, données de géolocalisation, etc.)³⁴. La collecte se fait soit directement à partir des premières transactions sur la plateforme même ou dès les premières manipulations de l'appareil ou de l'équipement par l'utilisateur ; soit indirectement par les données³⁵ reliées aux activités de l'utilisateur sur d'autres plateformes comme les réseaux sociaux.

« Les comportements sont révélés sur la base des données à caractère personnel récoltées au sein ou autour du dispositif. Dans un premier temps, il s'agit d'inciter le consommateur à déclarer un certain nombre d'éléments (données) à caractère objectif, comme l'âge, le sexe, l'identité civile, la composition du foyer, ou à caractère subjectif, comme des préférences de contenus, de genres, etc.) afin qu'il organise par lui-même une partie de sa consommation, et ce, dès l'entrée dans le dispositif, nécessitant généralement une inscription. [...] Dans un second temps, des données sont récoltées grâce à l'observation des parcours de consommation afin de déterminer des comportements attribuables au consommateur, à différents membres du foyer, etc.³⁶»

³² Wessbecher, L. (2018). « Netflix et les données personnelles : doit-on craindre pour la protection de notre vie privée ? », **France24**, 28 avril 2018, <https://www.france24.com/fr/20180428-netflix-donnees-personnelles-doit-on-craindre-protection-notre-vie-privee>

³⁴ En outre, des tendances montrent une augmentation des pratiques de monétisation de données personnelles par l'intermédiaire des plateformes de *data-brokers* et des systèmes de coffres-forts virtuels ou numériques où les internautes peuvent stocker ou vendre eux-mêmes leurs informations personnelles.

³⁵ Ces dernières étant systématiquement retracées et collectées à l'aide des cookies ou des balises Web installés ou préinstallés sur certaines applications, sur le navigateur de l'utilisateur, parfois à son insu.

³⁶ Farchy, J. ; Méadel, C.; Anciaux, A. (2017). « Une question de comportement. Recommandation des contenus audiovisuels et transformations numériques », *tic&société* [Online], Vol. 10, N° 2-3 | 2ème semestre 2016 - 1er semestre 2017, p.187, <https://journals.openedition.org/ticetsociete/2136>

La maîtrise et l'exploitation de ces données sont donc devenues aujourd'hui un facteur de différenciation essentiel dans la structuration de l'offre de services des plateformes numériques et dans la manière dont cette offre est adressée aux utilisateurs. Dans les domaines culturels et médiatiques, c'est ce facteur de différenciation qui détermine les positions concurrentielles entre les acteurs traditionnels et les nouveaux entrants que sont les plateformes numériques. Ces dernières se sont imposées très rapidement grâce à des innovations de produits et de services face aux acteurs traditionnels, car leurs produits et services sont développés sur la base de la fine connaissance des utilisateurs mêmes de ces services.

Pour Google, les avantages du marché de la télévision en continu ne se réduisent pas seulement à la programmation: connaître les habitudes des téléspectateurs permet de mieux cibler la publicité. "Pour eux, c'est une affaire de données", affirme Alan Wolk. "Ils obtiennent une masse d'informations sur les utilisateurs et ce qu'ils font, et ils peuvent ensuite faire des croisements avec leurs habitudes de navigation sur internet³⁷ ».

Ainsi, un service d'écoute en continu de musique comme Spotify exploite les données d'écoute de ses propres utilisateurs :

« Le retour à la maison se fera sur une chanson douce, le petit-déjeuner sera instrumental, le trajet en métro, passablement reggae. Et ainsi de transition en transition pour toute la journée (avec peut-être une pause de quelques minutes pour parler à quelqu'un). Ce « rêve » d'un algorithme extrêmement intelligent qui saurait vous recommander la parfaite liste de musique à tout moment du jour, en fonction de vos goûts, de vos activités, de votre humeur et d'une foule d'autres paramètres, ne tient plus tant de la science-fiction.»³⁸

Il en est de même pour toutes les plateformes proposant des services de classement, de hiérarchisation ou de recommandation de produits culturels en utilisant les notations des abonnés ou des acheteurs ayant utilisé le service.

³⁷ BFM Business, « Comment Amazon, Google et Netflix attaquent les offres TV des câblo-opérateurs? », 16 avril 2017. <http://bfmbusiness.bfmtv.com/entreprise/comment-amazon-google-et-netflix-attaquent-les-offres-tv-des-cablo-operateurs-1143774.html>

³⁸ Bourgault-Côté, G. (2017). Le pouvoir du code : Les algorithmes, une partition pleine de distorsions. » Le Devoir, 18 février 2017, <https://www.ledevoir.com/culture/musique/492020/musique-les-algorithmes-une-partition-pleine-de-distorsions> (Voir aussi : Tarvic, C. (2018). « Les algorithmes de recommandation musicale : ne bouge pas, on s'occupe de tout », <https://maze.fr/musique/05/2018/les-algorithmes-de-recommandation-musicale-ne-bougez-pas-on-soccupe-de-tout/>)

La plateforme Netflix a ainsi forgé son succès et sa réputation sur la précision de ses algorithmes³⁹ de recommandation personnalisée. Combinant l'analyse des données d'usage et de comportement à celle des métriques liées aux contenus, Netflix sait pratiquement tout de ses utilisateurs, notamment ce qu'ils regardent, la manière dont ils le regardent (via quel appareil ou terminal) et même l'intensité de ce qu'ils regardent.

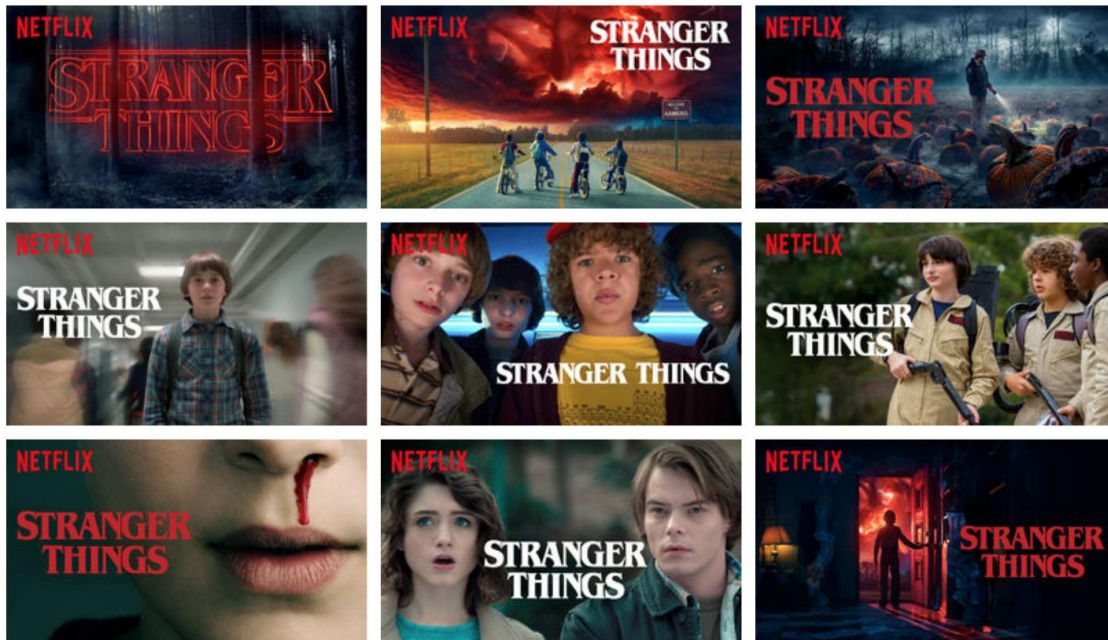
Ces données basées sur des analyses comportementales ont permis à Netflix d'identifier à l'échelle planétaire 1 300 communautés qui partagent les mêmes préférences, quelle que soit leur nationalité et de cibler encore plus finement les besoins de ses abonnés, en améliorant la personnalisation de l'expérience-utilisateur. Cette personnalisation permise par une catégorisation algorithmique entraîne une certaine pré-configuration des usages, sachant que les algorithmes de Netflix ainsi que les éléments de l'interface du service sont de plus en plus améliorés pour non seulement maintenir l'utilisateur captif sur la plateforme, mais aussi l'orienter vers des contenus pré-déterminés⁴⁰.

L'une des innovations récentes dans ce sens est la personnalisation des vignettes ou images d'illustration qui sont différentes en fonction des goûts de chaque utilisateur. (Voir sur la figure 4 de la page suivante, l'exemple de la série devenue culte, *Strangers Things*).

³⁹ « Il y a le *Personalized Video Ranker* (PVR, pour classement vidéo personnalisé) qui commande pour chaque profil un classement personnalisé selon les catégories... et la popularité des contenus (qui, elle, n'est pas personnalisée). Un autre algorithme qui détermine la recommandation et la page d'accueil du service est le *Top-N Video Ranker*, qui vise à fournir les meilleures recommandations pour chaque profil en se concentrant uniquement les vidéos les mieux classées, valorisant une « tête » de catalogue, c'est-à-dire d'une certaine manière les plus vues. Se combine aussi un autre algorithme de « tendances récentes », car les tendances à court terme sont un autre prédicteur puissant du visionnage. C'est un algorithme qui permet de faire entrer de la considération sociale dans les recommandations, liée à l'actualité par exemple. Un autre algorithme est le *Continue watching* qui consiste à recommander du contenu selon ce qui a déjà été visionné en prenant en compte le temps écoulé depuis le visionnage d'un épisode ou le point d'abandon des séries...L'algorithme « de similitude vidéo à vidéo », lui, active une recommandation du type : « parce que vous avez regardé » qui fait une recommandation depuis une des vidéos visionnées par un utilisateur. Celui-ci est un algorithme non personnalisé qui pointe vers des vidéos semblables calculées pour chaque vidéo du catalogue. Un autre algorithme encore assure la diversité de la page d'accueil et de la recommandation.» Guillaud, H. (2017). « Tout est recommandation » : comment Netflix s'est transformé », *Internet Actus*, 25 octobre 2017 <http://www.internetactus.net/2017/10/25/tout-est-recommandation-comment-netflix-sest-transforme/>

⁴⁰ Silva Mota Drumond, G. (2016). « La configuration des usages sur Netflix : le système de recommandation Cinematch et la représentation de l'utilisateur » Mémoire. Montréal (Québec, Canada), Université du Québec à Montréal, Maîtrise en communication. <https://archipel.uqam.ca/9253/>

Figure 4 : Les différentes vignettes pour la série *Stranger Things*. © Netflix



La performance des algorithmes développés par les plateformes numériques et la pertinence des résultats qu'ils proposent sont plus optimales si la plateforme dispose d'un grand nombre d'utilisateurs⁴¹. Ceci constitue d'ailleurs le fondement de la « loi » dite de Metcalf qui démontre que la valeur d'un réseau est proportionnelle au nombre de ses participants. En d'autres termes, plus le nombre d'utilisateurs d'une plateforme augmente, plus celle-ci suscite l'attrait et plus elle prend de la valeur tout en proposant des services de recommandation sur la base de l'analyse de données agrégées de sa masse d'utilisateurs.

« Un service de partage de vidéos comme YouTube préfère, à une intervention préalable sur les contenus, une recommandation contributive fondée sur l'agrégation des jugements des utilisateurs (likes, commentaires, etc.). [...] Dès lors, la puissance du modèle économique de la plateforme n'est pas liée à la seule force technique de son algorithme, mais à l'intensité et à la qualité des données collectées sur les utilisateurs eux-mêmes. La puissance de la recommandation personnalisée repose donc essentiellement sur les comportements de l'ensemble des utilisateurs scrutés en permanence par des opérateurs »⁴².

⁴¹ Voir : Colin, N.; Verdier, H. (2015). *L'âge de la multitude – Entreprendre et gouverner après la révolution numérique*, Armand Colin, 288 p.

⁴² Farchy, J. ; Méadel, C.; Anciaux, A. (2017). « Une question de comportement. Recommandation des contenus audiovisuels et transformations numériques », *tic&société* [Online], Vol. 10, N° 2-3 | 2ème semestre 2016 - 1er semestre 2017, p.181 <https://journals.openedition.org/ticetsociete/2136>

On considère que les plateformes numériques bénéficient d'externalités positives des réseaux liés à leur fonctionnement en écosystème. Elles disposent en effet d'un avantage concurrentiel important lié aux processus d'amélioration continue de la qualité et de l'utilité des services et produits proposés; ces processus étant induits par des effets de réseau direct déclenchés par l'accroissement même du nombre d'abonnés ou d'utilisateurs. En ce qui concerne les plateformes spécialisées dans l'intermédiation et proposant des services sur des marchés bifaces⁴³ ou à deux versants, elles fonctionnent plutôt sur le principe d'effets de réseaux indirects. Prenant l'exemple des plateformes de covoiturage ou d'une application comme Uber et celles de réservation de chambres d'hôtel (TripAdvisor, Booking), les effets de réseaux indirects se caractérisent par le fait que le bénéfice ou l'utilité que tirera une catégorie d'utilisateurs (clients/abonnés) des services de ces plateformes (exemples de personnes cherchant à se déplacer ou à être hébergées) dépend du nombre d'utilisateurs de l'autre catégorie (chauffeurs, hôtels) se situant du côté de l'offre et pouvant combler le besoin de la première catégorie d'utilisateurs⁴⁴.

Par ailleurs, des entreprises comme Apple tirent profit de nombreux effets de réseau indirects liés aux effets croisés de l'utilité de leur offre variée de produits et de services complémentaires qu'elle propose (applications mobiles et accessoires de tout genre comme les enceintes, les systèmes d'amplification, les casques audio, les housses, etc.). Concrètement, Apple met en place une stratégie de maîtrise de son écosystème de services et de produits en contrôlant et en interreliant ses équipements (iPhone, iPad,...) avec son système d'exploitation (iOS) et son magasin d'applications (App Store). Google a également un modèle de services en ligne qui repose sur l'ouverture, la compatibilité et la mise à disposition de son système d'exploitation (Android) à des fabricants de terminaux dans le but de pouvoir rendre à terme son moteur de recherche et ses produits incontournables et accessibles via tous les terminaux⁴⁵.

⁴³ Parent, R.; Chanal, V. (2009). Quels business models pour les plateformes Web 2.0. : les apports de la théorie des marchés bi-faces.. Conférence de l'AIMS, Jun 2009, Grenoble, France. p.1-27, 2009. <https://halshs.archives-ouvertes.fr/halshs-00455382/document>

⁴⁴ Voir Rochet, J.-Ch. et Tirole J. (2003). « Platform Competition in two-sided Markets », *Journal of the European Economic Association*, June 2003 1(4):990-1029 ; Perrot, A. (2018). « Données et règles de concurrence », Dans « *Enjeux numériques* », *Annales des Mines*, n°2, juin 2018, p. 35-37; ou encore Armstrong M. (2006). "Competition in Two-Sided Markets", *The RAND Journal of Economics*, vol. 37, n° 3 (Autumn, 2006), pp. 668-691.

⁴⁵ Voir à ce sujet le dernier rapport de l'ARCEP : ARCEP (2018), « Smartphones, tablettes, assistants vocaux...Les terminaux, maillon faible de l'ouverture d'Internet : Rapports sur leurs limites et actions à envisager ». https://www.arcep.fr/uploads/tx_gspublication/rapport-terminaux-fev2018.pdf

Ces logiques et ces nouveaux modèles d'affaires systémiques autour de l'exploitation des données sont à l'origine de phénomènes de concentration du marché et d'importants problèmes d'entorses aux règles de concurrence. D'abord, l'accumulation des données, surtout lorsqu'elles sont agrégées et bien structurées, par les plateformes numériques peut en effet entraîner des comportements anticoncurrentiels⁴⁶ et des abus liés à la position dominante engendrée par la détention exclusivement par une seule plateforme des données d'une multitude d'utilisateurs; ce qui confère à cette dernière un avantage concurrentiel exclusif par rapport aux autres. Il existe également des freins à la portabilité des données, notamment lorsque les utilisateurs et leurs données sont « verrouillés » dans les écosystèmes des grandes plateformes et des géants du numérique, ou dans les logiques propriétaires, fermés ou non interopérables de certains terminaux, équipements (tablettes, smartphones), applications ou systèmes d'exploitation.

L'effet de verrouillage se caractérise par des situations où les coûts de changement de technologie sont si élevés que l'utilisateur décide de ne pas changer le produit ou le service. Dans le cas des données, cette situation se produit par exemple lorsque l'utilisateur recourt fréquemment aux services d'une plateforme, d'une application, ou d'une technologie sur laquelle il a déjà stocké une quantité importante de données (musique, vidéos, photos, textos,...), avec un historique d'utilisation ou de navigation permettant de retracer des informations et données (identifiants, mots de passe, coordonnées bancaires,...) liées à certaines transactions récurrentes et qu'il se retrouve à ne plus vouloir changer de plateforme, de produit ou un équipement, comme son smartphone, à cause des coûts (*switching costs*) que pourrait engendrer un tel changement. Ces coûts peuvent être évalués aussi en termes de perte partielle ou totale de contrôle sur ses données et du temps qu'il devra prendre pour les reconstituer. Un autre problème concurrentiel est celui induit par les biais discriminatoires des algorithmes, notamment en ce qui a trait aux pratiques de différenciation de prix⁴⁷ en fonction des profils des utilisateurs.

⁴⁶ Google a notamment été accusé à plusieurs reprises au cours de ces dernières années d'abus de position dominante et fut sanctionné par la commission européenne pour avoir avantage, au détriment des autres comparateurs de prix, les services de son propre comparateur Google Shopping.

⁴⁷ « Selon le site de journalisme ProPublica, plusieurs grandes compagnies, dont Amazon, utilisent des algorithmes masquant le prix réel de leurs articles afin d'avantager leurs propres produits lors de recherches faites en ligne par les consommateurs. Des grandes chaînes disposent aussi d'algorithmes pour moduler leurs prix en fonction du code postal des consommateurs. Cela permet d'augmenter le prix dans les secteurs dépourvus de commerces similaires et de le réduire là où les consommateurs pourraient être tentés de se tourner vers la concurrence. » Paré, I. (2017). «Le pouvoir du code : Quelques algorithmes connus », *Le Devoir*, 18 février 2017, <https://www.ledevoir.com/societe/science/492016/quelques-algorithmes-connus>

4 Pistes de recommandations pour un encadrement de l'utilisation des données

Les constats, enjeux et tendances analysés dans les sections précédentes dressent un état des lieux sommaire et actuel des processus et dynamiques d'usage et d'exploitation des données à l'ère des plateformes et des nouveaux médias numériques, en mettant en exergue leurs implications pour les utilisateurs et les consommateurs en ligne de produits culturels canadiens et québécois.

il nous paraît nécessaire de conclure nos réflexions en identifiant des pistes de recommandations pour un meilleur encadrement de l'utilisation des données. Pour ce faire, nous ciblons quatre chantiers dont il faudrait tenir prioritairement compte eu égard au fait que la législation canadienne en matière de radiodiffusion et de télécommunications accuse un retard important sur le rythme d'adoption et de diffusion des innovations numériques. Ceci constitue d'ailleurs la cause de situations diverses et de manifestations concrètes d'iniquité, de concurrence déloyale, d'asymétrie réglementaire et même de menaces à la diversité des expressions culturelles au Québec et au Canada. Ces situations risquent de s'aggraver et de fragiliser davantage notre écosystème culturel, notamment si aucune mesure ou disposition appropriée n'est prise dans des délais raisonnables. Soulignons également qu'en matière de gestion de l'information et de données, il existe une double asymétrie technologique et cognitive. D'une part, nos capacités de gestion de l'information ne sont pas encore adaptées au numérique (sous-investissement en structuration des données) et d'autre part, nos systèmes d'information (y compris dans les administrations publiques) n'ont pas évolué depuis l'informatique, avec l'existence de silos freinant toute utilisation innovante de données descriptives et d'usage. Il est donc urgent de prendre les moyens nécessaires pour réduire cette asymétrie tout en modernisant les instruments législatifs et réglementaires qui, dans leur état actuel, permettent à de nouveaux types d'entreprises d'échapper à leur application.

Les quatre chantiers identifiés sont les suivants : 1) la Souveraineté numérique et la responsabilité en matière d'accès et de collecte des données d'intérêt public ; 2) la concurrence, l'innovation et l'accès aux services ; 3) la neutralité d'Internet et 4) la découvrabilité du contenu canadien et la promotion de la diversité des expressions culturelles.

Ces chantiers constituent des points de vigilance et des enjeux par rapport aux menaces que font peser les plateformes numériques dominantes sur la culture et sur les médias nationaux, de par leurs activités de captation et de contrôle des flux de données, de par le déséquilibre qu'elles occasionnent sur les marchés et les modèles d'affaires de nos industries culturelles, ainsi que de par leur intrusion dans la vie privée des consommateurs.

Au regard de leurs effets possibles, les 15 pistes de recommandation formulées ci-après doivent être perçues comme des principes essentiels, des propositions mesurées et souhaitables et des actions envisageables, s'inspirant de ce qui se fait dans d'autres pays (France, Allemagne, Royaume-Uni, Australie,...). Elles ne doivent aucunement être envisagées comme des solutions ultimes avec une prétention quelconque de déterminer les modes de gestion, régulation et valorisation des données personnelles. Ces recommandations visent, avant toute chose, à alimenter la réflexion dans le cadre du processus d'examen des lois en matière de radiodiffusion et de télécommunications canadiennes et des questions soulevées dans l'appel aux observations et dans le cadre de référence⁴⁸ sur lequel s'appuie le mandat du groupe d'experts présidé par Mme Yale.

4.1 Souveraineté numérique et responsabilité en matière d'accès et de collecte des données d'intérêt public

« Parce qu'il est important et urgent que l'essentiel de ces interactions numériques, de nos interactions numériques, reviennent dans l'espace public. Qu'elles y soient re-situées pour pouvoir mieux y être restituées. Et que s'y appliquent, aussi simplement qu'essentiellement et exclusivement, les seules lois régulant l'espace public de la démocratie ». Olivier Ertzscheid, maître de conférences en sciences de l'information

- 1) Adopter sur la base d'une régulation contraignante** des règlements obligeant des entreprises de radiodiffusion (nationales et étrangères) à fournir au CRTC ou à Statistique Canada des renseignements non seulement sur leurs émissions, leur situation financière, la conduite et la direction de leurs affaires, mais aussi sur les données descriptives et d'usage relatifs à la consommation du contenu canadien.

⁴⁸[https://www.ic.gc.ca/eic/site/110.nsf/vwapj/terms_of_reference_FR.pdf/\\$FILE/terms_of_reference_FR.pdf](https://www.ic.gc.ca/eic/site/110.nsf/vwapj/terms_of_reference_FR.pdf/$FILE/terms_of_reference_FR.pdf)

- 2) **Rendre publics par souci de transparence** (et dans le respect de la confidentialité accordée uniformément à toutes les entreprises de radiodiffusion) des rapports sur les informations transmises au sujet des comportements en ligne, en particulier les habitudes de consommation culturelle des clients et des abonnés canadiens (tout en respectant la vie privée de ceux-ci).
- 3) **Créer un cadre pour faciliter l'échange de l'information et de l'expertise relatives à la collecte des données** (incluant celle des agences statistiques gouvernementales) sur les pratiques culturelles numériques des Canadiens et la mise à disposition de ces données à des fins de recherche scientifique ou d'amélioration des politiques publiques.
- 4) **Prévoir une disposition législative obligeant la transparence et la loyauté⁴⁹** des fournisseurs de contenus (incluant les plateformes) et des fabricants d'équipements terminaux envers les utilisateurs finaux par rapport aux types de données qu'ils collectent et l'utilisation qui en est faite.

4.2 Concurrence, innovation et accès aux services

- 5) **Renouveler et mieux adapter les règles de concurrence actuelles** pour qu'elles soient capables de traiter les problèmes complexes liés aux données et aux algorithmes.

⁴⁹ On entend ici par loyauté, la nécessité pour les plateformes de ne pas faire autre usage (détourné à des fins de marketing ou publicitaires) des données des utilisateurs en dehors des usages dont l'utilisateur aurait été initialement ou préalablement informé par la plateforme. Voir quelques exemples de décrets adoptés en France en la matière :

Décret n° 2017-1434 du 29 septembre 2017 relatif aux obligations d'information des opérateurs de plateformes numériques :

<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000035720908&dateTexte=&categorieLien=id>

Décret n° 2017-1435 du 29 septembre 2017 relatif à la fixation d'un seuil de connexions à partir duquel les opérateurs de plateformes en ligne élaborent et diffusent des bonnes pratiques pour renforcer la loyauté, la clarté et la transparence des informations transmises aux consommateurs :

<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000035720925&dateTexte=&categorieLien=id>

Décret n° 2017-1436 du 29 septembre 2017 relatif aux obligations d'information relatives aux avis en ligne de consommateurs :

<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000035720935&dateTexte=&categorieLien=id>

- 6) **Réduire l'asymétrie d'information et rééquilibrer les relations** entre les plateformes et les professionnels de la culture et des médias dont l'activité repose sur l'utilisation de ces plateformes (incluant les éditeurs et développeurs de contenus, les gestionnaires de communautés et de stratégies d'audience, etc.).
- 7) **Mettre en place ou accroître des mesures permettant la « portabilité » des données** sur le modèle de la portabilité des numéros de téléphone destinée à faciliter le changement d'opérateur ou de plateforme et à fluidifier la concurrence, avec la possibilité pour les consommateurs d'apporter avec eux les données accumulées sur une plateforme.
- 8) **Prévoir des mesures législatives ou réglementaires spécifiques aux plateformes** ou à tout le moins une adaptation des règles de concurrence aux particularités de l'économie du partage et des marchés numériques bifaces, tout en tenant compte de la taille des plateformes, du caractère immatériel de leur activité, de l'existence d'effets de réseau tendant à la constitution de positions dominantes, de la détention de volumes considérables de données, ainsi que de tous les autres facteurs susceptibles de créer un environnement favorable à des pratiques et des comportements anticoncurrentiels. À ce titre, en Allemagne, une commission sur les monopoles qui s'est penchée sur les risques concurrentiels inhérents à la concentration des données, recommande de renforcer et moderniser les instruments législatifs et réglementaires existants plutôt que d'édicter de nouvelles dispositions⁵⁰.

4.3 Neutralité d'Internet

- 9) **Veiller à ce que les fournisseurs d'accès Internet acheminent les données** des diffuseurs de contenus vers les utilisateurs finaux, sans qu'ils en examinent le contenu, sans égard à la source et à la destination des données.

⁵⁰ German Monopolies Commission (Monopolkommission), Special Report No. 68: Competition policy: The challenge of digital markets (2015), http://www.monopolkommission.de/images/PDF/SG/s68_fulltext_eng.pdf

- 10) **S'assurer que dans un objectif d'intérêt général, les applications, normes et standards** liés à l'évolution, la croissance et à l'interopérabilité des infrastructures, réseaux et technologies d'Internet demeurent ouverts et sécurisés, respectent la liberté d'expression ainsi que le principe de l'innovation sans autorisation.
- 11) **Responsabiliser et engager de manière contraignante toutes les principales parties prenantes de l'Internet à respecter la prise en compte de l'éthique** dans le déploiement des technologies émergentes de l'Internet (l'Internet des objets, la 5G⁵¹, l'intelligence artificielle, la blockchain, l'infonuagique...) ainsi que dans le développement de nouveaux services et usages innovants favorisant la consommation de données toujours plus volumineuses.
- 12) **Soutenir le déploiement des infrastructures de points d'échange Internet et la localisation des serveurs et infrastructures de stockage de données au Canada** dans le but de favoriser le développement du trafic local et de limiter ainsi les transits via les Etats-Unis ou l'Europe ; et ce afin de garantir que les données de la population canadienne soient conservées au Canada, dans le respect des exigences de confidentialité prescrites par les normes, règles et lois d'ici.

4.4 Découvrabilité du contenu canadien et promotion de la diversité des expressions culturelles

« Pour utile qu'il soit, l'algorithme présente certains risques : outre bien sûr le profilage, l'enfermement dans ses propres goûts au détriment de la découverte, la tentation de manipulations destinées à favoriser les œuvres produites par l'éditeur, voire, perspective particulièrement inquiétante, aujourd'hui encore utopique, la menace d'une standardisation de la création qui serait dictée par les attentes du public. Ces risques pour la diversité culturelle peuvent toutefois être atténués si les algorithmes ne sont pas exclusivement centrés sur les données fournies par les consommateurs, mais inspirés par les caractéristiques des œuvres proposées, par un travail sur les

⁵¹ « La 5G ouvre déjà de nouvelles perspectives. En 2021, une connexion 5G générera près de cinq fois plus de trafic qu'une connexion 4G moyenne. En effet, la 5G décuplera les débits des réseaux (de l'ordre de 10 Gbit/s, voire 20 Gbit/s), raccourcira les temps de latence (sous la milliseconde) assurant une connectivité à tous les objets communicants, des téléphones intelligents (*smartphones*) aux voitures en passant par les robots industriels ou les capteurs de santé.» Extrait de : OIF (2018). Rapport 2018 sur l'état de la Francophonie numérique, p.19, <https://www.francophonie.org/IMG/pdf/rapport-2018-etat-francophonie-numerique.pdf>

*métadonnées associées à ces œuvres*⁵²». Allocution d'**Olivier Schrameck**, Président du Conseil Supérieur de l'Audiovisuel (CSA), au Forum de Tokyo. 2 décembre 2014.

- 13) **Imposer aux plateformes numériques la transparence des critères** utilisés pour le référencement, le classement, le traitement, la hiérarchisation et la recommandation du contenu canadien et les contraindre à intégrer dans ces critères des éléments favorables à la diversité et à la découvrabilité des contenus locaux/nationaux.

- 14) **Prendre des mesures afin de contraindre les plateformes et les entreprises de radiodiffusion numériques étrangères** à respecter des exigences minimales de recommandation et de mise en avant dans leur catalogue d'un quota minimum d'œuvres canadiennes et francophones.

- 15) **Accroître les mesures de normalisation et d'harmonisation des pratiques de référencement** des œuvres canadiennes (notamment les œuvres patrimoniales, les œuvres rares ou anciennes) afin de renseigner de façon adaptée et standardisée leur fiche numérique descriptive grâce à des métadonnées enrichies, exhaustives et de bonne qualité.

⁵² <https://www.csa.fr/Informer/Espace-presse/Interventions-publiques/Allocution-d-Olivier-Schrameck-au-Forum-de-Tokyo.-le-2-decembre-2014>