

**GOUVERNANCE DES DONNÉES D'USAGE
DANS LES BIBLIOTHÈQUES ET ARCHIVES :**

SONDAGE SUR LES PERCEPTIONS
DES USAGÈRES ET USAGERS DE BANQ

RAPPORT DE RECHERCHE

INRS – Urbanisation Culture Société
Chaire Fernand-Dumont sur la culture

Octobre 2023

Équipe de recherche

Chercheure principale
Nathalie Casemajor, INRS

Co-chercheur
Guillaume Sirois, Université de Montréal

Collaborateur
Guy Bellavance, INRS

Professionnelle de recherche statisticienne
Nathalie Vachon, INRS

Étudiantes et étudiants, postdoctorant
François R. Derbas Thibodeau
Alexandra Beauséjour-Villemare
Alexis Morin

Équipe de BAnQ

Nicole Goyer, Analyste d'affaires Web
Lucile Niort, Analyste en intelligence d'affaires
Louise Brunet, Agente de développement numérique

Avec la contribution de :
Jean-François Gauvin, Directeur de la stratégie numérique ; Viriya Tach, Chargée de mission - développement du numérique ; Marie-Michèle Hamel, Agente de développement numérique ; Élodie Zini, Chargée de pratiques UX-CX.

Montréal : INRS, 2023.

ISBN 978-2-89575-462-6

SOMMAIRE

1. INTRODUCTION	1
2. OBJECTIFS DE L'ENQUÊTE	1
3. ENJEUX	2
3.1. LA VALORISATION DES DONNÉES D'USAGE : UNE PRATIQUE EN EXPANSION	2
INTELLIGENCE D'AFFAIRE ET MUTUALISATION DES DONNÉES	2
ÉTAT DES LIEUX DANS LE SECTEUR DES BIBLIOTHÈQUES ET ARCHIVES	3
ÉVOLUTION DE LA RÉGLEMENTATION	5
3.2. QUELLES DONNÉES POUR QUELS NOUVEAUX SERVICES ?	6
LES DONNÉES COLLECTÉES PAR BANQ	6
L'OPINION DES USAGÈRES ET USAGERS	11
3.3. LA LITTÉRATIE DES DONNÉES PERSONNELLES	12
DÉFINITION	12
CADRE DE COMPÉTENCES	13
LES FOSSÉS DE LA VIE PRIVÉE	15
FACTEURS SOCIODÉMOGRAPHIQUES	16
3.4. INCIDENCES SUR LA VIE PRIVÉE DES ABONNÉS	21
LA VIE PRIVÉE DANS LES ENVIRONNEMENTS NUMÉRIQUES	21
DONNÉES PERSONNELLES ET DONNÉES SENSIBLES : DES INFORMATIONS À RISQUE	22
PERCEPTIONS ET ATTITUDES RELATIVES À LA GESTION DES DONNÉES PERSONNELLES	23
FACTEURS SOCIODÉMOGRAPHIQUES	26
3.5. LES CADRES ÉTHIQUES DE GOUVERNANCE DES DONNÉES D'USAGE	30
PRINCIPES RELATIFS À LA PROTECTION DE LA VIE PRIVÉE	30
SONDER LES USAGÈRES ET USAGERS POUR CONNAÎTRE LEUR POINT DE VUE	33
4. PROBLÉMATIQUE	34
4.1. QUESTIONS DE RECHERCHE	34
4.2. CONSTRUCTION DES DIMENSIONS D'ANALYSE	35
5. DÉMARCHE MÉTHODOLOGIQUE	41
5.1. POPULATION CIBLE	41
DESCRIPTION	41
ÉTUDES DE RÉFÉRENCE	41
5.2. PARAMÈTRES DU SONDAGE	43
QUESTIONNAIRE	43
PRÉ-TEST	43

RECRUTEMENT ET ADMINISTRATION DU SONDAGE	43
TRAITEMENT ET ANALYSE DES DONNÉES	44
5.3. ÉCHANTILLON	45
TAILLE ET VARIABLES DE CONTRÔLE	45
VARIABLES D'ANALYSE	45
TEST DE REDRESSEMENT DE L'ÉCHANTILLON	54
6. RÉSULTATS DÉTAILLÉS	55
6.1. LITTÉRATIE DES DONNÉES PERSONNELLES : MESURE DE SCORE ET DE NIVEAUX	55
MODE DE COLLECTE ET D'ANALYSE	55
NIVEAUX	57
SCORE PAR SOUS-DIMENSION	58
6.2. LITTÉRATIE DES DONNÉES PERSONNELLES : DÉTERMINANTS SOCIODÉMOGRAPHIQUES	65
MODÈLE D'ANALYSE	65
INFLUENCE SUR LE SCORE GLOBAL	65
INFLUENCE PAR SOUS-DIMENSION	67
6.3. INTÉRÊT POUR DE NOUVEAUX SERVICES : MESURE DE SCORE ET DE NIVEAUX	70
MODÈLE D'ANALYSE	70
SCORE GLOBAL	72
SCORE PAR SOUS-DIMENSION	74
6.4. INTÉRÊT POUR DE NOUVEAUX SERVICES : DÉTERMINANTS SOCIODÉMOGRAPHIQUES	78
MODÈLE D'ANALYSE	78
INFLUENCE DES VARIABLES SOCIODÉMOGRAPHIQUES PAR SOUS-DIMENSION	78
INFLUENCE DE LA LITTÉRATIE DES DONNÉES PERSONNELLES PAR SOUS-DIMENSION	79
6.5. PERCEPTION DE LA GESTION DES DONNÉES PERSONNELLES : MESURE DE SCORE ET TYPOLOGIE D'ATTITUDES	82
MODE DE COLLECTE ET D'ANALYSE	82
SCORE GLOBAL	84
ATTITUDES	85
SCORE PAR SOUS-DIMENSION	88
6.6. PERCEPTION AU REGARD DE LA PROTECTION DE LA VIE PRIVÉE : DÉTERMINANTS SOCIODÉMOGRAPHIQUES	96
MODÈLE D'ANALYSE	96
INFLUENCE DES VARIABLES SOCIODÉMOGRAPHIQUES PAR SOUS-DIMENSION	96
INFLUENCE DE LA LITTÉRATIE DES DONNÉES PERSONNELLES PAR SOUS-DIMENSION	97
6.7. RELATION ENTRE LITTÉRATIE, INTÉRÊT ET ATTITUDE	101
MODÈLE D'ANALYSE	101
LITTÉRATIE / INTÉRÊT	101
LITTÉRATIE / ATTITUDE	101
ATTITUDE / INTÉRÊT	102

7. SYNTHÈSE ET DISCUSSION DES RÉSULTATS	105
7.1. LITTÉRATIE DES DONNÉES PERSONNELLES	105
(QR 1A) UN NIVEAU PLUTÔT ÉLEVÉ, MAIS DES LACUNES IMPORTANTES	105
(QR 1B) UN NIVEAU QUI VARIE SURTOUT EN FONCTION DE L'ÂGE ET LA SCOLARISATION	106
7.2. INTÉRÊT POUR LES NOUVEAUX SERVICES FONDÉS SUR LES DONNÉES D'USAGE	108
(QR 2A) L'INTÉRÊT EST FORT POUR LA PLUPART DES SERVICES PROJETÉS	108
(QR 2B) CERTAINS ASPECTS DE L'INTÉRÊT SONT INFLUENCÉS PAR LA SCOLARISATION ET LE FAIT D'APPARTENIR À UNE MINORITÉ VISIBLE	109
(QR 2C) UNE LITTÉRATIE DES DONNÉES PERSONNELLES ÉLEVÉE FAIT BAISSER L'INTÉRÊT POUR LA MAJORITÉ DES NOUVEAUX SERVICES	111
7.3. PERCEPTION DE LA GESTION DES DONNÉES PERSONNELLES	112
(QR 3A) DES ATTITUDES TRÈS PARTAGÉES ENTRE PROTECTION ET OUVERTURE	112
(QR 3B) CERTAINS ASPECTS DE LA PERCEPTION SONT INFLUENCÉS PAR LA SCOLARISATION, L'ÂGE ET L'AFFILIATION AU GROUPE LGBTQ+	114
(QR 3C) UNE LITTÉRATIE DES DONNÉES PERSONNELLES ÉLEVÉE REND PLUS PROTECTEUR QUANT À LA GESTION DE SES DONNÉES PERSONNELLES	117
7.4. RELATION ENTRE LITTÉRATIE, INTÉRÊT ET ATTITUDE	118
(QR4) LES PERSONNES ABONNÉES QUI ONT PLUS DE LITTÉRATIE SONT PLUS PROTECTRICES ET MOINS INTÉRESSÉES PAR LES NOUVEAUX SERVICES	118
7.5. MODES D'ENCADREMENT DES DONNÉES D'USAGE IDENTIFIÉS PAR LES PERSONNES SONDÉES	119
8. PISTES D'ACTION ET DE RÉFLEXION	120
8.1. LITTÉRATIE DES DONNÉES PERSONNELLES	120
8.2. NOUVEAUX SERVICES FONDÉS SUR LES DONNÉES D'USAGE	121
8.3. PROTECTION DE LA VIE PRIVÉE	122
9. CONCLUSION ET NOUVELLES PISTES DE RECHERCHE	124
ANNEXES	126
1. QUESTIONNAIRE	126
2. TEST DE REDRESSEMENT DE L'ÉCHANTILLON	135
CONTRAINTES LORS DE LA DIFFUSION DU SONDAGE	137
DONNÉES DE RÉFÉRENCE SUR LA POPULATION CIBLE : LACUNES	137
3. LITTÉRATIE DES DONNÉES PERSONNELLES	140
4. INTÉRÊT POUR LE DÉVELOPPEMENT DE NOUVEAUX SERVICES	142
5. PERCEPTION RELATIVE À LA GESTION DES DONNÉES PERSONNELLES	143
BIBLIOGRAPHIE	144

LISTE DES TABLEAUX

Tableau 1 : Types de données recueillies par BAnQ à propos de ses usagères et usagers	7
Tableau 2 : Cadre de compétences en littératie des données personnelles retenu dans l'étude	14
Tableau 3 : Éléments analysés par dimension (1 à 3) et sous-dimension (A à L)	37
Tableau 4 : Dimension des caractéristiques sociodémographiques : variables primaires et choix de réponses possibles	38
Tableau 5 : Dimension des caractéristiques sociodémographiques : variables secondaires et choix de réponses possibles	39
Tableau 6 : Variable à fin de contrôle	40
Tableau 7 : Répartition par âge des personnes abonnées actives pour la période août-septembre 2022 (avec non-réponses)	42
Tableau 8 : Composition de l'échantillon selon l'âge	46
Tableau 9 : Composition de l'échantillon selon le genre	47
Tableau 10 : Composition de l'échantillon selon le niveau de diplomation	48
Tableau 11 : Composition de l'échantillon selon l'occupation	50
Tableau 12 : Composition de l'échantillon selon le secteur d'emploi	51
Tableau 13 : Composition de l'échantillon selon l'appartenance à une minorité visible	52
Tableau 14 : Composition de l'échantillon selon l'affiliation au groupe LGBTQ+	53
Tableau 15 : Littératie des données personnelles : questions du formulaire et mesures de score	56
Tableau 16 : Fréquences et pourcentage des réponses pour le score total de littératie des données personnelles	58
Tableau 17 : Identifier les données personnelles (A) : fréquence des réponses*	60
Tableau 18 : Connaissance de la nature et de l'étendue des données collectées (B) : fréquence des réponses	61
Tableau 19 : Données sensibles (C1) : fréquence des réponses	63
Tableau 20 : Anonymisation (C2) : fréquence des réponses	63
Tableau 21 : Rôle des tiers (C3) : fréquence des réponses	63
Tableau 22 : Réflexion critique sur la gouvernance des données personnelles (D) : distribution des réponses	64
Tableau 23 : Intérêt pour les nouveaux services : questions du formulaire et mesures de score	71
Tableau 24 : Distribution des réponses par niveau d'intérêt	73
Tableau 25 : Enrichir ses collections avec de nouveaux documents en fonction de la demande du public (sous-dimension E) : fréquence des réponses	76
Tableau 26 : Recommander des documents en fonction des habitudes et préférences (sous-dimension F) : fréquence des réponses	76
Tableau 27 : Recommander des activités culturelles en fonction des habitudes et préférences (sous-dimension G) : fréquence des réponses	77

Tableau 28 : Personnaliser l'expérience numérique sur le site Web de BAnQ (sous-dimension H) : fréquence des réponses	77
Tableau 29 : Perception relative à la protection des données personnelles : questions du formulaire et mesures de score	83
Tableau 30 : Perception relative à la gestion des données personnelles (sous-dimensions J et L) : distribution par attitude (détail des fréquences)	86
Tableau 31 : Perception relative à la gestion des données personnelles : distribution par attitude (sous- dimensions J et L) et croisement avec sous-dimensions K	87
Tableau 32 : Collecter seulement les données personnelles utiles aux services actuels (sous-dimension I) : fréquence des réponses	91
Tableau 33 : Recouper différentes données pour personnaliser les services (sous-dimension J) : fréquence des réponses	91
Tableau 34 : Conservation des données personnelles (sous-dimension K) : fréquence des réponses	92
Tableau 35 : Partager ses données anonymisées avec un ministère (sous-dimension L1) : fréquence des réponses	93
Tableau 36 : Partager ses données anonymisées avec une maison d'édition (sous-dimension L2) : fréquence des réponses	94
Tableau 37 : Partager ses données anonymisées avec une OBNL (sous-dimension L3) : fréquence des réponses	94
Tableau 38 : Partager ses données anonymisées avec des équipes de recherche universitaires (sous- dimension L4) : fréquence des réponses	95
Tableau 39 : Niveau d'intérêt pour le développement de nouveaux services, selon le niveau de littératie des données personnelles	102
Tableau 40 : Niveau de littératie des données personnelles, selon l'attitude relative à la gestion des données personnelles (fréquence et pourcentage)	103
Tableau 41 : Niveau d'intérêt pour le développement de nouveaux services, selon l'attitude relative à la gestion des données personnelles (fréquence et pourcentage)	104
Tableau 42 : Composition de l'échantillon comparé à la population de référence	136
Tableau 43 : Données de référence sur la population des abonnés (lacunes)	139
Tableau 44 - Identifier les données personnelles : score de la sous-dimension	140
Tableau 45 - Connaître la nature et l'étendue des données collectées : score de la sous-dimension	140
Tableau 46 - Comprendre les enjeux : score de la sous-dimension	141
Tableau 47 - Réfléchir à la gouvernance : score de la sous-dimension	141
Tableau 48 : Fréquences et pourcentage des réponses pour le score total de d'intérêt pour les nouveaux services	142
Tableau 49 : Fréquences et pourcentage des réponses pour le score total de perception relative à la gestion des données personnelles	143

LISTE DES FIGURES

Figure 1 : Distribution du score total de littératie des données personnelles par niveau (bas, moyen, élevé, très élevé)	57
Figure 2 : Littératie des données personnelles : distribution des fréquences par sous-dimension d'analyse	59
Figure 3 : Influence de l'âge et du niveau de scolarisation sur le score global de littératie des données personnelles	66
Figure 4 : Influence de l'âge et du niveau de scolarisation sur la capacité à identifier les données personnelles	68
Figure 5 : Influence de l'âge et du niveau de scolarisation sur la capacité à connaître la nature et l'étendue des données collectées par BAnQ sur ses abonné.e.s	68
Figure 6 : Influence de l'âge et du niveau de scolarisation sur la capacité à comprendre les enjeux des données personnelles	69
Figure 7 : Influence de l'âge et du niveau de scolarisation sur la capacité à réfléchir à la gouvernance des données personnelles	69
Figure 8 : Distribution par niveau du score d'intérêt pour les nouveaux services	72
Figure 9 : Intérêt pour le développement de nouveaux services : distribution des fréquences par sous-dimension d'analyse	75
Figure 10 : Intérêt pour le développement de nouveaux services : scores moyens par sous-dimension d'analyse	75
Figure 11 : Influence de l'âge sur l'intérêt pour enrichir les collections de BAnQ avec de nouveaux documents en fonction de la demande des publics	80
Figure 12 : Influence de l'âge et du score de littératie sur l'intérêt pour se faire recommander des documents en fonction de ses habitudes et préférences	80
Figure 13 : Influence de l'âge et du score de littératie sur l'intérêt pour se faire recommander des activités culturelles en fonction des habitudes et préférences	81
Figure 14 : Influence de l'âge et du score de littératie sur l'intérêt pour que son expérience numérique sur le site Web de BAnQ soit personnalisée	81
Figure 15 : Distribution globale du score de perception relative à la gestion des données personnelles (sans la sous-dimension I)	84
Figure 16 : Perception relative à la gestion des données personnelles (sous-dimensions J et L) : distribution par attitude (vue générale)	85
Figure 17 : Perception relative à la gestion des données personnelles : scores moyens par sous-dimension d'analyse	89
Figure 18 : Perception relative à la collecte et à l'exploitation des données personnelles (sous-dimensions I et J) : distribution des fréquences par sous-dimension d'analyse	90
Figure 19 : Conservation des données personnelles (sous-dimension K) : distribution des fréquences	92

Figure 20 : Perceptions relatives au partage des données personnelles anonymisées : distribution des fréquences par sous-dimension d'analyse (L1, L2, L3, L4)	93
Figure 21 : Influence du niveau de diplomation et de la littératie sur la perception quant au recoupement des données personnelles (sous-dimension J)	98
Figure 22 : Influence du niveau de diplomation et de la littératie des données personnelles sur la perception quant au partage des données personnelles avec un ministère québécois (sous-dimension L1)	99
Figure 23 : Influence du niveau de diplomation et de la littératie des données personnelles sur la perception quant au partage des données personnelles avec une maison d'édition (sous-dimension L2)	99
Figure 24 : Influence de l'âge, de l'affiliation au groupe LGBTQ+ et de la littératie des données personnelles sur la perception quant au partage des données personnelles avec une OBNL (sous-dimension L3)	100
Figure 25 : Influence de l'âge et du niveau de diplomation sur la perception quant au partage des données personnelles avec des équipes de recherche universitaires (sous-dimension L4)	100
Figure 26 : Synthèse des hypothèses concernant l'influence des variables sociodémographiques sur le score global de littératie des données personnelles	107
Figure 27 : Synthèse des variables qui influencent l'intérêt pour les nouveaux services	110
Figure 28 : Synthèse des variables qui influencent la perception relative à la gestion des données personnelles	114
Figure 29 : Synthèse des relations entre littératie des données personnelles, intérêt pour le développement de nouveaux services et attitude relative à la gestion des données personnelles	118

1. INTRODUCTION

Ce projet s'inscrit dans le cadre d'un partenariat de recherche entre une équipe de recherche de l'Institut national de la recherche scientifique (INRS) et de l'Université de Montréal puis Bibliothèque et Archives nationales du Québec (BAnQ).

Une première phase du projet (2020-2021) a produit une revue de la littérature pour synthétiser les travaux existants sur la gouvernance éthique des données d'usage, avec une considération particulière pour le secteur des bibliothèques et des archives et la perspective de leurs usagères et usagers : *La gouvernance des données d'usage : enjeux éthiques et perceptions des publics dans les bibliothèques et archives* (2021).

La deuxième phase de ce projet a pris la forme d'une enquête par sondage auprès des usagères et usagers de BAnQ. Le présent rapport expose les résultats de cette enquête menée en 2022. Sa conception et la rédaction des pistes de réflexions qui en sont tirées ont été co-construites dans un dialogue entre l'équipe de recherche et les membres du personnel de BAnQ participant au projet. L'équipe tient à remercier l'INRS et l'Observatoire des médiations culturelles pour leur contribution au financement de ce projet.

2. OBJECTIFS DE L'ENQUÊTE

L'objectif général de cette enquête est de connaître la perspective des personnes abonnées aux services de BAnQ concernant les pratiques de collecte, d'exploitation, d'archivage et de partage de données d'usage par l'institution. Ce faisant, le projet vise à contribuer au chantier de gouvernance des données de BAnQ : en éclairant la perspective des personnes abonnées, il offre des pistes de réflexion sur la gouvernance des données d'usage.

Plus précisément, l'enquête vise les objectifs spécifiques suivants :

- **Comprendre ce que savent les personnes abonnées aux services de BAnQ** à propos des pratiques de données d'usage de l'institution en mesurant leur littératie des données personnelles.
- **Mesurer l'intérêt des personnes abonnées pour le développement de nouveaux services** fondés sur la collecte et l'exploitation des données d'usage.
- **Mesurer leur perception des pratiques de BAnQ** concernant la gestion de leurs données personnelles (enjeux de protection de la vie privée).
- **Analyser les relations** entre les caractéristiques sociodémographiques des personnes abonnées, leur littératie des données personnelles, leur intérêt pour le développement de nouveaux services et leur perception en matière de gestion des données personnelles.

3. ENJEUX

3.1. La valorisation des données d'usage : une pratique en expansion

Au cours des dernières années, un nombre croissant d'organismes culturels québécois ont développé des projets numériques visant à récolter et analyser des données d'usage : il s'agit de données générées par l'activité des usagères et usagers lorsqu'ils se connectent à un service informatisé, par exemple l'historique d'emprunt d'un usager de bibliothèque. Couplées avec des informations qui renseignent sur le profil sociodémographique des usagères et usagers, ces données permettent de développer les connaissances sur les publics, leurs goûts et leurs comportements (Beaudouin et al., 2016 ; Tchéhouali et al., 2019). Ces données entrent dans la composition de nouveaux services liés à la découvrabilité et au développement de l'intelligence artificielle (Denis et Farchy, 2020).

Données d'usage

Les données d'usage sont constituées par les traces d'activité des utilisatrices et utilisateurs de services numériques, tel l'historique d'emprunt d'un usager de bibliothèque. Elles permettent de connaître le comportement des publics dans les environnements numériques, incluant leurs goûts et préférences en matière de contenus culturels.

Intelligence d'affaire et mutualisation des données

Les gouvernements fédéraux et provinciaux ont joué un rôle important dans le développement des initiatives en établissant des programmes de financement et d'accompagnement du personnel professionnel du secteur culturel. À l'échelle fédérale, le ministère du Patrimoine canadien et le Conseil des arts du Canada ont financé divers projets de mutualisation des données en culture, de littératie numérique et d'intelligence d'affaires. À l'échelle provinciale, le ministère de la Culture et des Communications du Québec (MCCQ) a joué un rôle particulièrement structurant à travers son premier Plan culturel numérique (2014-2021) : avec un budget total d'environ 110 millions de dollars, cette stratégie contenait plusieurs mesures visant spécifiquement les données sur les publics de la culture. Parmi elles, la mesure 111 consistait en un « Plan d'action concernant les données sur les contenus culturels québécois ». Le Ministère a également mis en place un réseau d'Agents de développement numérique (ADN) dont un des chantiers est le développement d'une « culture de la donnée ». Le milieu professionnel de la culture au Québec a lui-même impulsé plusieurs projets qui ont marqué le développement des données d'usage, notamment dans le secteur des arts de la scène (Place des Arts, Synapse C). **Certaines de ces approches sont orientées vers le marketing, dans une optique de segmentation des publics et de développement de marché, tandis que d'autres visent à**

générer de nouvelles formes de mutualisation des données sur le modèle de la fiducie (Culturepedia).

Ainsi, le développement de modèles de gouvernance des données d'usage **met en tension l'innovation économique** (nouveaux modèles d'affaires fondés sur la monétisation des données) **et l'innovation sociale** (nouvelles formes de gestion collective des ressources de données). Les modèles orientés vers la monétisation des données se sont développés en premier, sous l'influence des GAFAM¹. Le développement du profilage des audiences par les données est aussi alimenté par les approches marketing de segmentation des consommatrices et consommateurs culturels, orientées vers le développement de marché.

État des lieux dans le secteur des bibliothèques et archives

Dans le secteur des bibliothèques et archives, l'exploitation des données d'usage reste encore relativement nouvelle. La gestion des données informatisées fait partie des expertises maintenant bien ancrées dans les pratiques professionnelles des bibliothèques et archivistes, mais ces expertises se sont surtout développées du point de vue des métadonnées descriptives (informations qui décrivent le contenu et le contexte d'un document)² et les données d'usage restent peu exploitées à ce jour. **Bibliothèque et Archives nationales du Québec (BAnQ) s'est récemment engagée dans la direction d'une plus grande exploitation de ses données d'usage.** Dans sa *Stratégie numérique 2021-2028*, BAnQ identifie l'une de ses orientations en ces termes : « l'innovation est accélérée par l'exploitation des données et par l'intelligence artificielle » (voir encadré ci-après). Un chantier sur les données a par ailleurs été lancé en septembre 2020 avec pour mission, entre autres, de définir un cadre de gouvernance sur les données et de mettre en place un laboratoire d'innovation impliquant différents acteurs de l'écosystème numérique québécois.

¹ Acronyme formé par l'initiale des cinq entreprises Google, Apple, Facebook, Amazon et Microsoft.

² Pour plus de détails, voir Observatoire de la culture et des communications du Québec, *État des lieux sur les métadonnées relatives aux contenus culturels*, 2017.

Stratégie numérique 2021-2028, BAnQ (extraits)

Principe : Une institution intelligente

Orientation : l'innovation est accélérée par l'exploitation des données et par l'intelligence artificielle. L'intelligence artificielle et l'exploitation des données améliorent l'expérience des usagers et font de BAnQ une organisation plus performante et innovatrice.

Pistes de mise en œuvre :

- BAnQ fera une utilisation éthique et responsable de l'intelligence artificielle, en collaborant avec des acteurs du domaine.
- BAnQ fera la promotion des bonnes pratiques et deviendra la référence en intelligence d'affaires (gouvernance et valorisation des données).
- BAnQ mettra sur la force de l'infonuagique pour améliorer en continu la sécurité de l'information qu'elle conserve.
- BAnQ proposera aux abonnés des contenus reflétant leurs besoins et leurs intérêts grâce entre autres aux données sur le prêt de documents.

Face à l'enthousiasme pour le développement de nouveaux services fondés sur les données d'usage, **il existe une préoccupation croissante des citoyennes et citoyens et des organismes culturels eux-mêmes quant à la mise en place de pratiques de gouvernance pour mieux encadrer la collecte, l'exploitation et le partage de ces données.** La place omniprésente du numérique dans la vie quotidienne suscite de plus en plus d'inquiétude au regard de la protection des renseignements personnels (Véliz, 2021). Bien que la collecte et l'exploitation des données d'usage soient devenues l'épine dorsale des modèles d'affaires des géants du Web (Google, Facebook, Amazon), de telles activités comportent un ensemble de défis pour les organismes culturels québécois. En effet, des scandales largement médiatisés tels que celui de Cambridge Analytica ou encore la fuite des données des usagères et usagers des caisses Desjardins au Québec ont contribué à sensibiliser la population aux risques du partage des données personnelles. D'ailleurs, en 2021, une partie des services de BAnQ a dû fermer en urgence pour prévenir un risque de fuite de données causé par une faille de sécurité dans le logiciel des comptes de personnes abonnées (Gerbet, 2021). Cette cybermenace souligne l'urgence d'améliorer la gouvernance des données que BAnQ collecte sur ses usagères et usagers.

Évolution de la réglementation

BAnQ doit aussi s'adapter à l'évolution du paysage réglementaire sur les données numériques. Les législations provinciales et fédérales sur la protection des données personnelles se sont renforcées ces dernières années, avec la modernisation de la *Loi sur la protection des renseignements personnels* du Canada et la mise à jour de la *Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels au Québec* (Loi 25, 2022). La Commission d'accès à l'information du Québec a aussi mené une consultation sur les principes en intelligence artificielle relativement à la protection des renseignements personnels (CAIQ, 2021). De plus, en tant que société d'État, BAnQ est assujettie à la nouvelle *Loi sur la gouvernance et la gestion des ressources informationnelles des organismes publics et des entreprises du gouvernement* (Loi 95). Celle-ci instaure un nouveau cadre de gestion des données numériques gouvernementales. Elle établit des impératifs de partage de certaines données internes aux organismes avec le gouvernement québécois, ce qui soulève l'enjeu de l'anonymisation des données personnelles. En outre, la nouvelle *Loi sur l'intelligence artificielle et les données* (LIAD) en cours de préparation à l'échelle fédérale vise à encadrer la prise de décision et la formulation de recommandations fondées sur des modèles identifiés dans les données d'usage.

De ce contexte naît une double injonction pour BAnQ : profiter des possibilités qu'offrent les avancées technologiques pour développer de nouveaux services tout en protégeant la vie privée des personnes abonnées et la relation de confiance bâtie avec elles au fil des ans. Mais les usagers et usagers des bibliothèques sont rarement sondés au sujet de leur perception de ces pratiques. Ce projet d'enquête vise à combler ce manque en permettant de mieux comprendre la perspective des personnes abonnées.

3.2. Quelles données pour quels nouveaux services ?

Pratiques de données

Dans le contexte de cette étude, le terme générique de pratiques de données réfère à l'ensemble des activités institutionnelles de BAnQ relatives à la collecte, à la gestion, à l'exploitation, à l'archivage et au partage de données d'usage avec des tiers.

Les données collectées par BAnQ

BAnQ compile une masse croissante de données informatisées au sujet de ses usagères et usagers (voir Tableau 1). Sa plus grande source de données d'usage est l'historique d'emprunt associé au compte des personnes abonnées à la Grande Bibliothèque. Au moment de s'abonner aux services de BAnQ, les usagères et usagers remplissent un formulaire où ils fournissent des informations personnelles qui renseignent sur leur profil sociodémographique (telles le nom, l'adresse résidentielle, le sexe, l'âge). Par la suite, tous les emprunts de documents (qu'ils soient numériques ou analogiques) sont répertoriés dans le compte personnel des personnes abonnées. L'historique d'emprunt est donc une riche source d'information car il peut renseigner sur les habitudes de différents profils d'usagères et usagers. BAnQ collecte aussi d'autres types de données d'usage, telles les données d'utilisation du réseau sans-fil dans les lieux ouverts au public (quels équipements des personnes abonnées se connectent au réseau, où, quand et combien de temps) ou encore les données d'utilisation des postes multimédias dans les salles accessibles au public (identification de la personne abonnée, date et durée de la session).

L'ensemble est stocké dans un entrepôt de données qui contient plus de 100 millions de transactions collectées depuis la création de BAnQ. Seules certaines données collectées font l'objet d'une analyse statistique en vue du développement de nouveaux services : c'est notamment le cas des trois premiers caractères du code postal (RTA), de la date de naissance, du sexe et des données d'emprunt (par exemple le numéro de notice du document emprunté, le lieu de l'emprunt, le fait d'emprunter au comptoir ou en libre-service, la date et l'heure de l'emprunt et du retour).

BAnQ utilise par ailleurs des ensembles de données collectées par des fournisseurs tiers, dont elle n'est pas propriétaire. Ces données sont anonymisées par le fournisseur. Il s'agit en particulier de données de navigation sur les sites web de BAnQ qui sont fournies par Google Analytics (pages vues, liens sur lesquels les personnes cliquent, consultation des documents numériques patrimoniaux). BAnQ fait également usage des tableaux de bord fournis par les plateformes de médias sociaux numériques (Facebook, Instagram) qui renseignent sur les profils et les actions des internautes (anonymisés) qui interagissent avec les comptes de l'institution.

Tableau 1 : Types de données recueillies par BAnQ à propos de ses usagères et usagers

DONNÉES SOCIODÉMOGRAPHIQUES		DONNÉES DE GESTION DES ABONNEMENTS	
Prénom / nom*	Adresse courriel*	Numéro d'abonné	Ancienneté (date de première inscription)
Numéro de téléphone*	Adresse résidentielle*		
3 premiers caractères du code postal (RTA)	Ville / province / pays		
Date de naissance	Langue de communication (facultatif)*	Date de renouvellement d'abonnement (concerne tous services uniquement)	Type d'abonnement (tous services vs en ligne)
Sexe (jusqu'en 2022)*			
UTILISATION DU RÉSEAU SANS-FIL		UTILISATION/RÉSERVATION DE POSTES MULTIMÉDIAS	
Numéro d'abonné	Localisation du bâtiment (lieu d'utilisation)	Numéro d'abonné	Date / heure
Adresse de l'ordinateur / tablette / téléphone	Système d'exploitation / navigateur		
Date / heure de connexion	Durée de la session	Durée de la session	Utilisation du poste / réalisation
Modèle d'ordinateur / tablette / téléphone			

*Données qui sont uniquement enregistrées à des fins de gestion des opérations courantes mais qui ne font pas l'objet d'analyse

**Cette information n'est plus demandée dans le formulaire d'inscription depuis 2022.

DONNÉES D'USAGE

Données de circulation

<i>Emprunts physiques</i>		<i>Emprunts numériques</i>
Numéro d'abonné	Lieu de l'emprunt	Date et heure de l'emprunt
Méthode de l'emprunt (comptoir / libre-service)	Date et heure de l'emprunt	Nombre de documents empruntés
Date et heure du retour de l'emprunt	Retard de l'emprunt	Date et heure de réservation des documents
Numéro de notice du document emprunté	Montant à payer pour les retards	

Données de navigation sur les sites web (via Google Analytics, IP anonymisées)

Modèle d'ordinateur / tablette / téléphone	Système d'exploitation / navigateur	Langue d'utilisation
Date / heure de connexion	Durée moyenne des sessions	Pages vues
Certains liens sur lesquels les personnes cliquent (en fonction du paramétrage administrateur)	Adresse du site à partir duquel la personne est arrivée sur le site	Taux de rebond
Date / heure téléchargement / consultation des documents numériques patrimoniaux	Nom des documents numériques patrimoniaux téléchargés / consultés	Nombre total de téléchargements / consultations
Informations socio-démographiques agrégées (tranche d'âge, municipalité, etc.)		

Abonnement

Listes de documents enregistrés par l'abonné (ex. : favoris)*

*Données qui sont uniquement enregistrées à des fins de gestion des opérations courantes mais qui ne font pas l'objet d'analyse

Services potentiels

Pour BAnQ, l'exploitation des données d'usage est une opportunité de bonifier son offre de services aux usagères et usagers. Ces nouvelles avenues de développement de publics et de gestion organisationnelle sont présentées comme faisant partie d'une nouvelle « culture des données » qui offrirait un avantage concurrentiel aux organisations qui fondent leurs décisions sur l'analyse des données d'usage (Legoux, 2019). Les bénéfices attendus de l'analyse des données d'usage dans le secteur des bibliothèques et archives sont les suivants :

- *Identifier des profils d'usagères et usagers et des types de comportements.* BAnQ a collaboré avec des chercheuses et chercheurs pour analyser les données de sa Grande Bibliothèque. L'enquête a permis d'identifier trois types de personnes qui empruntent (les « numériques », les « physiques » et les « hybrides ») en croisant ces profils d'emprunt avec des informations sociodémographiques et des informations sur les types de livres empruntés (Lapointe et al., 2022). Dans une démarche similaire, la Bibliothèque nationale de France a analysé les données de navigation de son site web pour identifier des profils d'usagères et usagers (Rollet et al., 2017) et repérer des types d'usage et d'engagement avec les contenus (Beaudouin et al., 2016).
- *Personnaliser l'expérience de navigation sur le site web.* Les données d'usage permettent de mieux comprendre les parcours des publics sur les sites web des institutions de façon à mettre en adéquation l'offre avec les différents profils d'usage (Nouvellet et al., 2017 ; d'Alché-Buc et al., 2017).
- *Augmenter la découvrabilité des contenus en faisant des recommandations ciblées.* La connaissance des profils et des comportements des usagères et usagers permettrait de leur recommander des documents de la collection en fonction de leurs habitudes et de l'analyse des comportements des groupes d'usagères et usagers qui leurs ressemblent. Des algorithmes de recommandation des contenus pourraient ainsi être déployés sur le site web de BAnQ pour augmenter la découvrabilité de certains contenus (par exemple, suggérer d'autres documents similaires à ceux déjà empruntés) ou leur faire parvenir par courriel lors d'un rappel d'échéance d'emprunt (Beck, 2018).
- *Développer la collection en fonction de la demande des publics.* Les données d'usage permettent de connaître les différents niveaux de popularité des contenus des collections et de repérer des effets de « longue traîne » (Benghozi et Benhamou, 2008). Une hypothèse envisagée par BAnQ est de fonder une partie de l'acquisition de nouveaux documents sur l'analyse de l'historique d'emprunt des personnes abonnées.
- *Mutualiser des données avec d'autres organismes.* Le partage de données anonymisées entre divers organismes partenaires est une occasion de mieux comprendre les pratiques culturelles des usagères et usagers. Il fournit une analyse plus fine des pratiques « de niche » qui pourrait compléter les informations générales tirées des grandes enquêtes nationales sur les pratiques culturelles (Hanquinet et al., 2019). Le partage de données

entre organismes culturels pourrait aussi permettre de développer des services de recommandation croisée ouvrant sur des offres culturelles complémentaires.

Plusieurs modes d'exploitation des données d'usage soulèvent toutefois un certain nombre de critiques.

- Les travaux académiques sur le sujet soulignent que *les données d'usage ne donnent pas une représentation complète et objective des publics* (Gillespie, 2014). Il s'agit plutôt d'une construction statistique partielle et la compréhension qu'elle donne des pratiques culturelles est limitée et peut comporter des biais (Gitelman, 2013). Fonder les décisions de développement de nouveaux services uniquement sur l'analyse des données d'usage ferait courir le risque que ces services ne soient pas adaptés aux besoins des publics. Le fait que des décisions ayant un impact sur l'intérêt collectif soient fondées uniquement ou principalement sur l'analyse de données d'usage (par corrélations de données et modélisations de comportements) a été qualifié de « gouvernamentalité algorithmique » (Rouvroy et Berns, 2013). Cette approche est critiquée pour son manque de débat public sur l'instrumentalisation des comportements quotidiens (Elmer, 2003 ; Cardon, 2015 ; Beer, 2015 ; Badouard et al., 2016), d'autant plus dans un contexte de développement accéléré et peu régulé de l'intelligence artificielle.
- Par ailleurs, certaines utilisations des algorithmes de recommandation fondés sur l'analyse des données d'usage pourraient *contrevenir à la mission des bibliothèques et archives publiques de fournir un accès libre et non discriminant à une information plurielle*. La recommandation peut enfermer les usagères et usagers dans des profils qui déterminent l'offre qui leur est proposée (Dilhac et al., 2018), voire reproduire des inégalités et des discriminations (Pasek et al., 2019 ; Apprich et al., 2018). Pré-formuler des choix peut avoir pour conséquence de renforcer la consultation des contenus déjà les plus populaires, et donc aller à l'encontre de la découvrabilité des documents plus nichés. Ce risque s'applique aussi à l'utilisation des données d'usage pour acquérir de nouveaux documents dans la collection.
- Enfin, les *risques posés par la collecte et l'analyse des données personnelles en lien avec la protection de la vie privée* sont largement documentés (Bennet et Parsons, 2013 ; Van Dijck, 2014). Dans les sections suivantes, nous reviendrons plus spécifiquement sur les enjeux liés à la protection de la vie privée, qui est une dimension centrale de notre enquête.

À ce jour, l'exploitation des données d'usage à BAnQ est en phase de développement.

Outre l'analyse statistique pour dresser un portrait des emprunts (Lapointe et al., 2022), l'institution a jusqu'ici concentré ses efforts sur un chantier de gouvernance des données : mise en place d'une politique de valorisation et de partage des données, création de la fonction de gestionnaire des données institutionnelles, production d'une feuille de route sur l'IA. Cette étape est un préalable à l'éventuelle autorisation de prototypes d'exploitation des données ou de partage de données avec des tiers (données ouvertes ou accès contractuel). BAnQ travaille

également à la création d'un « comptoir de données » à usage interne pour produire des tableaux de bord (via un logiciel tel Microsoft Power BI) dont l'un des objectifs est de pouvoir suivre quotidiennement et dynamiquement l'évolution de l'abonnement (nombre de personnes abonnées actives, répartition par âge ou par région, etc.). La *Stratégie numérique de BAnQ (2021-2028)* inclut par ailleurs le fait de consulter les usagères et usagers au sujet de la conception de nouveaux services, ce qui mène à s'interroger sur le point de vue des usagères et usagers concernant la gouvernance des données.

L'opinion des usagères et usagers

L'opinion des usagères et usagers au sujet du développement de nouveaux services fondés sur les données est relativement peu documentée.

- *Un sondage de l'Association des bibliothèques publiques du Québec* auprès de 5803 usagères et usagers des bibliothèques (ABPQ, 2021) suggère que les cinq services les plus populaires auprès des personnes sondées pour améliorer l'offre de services numériques en bibliothèque sont les suivants : offrir des suggestions de lecture basées sur l'historique d'emprunt et les préférences de lecture (48,5% des personnes sondées), avertir lorsque de nouveaux documents correspondant aux intérêts enregistrés sont ajoutés à la collection (40,3%), avertir lorsque la suite d'une série est disponible (romans, BDs, films, séries télévisées, etc. ; 39,8%), améliorer l'expérience de recherche dans le catalogue (34,9%) et faciliter l'accès à des livres numériques ou audionumériques (31%). Toutefois, le sondage de l'ABPQ ne permet pas de croiser ces résultats avec les profils sociodémographiques des personnes sondées, ni avec leur littératie des données et leur attitude relative à la protection des données personnelles.
- Par ailleurs, *une consultation publique a été menée en 2020 par BAnQ auprès des citoyennes et citoyens* pour recueillir leur opinion au sujet des grandes orientations de sa nouvelle Stratégie numérique. D'une part, les commentaires publiés sur le site de consultation par les personnes sondées suggèrent qu'elles ont davantage d'intérêt pour l'amélioration des services existants (notamment l'optimisation des résultats de recherche dans les bases de données et la navigation sur le site web) et que les services innovants tels un robot conversationnel ne sont pas identifiés comme une priorité par les personnes sondées. Une critique écologique est également adressée à l'utilisation de l'intelligence artificielle, au motif qu'il s'agit d'un outil énergivore. D'autre part, parmi les commentaires, ceux qui dénotent une forte littératie des données (bonne connaissance des sources de données et des applications possibles) tendent aussi à être plus favorables au développement de nouveaux services fondés sur les données. Le taux relativement faible de participation à cette consultation (137 personnes sondées pour 69 commentaires) ne permet toutefois pas de généraliser ces résultats.

Notre enquête vise à combler ces lacunes en offrant un portrait plus complet de l'intérêt des personnes abonnées aux services de BAnQ pour le développement de nouveaux services fondés sur l'analyse des données d'usage. Il est également nécessaire de mieux comprendre comment

cet intérêt peut changer en fonction des différents groupes d'usagères et usagers et de leur niveau de littératie des données personnelles – des aspects non documentés à ce jour. Dans le cadre de notre enquête, nous formulons donc les questions de recherche suivantes :

Quel est l'intérêt des personnes abonnées pour les services qui pourraient être développés à leur intention grâce à ces pratiques de données ?

Cet intérêt varie-t-il en fonction des caractéristiques sociodémographiques des personnes abonnées ?

3.3. La littératie des données personnelles

Puisque notre enquête vise à sonder les usagères et usagers au sujet des politiques institutionnelles de BAnQ relatives aux données personnelles, il est important de savoir ce qu'ils en comprennent afin de contextualiser leurs réponses. Que savent les personnes abonnées aux services de BAnQ à propos des pratiques de données d'usage de l'institution ? Lorsqu'elles consentent à certains services utilisant leurs données, sont-elles conscientes des implications ? Pour étudier cet enjeu, nous adoptons l'angle de la littératie des données personnelles du point de vue des usagères et usagers.

Définition

Littératie des données personnelles

Cette notion réfère aux connaissances et compétences utiles pour comprendre et utiliser l'information relative à la gestion de ses données personnelles dans les environnements numériques. Elle se situe au point de rencontre entre la littératie numérique et la littératie de la vie privée. Les compétences visées par cette littératie sont à la fois opérationnelles (utiles pour agir) et critiques (utiles pour discerner les risques) : elles concernent la capacité à identifier la nature des données personnelles, à comprendre leurs enjeux et à réfléchir à leur gouvernance en contexte numérique (Casemajor et al., 2021).

La littératie des données personnelles cible les enjeux spécifiques en lien avec la protection des données personnelles dans les environnements numériques.

- Elle est donc plus étroite que la *littératie numérique* qui porte sur les connaissances et les capacités liées aux environnements numériques dans leur ensemble (Hargittai, 2005 ; Hobbs, 2010 ; Ferrari et Punie, 2013 ; Pangrazio et al., 2020).

- Elle est aussi plus précise que la *littératie des données* qui concerne les compétences informationnelles, statistiques et techniques liées aux environnements de données sans cibler les enjeux spécifiques à la vie privée dans les données personnelles (Bonikowska et al., 2019 ; Pangrazio et Sefton-Green, 2020 ; Fotopoulou, 2020).
- Sa définition inclut plusieurs aspects de la *littératie de la vie privée (privacy literacy ; Wissinger, 2017 ; Stoilova et al., 2020)* tout en focalisant sur les enjeux spécifiques que posent le numérique relativement à la vie privée. En ce sens, elle est proche de la notion de « littératie de la vie privée en ligne » (Trepte et al., 2015). Mais elle s'en distingue car les enjeux relatifs à la vie privée posés par les pratiques de données de BAnQ ne sont pas circonscrits par les activités en ligne (et encore moins aux médias sociaux, un angle privilégié par de nombreux travaux) : par exemple, le fait d'emprunter un livre sur place au comptoir de la bibliothèque est une des principales sources de données pour l'institution.

À partir de cet angle conceptuel, nous posons la question de recherche suivante :

Quel est le niveau de littératie des données personnelles des personnes abonnées aux services de BAnQ ?

Cadre de compétences

Les recherches antérieures ont défini des cadres de compétences spécifiques à la littératie des données personnelles (Pangrazio et al., 2019, 2020). On peut en dégager deux principales dimensions : opérationnelle et critique.

- La dimension des *compétences opérationnelles* concerne la capacité à identifier les données personnelles (ce que sont ces données), la capacité à comprendre les processus de « mise en données » ou cycle de vie des données (comment elles sont générées, exploitées, partagées et archivées) et la capacité à utiliser les données (analyse, applications, création).
- La dimension des *compétences critiques* inclut la pensée critique (capacité à questionner les processus, évaluation des décisions fondées sur les données), la réflexivité (évaluation des impacts éthiques de ces données, notamment sur la vie privée), la capacité à contrôler les paramètres de sa vie privée en ligne (ou à adopter des tactiques de contournement) et la capacité à envisager des formes de gouvernance des données.
- La mesure de niveaux de compétences en littératie se fait généralement en distinguant des ensembles de compétences classées par *ordre de complexité croissante* (Kirsch et Lennon, 2005).
- Pangrazio et Selwyn (2019) ont regroupé les capacités liées à la littératie des données personnelles en cinq domaines : « (1) *identification des données*, (2) *compréhension des données*, (3) *réflexivité des données*, (4) *utilisation des données* et (5) *tactique de données* ». L'utilisation des données par les usagers et usagers eux-mêmes (usages et tactiques) ne faisant pas partie de notre étude, nous avons seulement retenu les trois premières dimensions pour bâtir notre enquête.

- Dans le choix de nos dimensions d'analyse, nous nous sommes également appuyés sur l'échelle de littératie de la vie privée en ligne (*online privacy literacy scale – OPLIS*) définie par Trepte et al. (2015). Elle distingue cinq dimensions : « (1) *connaissance des pratiques des organisations, institutions et fournisseurs de services en ligne* ; (2) *connaissance des aspects techniques de la protection de la vie privée et des données en ligne* ; (3) *connaissance des lois et des aspects juridiques de la protection des données en ligne* [...] ; (4) *connaissance des directives [...] sur la protection de la vie privée et des données* ; et (5) *connaissance des stratégies des utilisateurs pour la régulation individuelle de la vie privée* ». Étant donné que notre enquête porte sur les pratiques de données d'une institution spécifique (BAnQ), nous avons concentré l'analyse sur les deux premières dimensions de cette échelle : connaissance des pratiques de BAnQ et connaissance des aspects techniques de la protection des données personnelles en ciblant les enjeux touchant BAnQ (anonymisation, contrôle des données par des tiers fournisseurs de service tel Google Analytics).
- Le Tableau 2 résume le cadre de compétences en littératie des données personnelles retenu dans notre étude et adaptées au contexte de BAnQ, en les ordonnant par degré de complexité croissant.

Tableau 2 : Cadre de compétences en littératie des données personnelles retenu dans l'étude

Complexité	Dimension	Capacité
-	Identification	1. Identifier les données personnelles (en général). 2. Identifier les données personnelles produites par BAnQ (ex. l'historique d'emprunt lié au compte d'abonné).
↓	Compréhension	3. Comprendre les implications des données personnelles (en lien avec les données sensibles, l'anonymisation et le contrôle des données par des tiers fournisseurs de service).
+	Réflexion critique	4. Réflexivité sur les données personnelles (enjeux de la gouvernance des données).

L'approche de la littératie des données personnelles par l'angle d'un cadre de compétence comporte néanmoins des limites. Choi (2022) souligne que ces compétences doivent être étudiées en prenant en compte le contexte et la situation d'usage des environnements numériques. De plus, au-delà de la définition d'un cadre de compétences à acquérir par les individus pour valoriser les données au sein de leur organisation, les études critiques de données (*critical data studies*) soulignent l'importance de la dimension politique et de pédagogie critique pour que les usagères et usagers puissent participer à l'imagination des infrastructures de données qui les concernent en les faisant concorder avec leurs intérêts (Gray et al., 2018 ; Ruppert, 2018 ; Markham, 2019).

Les fossés de la vie privée

Les travaux sur la littératie des données personnelles invitent à considérer l'existence d'un « fossé de la vie privée » voire des « inégalités de vie privée » qui divisent les usagères et usagers selon leurs aptitudes et leurs caractéristiques sociodémographiques (*privacy divide*, Papacharissi, 2010 ; *privacy inequality*, Litt, 2013). Concernant les usagères et usagers de BAnQ, quelques données existent sur leur littératie numérique (50% d'entre eux se déclarent peu à l'aise avec les interfaces numériques selon Zini, 2020) mais pas sur la littératie des données personnelles. De plus, à ce jour, il n'existe que peu d'études sociodémographiques portant spécifiquement sur la notion de littératie des données personnelles. Mais les nombreuses recherches antérieures sur les fossés numériques, sur la littératie numérique, la littératie des données et la littératie de la vie privée en ligne laissent supposer qu'elle ne serait pas partagée de manière égale dans tous les groupes de population.

Les facteurs les plus documentés à ce jour sont l'âge, le niveau d'éducation, le genre et l'ethnicité. L'effet du statut économique, de l'occupation, du statut d'immigration et de l'orientation sexuelle sont documentés dans une moindre mesure (voir section suivante). Selon les travaux existants, **les facteurs qui ont un effet plus significatif sur la littératie des données personnelles sont l'âge et le niveau d'éducation.** Compte tenu des résultats contradictoires des précédents travaux sur l'effet de ces principaux facteurs, et vu le faible nombre d'études existantes sur les autres facteurs sociodémographiques, il est pertinent d'étudier leurs effets potentiels dans le cadre de notre enquête.

L'étude des facteurs sociodémographiques est d'autant plus importante que **les travaux existants mettent en évidence une dynamique d'accumulation des inégalités** (*digital inequality stack* ; Robinson et al., 2020) : lorsque des inégalités de base existent aux niveaux inférieurs de cette accumulation (telles que les conditions de vie et l'accès à l'Internet), elles se traduisent souvent par des inégalités à des niveaux plus élevés comme dans les usages et les compétences numériques (Denouël et Granjon, 2011 ; Park, 2013b). Ainsi, Epstein et al. (2020) suggèrent que les inégalités en termes de littératie numérique ont des effets cumulatifs sur la littératie de la vie privée en ligne.

Plusieurs lacunes demeurent cependant dans les connaissances actuelles. Premièrement, la littératie de la vie privée a surtout été étudiée dans le contexte particulier des médias

socionumériques (surtout Facebook) et moins souvent dans des environnements numériques plus larges. Le contexte des institutions culturelles (telle que BAnQ) est très peu étudié à ce jour. Les institutions publiques ayant fait l'objet d'études sont plutôt les établissements scolaires (boyd and Hargittai, 2010 ; Gebre, 2018). Deuxièmement, la majorité des études ont été menées aux États-Unis et très peu concernent le Canada. **En étudiant le cas des personnes abonnées de BAnQ, notre enquête vise à améliorer les connaissances sur le contexte d'une institution publique canadienne.**

Facteurs sociodémographiques

Âge

Concernant la variable de l'âge, une première série de travaux tend à montrer que **les groupes les plus âgés ont un usage moindre d'Internet et un niveau de compétence plus bas** en littératie numérique.

- En termes d'usage d'Internet, les personnes âgées de 65 ans et plus (et plus encore celles âgées de 80 ans et plus) sont toujours moins susceptibles d'utiliser l'Internet comparé aux groupes d'âge plus jeunes (Robinson et al., 2020). Certaines personnes âgées éprouvent également des défis particuliers relatifs au maintien de la sécurité en ligne et à la compréhension de l'évolution des interfaces des TIC (Robinson et al., 2020).
- En ce qui concerne le niveau de compétences relatif à l'usage d'Internet, Hargittai (2002) conclut que l'âge y est négativement associé (plus les personnes sont âgées, moins elles sont compétentes), et plusieurs travaux concluent même que l'âge est le principal facteur contribuant à ce niveau de compétence (Van Deursen et Van Dijk, 2011; Park, 2013a).

Toutefois, certains travaux récents nuancent ces résultats en suggérant que **les personnes âgées ne sont pas un groupe homogène** de « retardataires » en termes d'adoption des nouvelles technologies (*laggards* selon le terme de Rogers, 1995).

- Ainsi, Hargittai et Dobransky (2017) montrent qu'il existe une diversité de compétences et d'usages de l'Internet chez les personnes âgées (au niveau des compétences, les différences sont toutefois moins prononcées que pour les usages).
- De même, les résultats d'une étude menée au Canada par Quan-Haase et al. (2018) montrent que les personnes âgées n'ont pas toutes un faible niveau de compétences.
- Oh et al. en concluent que « **les divisions basées sur l'âge sont trop simplistes et peu claires pour expliquer la fracture numérique** » (2021: 10) : d'autres facteurs tels que l'éducation, le revenu, les étapes de la vie et la socialisation doivent aussi être pris en compte.

Une deuxième série de travaux ont mesuré **les effets de l'âge sur les connaissances et compétences relatives à la vie privée** dans les environnements numériques.

- Li et al. (2018) concluent que chez les groupes socialement défavorisés, l'âge est le seul facteur sociodémographique significativement lié aux compétences en matière de

protection de la vie privée numérique : **les personnes plus jeunes ont un niveau plus élevé de compétences**. Ces résultats vont dans le même sens que les travaux de Litt (2013) qui montrent que les personnes plus âgées sont moins susceptibles que les plus jeunes d'utiliser des stratégies technologiques de protection de leur vie privée sur les réseaux socionumériques. Les résultats de Park (2013b) nuancent toutefois l'effet de l'âge sur les compétences : dans son étude, l'âge n'était pas un déterminant constant des compétences.

- Dans une perspective inverse, deux études ciblant davantage les connaissances (et non seulement les compétences) relatives à la protection de la vie privée via le cadre de compétences « *OPL online privacy literacy* » formalisé par Trepte (2015) ne trouvent pas d'association significative entre les différents groupes d'âge (Kezer et al., 2016), voire trouvent une association inverse à ce qui était attendu : dans l'étude d'Epstein et al. (2020), l'âge est positivement associé à la littératie de la vie privée en ligne (les groupes les plus âgés ont un meilleur score). Les auteurs expliquent cette différence par le fait que le cadre OPL inclut des connaissances relatives au cadre légal de la protection des données personnelles, et que ces connaissances générales seraient moins présentes chez les groupes les plus jeunes. Toutefois, à l'inverse, dans une étude de Park (2013b) qui n'utilisait pas le cadre OPL, mais intégrait tout de même des connaissances sur le cadre légal étatsunien, « les utilisateurs les plus jeunes étaient plus susceptibles d'obtenir un score correct pour au moins trois éléments de connaissance », incluant un item sur le cadre légal (Park, 2013b). Ce résultat est donc contradictoire avec l'hypothèse d'Epstein.
- Par ailleurs, si les travaux sur les compétences relatives à la vie privée tendent à montrer que les plus jeunes les maîtrisent davantage dans le contexte des réseaux sociaux où ils sont plus actifs en partageant davantage de contenus personnels, Pangrazio et Selwyn (2018) montrent qu'ils ont des connaissances générales limitées sur la nature et les conséquences des données personnelles (ex. effets de la géolocalisation, cycle de vie des données).

Pour conclure, **en dehors des usages généraux de l'Internet et du cas particulier des réseaux socionumériques, l'influence de la variable de l'âge sur la littératie des données personnelles n'est pas claire.**

Éducation (niveau de diplomation)

Les travaux sur la fracture numérique montrent de manière relativement constante que **le niveau de diplomation a un impact important sur les usages d'Internet.**

- Au Québec, les personnes diplômées universitaires et les personnes diplômées du collégial sont plus nombreuses à se déclarer à l'aise sur Internet (65% et 51%) que les personnes diplômées du secondaire (23%) (MCCQ, 2013). Dans une étude menée au Canada, Haight et al. (2014) ont fait ressortir que les personnes diplômées universitaires effectuent davantage d'activités en ligne que les personnes n'ayant pas terminé leurs études secondaires.
- Diverses études fondées sur des tests de performance (et non sur les déclarations des enquêtés) vont dans le même sens : elles suggèrent que le niveau d'éducation est un

facteur prédictif des compétences techniques, informationnelles et stratégiques sur Internet (Van Deursen et Van Dijk, 2011 ; Park, 2013a).

Les résultats sont **plus mitigés en ce qui concerne l'effet du niveau de diplomation sur les compétences et les connaissances liées à la protection de la vie privée en ligne.**

- Plusieurs travaux qui ont mesuré les compétences opérationnelles indiquent qu'il n'y a pas de relation significative entre le niveau d'éducation et les connaissances liées à l'utilisation de stratégies pour protéger sa vie privée en ligne.
- À l'inverse, les études qui portent sur les connaissances liées à la protection de la vie privée en ligne mettent en évidence que le niveau d'éducation a un impact sur la compréhension des politiques d'appropriation et de transfert des données (Park, 2013a et 2013b) et sur la littératie de la vie privée en ligne (Epstein et al., 2020).

Genre

L'impact du genre sur les compétences numériques a été largement étudié, mais les résultats de ces travaux sont contradictoires.

- Une première série de travaux sur la **fracture numérique** menée dans les années 2000 et 2010 tend à montrer que les femmes étaient moins actives sur Internet (Haight et al., 2014) et qu'elles avaient de moindres compétences d'usage d'Internet (Gui, 2007 ; Park 2013a).
- Cependant, d'autres travaux ne détectent pas d'effet du genre (Van Deursen et Van Dijk, 2011) ou concluent à une disparition voire à une inversion de l'effet du genre sur l'usage d'Internet et les compétences qui y sont liées (Ritzhaupt et al., 2013 ; Robinson et al., 2020).

Les travaux sur les compétences et connaissances relatives à la vie privée dans les environnements numériques sont tout autant divergents.

- En ce qui concerne les **compétences** d'utilisation des outils de protection de la vie privée sur les médias sociaux, boyd et Hargittai (2010) ne trouvent pas d'effet du genre sur le changement des paramètres de la vie privée, tandis que Litt (2013) indique que les hommes utilisent des outils moins diversifiés que les femmes. Hargittai et Litt (2013) identifient quant à elles un effet du genre sur la gestion des paramètres seulement chez les utilisateurs qui ont le moins de compétences numériques.
- Du côté des connaissances relatives à la vie privée en ligne, les résultats sont aussi très partagés : certaines études font ressortir que les femmes ont des connaissances moindres que les hommes (Park 2013a, 2013b ; Epstein et al., 2020), mais à l'inverse, d'autres montrent que le niveau de littératie de la vie privée sur les médias sociaux est plus élevé chez les femmes que chez les hommes.

En conclusion, les travaux précédents ne permettent pas de conclure à une direction claire de l'effet du genre sur la littératie numérique et sur la littératie de la vie privée.

Statut économique

- Les travaux sur la fracture numérique montrent que **les groupes économiquement privilégiés ont des usages d'Internet plus étendus et plus diversifiés** (Haight et al., 2014 ; Robinson et al., 2020).
- Concernant les enjeux de protection de la vie privée en ligne, les études existantes ne décèlent pas d'impact du statut économique sur les *compétences* d'usage des outils de protection (Litt, 2013 ; Park, 2013b). Mais à l'inverse, **en termes de connaissances relatives à la vie privée en ligne**, Park (2013a, 2013b) conclut que **le niveau de revenu a un impact significatif**.

Occupation et secteur d'emploi

- **Les effets de l'occupation sur la littératie de la vie privée en ligne sont peu documentés.** Li et al. (2018) n'ont pas trouvé d'association entre le statut d'emploi et le niveau de littératie de la vie privée en ligne. Toutefois, certains travaux sur la fracture numérique suggèrent un lien entre cette variable et l'usage d'Internet : les personnes actives (à l'emploi) utilisent davantage Internet que les personnes sans emploi ou retraitées selon Van Deursen et Van Dijk (2011), tandis que les personnes aux études sont le groupe qui utilise le plus les réseaux sociaux numériques (Haight et al., 2014).
- De plus, **il est possible d'imaginer un effet du secteur d'emploi sur la littératie des données personnelles.** Le fait d'être employé dans un secteur davantage lié à un travail de bureautique pourrait augmenter les chances d'avoir un niveau de littératie des données personnelles élevé. Notre enquête vise à documenter cette variable peu étudiée à ce jour.

Ethnicité

- Divers travaux ont documenté les effets de l'ethnicité sur la littératie numérique, mais **les résultats ne sont pas toujours concordants en ce qui concerne la littératie de la vie privée en ligne.**
- Un clivage entre les groupes afro-américain et caucasien a été documenté à plusieurs reprises aux États-Unis. Une enquête de Ritzhaupt et al. (2013) en milieu scolaire aux États-Unis a ainsi montré que les élèves blancs participant à leur enquête avaient un niveau plus élevé de littératie numérique. Par ailleurs, Hargittai et Litt (2013) concluent que les personnes caucasiennes sont beaucoup plus susceptibles que les autres groupes d'avoir utilisé des paramètres de protection de la vie privée sur les réseaux sociaux numériques lors d'une recherche d'emploi. En parallèle, plusieurs études font ressortir une association négative entre le fait de s'identifier comme afro-américain et le niveau de connaissance et de compétence relatives à la protection de la vie privée en ligne (Park et Jang, 2014 ; Epstein et al., 2020).

- Toutefois, une autre série de travaux contredit ces résultats : les études menées par Litt (2013), Li et al. (2018) et Choi (2022) ne trouvent pas d'association entre l'ethnicité et le niveau de littératie de la vie privée en ligne.
- Dans le contexte canadien caractérisé par le multiculturalisme et la présence de populations autochtones, il est pertinent de s'intéresser à l'ethnicité pour analyser les effets possibles de cette variable sur la littératie des données personnelles. Mais nous proposons plutôt d'adopter la catégorie de « minorité visible » pour limiter la collecte d'information durant l'enquête, en supposant que le fait d'appartenir à une minorité visible est un facteur déterminant de l'expérience de marginalisation sociale associée à l'appartenance à certaines ethnicités.

Statut d'immigration

- **Cette variable est peu documentée dans les travaux existants.** Du point de vue des usages d'Internet, l'étude de Haight et al. a fait ressortir que les personnes immigrantes les plus récentes au Canada ont des taux d'accès plus faibles à l'Internet, mais celles qui sont en ligne « ont des niveaux d'activité en ligne significativement plus élevés que les résidents nés au Canada et les immigrants antérieurs » (2014).

Orientation sexuelle

- **Cette variable est peu documentée sous l'angle spécifique de la littératie des données personnelles.** Mais il existe un corpus de travaux sur les risques d'atteinte à la vie privée des minorités sexuelles en ligne et sur leur adoption de comportements de protection (Robinson et al., 2020).

Pour mieux comprendre l'existence de fossés dans la littératie des données personnelles des usagères et usagers de BAnQ, nous formulons la question de recherche suivante :

Le niveau de littératie des données personnelles des personnes abonnées aux services de BAnQ varie-t-il en fonction de leurs caractéristiques sociodémographiques ?

3.4. Incidences sur la vie privée des abonnés

La vie privée dans les environnements numériques

La définition de la vie privée fait débat : ses contours restent difficiles à saisir, d'autant plus que les technologies numériques en remodèlent les conditions (Margulis, 2003 ; Westin, 2003). D'un point de vue juridique, la vie privée est définie sous l'angle de sa protection, qui fait partie des « droits fondamentaux » au Canada (CPVPC, 2019). En ce sens, la protection de la vie privée est entendue comme la capacité à déterminer quelles informations communiquer ou soustraire au regard d'autrui, selon quelles modalités, à quel moment et pour quelle durée (Westin, 1967). D'un point de vue sociologique et communicationnel, la gestion des frontières de la vie privée est un processus fondé sur des règles d'ouverture et de fermeture arbitrées tant au niveau individuel que collectif (tel que modélisé par Petronio, 2002, dans le cadre *communication privacy management* – CPM). La négociation de ces frontières et la manière d'en déterminer les paramètres de protection (*privacy calculus*) dépendent de plusieurs facteurs : les environnements de pratique, les interactions sociales, mais aussi les époques et les contextes nationaux (Altman, 1975, 1977).

Depuis la fin des années 1990, le développement accéléré des systèmes numériques et des dispositifs de surveillance ont profondément remanié la négociation des frontières entre vie privée et vie publique. Les appareils et services connectés à l'Internet, et notamment les réseaux sociaux, ont instauré de nouvelles possibilités de partage d'information dans des cercles de plus en plus étendus, et selon des modalités de plus en plus opaques. Si bien que la notion de vie privée a été reconceptualisée pour mieux saisir les nouveaux paramètres de cette « vie privée en réseau » (*networked privacy* ; Marwick et boyd, 2014 ; Marwick, 2023). La myriade de systèmes techniques qui collectent des données multiplie les possibilités de circulation et de réutilisation de l'information personnelle, reconfigurant en profondeur les circuits de gestion de ces flux d'information. L'intégrité de cette information doit donc être examinée dans chaque situation spécifique d'utilisation, en examinant si les valeurs et les intérêts définis dans les contextes d'origine de production des données sont respectés lorsque les données sont mises en circulation et réutilisées dans d'autres contextes (*contextual integrity*, Nissenbaum, 2009).

Bien qu'une plus grande circulation de l'information personnelle puisse apporter certains bénéfices, de nombreuses études font aussi état de risques accrus d'atteinte à la vie privée liés à la collecte de données par les services numériques (Véliz, 2021). Les approches critiques de l'économie numérique ont associé ces pratiques à la montée d'un capitalisme de surveillance (Zuboff, 2019) et d'un capitalisme de données (West, 2019). Les débats actuels sur les systèmes d'intelligence artificielle (IA) ont encore accentué les préoccupations liées à la vie privée dans les environnements numériques. En effet, l'entraînement de modèles algorithmiques à partir de renseignements personnels peut permettre le profilage, l'inférence de certains renseignements au sujet des personnes et la prédiction de certains de leurs comportements (Chun, 2021). De plus, la prise de décision automatisée peut reproduire des inégalités et des formes de discrimination (Eubanks, 2018). C'est dans ce contexte mouvant et traversé de controverses

qu'évoluent les bibliothèques et archives qui souhaitent collecter et exploiter des données sur leurs usagers.

Données personnelles et données sensibles : des informations à risque

Pour les bibliothèques et archives, la collecte et l'exploitation des données sur leurs usagères et usagers comporte certains risques d'atteinte à la vie privée. Ces risques sont notamment liés à la présence d'informations personnelles dans les données collectées, voire de données sensibles, qui posent un risque accru. C'est le cas de **l'historique d'emprunt d'une personne abonnée à la bibliothèque qui peut potentiellement révéler des informations sensibles** sur cette personne (Samarati, 2001). Une difficulté de gestion de ces données vient du fait que malgré toutes les précautions, **l'anonymisation des données personnelles ne peut pas être garantie hors de tout doute**, car il n'existe pas de technique invulnérable qui supprimerait tout moyen de reconstituer un lien entre une donnée et une personne. La confidentialité des données personnelles est donc un enjeu majeur dans le contexte de collecte et d'exploitation des données d'usage.

Données personnelles

Sont considérées personnelles les données qui se rapportent à une personne physique en permettant de l'identifier. Par exemple, l'historique d'emprunt de documents d'une personne abonnée à la bibliothèque ayant un compte enregistré à son nom contient des informations personnelles. Cet historique est constitué par le croisement de deux types de données : des données d'usage qui décrivent les transactions (ex. date, lieu, durée de l'emprunt) et des données descriptives des documents empruntés (ex. titre, auteur). L'identification de la personne peut être directe (par son nom) ou indirecte (par recoupement de plusieurs autres informations, ce qui rend la confidentialité difficile à garantir).

Données sensibles

Les données sensibles sont un type particulier de données personnelles qui, lorsque révélées, peuvent contraindre l'exercice des libertés fondamentales de l'individu identifié, le rendre vulnérable à la discrimination ou compromettre son intimité. Les données sensibles peuvent avoir trait, entre autres, à l'état de santé, à l'ethnicité, aux convictions religieuses ou philosophiques ou à l'orientation sexuelle d'un individu.

Le défi de **trouver un équilibre entre fournir un accès libre aux collections et protéger les données personnelles** des usagères et usagers existe depuis longtemps dans le secteur des bibliothèques et archives (Sillitoe, 1998 ; Iacovino et Todd, 2007). Des pressions ont déjà été exercées par des services de renseignement étatiques pour obtenir des informations sur des

personnes abonnées à des fins de sécurité publique (Boisselet, 2018). Mais les nouvelles applications numériques qui systématisent la collecte des données d'usage soulèvent de nouveaux enjeux (Mathson et al., 2007 ; Larson, 2019). Ainsi, la mise en place dans les années 2000 de points d'emprunt en libre-service en bibliothèque a fait l'objet de critiques relatives à l'atteinte de la vie privée des usagères et usagers (Bowers, 2008). Par ailleurs, la collecte de données en ligne dépasse généralement ce qui est annoncé dans les politiques de confidentialité des bibliothèques – quand celles-ci existent et qu'elles sont publiques – et l'information y est rarement intelligible (Affonso et al., 2018). **Il existe donc une asymétrie d'information** (Ytre-Arne et al., 2019) entre les bibliothèques et leurs usagères et usagers concernant la collecte des données d'usage. Il est donc nécessaire d'enquêter sur **la spécificité du contexte de BAnQ** concernant les rapports de pouvoir entre l'organisme et ses usagères et usagers, sachant que les données d'usage collectées par cette institution gouvernementale oscillent entre bien public et bien individuel.

Ce défi est d'autant plus grand qu'une partie des données utilisées par l'institution pour connaître ses publics et leurs pratiques est gérée par **des tiers fournisseurs de service** (plateformes de médias sociaux, Google Analytics, services de prêt en ligne) qui ne sont pas tous soumis aux lois québécoises et canadiennes de protection des renseignements personnels. De plus, les grandes plateformes dont BAnQ dépend pour fournir plusieurs services en ligne ont des modèles d'affaires fondés sur l'exploitation des données personnelles des usagères et usagers. Ces modèles commerciaux ont fait l'objet de nombreuses critiques au motif que leurs activités de traçage et de profilage **ne respectent pas la protection de la vie privée, qu'ils peuvent revendre les données à des tiers et qu'ils font courir un risque de surveillance** (Bennett et Parsons, 2013 ; Van Dijck, 2014).

Ainsi, la collecte, l'analyse et l'archivage responsable de ces données constituent un défi majeur pour BAnQ. Étant donné l'asymétrie d'information entre les bibliothèques et leurs usagères et usagers, il est important de prendre en compte les perceptions de ces derniers. Or celles-ci sont mal connues. **Nous posons donc la question suivante :**

Quelle est la perception des personnes abonnées aux services de BAnQ relativement à la gestion de leurs données personnelles par l'institution ? Comment caractériser leur attitude ?

Perceptions et attitudes relatives à la gestion des données personnelles

Notre projet vise à étudier les perceptions relatives à la gestion des données personnelles par BAnQ. Il s'inscrit au prolongement de nombreux travaux qui analysent les aspects sociaux de la vie privée en contexte numérique. Outre l'étude des perceptions (qui inclut la caractérisation d'attitudes et de préoccupations), d'autres enquêtes se sont plutôt penchées sur les comportements relatifs à la vie privée. Il s'agit d'un aspect connexe à ceux que nous étudions dans le cadre de notre propre projet, plutôt centré sur les préoccupations et les attitudes.

L'étude des perceptions relatives à la vie privée s'intéresse aux dispositions exprimées par les individus et les groupes en termes de degré de favorabilité ou de non-favorabilité face à certaines pratiques ou situations qui concernent leurs informations personnelles. Ces perceptions varient en fonction des contextes d'usage.

- Les travaux existants invitent à distinguer le contexte des interactions dites « horizontales », entre usagères et usagers d'un réseau social comme Facebook, et les **interactions dites institutionnelles ou « verticales »**, entre un organisme et ses usagères et usagers (Raynes-Goldie, 2010). C'est dans ce dernier cas de figure que se situent les interactions entre BAnQ et ses usagères et usagers. Plusieurs études ont montré que les perceptions des usagères et usagers varient fortement entre ces deux contextes : le risque perçu est plus grand en ce qui concerne les interactions sociales (menace de la part des autres usagères et usagers) et moins grand en ce qui concerne les institutions (Young et Quan-Haase, 2013).
- Par ailleurs, **les perceptions tendent à être différentes en fonction des types d'organisations** qui collectent et exploitent les données. Ainsi, les organismes gouvernementaux et paragouvernementaux tendent à susciter davantage confiance que les entreprises (Phoenix SPI, 2021). Toutefois, **l'enchevêtrement des pratiques des institutions publiques et des entreprises** qui leur fournissent des services (Microsoft, Amazon, Google) tend à susciter de la confusion et des préoccupations chez les usagères et usagers (Kennedy et al., 2022).
- Parmi les secteurs culturels, ceux de la **musique et de l'audiovisuel** sont les plus étudiés du point de vue de la perception de la vie privée en ligne, car ce sont dans ces secteurs que se sont d'abord développés les systèmes de profilage des usagères et usagers et de recommandation algorithmique de contenus. Dans une enquête menée par Tchéhouali (2021) au Canada, plus de la moitié des personnes sondées (61%) déclare avoir confiance quant à l'utilisation de l'historique d'écoute pour suggérer du contenu personnalisé sur les plateformes audiovisuelles et audio. Toutefois, une grande majorité de personnes sondées (81%) envisage d'augmenter son contrôle sur les données personnelles partagées avec les plateformes audiovisuelles et audio en ligne.
- Dans le secteur des bibliothèques et archives, les enquêtes précédentes ont plutôt porté sur la **perception des bibliothécaires** (Vitak, 2018 ; Tummon et McKinnon, 2018 ; Travis et Ramirez, 2020) ou sur le contexte des bibliothèques universitaires (Travis et Ramirez, 2020), mais très peu sur la perception des usagères et usagers en bibliothèque publique. La compréhension des perceptions des usagères et usagers de bibliothèque publique est donc très lacunaire.
- En ce qui concerne les enquêtes plus générales sur les perceptions relatives à la vie privée en ligne, **au Québec**, à quelques exceptions près (Dilhac et al., 2018 ; Azzaria et Zimmermann, 2019), elles sont limitées à des observations parcellaires. Quant aux enquêtes qui concernent le Canada, elles ont surtout porté sur les **préoccupations** relatives à la vie privée en général (Phoenix SPI, 2021 ; Quan-Haase et Ho, 2020). Elles montrent une augmentation du niveau de préoccupation des citoyennes et citoyens (ACEI, 2020).
- Enfin, dans une enquête menée au Royaume-Uni sur les perceptions relatives à l'usage des données personnelles par différents services publics, Kennedy et al. (2022) ont trouvé

que les usagères et usagers de ces services « peuvent reconnaître les bénéfices de l'utilisation des données, tout en exprimant certaines inquiétudes à ce sujet » (p. 18). Ces résultats montrent que **les perceptions peuvent être ambiguës, teintées à la fois d'intérêt et de préoccupation**.

- Cette ambiguïté est au cœur d'une série de travaux sur le **paradoxe de la vie privée**, soit une contradiction entre les préoccupations de confidentialité déclarées par les usagers et leurs comportements généralement peu sécuritaires (Barnes, 2006 ; Barth et al., 2019). Ce phénomène a également été abordé sous l'angle du *cynisme de la vie privée* (Hoffmann et al., 2016 ; Lutz et al., 2020), de l'*apathie de la vie privée* (Hargittai et Marwick, 2016), de la *fatigue de la vie privée* (Choi et al., 2018), de la *résignation numérique* (Draper et Turow, 2019) et du *réalisme de la surveillance* (Dencik et Cable, 2017).

La construction de typologies d'attitudes permet de saisir plus en détail la combinaison des perceptions. En effet, la notion d'attitude renvoie à un ensemble de traits caractéristiques de schémas particuliers de perception. Il existe plusieurs typologies d'attitudes relatives à la vie privée, la plus influente étant celle de Westin.

- Westin a été l'un des premiers à proposer une typologie des attitudes relatives à la vie privée dans le cadre de sondages menés aux États-Unis. Son modèle distingue trois types d'attitudes. Premièrement, les **fondamentalistes** de la protection de la vie privée sont partisans d'une protection élevée de la vie privée ; ils rejettent l'idée que l'exploitation des données apportera des bénéfices individuels et collectifs et ils sont en faveur de règles et de lois de protection de la vie privée. Deuxièmement, les **insoucieux** sont « généralement prêts à fournir leurs informations personnelles aux entreprises et aux pouvoirs publics et rejettent ce qu'ils considèrent comme un affolement injustifié autour de la protection de la vie privée » (Westin, 2003 : 445). Troisièmement, les **pragmatiques** de la protection de la vie privée « examinent les avantages de la collecte et de l'utilisation des données pour eux ou pour la société, veulent connaître les risques pour la vie privée et la manière dont les organisations se proposaient de les contrôler, puis décident s'ils doivent faire confiance à l'organisation ou demander un contrôle juridique » (Westin, 2003 : 445).
- Entre 1995 et 2001, la distribution des réponses dans les enquêtes de Westin a significativement évolué : **la part des insoucieux a drastiquement chuté de 20% à 8%, tandis que celle des fondamentalistes a augmenté de 25% à 34%**. La part des pragmatistes s'est relativement maintenue, passant de 55% à 58% (Westin, 2003). Dans une étude menée dix ans plus tard, toujours aux USA, Motiwalla et Li (2016) ont plutôt trouvé que **les pragmatiques représentaient le plus grand groupe avec 69%**, tandis que 25% étaient fondamentalistes et seulement 6% étaient insoucieux. Ces résultats suggérant que la catégorie des pragmatiques est celle qui rassemble le plus de personnes sondées concordent avec une enquête menée par Sheehan (2002).
- Elueze et Quan-Haase (2018) ont revisité la typologie de Westin dans une enquête canadienne sur les préoccupations des plus de 65 ans relativement à la vie privée. Elles ont retravaillé la typologie de Westin pour proposer six catégories d'attitudes : les *fondamentalistes* (13%), les *pragmatiques intenses* (15%), les *pragmatiques détendus* (42%), les *marginaleusement soucieux* (25%) et les *experts cyniques* (5%). Ces résultats

suggèrent que **les pragmatiques représentent la catégorie la plus fréquente avec un total de 57%**.

- Enfin, dans une enquête menée au Royaume-Uni, Kennedy et al. (2022) ont utilisé un découpage différent de celui de Westin. À partir d'une analyse de classes latentes, elles ont identifié quatre types d'attitude relatives à la vie privée dans des services publics numériques : les *critiques* représentaient 36% de leur échantillon, les *prudents* 34%, les *neutres* 13%, et ceux *en accord* 17%. Dans leurs résultats, **une grande majorité (70%) des personnes sondées tend à juger négativement les pratiques de données des institutions publiques**.
- Dans le cadre de notre étude, nous utiliserons une typologie plus ouverte pour nous permettre d'explorer les attitudes relatives aux pratiques de données de BAnQ. **Nous retenons deux polarités : une attitude *protectrice* caractérisée par le fait de vouloir limiter les pratiques de collecte, d'exploitation, de conservation et de partage de ses données personnelles, et une attitude *ouverte* caractérisée par le fait d'être favorable à ces pratiques.**

Facteurs sociodémographiques

Pour mieux comprendre la distribution de ces attitudes parmi les abonnés de BAnQ, il convient de prendre en compte l'effet des facteurs sociodémographiques sur la perception de la gestion des données personnelles en contexte numérique. Les études existantes montrent que **les principaux prédictors sociodémographiques de ces perceptions sont la scolarité (niveau de diplomation), l'âge et le niveau de littératie numérique**. Certaines études suggèrent également que l'appartenance à un groupe minorisé (LGBTQ+, minorités visibles) ou marginalisé (itinérance) peut également affecter ces perceptions.

Âge

- Le facteur de l'âge est identifié comme étant **très significatif** par la plupart des études que nous avons consultées, et **les préoccupations tendent à augmenter en fonction de l'âge**. Ainsi au Canada, selon Phoenix SPI (2021), les personnes de 55 ans et plus sont les plus préoccupées par la protection de leur vie privée : elles sont 38% à se déclarer énormément préoccupées contre 20% chez le groupe de 16 à 24 ans, soit près du double. Cette tendance est corroborée par une étude québécoise réalisée en 2019 qui rapporte que plus les personnes sondées sont âgées, plus elles tendent à prioriser la confidentialité et la protection de leur vie privée en contexte numérique (Azzaria et Zimmermann, 2019). Une enquête de McStay (2020) par sondage auprès de 2000 personnes en Grande-Bretagne va dans le même sens. Azzaria et Zimmermann (2019) soulignent par ailleurs que les écarts de perception liés à l'âge seraient davantage liés au cycle de vie plutôt qu'à des différences générationnelles.
- Toutefois, une enquête qualitative menée au Canada par Elueze et Quan-Haase (2018) auprès de personnes de 65 ans et plus invite à **relativiser l'idée que les groupes les plus âgés ont une attitude plus protectrice** de la vie privée. Les résultats montrent que les personnes sondées ne sont pas un groupe homogène de « fondamentalistes » de la

protection de la vie privée. Dans leur échantillon (n=40), seule une petite minorité montre une attitude « fondamentaliste » tandis qu'un grand nombre de personnes sondées ont une attitude « marginalement préoccupée ».

Niveau de diplomation

- Les résultats des études existantes ne permettent pas de conclure à un effet univoque du niveau de diplomation sur les perceptions relatives à la vie privée. Il apparaît que **son effet dépend des contextes d'utilisation des données personnelles**.
- Certaines études tendent à montrer que plus une personne est diplômée, plus elle est susceptible d'être préoccupée ou protectrice en ce qui concerne le partage de ses données personnelles. Ainsi, l'enquête de Phoenix SPI (2019) menée au Canada montre que plus on est diplômé, **plus on tend à refuser de fournir ses renseignements personnels** à une organisation ou à une entreprise. En corollaire, les groupes moins diplômés sont plus susceptibles de faire confiance aux entreprises en ce qui concerne le respect de leur droit à la vie privée comparé aux personnes sondées titulaires d'un diplôme universitaire (Phoenix SPI, 2021). Ces résultats vont dans le sens des travaux de Blank et al. (2014) et Sheehan (2002) qui suggèrent que les personnes ayant un niveau d'éducation plus élevé ont des niveaux plus élevés de préoccupation relative à la vie privée.
- À l'inverse, **dans certains contextes, les groupes les plus diplômés apparaissent plus susceptibles d'être ouverts à l'exploitation de leurs données personnelles et moins préoccupés** par les implications de ces pratiques au regard de la vie privée. Par exemple, ils sont plus susceptibles de vouloir communiquer des renseignements personnels en échange de rabais ou d'incitatifs, mais aussi d'accepter que le gouvernement recueille et utilise des renseignements personnels dans le cadre de ses activités de renseignement (Phoenix SPI, 2021). Une étude de Zukowski et Brown (2007) va dans le même sens : ses résultats indiquent que plus le niveau d'éducation est élevé, moins les personnes sondées sont préoccupées par la protection de leur vie privée. Dans une perspective connexe, les résultats de Bergström (2015) indiquent que la recherche d'information, le traitement des courriels et l'utilisation de cartes de débit en ligne suscitent plus d'inquiétudes auprès des individus avec moins d'éducation.
- En définitive, **il est difficile d'établir des comparaisons entre les études existantes** car la variable de la scolarisation y est traitée selon des catégories non uniformes et parce que les contextes d'utilisation des données étudiés diffèrent.

Littératie de la vie privée

- Les travaux qui portent spécifiquement sur la relation entre la littératie de la vie privée et l'attitude protectrice de la vie privée **tendent à montrer une association positive entre les deux**, bien que certaines divergences puissent être observées entre les études existantes.

- Ainsi, selon Kennedy et al. (2022), plus les usagères et usagers ont conscience de l'utilisation qui est faite de leurs données, plus ils tendent à adopter une attitude critique ou prudente au regard de la protection de leur vie privée.
- Hoffmann et al. (2016) concluent également à une association positive entre la conscience des risques relatifs à la vie privée et les préoccupations relatives à la protection de la vie privée. Par ailleurs, ils suggèrent que plus le niveau de conscience du risque est élevé, plus les utilisatrices et utilisateurs ont de chances d'adopter une attitude de résignation au regard des risques encourus.
- Par contre, une étude de Lutz et al. (2020) portant sur le cynisme de la vie privée conclut plutôt qu'il n'y a pas d'association entre les préoccupations relatives à la vie privée et la littératie numérique. Toutefois, cette étude ne portait pas spécifiquement sur la littératie de la vie privée en ligne mais sur la littératie numérique plus généralement.
- D'autres travaux, plus éloignés de notre objet de recherche, ont aussi étudié la relation entre les comportements de protection de la vie privée en ligne et la littératie numérique ou la littératie de la vie privée. Park (2013a) et Bartsch et Dienlin (2016) concluent à une association positive entre les deux, tandis que Barth et al. (2019) n'en trouvent pas.

Genre

- D'après les recherches antérieures, les femmes accordent une plus grande importance à la protection de la vie privée. Cela s'explique par le fait qu'elles sont exposées à davantage de conséquences négatives en cas de bris de confidentialité (Hoy et Milne, 2010). L'enquête de Tchéhouali (2021) qui s'est intéressée à l'utilisation de l'historique des écoutes de musique sur Spotify va dans le même sens : ses résultats suggèrent que les hommes « semblent moins préoccupés par l'encadrement des données personnelles », qu'« ils sont davantage à estimer avoir consenti librement à ce que leur historique soit utilisé » et qu'ils sont « plus confiants quant à l'utilisation de l'historique d'écoute pour leur suggérer du contenu personnalisé » (Tchéhouali, 2021 : 56).
- Par ailleurs, les études concernant les comportements (et non les perceptions) de protection de la vie privée en ligne mènent à des résultats divergents, comme le soulignent Frener et Trepte (2022) : certaines ne constatent pas de différence entre les genres (Malik et al., 2016), tandis que d'autres montrent que les femmes prennent plus de précautions (Fogel et Nehmad, 2009), et d'autres encore que les hommes adoptent des pratiques plus protectrices (Sheehan, 1999). Toutefois, un certain nombre d'études qui concernent la vie privée en ligne ont été menées dans le contexte des réseaux sociaux, et ne s'appliquent pas nécessairement au contexte d'une institution culturelle.

Groupes minorisés

- Plusieurs travaux suggèrent que l'appartenance à certains groupes minorisés peut influencer sur la perception et sur les attitudes relatives à la vie privée en ligne.

- La littérature existante suggère que **l'appartenance à un groupe marginalisé ou racisé constitue un facteur d'influence**. Une étude menée par Marwick et al. (2017) a examiné la perception du partage d'informations en ligne chez de jeunes adultes ayant un faible statut socio-économique. La majorité des participantes et participants étaient issus de groupes racisés et avaient déjà eu des expériences négatives avec les services policiers et de la surveillance dans l'espace public. Les résultats ont révélé que plusieurs d'entre eux étaient conscients des risques liés au partage de données en ligne en raison de leur expérience de surveillance par la police (Marwick et al., 2017).
- En outre, des études antérieures suggèrent que **l'affiliation au groupe LGBTQ+ expose les individus à de plus grands risques d'atteinte à la vie privée** (Robinson et al., 2020). Ces individus sont davantage susceptibles de faire l'objet d'un rejet, voire de violence, lorsqu'ils divulguent leur orientation sexuelle. Pour pallier ces risques, ils mettent en œuvre diverses stratégies de gestion de leur identité en ligne, dont le paramétrage de la confidentialité et de sécurité sur les réseaux sociaux et la restriction du partage des contenus liés à leur identité sexuelle (McConnell et al., 2018 ; Duguay, 2016 ; Taylor et al., 2014). Dans le contexte des services de bibliothèque, plusieurs travaux ont montré les enjeux spécifiques de divulgation d'informations personnelles par les usagères et usagers LGBTQ+ lorsque celles-ci révèlent leur identité sexuelle (Mathson et Hancks, 2007 ; Curry, 2005).
- **Enfin, certains groupes minoritaires avec des besoins spécifiques peuvent être plus ouverts à l'exploitation de leurs données** car ils perçoivent des bénéfices dans la création de nouveaux services adaptés. Ainsi, Kennedy et al. (2022) ont constaté dans leur enquête que les personnes sondées en situation de handicap étaient plus favorables à la réutilisation et au partage des données de santé à des fins de recherche.
- Toutefois, Guyan (2022) met en garde contre les initiatives qui visent à accumuler toujours plus de données pour corriger des biais et des inégalités de service ou de représentation dans les données, car elles peuvent causer des préjudices aux communautés qu'elles prétendent aider.

En conclusion, peu de travaux existent sur la perception de la vie privée numérique des usagères et usagers dans le contexte des bibliothèques publiques. Afin d'approfondir la compréhension des écarts existants dans les perceptions et les attitudes des utilisateurs de BAnQ relativement à la protection de la vie privée, nous posons les questions de recherche suivantes :

La perception des personnes abonnées aux services de BAnQ relativement à la gestion de leurs données personnelles par l'institution varie-t-elle en fonction des caractéristiques sociodémographiques des personnes abonnées ?

Cette perception varie-t-elle en fonction du niveau de littératie des données personnelles des personnes abonnées ?

3.5. Les cadres éthiques de gouvernance des données d'usage

Gouvernance des données d'usage

Nous définissons la gouvernance institutionnelle des données d'usage comme la mise en œuvre de processus décisionnels qui incluent la perspective et l'engagement de multiples actrices et acteurs issus de groupes diversifiés, la mise en œuvre d'un encadrement des données d'usage et la mise en débat des pratiques liées à ces données au sein de l'institution et dans la sphère publique.

Une précédente revue de littérature menée par notre équipe (Casemajor et al., 2021) a fait ressortir que les principaux enjeux de gouvernance des données d'usage soulevés concernent :

- *la représentation des publics*, car les données d'usage ne sont pas des traces objectives de l'activité humaine ; elles peuvent comporter des manques ou des biais, et construisent ainsi une image incomplète des publics et de leurs pratiques;
- *la protection de la vie privée des usagers et usagères*, étant donné la confidentialité des données personnelles qui doit être préservée, et les craintes liées à la surveillance numérique.
- *la diversité des contenus prescrits et la sérendipité* (ou libre exploration) dans la découverte de nouveaux contenus culturels, sachant que la recommandation algorithmique offre au contraire des choix déjà présélectionnés.

Principes relatifs à la protection de la vie privée

Concernant la **protection de la vie privée**, nous avons retenu ci-dessous les principes récurrents identifiés par les études existantes (voir par exemple le guide du Laboratoire de Cyberjustice, 2019). Il est à souligner que la Loi modernisant des dispositions législatives en matière de protection des renseignements personnels dans le secteur privé (Loi 25) entrée en vigueur en 2022 au Québec pose une obligation réglementaire sur plusieurs de ces principes. Il s'agit d'un rattrapage puisque l'Union Européenne a consacré dès 2016 plusieurs de ces principes dans son Règlement général sur la protection des données (RGPD). Si la Loi 25 adoptée a été bâtie sur des principes similaires à ceux du RGPD, il y existe néanmoins certaines différences entre les deux.

- **La collecte minimale de données personnelles** : la vulnérabilité intrinsèque des données personnelles conduit de nombreux analystes à recommander un principe de collecte de données minimale (Gambs, 2020). Aussi nommé principe de minimisation des

données (ou principe de nécessité), il restreint la collecte et le traitement de données personnelles à celles qui sont « adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées » (Pierre, 2023). Ce principe a été adopté dans le texte du RGPD. Dans la Loi 25, bien que la formulation adoptée diffère de celle utilisée dans le RGPD, l'article 105 de la Loi 25 stipule également que « la personne qui recueille des renseignements personnels sur autrui ne doit recueillir que les renseignements nécessaires aux fins déterminées avant la collecte » (LQ 2021, c 25, p.37).

- **La protection de la vie privée par défaut** (« *privacy by default* » ou « *privacy by design* » ; Cavoukian, 2006) : selon cette approche, offrir aux usagères et usagers l'option d'« opter pour » (*opt-in*) plutôt que l'option de refuser (*opt-out*) permet aux participantes et participants de consentir activement en choisissant de prendre part (ou non) à certains modes de collecte et de valorisation des données (Roberts et al., 2016). Le RGDP européen a rejeté dès 2016 le modèle de l'option de refus (*opt out*) pour les cookies et autres traceurs sur Internet, au motif qu'un tel modèle présuppose sans fondement un consentement par défaut des usagères et usagers (Boisson, 2021). Le modèle privilégié par l'Union européenne est plutôt celui du choix d'accepter (*opt in*) qui part du principe que le choix par défaut est celui du refus et que les usagères et usagers doivent accorder leur consentement. Ainsi, le fardeau du consentement incombe aux organisations collectant les données et non aux usagers.

En résumé, les approches « *opter pour* » et « *opter contre* » peuvent se comprendre de la manière suivante :

- L'approche « *opter contre* » mobilise la notion de consentement passif (ou tacite), tel que dans le proverbe populaire « qui ne dit mot consent ».
- L'approche « *opter pour* » fait appel à la notion du consentement actif et nécessite l'accord explicite de l'utilisateur : « sans oui, c'est non ».

Au Québec, la nouvelle Loi 25 introduit une obligation de consentement actif (*opt in*) dans le cadre de la collecte de renseignements personnels lorsque ces données visent à être exploitées par des technologies d'identification, de localisation ou de profilage des usagères et usagers (données comportementales) (CAIQ, 2023).

- **La transparence** : la Loi 25 oblige les organisations à agir face à quelconque incident pouvant mettre en danger la confidentialité des renseignements personnels qu'elles collectent ou exploitent. Si cet incident présente un risque de préjudice sérieux, l'organisation devra déclarer cet incident à la CAI (Commission d'accès à l'information). Cependant, aucune disposition n'est prévue vis-à-vis du délai exigé pour informer les usagères et usagers dont les données auraient pu être compromises par de tels incidents.

Somme toute, la Loi 25 implique un niveau de transparence et de responsabilité plus élevé de la part des organisations relativement au contexte antérieur.

- **Le droit à l'oubli** : il concerne l'obligation d'effacer des renseignements personnels après une certaine durée. La troisième phase de l'application de la Loi 25 (prévue pour 2024) prévoit l'obligation pour les organisations d'avoir mis en place les mécanismes nécessaires pour pouvoir accorder à leurs usagères et usagers le droit à l'oubli. Ce droit à l'oubli est défini comme le droit de l'usagère ou usager à exiger la cessation de la diffusion de ses renseignements personnels, la désindexation des hyperliens permettant d'accéder à ces renseignements, voire la destruction de leurs données personnelles. Cependant, la destruction de ces renseignements n'est pas automatique : il faut notamment que la finalité pour laquelle elles ont été recueillies ait été accomplie et/ou qu'il existe une raison légitime pour l'organisation de les conserver (en format anonymisé).

Certains défis sont propres au développement des systèmes d'intelligence artificielle (SIA) et sont encore peu régulés.

- Les SIA, malgré leur potentiel d'innovation, comportent des risques importants en ce qui concerne la protection des données personnelles, voire des droits de la personne (CAIQ, 2021). C'est que pour « entraîner » les systèmes d'intelligence artificielle (SIA), une quantité considérable de données - dont plusieurs de nature personnelle - est nécessaire (Gauthier, 2023). En outre, les données utilisées par les SIA peuvent être de qualité moindre, incomplètes ou biaisées, entraînant des risques réels lorsque les SIA sont utilisés pour prendre des décisions relatives à des populations ou des modèles d'intervention appliqués sur des êtres humains, sur lesquels les impacts sont bien réels (HCDH, 2021).
- Le gouvernement canadien propose un début de réglementation des SIA avec la *Loi sur l'intelligence artificielle et les données* (LIAD) qui devrait entrer en vigueur en 2025. La LIAD vise à réglementer la conception des SIA, leur mise à disposition et la gestion de leurs opérations (Institut Fasken, 2022). Elle vise à augmenter le niveau de transparence des acteurs impliqués en exigeant d'eux qu'ils divulguent des informations permettant aux usagères et usagers de mieux comprendre à quoi servent l'utilisation de leurs données personnelles, par exemple en décrivant la manière dont celles-ci sont destinées à être utilisées et le ou les types de contenus que les SIA sont censés produire à partir de ces mêmes données (Jarvie et al., 2022).

Sonder les usagères et usagers pour connaître leur point de vue

Au terme de cette revue des connaissances existantes sur les données d'usage, on a pu faire état d'enjeux importants dans le contexte des bibliothèques et archives, tels que ceux reliés à l'implantation de services pour les usagères et usagers dont la mise en œuvre implique la collecte des données d'usage, ou encore l'**asymétrie** de l'information transmise entre les organismes et leurs publics vis-à-vis de ces mêmes données. Il apparaissait donc essentiel d'**évaluer les connaissances des** usagères et usagers d'organismes culturels vis-à-vis les données d'usage, mais aussi de leur permettre de nommer leurs préoccupations et de leurs besoins vis-à-vis de la protection de leur vie privée pour jeter les bases d'une collaboration bidirectionnelle entre les organismes culturels et leurs publics quant aux directions futures des activités de collecte, d'analyse, du partage, d'archivage et d'utilisation des données d'usages. Un rééquilibrage des pouvoirs passe par une meilleure gouvernance des données d'usage, qui suppose de consulter les publics afin de connaître leurs perceptions et leurs attentes à ce sujet.

Ainsi, pour mettre en place des pratiques éthiques et durables de gouvernance des données, il est important de sonder les usagères et usagers au sujet de leurs préoccupations et de leurs besoins relativement à la protection de la vie privée (Hartman 2020). Ce projet répond au besoin crucial de mieux connaître le point de vue des différents groupes d'usagères et usagers de BAnQ afin de les inclure dans la définition de ce qui constitue de bonnes pratiques de collecte, d'analyse et d'archivage des données. Considérant la place centrale de l'historique d'emprunt dans la stratégie de BAnQ, l'enquête porte sur le point de vue des personnes abonnées (plutôt que des usagères et usagers au sens large).

4. PROBLÉMATIQUE

4.1. Questions de recherche

Les questions de recherche formulées à partir de l'examen des travaux existants sont les suivantes.

Littératie des données personnelles

- *QR 1a : Quel est le niveau de littératie des données personnelles des personnes abonnées aux services de BANQ ?*
- *QR 1b : Ce niveau de littératie des données personnelles varie-t-il en fonction des caractéristiques sociodémographiques des personnes abonnées ?*

Intérêt pour les nouveaux services

- *QR 2a : Quel est l'intérêt des personnes abonnées pour les services qui pourraient être développés à leur intention grâce aux données personnelles ?*
- *QR 2b : Cet intérêt varie-t-il en fonction des caractéristiques sociodémographiques des personnes abonnées ?*
- *QR 2c : Cet intérêt varie-t-il en fonction du niveau de littératie des données personnelles des personnes abonnées ?*

Perception des pratiques de gestion des données personnelles

- *QR 3a : Quelle est la perception des personnes abonnées relativement à la gestion de leurs données personnelles par l'institution ? Comment caractériser leur attitude ?*
- *QR 3b : Cette perception varie-t-elle en fonction des caractéristiques sociodémographiques des personnes abonnées ?*
- *QR 3c : Cette perception varie-t-elle en fonction du niveau de littératie des données personnelles des personnes abonnées ?*

Relations entre littératie, intérêt et attitude

- *QR 4a : Quelles sont les relations d'influence entre la littératie des données personnelles, l'attitude au regard de la gestion des données personnelles et l'intérêt pour les nouveaux services fondés sur les données ?*

4.2. Construction des dimensions d'analyse

Pour mesurer ces différents éléments et analyser leurs relations, l'enquête a été menée sous la forme d'un sondage des personnes abonnées aux services de BAnQ. L'étude a porté sur 25 variables organisées en 4 dimensions (voir Tableaux 3 et 4). Elles ont été définies sur la base des études existantes et adaptées au contexte spécifique de BAnQ.

1. **Littératie des données personnelles.** Cette dimension est composée de 4 sous-dimensions (A, B, C, D ; 6 variables) associées à des degrés croissants de capacités relatives aux données personnelles. Le premier degré est d'*identifier les données personnelles (A)* de façon générale (être capable de repérer une donnée personnelle parmi différents types de données, par exemple « l'historique de mes emprunts à la bibliothèque ». Le deuxième degré est de *connaître les pratiques actuelles de BAnQ (B)* relativement à la collecte de données personnelles (capacité à identifier parmi différents types de données lesquelles sont effectivement collectées par BAnQ, par exemple « la liste des documents favoris enregistrés dans mon dossier d'abonné de BAnQ »). Le troisième degré est de *comprendre les risques des données personnelles (C)* relativement à plusieurs enjeux : les données sensibles (C1) (capacité à savoir qu'à partir de l'historique d'emprunt de documents, on peut potentiellement déduire des renseignements sur l'état de santé, l'orientation politique ou l'orientation sexuelle d'une personne abonnée) ; l'anonymisation des données (C2) (capacité à savoir qu'il peut y avoir des moyens indirects de relier des informations à une personne même après l'anonymisation des données) ; et le contrôle de certaines données par des fournisseurs de services (C3) tels que Google Analytics ou Facebook (capacité à savoir que ces fournisseurs ont leurs propres règles de protection des données, hors du contrôle de BAnQ). Le quatrième degré est de *développer une réflexion critique sur la gouvernance des données personnelles (D)* (soit la capacité à identifier des initiatives pouvant contribuer à encadrer l'utilisation des données personnelles par BAnQ, par exemple donner la possibilité de refuser que ses données soient collectées).
2. **Intérêt pour le développement de nouveaux services fondés sur les données d'usage.** Cette dimension est composée de 4 sous-dimensions (E, F, G, H ; 4 variables) portant sur la perception des personnes abonnées au sujet des principaux services que BAnQ envisage de développer à court ou moyen terme à partir de l'exploitation des données d'usage. Ces services sont : *l'enrichissement des collections (E)* de la Grande bibliothèque en fonction de la demande ; *la recommandation de documents (F)* lors de recherche dans les bases de données de BAnQ en s'appuyant sur l'historique d'emprunt ; et *la personnalisation de l'expérience de navigation sur le site web de BAnQ (G)* en fonction du profil de la personne abonnée. S'y ajoute *la recommandation d'activités culturelles hors BAnQ (H)*, un service non planifié par BAnQ à ce jour, mais qui pourrait découler de la mutualisation des données d'usage avec d'autres organismes culturels.
3. **Perception en matière de protection de la vie privée.** Cette dimension se décline en 4 sous-dimensions (I, J, K, L ; 7 variables). Elle évalue le degré de favorabilité des

personnes abonnées relativement aux pratiques actuelles ou futures de BAnQ en matière de : *collecte de données personnelles (I)* (accord ou désaccord avec une collecte limitée aux besoins des services actuels) ; *exploitation des données personnelles (J)* (accord ou désaccord avec le fait de recouper différentes données au sujet des personnes abonnées pour personnaliser les services, par exemple analyser l'historique d'emprunt selon l'âge) ; *archivage par BAnQ des données personnelles anonymisées (K)* (accord ou désaccord avec le fait d'archiver ces données et avis sur la durée de conservation maximum) ; *partage des données personnelles anonymisées (L)* avec différents tiers en illustrant leurs usages possibles, tels un ministère du gouvernement québécois pour concevoir des politiques publiques (L1), une maison d'édition pour lui permettre de mieux cibler ses clients (L2), un OBNL pour faire de la sensibilisation à la lecture (L3) et des équipes de recherche universitaires, afin d'étudier les pratiques de lecture de la population québécoise (L4).

4. **Caractéristiques sociodémographiques des personnes abonnées.** Cette dimension se compose de deux séries de variables (8 variables en tout) : des variables dites *primaires* dont l'influence sur la littératie des données et sur les perceptions relatives à la vie privée ont déjà été bien documentées par de précédentes études (genre, âge, niveau de diplomation, affiliation au groupe LGBTQ+ ; Tableau 4) ; et des variables dites *secondaires*, dont l'influence est encore peu documentée dans la littérature académique (occupation principale, secteur d'emploi, revenu du ménage, situation de minorité visible ; Tableau 5). Nous avons initialement prévu d'inclure trois autres variables secondaires : le statut de citoyenneté, l'appartenance à un groupe autochtone et le revenu du ménage. Mais un trop faible nombre de personnes sondées se sont déclarées en permis de séjour temporaire et appartenant à un groupe autochtone, ce qui a empêché d'inclure ces variables dans l'analyse. La question sur le revenu du ménage a suscité un nombre trop élevé de non-réponses (29%) pour que cette variable puisse être incluse, et par ailleurs, le revenu individuel aurait été plus utile pour le mettre en rapport avec le niveau de littératie qui se situe à l'échelle individuelle. Enfin, une dernière série de variables a été récoltée à fin de contrôle de l'échantillon, sans être incluse dans l'analyse : l'identité de genre (les travaux existants montrent que cette variable n'est pas déterminante pour notre problématique) et la région de résidence (variable documentée par BAnQ dans ses études de référence internes).

Tableau 3 : Éléments analysés par dimension (1 à 3) et sous-dimension (A à L)

1. Littératie des données personnelles

- A. Identifier les données personnelles (de façon générale)
- B. Connaître la nature et l'étendue des données collectées par BAnQ au sujet des abonné.e.s
- C. Comprendre les enjeux des données personnelles relativement :
 C1) aux données sensibles,
 C2) à l'anonymisation des données personnelles,
 C3) au rôle des tiers dans la collecte et le partage des données
- D. Développer une réflexion critique sur la gouvernance des données personnelles

2. Intérêt pour le développement de nouveaux services fondés sur les données d'usage

- E. Intérêt pour enrichir les collections de BAnQ avec de nouveaux documents en fonction de la demande des publics
- F. Intérêt pour se faire recommander des documents en fonction de ses habitudes et préférences
- G. Intérêt pour se faire recommander des activités culturelles en fonction des habitudes et préférences
- H. Intérêt pour que son expérience numérique sur le site Web de BAnQ soit personnalisée

3. Position relative à la protection de la vie privée

- I. Concernant la collecte des données d'usage
- J. Concernant l'exploitation des données d'usage
- K. Concernant l'archivage des données d'usage
- L. Concernant le partage des données d'usage avec :
 L1) un ministère du gouvernement québécois, pour concevoir des politiques publiques
 L2) une maison d'édition, pour lui permettre de mieux cibler ses clients
 L3) un OBNL, pour faire de la sensibilisation à la lecture
 L4) des équipes de recherche universitaires, afin d'étudier les pratiques de lecture de la population québécoise

Tableau 4 : Dimension des caractéristiques sociodémographiques : variables primaires et choix de réponses possibles

Genre		
Homme	Femme	Autre (non-binaire)
Âge		
18-24 ans	45-54 ans	74-85 ans
25-34 ans	55-64 ans	85 ans et +
35-44 ans	65-74 ans	
Niveau de diplomation		
Aucun diplôme, certificat ou grade	Certificat ou diplôme d'apprenti, d'une école de métiers, d'un collège, d'un cégep ou d'un autre établissement non universitaire	Baccalauréat
Diplôme d'études secondaires ou attestation d'équivalence (DES)	Diplôme ou grade universitaire inférieur au baccalauréat (certificat, mineure, majeure)	Diplôme ou grade universitaire supérieur au baccalauréat (DESS, maîtrise, doctorat)
Affiliation au groupe LGBTQ+		
Oui	Non	

Tableau 5 : Dimension des caractéristiques sociodémographiques : variables secondaires et choix de réponses possibles

Occupation principale		
Travailleur.euse rémunéré.e (incluant congé de maladie, parental, etc.)	Travailleur.euse à son compte	Aux études
Sans emploi ou à la recherche d'emploi	Au foyer (s'occupe des enfants, mais non retraité ni en congé)	Retraité.e (Statistiques Canada)
Secteurs d'emploi		
Membres des corps législatifs et cadres supérieur.e.s	Affaires, finance et administration	Sciences naturelles et appliquées et domaines apparentés
Secteur de la santé	Enseignement, droit et services sociaux, communautaires et gouvernementaux	Arts, culture, sports et loisirs
Vente et services	Métiers, transport, machinerie et domaines apparentés	Ressources naturelles, agriculture et production connexe
Fabrication et services d'utilité publique		
Niveau de revenu du ménage		
Moins de 30 000 \$	30 000 \$ - 49 999 \$	50 000 \$ - 69 999 \$
70 000 \$ - 89 999 \$	90 000 \$ - 124 999 \$	125 000 \$ - 199 999 \$
200 000 \$ ou plus		

Appartenance à une minorité visible	
Oui	Non
Appartenance à un groupe autochtone	
Oui	Non
Statut de citoyenneté	
Citoyen.ne canadien.ne	Résident.e permanent.e
Permis temporaire (travail, études)	Réfugié.e.s

Tableau 6 : Variable à fin de contrôle

Région de résidence		
Abitibi-Témiscamingue	Bas-Saint-Laurent	Capitale-Nationale
Centre-du-Québec	Chaudière-Appalaches	Côte-Nord
Estrie	Gaspésie-Îles-de-la-Madeleine	Lanaudière
Laurentides	Laval	Mauricie
Montréal	Montréal	Nord-du-Québec
Outaouais	Saguenay-Lac-Saint-Jean	Autre province ou autre pays

5. DÉMARCHE MÉTHODOLOGIQUE

L'enquête a pris la forme d'un sondage par questionnaire auto-administré en ligne et sur place à la Grande Bibliothèque de BAnQ. Elle ciblait les personnes adultes abonnées aux services de BAnQ. Le sondage a permis de récolter 1430 réponses analysables. L'échantillon est non-probabiliste : il n'est pas strictement représentatif de la population des personnes abonnées aux services de BAnQ, néanmoins sa taille et sa diversité fournissent un portrait significatif des connaissances et des perceptions des personnes abonnées actives au moment du sondage et intéressées par le sujet de la gouvernance des données d'usage.

5.1. Population cible

Description

La population cible du sondage est formée par les personnes abonnées aux services de BAnQ âgées de 18 ans et plus. Les personnes sondées devaient être en mesure de répondre à un sondage en français³.

Plutôt que les usagères et usagers au sens large (toute personne qui utilise les services de BAnQ), l'enquête cible les personnes abonnées car ce sont elles qui sont le plus concernées par les pratiques de collecte de données (via l'historique d'emprunt notamment). Il y a deux types de personnes abonnées : celles qui sont abonnées à tous les services (incluant les emprunts sur place à la Grande Bibliothèque), et celles qui sont abonnées aux services numériques seulement. Les deux catégories sont incluses dans la population cible. Les usagères et usagers des archives et de la collection nationale (essentiellement des individus œuvrant en recherche ou en situation professionnelle) ne sont pas ciblés. Parmi les personnes abonnées, il convient aussi de distinguer celles qui sont actives (ayant fait une action d'utilisation des services dans la dernière année) et celles qui sont non actives.

Études de référence

La littérature montre que les usagères et usagers de BAnQ sont très diversifiés (Beauchemin et al., 2020), allant des jeunes aux personnes âgées, incluant francophones, anglophones, personnes nouvellement arrivées et Autochtones, en situation d'étude, de pratique amateur, d'activité professionnelle ou d'itinérance. En ce qui a trait au volet Archives, une majorité d'utilisatrices et utilisateurs sont des spécialistes, en particulier des chercheuses et chercheurs

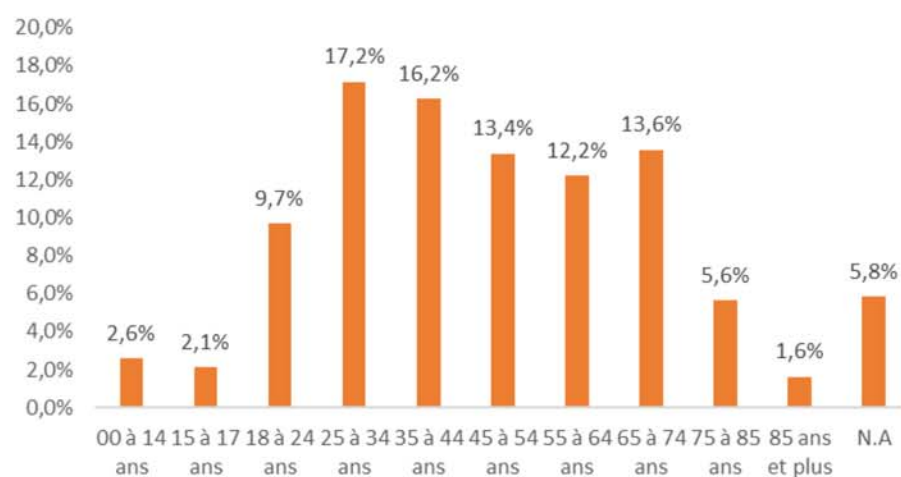
³ Conformément à la politique institutionnelle de BAnQ, l'ensemble du contenu du site web de l'institution est rédigé en français.

et des généalogistes (professionnels et amateurs) (Casemajor, 2009). En termes de fréquentation, en 2018, la Grande Bibliothèque comptait 2 166 734 visites physiques annuelles et 7 094 773 visites virtuelles annuelles (ISQ, 2020). Certains groupes d'usagères et usagers utilisent les services de BAnQ sans abonnement, comme les usagères et usagers ponctuels ou les personnes en situation d'itinérance. Aucune statistique n'est toutefois disponible au sujet des usagères et usagers qui ne sont pas abonnés.

BAnQ compile des statistiques sur les personnes abonnées grâce aux données récoltées via son formulaire d'abonnement. Au moment du sondage (août 2022), les personnes abonnées (tous types d'abonnement, enfants et adultes) représentaient environ **405 000 personnes**⁴. Trois variables sont documentées par BAnQ sur ses abonnées et abonnés : l'âge, le sexe et la région d'habitation (portrait sociodémographique de mars 2022). Comparé à la population du Québec, la population des personnes abonnées aux services de BAnQ présente une certaine sous-représentation des hommes. Une grande part des personnes abonnées est québécoise, avec près de 40% d'entre elles établies dans la région de Montréal.

Quant aux **personnes abonnées actives** (ayant accédé au moins une fois aux services dans la dernière année précédant le sondage), elles représentaient environ 195 000 personnes (enfants et adultes) au moment du sondage. La tranche des 25 à 44 ans représente une part importante des personnes abonnées actives, soit environ le tiers (33,4%). Les 44 à 74 ans représentent l'autre part essentielle des personnes abonnées actives (39,2%). Après 75 ans, on note une diminution importante de la part des personnes abonnées actives (Tableau 7). Bien qu'il soit possible d'imaginer qu'une grande part des personnes sondées soient abonnées actives, certaines ont pu être rejointes via les médias sociaux ou l'infolettre de BAnQ sans avoir été actives dans les services de BAnQ durant la dernière année.

Tableau 7 : Répartition par âge des personnes abonnées actives pour la période août-septembre 2022 (avec non-réponses)



⁴ En mars 2022, soit cinq mois avant le sondage, la répartition était de 31% pour les abonnements tous services et 69% pour les abonnements à distance.

5.2. Paramètres du sondage

Questionnaire

Le questionnaire (voir Annexe 1) comportait 23 questions de formats variés : choix multiples ; vrai ou faux ; choix dans échelle de Likert en cinq points indiquant si les personnes sondées sont d'accord ou non avec une série d'affirmations ; question ouverte en fin de sondage (commentaires libres). Il était découpé en 5 sections : « préambule » (identification préalable) ; « connaître vos données » (littératie des données personnelles) ; « votre point de vue » (intérêt pour le développement de nouveaux services et perception au regard de la protection de leur vie privée) ; « votre profil » (profil sociodémographique) ; et « commentaires ». Après avoir répondu à chaque question de la section sur la littératie des données personnelles, les personnes sondées voyaient s'afficher à l'écran la bonne réponse avec une phrase explicative. La progression dans le questionnaire était donc informée par une connaissance éclairée des réponses aux questions passées. Une question sur les données ouvertes a été ajoutée par BAnQ pour répondre à un besoin ciblé de connaissances sur ce sujet. Le sondage a été programmé sur le logiciel SurveyGizmo d'Alchemer.

Pré-test

Un pré-test du sondage a été conduit auprès de 30 personnes entre le 19 et le 23 mai 2022. Le dispositif visait à recréer le plus fidèlement possible les conditions de participation au sondage subséquent. Vingt personnes ont été recrutées à la Grande Bibliothèque par des membres de l'équipe ; elles ont répondu sur deux postes informatiques dédiés sur place. Dix personnes ont été recrutées via une liste de contacts personnels des membres de l'équipe ; elles ont répondu sur leur ordinateur personnel.

La phase de pré-test a permis de recueillir des informations sur la formulation des questions ; la qualité de l'interface ; la durée de réponse ; les questions et inquiétudes éventuellement soulevées. Une compensation leur a été offerte pour leur participation, sous la forme d'un objet de la boutique de BAnQ.

Recrutement et administration du sondage

Le sondage a été diffusé aux personnes abonnées à BAnQ via l'infolettre institutionnelle, via son site web et ses médias sociaux entre le 15 août et le 11 septembre 2022 (pour une période totale de 28 jours). Le tirage au sort d'un prix (une liseuse) a été organisé pour encourager la participation. Le comité éthique de l'INRS a délivré un certificat éthique pour l'enquête par sondage.

Le format du sondage en ligne assure une bonne visibilité auprès d'une large diversité de personnes abonnées à travers les différentes régions du Québec. Il a l'avantage de rejoindre rapidement et à moindre coût les personnes abonnées. Mais ce mode de collecte de données a tendance à sous-représenter les populations moins à l'aise avec les outils numériques (Gingras et Belleau, 2015). Selon l'expérience de sondage de l'équipe de BAnQ, le groupe des 55-75 ans tend à être sur-représenté parmi les personnes sondées en ligne. Par ailleurs, selon les précédentes expériences de sondage de BAnQ, les personnes sondées sur place à la Grande Bibliothèque tendent à avoir un niveau de scolarisation moins élevé que les personnes sondées en ligne.

Deux modes complémentaires d'administration du sondage ont été utilisés : en ligne et sur place à la Grande Bibliothèque.

Mode en ligne

L'invitation à participer au sondage a été diffusée auprès des personnes abonnées aux services de BAnQ par les voies suivantes : une bannière d'information cliquable au sujet du sondage a été publiée sur la page d'accueil du site web de BAnQ ; une affichette assortie d'un court message a été publiée sur la page Facebook de BAnQ (57 600 abonnés) et sur son compte Twitter (36 200 abonnés) ; la bannière a aussi été publiée dans la signature des courriels des employés de BAnQ.

Mode sur place à la Grande Bibliothèque

Les usagers et usagers abonnés étaient invités à répondre au sondage sur deux postes informatiques dédiés au sondage dans l'espace d'accueil au rez-de-chaussée de la Grande Bibliothèque. BAnQ a diffusé une affiche d'information dans le hall d'accueil et dans les ascenseurs. Un panneau imprimé a aussi été installé à côté des ordinateurs dédiés au sondage au rez-de-chaussée.

Traitement et analyse des données

Les réponses au sondage ont été collectées et archivées par l'équipe de BAnQ. Seule une version anonymisée des données a été transmise à l'équipe de recherche. Après traitement et validation, les données ont été analysées via le logiciel SAS pour :

- mesurer des scores d'échelle et des niveaux pour les trois dimensions ;
- identifier les relations entre les scores d'échelles ou niveaux et les caractéristiques sociodémographiques, à l'aide d'un modèle de régression logistique ordinaire (ce modèle permet de mesurer l'effet net de certaines caractéristiques sur les probabilités d'avoir un intérêt élevé pour chacune des sous-dimensions, par rapport à un groupe de référence) ;
- identifier les relations entre les dimensions, toujours à l'aide d'un modèle de régression logistique ordinaire.

5.3. Échantillon

Taille et variables de contrôle

Le sondage a permis de récolter **1430 réponses** analysables (personnes sondées ayant consenti, de plus de 18 ans, abonnées et non employées de BAnQ⁵). De par la structure du sondage (volontaire, auto-administration en ligne et sur place à la Grande Bibliothèque), l'échantillon est non-probabiliste. Il n'est donc pas strictement représentatif et comporte certains biais de sélection et biais de couverture. Par exemple, s'auto-excluent celles et ceux que la question n'intéresse pas.

Le type d'abonnement et la localisation ont été collectés afin d'évaluer la composition de l'échantillon en la comparant avec la population de personnes abonnées telle que documentée par BAnQ dans ses statistiques internes (informations collectées au moment de l'abonnement). Ces variables de contrôle ne sont pas prises en compte dans l'analyse.

- Environ la moitié des personnes sondées sont abonnées à tous les services de BAnQ, l'autre moitié étant abonnée aux services en ligne seulement. Les deux groupes ont été inclus dans l'analyse sans distinction.
- Une proportion de 40% des personnes sondées ont accédé au sondage depuis la région administrative de Montréal. Ce nombre est proche des 39% de la population des personnes abonnées vivant sur l'île de Montréal⁶ tel que documenté par BAnQ.

L'examen des variables de contrôle nous permet donc de juger satisfaisante la composition de notre échantillon sur les deux points. Bien que notre échantillon non-probabiliste ne soit pas représentatif de la population générale des personnes abonnées aux services des bibliothèques de BAnQ ni de la population québécoise, sa taille et la diversité des profils des personnes sondées fournissent néanmoins des indications significatives sur leurs connaissances et leurs perceptions en matière de gouvernance des données personnelles.

Variables d'analyse

Neuf variables sociodémographiques ont été collectées. De ce nombre, **six sont exploitables** : l'âge, le niveau de diplomation, l'occupation principale dans la dernière année, le secteur d'emploi, le statut de minorité visible et l'affiliation au groupe LGBTQ+.

⁵ 108 réponses de membres du personnel de BAnQ ont par ailleurs été collectées. À ce stade, en cohérence avec les objectifs initiaux, seules les réponses des personnes abonnées non employées ont été analysées. En effet les membres du personnel constituent un groupe à part (du point de vue de sa démographie et de sa littératie) qui pourra éventuellement être analysé dans une phase ultérieure de la recherche.

⁶ Les deux mesures sont toutefois différentes : les résultats du sondage documentent le lieu d'où on se connecte au sondage en ligne, tandis que les statistiques internes documentent le lieu de résidence. Malgré tout, la proximité de ces deux mesures donne une indication sur la localisation géographique des personnes abonnées.

Âge

Notre échantillon présente une **forte surreprésentation des groupes plus âgés** (65,5% de 55 à 84 ans et plus) comparé à la population des personnes abonnées (27,9%), soit plus du double (Tableau 8). En parallèle, les classes d'âge les plus jeunes sont sous-représentées dans notre échantillon (21,1% de 18-44 ans) comparé à la population des personnes abonnées (39,6%). La surreprésentation des groupes plus âgés (et notamment les personnes retraitées qui ont davantage de temps libre pour répondre à des sondages) est un biais courant des sondages en ligne auto-administrés, tel que documenté dans la littérature académique (Gingras et Belleau, 2015). La période d'administration du sondage (durant l'été) peut aussi avoir eu pour effet de limiter la participation des jeunes, qui tendent à fréquenter davantage la Grande bibliothèque durant la période scolaire.

Tableau 8 : Composition de l'échantillon selon l'âge

Âge	Échantillon de l'étude (2022)	Abonné.e.s de BAnQ (mars 2022)	Population du Québec (Statistique Canada, 2021) ⁷
		18 ans et plus (N = 592 430)	18 ans et plus (N = 6 850 675)
18-24 ans	3,6	10,5	9,2
25-34 ans	7,8	15,7	15,5
35-44 ans	9,7	13,4	16,4
45-54 ans	12,4	11,6	15,2
55-64 ans	23,6	11,3	18,1
65-74 ans	31,7	11,5	14,6
75-84 ans	10,5	5,1	7,9
85 ans et plus	0,8	1,4	3,1
n.d.	/	19,4	/

⁷ Statistique Canada (2022), Profil du recensement, Recensement de la population 2021, <https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/index.cfm?Lang=F>

Genre

En ce qui concerne le genre, notre échantillon présente une certaine **surreprésentation des hommes et une sous-représentation des femmes** mais une incertitude demeure relativement aux non-réponses dans l'enquête de référence de BAnQ (Tableau 9). La proportion d'hommes de notre échantillon (49,4%) est sensiblement plus élevée que dans la population des personnes abonnées (36,9%) ; elle est plutôt similaire à la population du Québec (49,4%). On note par ailleurs une sous-représentation des femmes dans notre échantillon (44,9%) comparé à la population des personnes abonnées (52,7%). Toutefois, la proportion élevée de non-réponses à la question du genre dans l'enquête de référence de BAnQ (10,4%) rend incertaine la lecture de cette variable dans la composition de l'échantillon.

Tableau 9 : Composition de l'échantillon selon le genre

	Échantillon de l'étude (2022)	Abonné.e.s de BAnQ (mars 2022)	Population du Québec (ISQ, 2021) ⁸
Genre		18 ans et plus (N = 592 430)	15 ans et plus (N = 4 977 595)
Homme	49,4%	36,9%	49,24%
Femme	44,9%	52,7%	50,53%
Autre	0,9%	/	0,23%
n.d.	4,8%	10,4%	

⁸ Institut de la statistique du Québec, *Vitrine statistique*. Résultats adaptés par l'ISQ à partir du Recensement de 2021 de Statistique Canada.

Niveau de diplomation

BAnQ ne documente pas cette variable à l'interne. Comparé à la population du Québec, l'échantillon présente une forte surreprésentation des **personnes diplômées universitaires** (67,5% contre 27,2%) (Tableau 10). Le groupe qui possède un diplôme ou grade universitaire **supérieur au baccalauréat** est particulièrement surreprésenté (31,4% dans l'échantillon, 8,9% dans la population du Québec). Parallèlement, les personnes non-diplômées sont fortement sous-représentées dans l'échantillon (1,3%) comparé à la population du Québec (18,2%). Cet écart pourrait en partie s'expliquer par le fait que le groupe des 15-17 (majoritairement non diplômés) ne fait pas partie de notre échantillon. Par ailleurs, on peut supposer que cet écart reflète le fait que la communauté étudiante universitaire fréquente généralement plus que les autres groupes la Grande bibliothèque dans le cadre de travaux scolaires.

Tableau 10 : Composition de l'échantillon selon le niveau de diplomation

Niveau de diplomation	Échantillon de l'étude (2022)	Population du Québec (Statistique Canada, 2021)
		15+ ans (N = 6 918 730)
Aucun diplôme, certificat ou grade	1,3%	18,2%
Diplôme d'études secondaires ou attestation d'équivalence	11,2%	21,4%
Certificat ou diplôme d'apprenti, d'une école de métiers, d'un collège, d'un cégep ou d'un autre établissement non universitaire	20,1%	33,2%
Diplôme ou grade universitaire inférieur au baccalauréat (certificat, mineure, majeure)	11,0%	3,7%
Baccalauréat	25,1%	14,6%
Diplôme ou grade universitaire supérieur au baccalauréat	31,4%	8,9%

Quelques statistiques produites dans d'autres contextes nous permettent de penser que la surreprésentation des personnes diplômées universitaires parmi les personnes abonnées aux services des bibliothèques publiques est courante. Un sondage sur la notoriété de BAnQ, commandé par l'institution et réalisé par Léger Marketing en 2021 (inédit)⁹, s'appuie sur un échantillon représentatif de 1001 Québécoises et Québécois parlant français ou anglais et âgés de 18 ans ou plus. Le sondage a permis de déterminer que :

- 4 personnes sondées sur 10 (41%) affirment connaître BAnQ, ne serait-ce que pour en avoir entendu parler. La notoriété de BAnQ est notamment plus forte chez les personnes ayant un diplôme de niveau universitaire (61%).
- Le taux de fréquentation des lieux et de participation aux différentes activités de BAnQ est plus élevé chez les Québécoises et Québécois détenant un diplôme universitaire.

De plus, un sondage sur les habitudes des usagères et usagers des bibliothèques de Montréal, commandé par la Ville et réalisé par Léger Marketing en 2015 (inédit)¹⁰, s'appuie par ailleurs sur un échantillon représentatif de 600 personnes sondées. De ces usagères et usagers rencontrés sur le terrain de trois bibliothèques publiques :

- 27% détenaient un diplôme d'études primaire ou secondaire ;
- 16% détenaient un diplôme d'études collégial ;
- 57% détenaient un diplôme d'études universitaires.

Par ailleurs, l'*Enquête sur les pratiques culturelles de Français*, menée en 2018 auprès d'un échantillon représentatif de 9200 personnes, permet de constater à travers ses diverses éditions depuis 1988 l'évolution et le maintien de la fréquentation supérieure des bibliothèques publiques par les personnes sondées détenant des études universitaires (BAC ou équivalent et études supérieures). Chacune des 4 éditions menées ainsi de 1988 à 2018 révèle un pourcentage plus élevé d'usagères et usagers détenant un diplôme d'études supérieures, par ailleurs surreprésentés relativement à la population générale.

Enfin, en termes de scolarisation, les données de l'Institut de la statistique du Québec (2020) montrent que les femmes obtiennent en plus grande proportion un diplôme universitaire. On peut donc supposer, étant donné la surreprésentation des femmes dans la population de BAnQ, que **les personnes diplômées universitaires sont aussi surreprésentées** parmi les personnes abonnées aux services de BAnQ.

Ces statistiques concernant BAnQ, les bibliothèques publiques de Montréal et les bibliothèques publiques françaises tendent toutes à confirmer que **parmi les personnes abonnées aux bibliothèques publiques, les personnes détenant un diplôme d'études supérieures sont surreprésentées relativement à la population générale. Notre échantillon concorde sur ce point.**

⁹ Transmis directement à l'équipe de recherche par BAnQ.

¹⁰ Transmis directement à l'équipe de recherche par BAnQ.

Occupation

Notre échantillon présente **un fort taux de personnes retraitées (près de la moitié) et un très faible taux de personnes aux études** (5,5% (Tableau 11)). Cette variable n'est pas documentée à l'interne par BAnQ, mais nous pouvons penser qu'il s'agit d'un biais de notre échantillon concomitant à celui identifié sur la variable de l'âge (forte sous-représentation des jeunes). Encore une fois, la période d'administration du sondage (à la fin de l'été, hors période scolaire) a probablement nuit au recrutement de personnes aux études. En effet, selon une étude interne de BAnQ, les visites en personne et l'utilisation des espaces de la Grande Bibliothèque et de la Bibliothèque nationale (18%) sont particulièrement plus élevés chez les personnes aux études (BAnQ / Léger Marketing, 2021). Toutefois, les personnes qui fréquentent les espaces de BAnQ ne sont pas nécessairement des personnes abonnées.

Tableau 11 : Composition de l'échantillon selon l'occupation

	Échantillon de l'étude (2022)
Occupation principale dans la dernière année	
Au foyer	0,7%
Aux études	5,5%
Retraité	47,8%
Sans emploi ou à la recherche d'emploi	1,7%
Travailleur.euse autonome	8,6%

Secteur d'emploi

Il n'existe pas de comparatif avec la population des personnes abonnées aux services de BAnQ, cette variable n'étant pas documentée dans les statistiques internes de BAnQ. Comparé à la population du Québec (Tableau 12), l'échantillon présente une forte surreprésentation des

secteurs « Enseignement, droit et services sociaux, communautaires et gouvernementaux » (26,6%) et **« Arts, culture, sports et loisirs »** (12,9%). En parallèle, l'échantillon présente une forte sous-représentation des secteurs « Vente et services » (7,2%) et « Métiers, transport, machinerie et domaines apparentés » (4,5%). On peut supposer que les métiers liés aux secteurs surreprésentés (notamment en enseignement, art et culture) nécessitent de consulter plus fréquemment les ressources documentaires des bibliothèques.

Tableau 12 : Composition de l'échantillon selon le secteur d'emploi

Secteur d'emploi	Échantillon de l'étude (2022)	Population du Québec (Statistique Canada, 2021) 15+ ans (N = 6 918 730)
Membres des corps législatifs et cadres supérieur.e.s	1,4%	1.8%
Affaires, finance et administration	13,1%	17.6%
Sciences naturelles et appliquées et domaines apparentés	4,6%	8.1%
Secteur de la santé	7%	8.1%
Enseignement, droit et services sociaux, communautaires et gouvernementaux	26,6%	12.6%
Arts, culture, sports et loisirs	12,9%	3.3%
Vente et services	7,2%	24%
Métiers, transport, machinerie et domaines apparentés	4,5%	15.8%
Ressources naturelles, agriculture et production connexe	1,7%	2%
Fabrication et services d'utilité publique	3,6%	4.7%
n.a	17,2%	2%

Minorité visible

Les personnes se déclarant appartenir à une minorité visible représentent **6,4% de l'échantillon**¹¹. Cette variable n'est pas documentée à l'interne par BAnQ. Parmi la population du Québec, les minorités visibles représentent 16,1%, soit un taux supérieur à notre échantillon (Tableau 13). Toutefois, il n'existe que très peu de données sur la fréquentation des bibliothèques publiques par les minorités visibles au Québec. Certaines études existent aux États-Unis, mais elles ne sont pas transposables dans le contexte québécois étant données les importantes différences populationnelles, sociales et historiques entre les deux territoires. En ce qui concerne le contexte canadien, les recherches existantes indiquent que « les jeunes racialisés participent moins aux programmes et profitent moins des services des bibliothèques publiques » que les autres jeunes (Matthews, 2021 : 107 ; Kumasi et Hughes-Hassell, 2017). **Il est donc possible de penser que la sous-représentation des minorités visibles dans notre échantillon reflète une tendance à la plus faible représentation de ce groupe parmi les personnes abonnées aux services de BAnQ.**

Tableau 13 : Composition de l'échantillon selon l'appartenance à une minorité visible

	Échantillon de l'étude (2022)	Population du Québec (Statistique Canada, 2021)
Minorité visible		15+ ans (N = 8 308 480)
Non	93,6%	83,9%
Oui	6,4%	16,1%

¹¹ Cette question a généré 11% de non-réponses.

Groupe LBGTQ+

Les personnes se déclarant affiliées au groupe LBGTQ+ représentent **10% de l'échantillon**¹². Là encore, cette variable n'est pas documentée à l'interne par BAnQ. Au Canada, le groupe LBGTQ+ représente 4% de la population (Tableau 14). Notre échantillon présente donc un taux plus fort de représentation de ce groupe comparé à la population du Québec. Toutefois, en l'absence de statistiques sur la fréquentation des bibliothèques publiques par ce groupe, il est difficile de savoir si cette surreprésentation dans notre échantillon reflète une tendance parmi les personnes abonnées aux services de BAnQ.

Tableau 14 : Composition de l'échantillon selon l'affiliation au groupe LBGTQ+

	Échantillon de l'étude (2022)	Population du Canada (Statistique Canada, 2018)
Minorité sexuelle		15+ ans
Non	90%	96%
Oui	10%	4%

¹² Cette question a généré 13% de non-réponses.

Test de redressement de l'échantillon

Malgré les lacunes dans les données de référence, nous avons testé le redressement de l'échantillon basé sur les variables de l'âge et du sexe (voir Annexe 2). À première vue, le redressement n'a **pas d'impact sur les scores globaux d'intérêt ou de perception de la vie privée**. On observe un **léger effet sur le score moyen de littératie** des données personnelles, qui augmente de 8,0 à 8,5. Cet écart va dans le même sens que les résultats avant redressement (ces résultats montrent que le score de littératie est plus élevé chez les jeunes).

Par contre, le redressement produit un **effet collatéral sur d'autres variables d'analyse**. Les jeunes de 18-34 ans étant peu nombreux dans l'échantillon, le redressement implique de donner beaucoup de poids aux réponses de ce groupe. Leurs autres caractéristiques sociodémographiques s'en trouvent exacerbées. C'est notamment le cas de la variable de l'affiliation au groupe LGBTQ+ : elle passe chez ce groupe de 10% (déjà passablement élevé) à 17%. Il est raisonnable de penser que ce redressement du point de vue de l'âge et du genre introduit un autre biais du point de vue du groupe LGBTQ+. De plus, on aurait pu s'attendre à ce que le redressement fasse baisser un peu la surreprésentation des personnes diplômées universitaires, ce qui n'est pas le cas, au contraire celles-ci voient leur proportion passer de 67% à 70%.

Bilan du test de redressement

Le redressement a posteriori de notre échantillon produirait les effets suivants :

- pas d'impact sur les scores globaux d'intérêt ou de perception ;
- un faible impact sur le score moyen de littératie des données personnelles (qui va dans le même sens que les résultats avant redressement) ;
- l'introduction possible d'une nouvelle série d'erreurs sur les variables de l'âge et du genre, étant donné les lacunes des données de référence sur ces deux variables ;
- l'introduction probable d'un nouveau biais sur la variable de l'affiliation au groupe LGBTQ+.

Étant donné que la somme des inconvénients dépasse celle des bénéfices, **il apparaît contre-indiqué de procéder au redressement de l'échantillon.**

6. RÉSULTATS DÉTAILLÉS

6.1. Littératie des données personnelles : mesure de score et de niveaux

Mode de collecte et d'analyse

Chaque sous-dimension de la littératie des données personnelles (A, B, C, D ; Tableau 15) a été mesurée par une question, à l'exception de la sous-dimension C (comprendre les enjeux) qui a été mesurée par trois questions pour distinguer les enjeux relatifs aux données sensibles (C1), à l'anonymisation des données personnelles (C2), et au rôle des tiers dans la collecte et le partage des données (Cc3). Les questions des sous-dimensions A, B et D se présentaient sous un format de questions à choix multiple (3 bonnes réponses et 1 à 2 mauvaises réponses). Les questions de la dimension C se présentaient sous le format vrai ou faux. L'option « je ne sais pas » était proposée à chaque question.

La littératie des données personnelles a été mesurée par un score attribué à chaque participante ou participant en fonction de sa capacité à choisir les bonnes réponses aux questions posées. Chaque sous-dimension a été mesurée par un score de 0 à 3 points, pour un score total entre 0 et 12 points. Un score total de 0 points correspond au plus faible niveau de littératie des données personnelles (pour une personne sondée qui n'aurait choisi aucune bonne réponse). Un score de 12 points correspond à l'inverse au plus haut niveau de littératie des données personnelles (pour une personne sondée qui aurait choisi toutes les bonnes réponses). Voir le Tableau 16 pour le détail des fréquences de réponses sur le score total. Sur la base de ces résultats et du découpage par quartile, les scores totaux ont ensuite été répartis par niveaux de littératie des données personnelles : niveau très bas (0 à 3 points), bas (4 à 6 points), moyen (7 à 8 points) et élevé (9 à 10 points) et très élevé (11 à 12 points).

Tableau 15 : Littératie des données personnelles : questions du formulaire et mesures de score

A		Identifier les données personnelles (de façon générale)
(Q1)	0-3 pts	<i>À votre avis, quelles informations parmi les suivantes constituent des données personnelles ?</i>
B		Connaître la nature et l'étendue des données collectées par BAnQ au sujet des abonné.e.s
(Q2)	0-3 pts	<i>Selon vous, lesquelles de ces informations vous concernant sont actuellement collectées par BAnQ ?</i>
C		Comprendre les enjeux des données personnelles relativement :
(Q3)	0-1 pt	C1) aux données sensibles <i>Vrai ou faux ? À partir de votre historique d'emprunt de documents, on peut déduire des renseignements sur votre état de santé, votre orientation politique ou votre orientation sexuelle.</i>
(Q4)	0-1 pt	C2) à l'anonymisation des données personnelles <i>Vrai ou faux ? Quand BAnQ rend anonymes vos données personnelles, il n'existe aucun moyen de vous relier à nouveau à l'information en question.</i>
(Q5)	0-1 pt	C3) au rôle des tiers dans la collecte et le partage des données <i>BAnQ utilise divers services Web (ex. : Google Analytics, Facebook) qui collectent également des données à votre sujet. Est-il vrai ou faux que ces fournisseurs de services Web sont soumis aux mêmes règles que BAnQ concernant la protection de vos données ?</i>
D		Développer une réflexion critique sur la gouvernance des données personnelles
(Q6)	0-3 pts	<i>À votre avis, quelles initiatives pourraient le mieux contribuer à encadrer l'utilisation de vos données personnelles par BAnQ ?</i>
Score total de 0 à 12 points (0 = niveau très bas ; 12 = niveau très élevé)		

Niveaux

Le score moyen est de 8 (sur 12) et le score médian se situe également à 8, ce qui signifie que la moitié des personnes sondées ont un score de littératie des données personnelles inférieur à 8 et l'autre moitié obtient un score supérieur à 8.

On observe à la Figure 1 une distribution des fréquences décalée vers les niveaux moyen à élevé. Elle culmine au score de 10 points, au niveau élevé (voir le Tableau 16 (1-total) pour la distribution détaillée des fréquences).

La moitié des personnes sondées (53%) se situe entre les niveaux très bas à moyen. L'autre moitié (47,6%) se situe entre les niveaux élevés à très élevés. Seules 5,8% des personnes sondées ont un niveau très bas, mais près du quart (26,1%) se situe entre les niveaux bas à très bas. Une majorité des personnes sondées, soit près de 60%, se situent entre les niveaux moyen (26,9%) à élevé (32,7%). Le niveau très élevé rassemble 14,9% des personnes sondées.

Figure 1 : Distribution du score total de littératie des données personnelles par niveau (bas, moyen, élevé, très élevé)

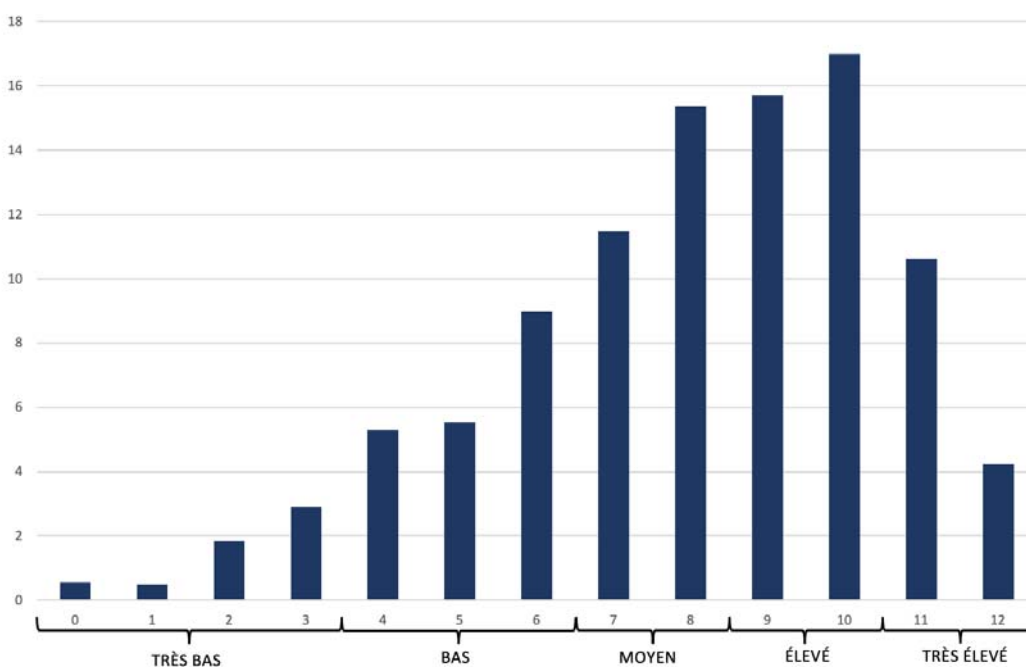


Tableau 16 : Fréquences et pourcentage des réponses pour le score total de littératie des données personnelles

Niveau de littératie	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé	Pctage groupé
0 (très bas)	8	0,6	8	0,6	
1	7	0,5	15	1,1	
2	26	1,8	41	2,9	5,8
3	41	2,9	82	5,8	
4 (bas)	75	5,3	157	11,1	
5	78	5,5	235	16,6	20,3
6	127	9,0	362	25,6	
7 (moyen)	162	11,5	524	37,1	26,9
8	217	15,4	741	52,4	
9 (élevé)	222	15,7	963	68,2	
10	240	17,0	1203	85,1	32,7
11 (très élevé)	150	10,6	1353	95,8	14,9
12	60	4,3	1413	100,0	

Fréquence manquante = 17

Score par sous-dimension

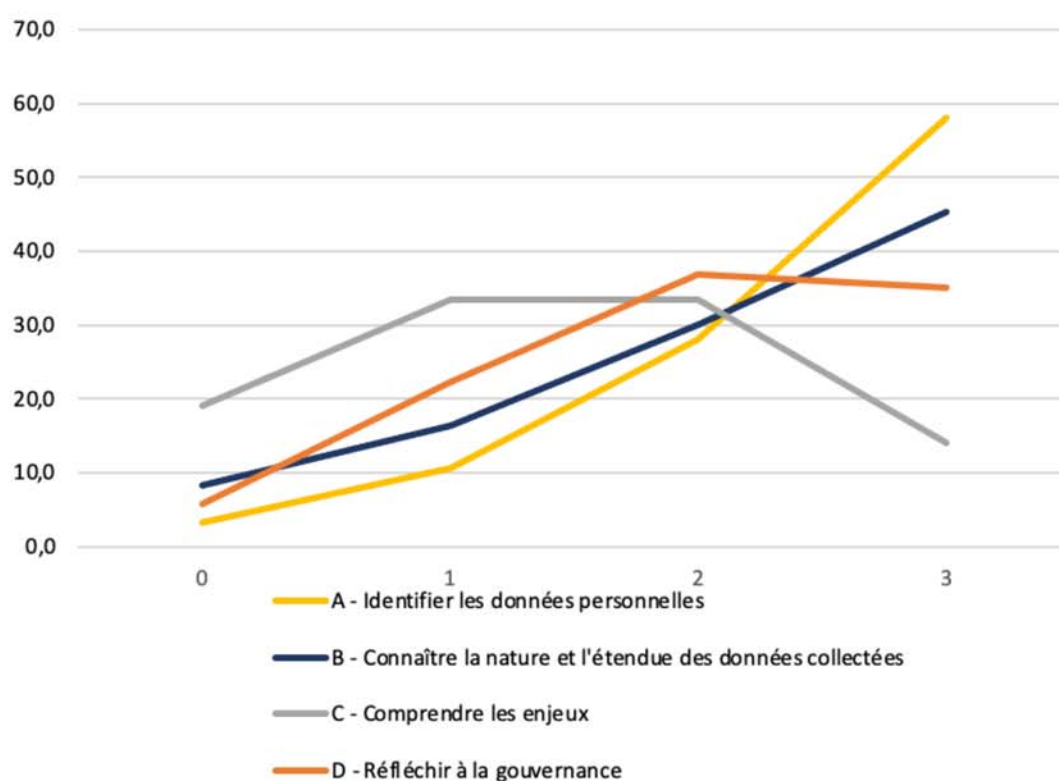
Remarques générales

Sans surprise, **le score tend à décroître au fur et à mesure que le degré de difficulté des questions augmente** : le score est meilleur à la dimension A, dont le degré de difficulté est plus bas, qu'à toutes les autres dimensions, qui impliquent de maîtriser des connaissances plus avancées (Figure 2). Voir les Tableaux 17 à 22 pour le détail des fréquences de réponses sur le score total.

Les scores sont globalement très bons sur les deux premières dimensions (A-Identifier les données personnelles et B-Connaître la nature et l'étendue des données collectées). La fréquence y augmente progressivement vers les plus hauts scores : seules 3,2% et 8,2% des personnes sondées ont un score de 0 sur 3, alors qu'environ la moitié des personnes sondées atteignent le plus haut score (58% pour A et 45,2% pour B atteignent un score de 3 sur 3).

Les sous-dimensions C (Comprendre les enjeux) et D (Réfléchir à la gouvernance) se démarquent avec une fréquence plus basse sur le score le plus élevé (seules 14,1% et 35,1% des personnes sondées ont un score de 3 sur 3, pour 33,4% et 36,9% qui ont un score de 2 sur 3). Ainsi, il a été plus difficile pour les personnes sondées de trouver toutes les bonnes réponses à ces deux sous-dimensions qui demandent de mobiliser des connaissances plus complexes. Cette tendance est davantage marquée pour **la sous-dimension C, qui présente une distribution des réponses différente des autres** : la manière dont cette sous-dimension est structurée dans l'enquête (elle comprenait un ensemble de trois questions, alors que les autres sous-dimensions ne comportaient qu'une question) a possiblement causé cette différence de résultats : les chances de commettre au moins une erreur dans cette dimension étaient plus élevées.

Figure 2 : Littératie des données personnelles : distribution des fréquences par sous-dimension d'analyse



Identifier les données personnelles (sous-dimension A)

Le nom et l'adresse postale ont bien été identifiés par la grande majorité des personnes sondées (87%) comme des données personnelles. Mais la proportion de bonnes réponses baisse en ce qui concerne **l'historique des emprunts à la bibliothèque** : 67% des personnes sondées l'identifient comme une donnée personnelle, soit nettement moins que les items précédents, mais tout de même dans une large majorité (plus des deux tiers des personnes sondées).

Concernant les mauvaises réponses, « le mot de passe du réseau Internet sans fil du café où je me branche » est incorrectement identifié comme une donnée personnelle par près d'un tiers des personnes sondées (32%). Cette erreur est beaucoup plus fréquente que le fait d'identifier le nom de la PDG de BAnQ comme une donnée personnelle (5%).

Tableau 17 : Identifier les données personnelles (A) : fréquence des réponses*

Q1- À votre avis, quelles informations parmi les suivantes constituent des données personnelles ?

Mon nom	1249	87%
Mon adresse postale	1245	87%
L'historique de mes emprunts à la bibliothèque	962	67%
Le mot de passe du réseau Internet sans fil du café où je me branche	459	32%
Le nom de la PDG de BAnQ	73	5%
NSP/NRP (Ne se prononce pas / Ne répond pas)	31	2%

**En bleu, les bonnes réponses ; en jaune, les mauvaises réponses.*

Connaître la nature et l'étendue des données collectées (sous-dimension B)

La structure des réponses à cette question (Tableau 18) est relativement proche de celle de la dimension A, bien que le nombre de non-réponses soit plus important ici (9% comparé à 2% pour A). Les propositions « **la municipalité où j'habite** » et « **la liste des documents favoris enregistrés dans mon dossier d'abonné de BAnQ** » ont bien été identifiées par la grande majorité des personnes sondées (83%) comme des données collectées par BAnQ. La proportion de bonnes réponses baisse en ce qui concerne « **le contenu des messages que j'envoie au service d'aide à distance de BAnQ** » : 64% des personnes sondées l'identifient comme une donnée collectée par BAnQ. Extrêmement peu de personnes sondées choisissent la mauvaise réponse « Le contenu des messages que j'envoie à mes amis en étant connecté au réseau Internet sans fil de la Grande Bibliothèque ». L'absence d'une deuxième mauvaise réponse dans cette sous-dimension a possiblement généré un biais dans les résultats du score, en offrant moins de chances de se tromper.

Tableau 18 : Connaissance de la nature et de l'étendue des données collectées (B) : fréquence des réponses

Q2- Selon vous, lesquelles de ces informations vous concernant sont actuellement collectées par BAnQ ?

La municipalité où j'habite	1183	83%
La liste des documents favoris enregistrés dans mon dossier d'abonné de BAnQ	1182	83%
Le contenu des messages que j'envoie au service d'aide à distance de BAnQ	914	64%
Le contenu des messages que j'envoie à mes amis en étant connecté au réseau Internet sans fil de la Grande Bibliothèque	13	1%
NSP/NRP	126	9%

Comprendre les enjeux des données personnelles (sous-dimension C)

La sous-dimension C comprenait trois questions en format vrai ou faux. **Deux des trois questions suscitent une majorité de mauvaises réponses ou de non-réponses.**

C'est à la question sur le rôle des services tiers que le taux de bonnes réponses est le plus haut (Tableau 21) : près du deux-tiers des personnes sondées (61%) ont correctement répondu que les services tiers utilisés par BAnQ (ex. Google Analytics, Facebook) ne sont pas soumis aux mêmes règles que BAnQ concernant la protection des données personnelles.

La question sur les données sensibles (Tableau 19) **suscite un écart très mince entre le taux de bonnes et de mauvaises réponses.** Ainsi, 48% des personnes sondées ont correctement identifié qu'il est possible de déduire des renseignements personnels (état de santé, orientation politique ou orientation sexuelle) à partir de l'historique d'emprunt de documents. Toutefois un nombre très important des personnes sondées (42%) a choisi la mauvaise réponse à cette question. Seul un écart de 6% sépare les bonnes et les mauvaises réponses. Si on ajoute les non-réponses aux mauvaises réponses, **une courte majorité de personnes sondées (51%) n'a pas correctement répondu à cette question.** Certaines personnes sondées ont mentionné lors du pré-test qu'elles empruntent parfois des documents dans le cadre de certaines activités prescrites (ex. cadre scolaire) et non pour des motifs personnels, ce qui a pu créer de la confusion autour de la question C1. Toutefois l'échantillon contient peu de personnes aux études ; ces cas de figure seraient donc en nombre restreint. Mais il est important de noter que les personnes abonnées peuvent aussi emprunter des documents pour quelqu'un d'autre.

Enfin, la question sur l'anonymisation des données (Tableau 20) **génère le plus haut taux de mauvaises réponses** : 42% des personnes sondées estiment de manière erronée qu'il n'y a aucun moyen de relier une usagère ou un usager à des informations anonymisées. Cette question demande des connaissances très pointues. Seul un tiers des personnes sondées (33%) a choisi la bonne réponse. Si l'on ajoute les non-réponses aux mauvaises réponses, on observe qu'une forte majorité (67%) des personnes sondées n'ont pas correctement répondu à la question.

Tableau 19 : Données sensibles (C1) : fréquence des réponses

Q3. À partir de votre historique d'emprunt de documents, on peut déduire des renseignements sur votre état de santé, votre orientation politique ou votre orientation sexuelle

Faux	604	42%
Vrai	692	48%
NSP/NRP	134	9%

Tableau 20 : Anonymisation (C2) : fréquence des réponses

Q4. Quand BAnQ rend anonymes vos données personnelles, il n'existe aucun moyen de vous relier à nouveau à l'information en question

Faux	471	33%
Vrai	595	42%
NSP/NRP	364	25%

Tableau 21 : Rôle des tiers (C3) : fréquence des réponses

Q5. BAnQ utilise divers services Web (ex. Google Analytics, Facebook) qui collectent également des données à votre sujet. Est-il vrai ou faux que ces fournisseurs de services Web sont soumis aux mêmes règles que BAnQ concernant la protection de vos données ?

Faux	873	61%
Vrai	304	21%
NSP/NRP	253	18%

Développer une réflexion critique sur la gouvernance des données personnelles (sous-dimension D)

Dans la sous-dimension relative à la réflexion critique, **les personnes sondées ont majoritairement choisi des bonnes réponses, autrement dit des solutions pertinentes pour la gouvernance des données personnelles** (Tableau 22). Une très forte majorité (81%) est d'avis que le fait de « **donner la possibilité de refuser que [s]es données soient collectées** » pourrait contribuer à encadrer l'utilisation des données personnelles par BAnQ. En deuxième position, le fait d'« **informer précisément sur les données collectées et leur utilisation** » a été choisi par près des trois quarts des personnes sondées (73%). En troisième position, avec une fréquence beaucoup plus basse, le fait de « faire appel à un organisme de réglementation indépendant pour surveiller ce qui est fait avec [l]es données » a recueilli 46% des réponses. Environ un quart des personnes sondées (26%) ont choisi une réponse non pertinente (« Publier un guide d'utilisation du catalogue en ligne de BAnQ »). L'absence d'une deuxième mauvaise réponse dans cette sous-dimension a possiblement généré un biais dans les résultats du score, en offrant moins de chances de se tromper.

Tableau 22 : Réflexion critique sur la gouvernance des données personnelles (D) : distribution des réponses

Q6- À votre avis, quelles initiatives pourraient le mieux contribuer à encadrer l'utilisation de vos données personnelles par BAnQ ?

Me donner la possibilité de refuser que mes données soient collectées	1165	81%
M'informer précisément sur les données collectées et leur utilisation	1041	73%
Faire appel à un organisme de réglementation indépendant pour surveiller ce qui est fait avec mes données	652	46%
Publier un guide d'utilisation du catalogue en ligne de BAnQ	378	26%
NSP/NRP	72	5%

6.2. Littératie des données personnelles : déterminants sociodémographiques

Modèle d'analyse

Les analyses ont été réalisées à l'aide d'un modèle de régression logistique ordinaire, appliqué dans un premier temps sur le score global de littératie, puis par sous-dimensions. Ce modèle permet de mesurer l'effet net de certaines caractéristiques sur les probabilités d'avoir un score global de littératie élevé, par rapport à un groupe de référence (entre parenthèses). Pour les analyses par sous-dimension, ce même modèle a été utilisé en utilisant une échelle de 0 à 3, selon le nombre de bonnes réponses identifiées (0 correspondant à aucune bonne réponse et 3, toutes les bonnes réponses). Nous avons modélisé les probabilités cumulées d'avoir un maximum de bonnes réponses.

Influence sur le score global

Seules les variables de l'âge et de la scolarisation ont une influence déterminante sur le score global de littératie des données personnelles (Figure 3).

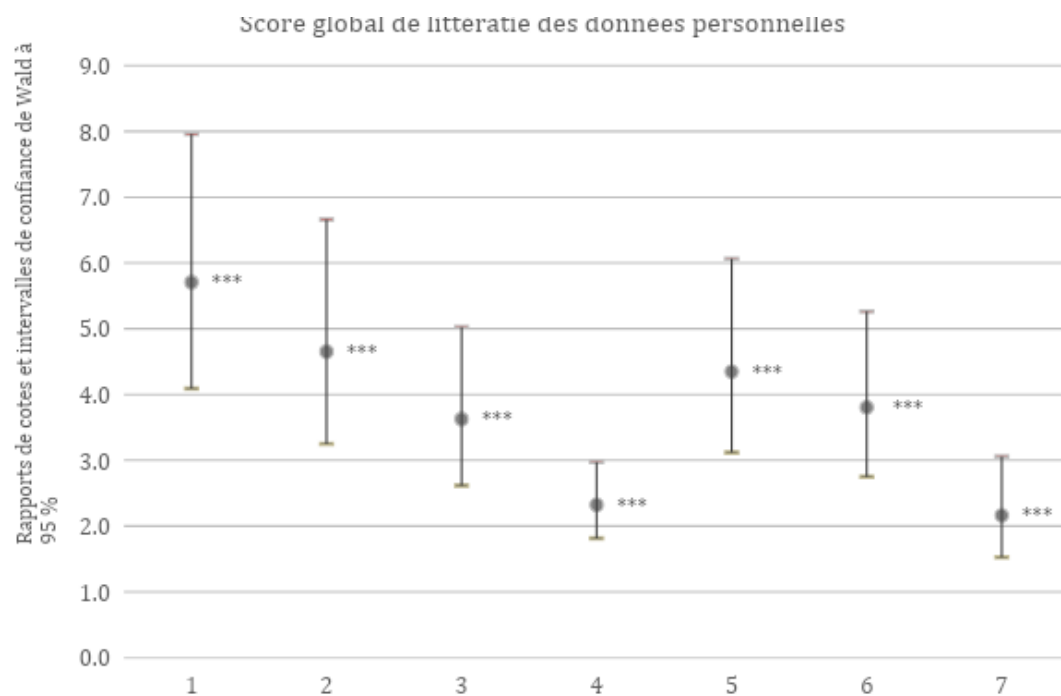
Les groupes d'âge **les plus jeunes ont entre deux et six fois plus de chances d'obtenir un score global élevé** en littératie par rapport aux personnes âgées de 65 ans et plus.

Détenir un **diplôme d'études postsecondaires** accroît les chances d'avoir un score élevé par rapport à celles et ceux n'ayant aucun diplôme ou un diplôme d'études secondaires uniquement. Des **études universitaires** augmentent les chances d'avoir un meilleur score en littératie (environ quatre fois plus), alors qu'un diplôme postsecondaire, mais non universitaire, les double.

Le fait d'appartenir au **groupe LGBTQ+** a une influence seulement sur une des sous-dimensions (celle relative aux enjeux des données, en particulier les enjeux relatifs aux données sensibles).

Les autres variables testées (genre, occupation, secteur d'emploi, minorité visible) n'ont pas d'influence significative.

Figure 3 : Influence de l'âge et du niveau de scolarisation sur le score global de litt ratie des donn es personnelles



* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Influence par sous-dimension

Si on considère individuellement les **différentes dimensions qui composent le score de littératie** (capacité à identifier les données personnelles, à connaître la nature et l'étendue des données collectées par BAnQ, à comprendre leurs enjeux et à réfléchir à leur gouvernance en contexte numérique) : **pour toutes ces dimensions**, l'âge et le niveau de scolarité jouent un rôle, avec plus ou moins d'ampleur selon la dimension, mais toujours dans le même sens (Figures 4 à 7). C'est-à-dire que **plus on est jeune ou plus on a un haut niveau de scolarité, meilleures sont nos chances d'obtenir un score élevé, en termes de nombre de bonnes réponses, sur chacune des dimensions.**

Le niveau de scolarisation a un impact particulièrement fort sur la **capacité à identifier les données personnelles** (Figure 4) : les personnes qui ont un diplôme universitaire ont entre 3 et 4 fois plus de chances et celles qui détiennent un diplôme postsecondaire (mais non universitaire) ont deux fois plus de chances d'avoir un score élevé à cette dimension, comparé aux personnes non diplômées ou aux personnes diplômées d'études secondaires uniquement.

L'âge a une influence particulièrement importante sur la **compréhension des enjeux** des données personnelles (Figure 6) : les 18-34 ans et les 35-44 ans ont environ quatre fois plus de chance d'avoir un score élevé à cette dimension comparé aux 65 ans et plus.

Le fait d'être affilié au groupe **LGBTQ+** a une influence sur la **capacité des personnes sondées à comprendre les enjeux de leurs données personnelles** : le fait de s'identifier à une minorité sexuelle augmente les chances (une fois et demie) d'avoir un score élevé sur cette dimension par rapport aux personnes qui ne s'identifient pas à une minorité sexuelle. Cet impact est particulièrement marqué en ce qui concerne les **enjeux relatifs aux données sensibles** (les autres enjeux abordés dans cette sous-dimension étant l'anonymisation des données et leur partage avec des tiers).

Figure 4 : Influence de l'âge et du niveau de scolarisation sur la capacité à identifier les données personnelles

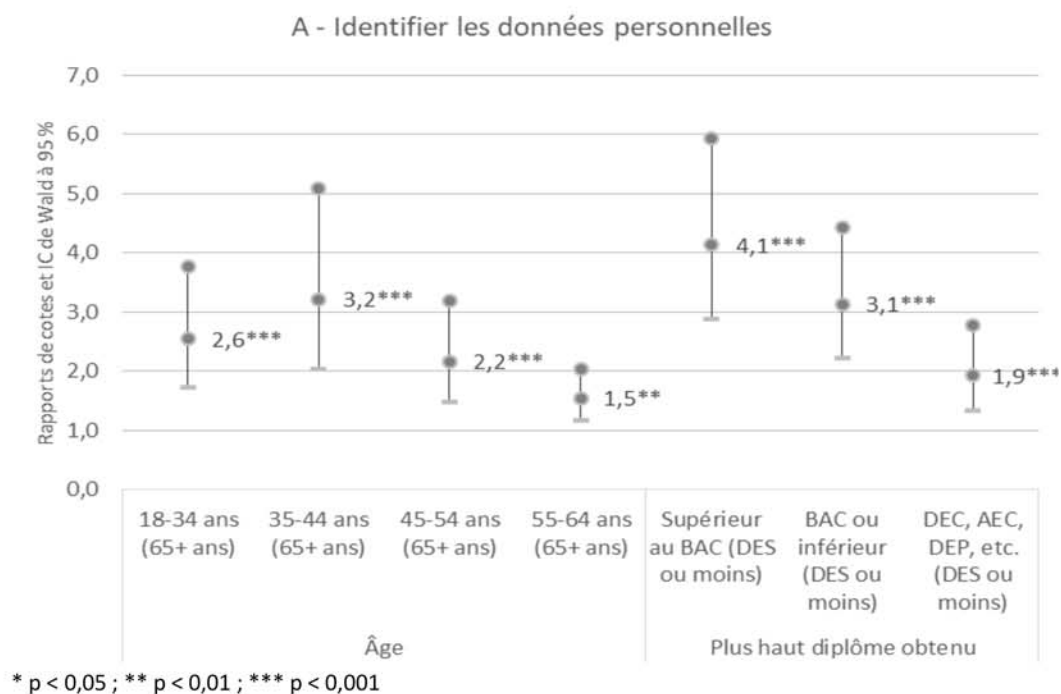


Figure 5 : Influence de l'âge et du niveau de scolarisation sur la capacité à connaître la nature et l'étendue des données collectées par BAnQ sur ses abonné.e.s

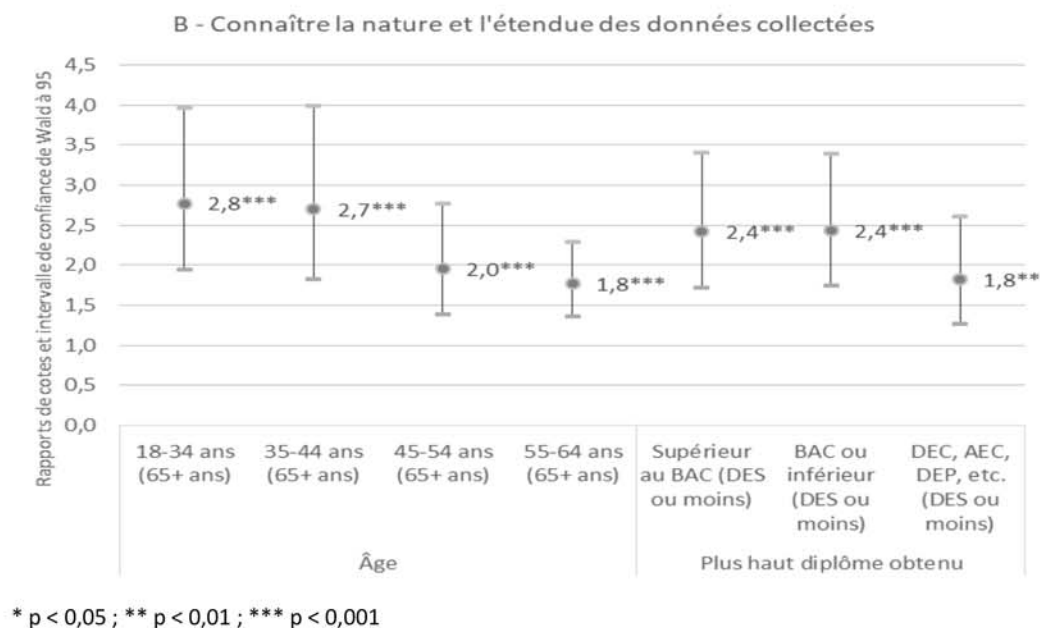
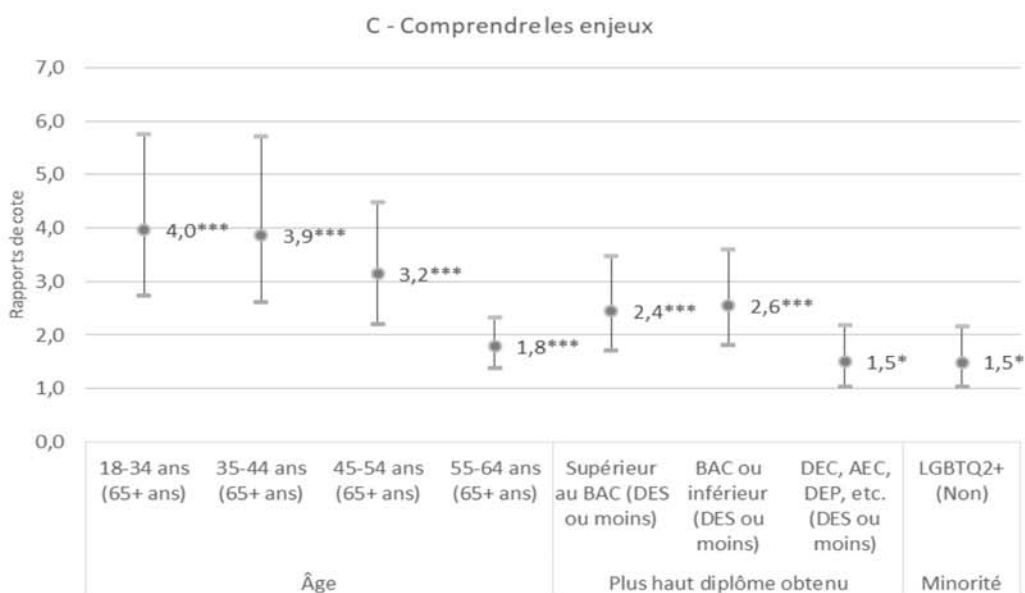
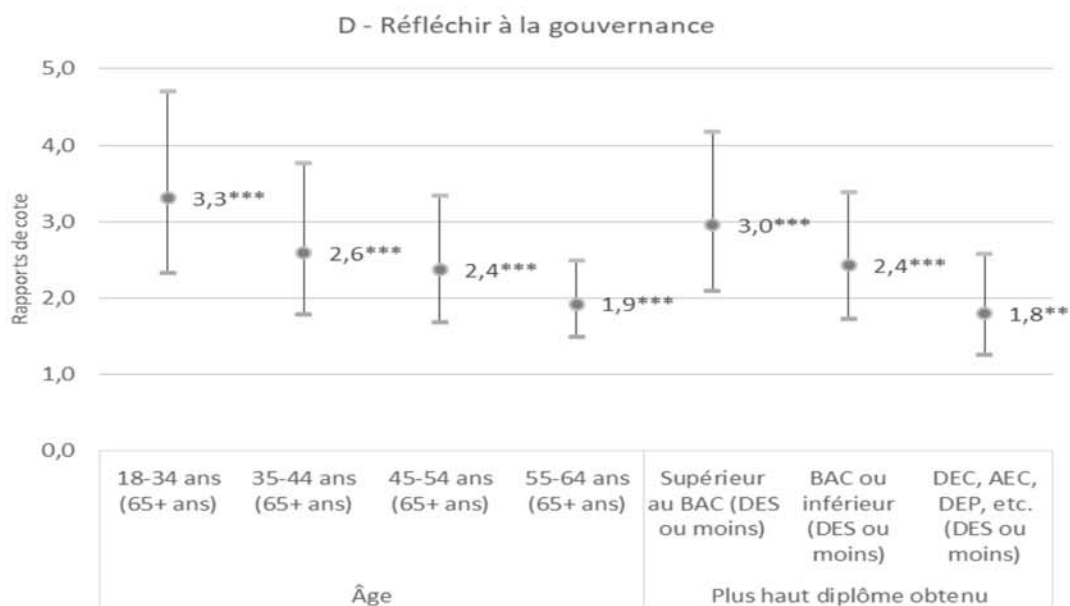


Figure 6 : Influence de l'âge et du niveau de scolarisation sur la capacité à comprendre les enjeux des données personnelles



* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Figure 7 : Influence de l'âge et du niveau de scolarisation sur la capacité à réfléchir à la gouvernance des données personnelles



* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

6.3. Intérêt pour de nouveaux services : mesure de score et de niveaux

Modèle d'analyse

La dimension de l'intérêt pour le développement de nouveaux services rassemble quatre sous-dimensions (voir Tableau 23) associées à différentes formes d'exploitation des données d'usage envisagées par BAnQ : (E) pour enrichir la collection avec de nouveaux documents en fonction de la demande du public, (F) pour recommander des documents en fonction des habitudes et préférences des personnes abonnées, (G) pour leur recommander d'autres offres culturelles en dehors de l'institution en fonction des habitudes et préférences, (H) pour personnaliser leur expérience de navigation sur le site web.

Cette dimension a été mesurée par une question sous la forme d'une échelle de Likert à 5 points : les personnes sondées étaient amenées à se prononcer sur chaque possibilité d'exploitation de leur historique d'emprunt en se positionnant comme Fortement d'accord, Plutôt d'accord, Ni d'accord ni en désaccord, Plutôt en désaccord, Fortement en désaccord ou Ne sait pas.

Un score a été attribué à chaque personne en fonction de ses choix de réponse. Chaque sous-dimension a été mesurée par un score de 1 à 5 points, pour un score total entre 4 points (intérêt très bas) et 20 points (intérêt très élevé). Voir en Annexe 4 le Tableau 48 pour le détail des fréquences de réponses sur le score total. Ce score a été utilisé pour analyser l'influence des variables sociodémographiques.

Les fréquences de réponses ont aussi été réparties par niveaux d'intérêt pour les nouveaux services. Pour cela, nous avons organisé les réponses aux quatre sous-dimensions (E, F, G, H) en une échelle de -4 à 4. Les niveaux d'intérêt ont été découpés de la manière suivante : très bas (-4 à -3), bas (-2 à -1), moyen (0 à 1), élevé (2 à 3) et très élevé (4). Étant donné la distribution des réponses plus abondantes dans les échelles élevées, nous avons estimé préférable de discriminer davantage aux échelles élevées pour mieux distinguer les nuances de niveau d'intérêt.

Tableau 23 : Intérêt pour les nouveaux services : questions du formulaire et mesures de score

(Q8) Veuillez indiquer dans quelle mesure vous êtes d'accord ou non avec chacune des affirmations suivantes. BAnQ devrait pouvoir utiliser mon historique d'emprunt pour :

E	Enrichir ses collections avec de nouveaux documents en fonction de la demande des publics
1-5 pts	<i>Fortement d'accord - Plutôt d'accord - Ni d'accord ni en désaccord - Plutôt en désaccord - Fortement en désaccord</i>
F	Recommander des documents en fonction des habitudes et préférences
1-5 pts	<i>Fortement d'accord - Plutôt d'accord - Ni d'accord ni en désaccord - Plutôt en désaccord - - Fortement en désaccord</i>
G	Recommander des activités culturelles en fonction des habitudes et préférences
1-5 pts	<i>Fortement d'accord - Plutôt d'accord - Ni d'accord ni en désaccord - Plutôt en désaccord - Fortement en désaccord</i>
H	Personnaliser l'expérience numérique sur le site Web de BAnQ
1-5 pts	<i>Fortement d'accord - Plutôt d'accord - Ni d'accord ni en désaccord - Plutôt en désaccord - Fortement en désaccord</i>

Score total de 4 à 20 points (4 = intérêt très bas ; 20 = intérêt très élevé)

Score global

La distribution des réponses est décalée vers les niveaux d'intérêt élevés et elle culmine au niveau très élevé (Figure 8). **Les deux tiers des personnes sondées (66,8%) déclarent un intérêt élevé à très élevé.** Un tiers (34,3%) a un intérêt moyen à bas.

Très peu de personnes sondées déclarent un niveau d'intérêt très bas (5,9%). Les niveaux bas (11,2%) et moyen (16,9%) recueillent des taux de réponse relativement proches. Tandis que les niveaux d'intérêt élevé (31,9%) et très élevé (34,9%) recueillent une large part des taux de réponse.

Toutefois, les résultats des sous-dimensions ne sont pas homogènes. Il faut donc aussi considérer la distribution des réponses par sous-dimension (Tableau 24).

Figure 8 : Distribution par niveau du score d'intérêt pour les nouveaux services

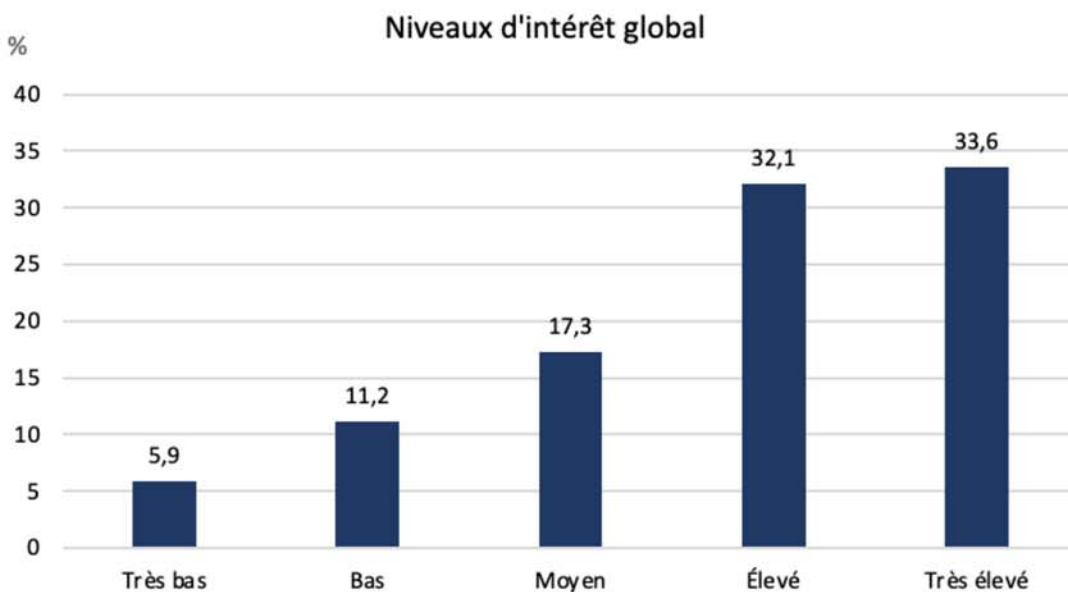


Tableau 24 : Distribution des réponses par niveau d'intérêt

	Score de niveau	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé	Pourcentage groupé
Très bas	-4	51	4,3	51	4,3	5,9
	-3	19	1,6	70	5,9	
Bas	-2	95	7,9	165	13,8	11,2
	-1	39	3,3	204	17,0	
Moyen	0	93	7,8	297	24,8	17,3
	1	114	9,5	411	34,3	
Élevé	2	186	15,5	597	49,9	32,0
	3	198	16,5	795	66,4	
Très élevé	4	402	33,6	1197	100,0	33,6

Fréquence manquante = 233

Score par sous-dimension

On observe une **majorité (ou quasi) d'avis favorables sur toutes les propositions.**

La proposition d'enrichir les collections de BAnQ grâce à l'historique d'emprunt est celle qui recueille le plus d'avis favorables, avec 85,2% des personnes sondées qui sont fortement ou plutôt d'accord avec cette utilisation (Tableau 25). Cependant, la formulation de cette question a possiblement créé un biais : le choix du terme « d'enrichir » la collection induit une convocation positive. Nous l'avons donc retirée pour la suite de l'analyse.

En deuxième position se trouve la proposition de recommander des documents, suivie de près par le fait de personnaliser l'expérience sur le site web, et en dernière position, la proposition de recommander d'autres activités culturelles (Figures 9 et 10).

La proposition qui recueille le moins d'avis favorables est celle de recommander d'autres offres culturelles en dehors de BAnQ, avec un nombre important d'avis neutres (28,7% sont ni d'accord ni en désaccord), mais elle recueille tout de même l'avis favorable de près de la moitié des personnes sondées. La proposition de personnaliser l'expérience numérique sur le site Web de BAnQ recueille aussi un nombre important d'avis neutres (22,1% sont ni d'accord ni en désaccord), mais une majorité d'avis favorables.

Les autres utilisations de l'historique d'emprunts, soit pour recommander des documents (Tableau 26), recommander des activités culturelles (Tableau 27) ou personnaliser l'expérience numérique sur le site Web de BAnQ (Tableau 28), **connaissent toutes un fléchissement** dans la proportion de personnes sondées fortement d'accord par rapport à plutôt d'accord.

Il n'en reste pas moins que **la presque majorité des personnes sondées sont favorables ou très favorables à ces utilisations** de leur historique d'emprunts (67,3% pour recommander des documents ; 55,3% pour personnaliser l'expérience numérique sur le site Web ; 48,5% pour recommander des activités culturelles).

Figure 9 : Intérêt pour le développement de nouveaux services : distribution des fréquences par sous-dimension d'analyse

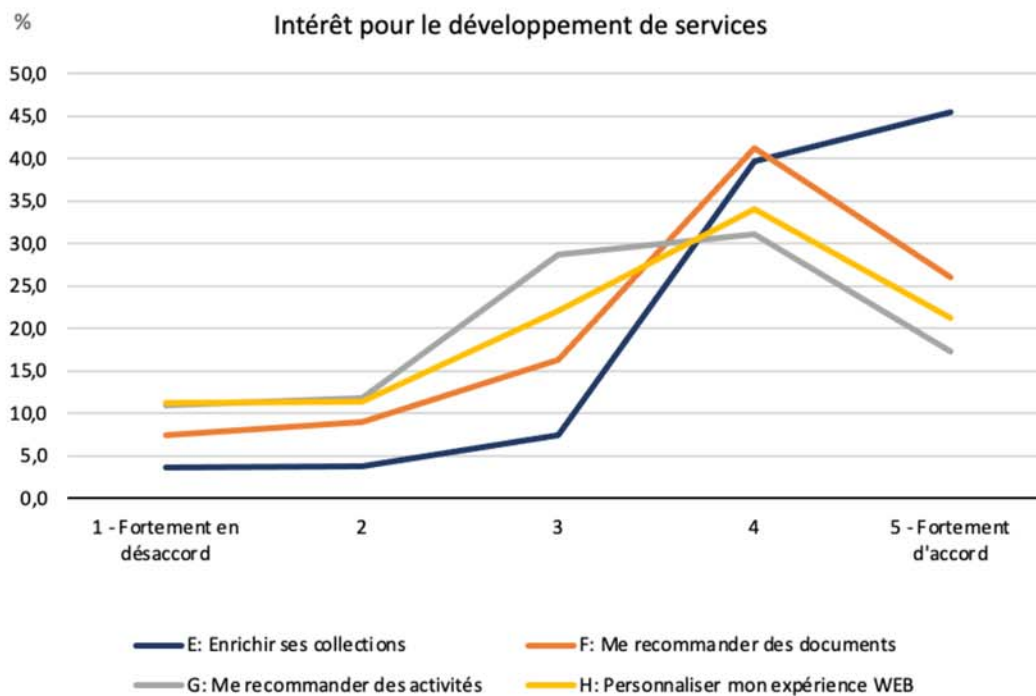


Figure 10 : Intérêt pour le développement de nouveaux services : scores moyens par sous-dimension d'analyse

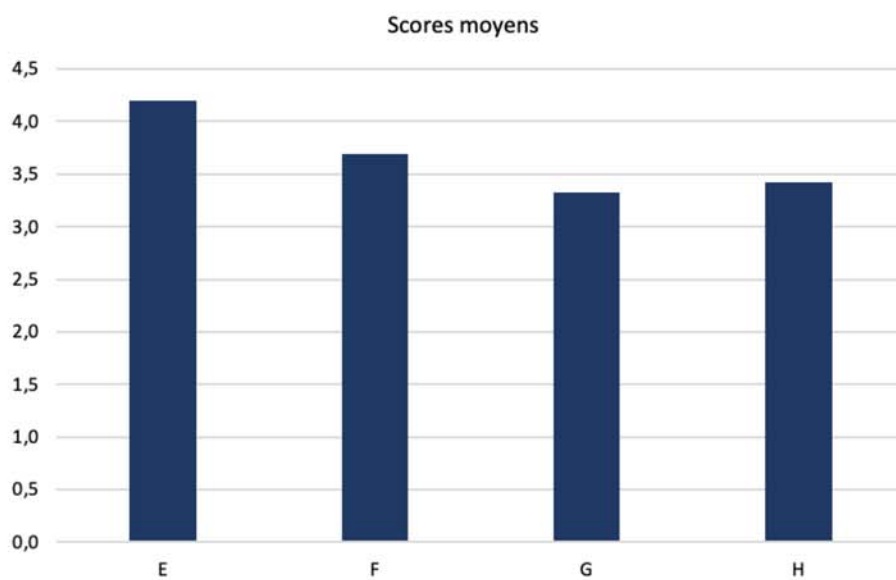


Tableau 25 : Enrichir ses collections avec de nouveaux documents en fonction de la demande du public (sous-dimension E) : fréquence des réponses

	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
1 - Fortement en désaccord	50	3,6	50	3,6
Plutôt en désaccord	53	3,8	103	7,5
Ni d'accord ni en désaccord	102	7,4	205	14,8
Plutôt en accord	548	39,7	753	54,5
5 - Fortement d'accord	629	45,5	1382	100,0

Fréquence manquante = 48 (3%)

Tableau 26 : Recommander des documents en fonction des habitudes et préférences (sous-dimension F) : fréquence des réponses

	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
1 - Fortement en désaccord	97	7,5	97	7,5
Plutôt en désaccord	116	8,9	213	16,4
Ni d'accord ni en désaccord	212	16,3	425	32,7
Plutôt en accord	536	41,3	961	74,0
5 - Fortement d'accord	338	26,0	1299	100,0

Fréquence manquante = 131 (9%)

Tableau 27 : Recommander des activités culturelles en fonction des habitudes et préférences (sous-dimension G) : fréquence des réponses

	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
1 - Fortement en désaccord	138	11,0	138	11,0
Plutôt en désaccord	148	11,8	286	22,8
Ni d'accord ni en désaccord	361	28,7	647	51,5
Plutôt en accord	391	31,1	1038	82,6
5 - Fortement d'accord	218	17,4	1256	100,0

Fréquence manquante = 174 (12%)

Tableau 28 : Personnaliser l'expérience numérique sur le site Web de BAnQ (sous-dimension H) : fréquence des réponses

	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
1 - Fortement en désaccord	141	11,3	141	11,3
Plutôt en désaccord	142	11,3	283	22,6
Ni d'accord ni en désaccord	277	22,1	560	44,7
Plutôt en accord	426	34,0	986	78,7
5 - Fortement d'accord	267	21,3	1253	100,0

Fréquence manquante = 177 (12%)

6.4. Intérêt pour de nouveaux services : déterminants sociodémographiques

Modèle d'analyse

Les quatre sous-dimensions étant mesurées de la même façon, sous forme d'échelle d'accord (de 1 à 5), nous avons encore une fois utilisé le modèle de régression logistique ordinaire. Rappelons que ce modèle permet de mesurer l'effet net de certaines caractéristiques sur les probabilités d'avoir un intérêt élevé pour chacune des sous-dimensions, par rapport à un groupe de référence (entre parenthèses).

L'influence des déterminants sociodémographiques et de la littératie des données personnelles n'a pas été calculée sur le score *global* d'intérêt pour les nouveaux services pour deux raisons : les niveaux sont trop nombreux (entre 4 et 20) et les résultats des sous-dimensions ne sont pas homogènes. L'influence a plutôt été calculée sur chaque sous-dimension.

Influence des variables sociodémographiques par sous-dimension

Les variables sociodémographiques qui ont une influence sur le niveau d'intérêt pour les nouveaux services sont : le niveau de **scolarisation** (influence sur deux sous-dimensions) et le fait d'appartenir à une **minorité visible** (influence sur une sous-dimension).

Détenir un diplôme universitaire diminue les chances d'avoir un fort intérêt à se faire recommander des documents ou pour personnaliser son expérience sur le site Web de BAnQ (voir Figure 12 et Figure 14). Par contre, le niveau de diplomation n'a pas d'incidence sur l'intérêt pour l'enrichissement des collections en fonction de la demande des publics et pour la recommandation d'activités culturelles.

Faire partie d'une minorité visible double les chances (2,3) d'être favorable à se voir recommander des activités culturelles (voir Figure 13).

Nous avons aussi relevé que l'âge a une influence sur le fait d'être favorable à l'utilisation de l'historique d'emprunts afin d'enrichir les collections de BAnQ (avoir entre 35 et 54 ans diminue les chances d'y être favorable) mais cette sous-dimension a été retirée de l'analyse en raison de sa formulation biaisée.

Les autres variables testées (genre, occupation, secteur d'emploi, affiliation au groupe LGBTQ+) n'ont pas d'influence significative sur l'intérêt pour le développement de nouveaux services.

Influence de la littératie des données personnelles par sous-dimension

Les personnes sondées ont répondu aux questions portant sur l'intérêt pour les nouveaux services après avoir été informées des bonnes réponses aux questions portant sur la littératie des données personnelles. Ainsi leur perception concernant les nouveaux services a pu être éclairée par une meilleure compréhension de la nature, de la portée et des enjeux des données personnelles.

Les résultats montrent que **le niveau de littératie des données personnelles a un impact** sur la plupart (trois sur quatre) des sous-dimensions de l'intérêt pour le développement de nouveaux services.

Un **score élevé en littératie des données personnelles diminue les chances d'avoir un fort intérêt** pour le développement de nouveaux services sur **trois des quatre sous-dimensions mesurées** : se faire recommander des documents (voir Figure 12) ; personnaliser son expérience sur le site Web (voir Figure 14) ; et se voir recommander des activités culturelles (voir Figure 13). Par exemple, une augmentation de 5 points dans le score global de littératie des données personnelles représente une baisse de 30% à 40% de l'intérêt selon la sous-dimension (un facteur de 0,6 pour les sous-dimensions F et H et 0,7 pour la sous-dimension G).

Le fait d'enrichir les collections de BAnQ en fonction de la demande du public est la seule sous-dimension où le score de littératie n'a pas d'influence (Figure 11), probablement parce que la très grande majorité des répondants sont favorables ou neutres (93%).

Figure 11 : Influence de l'âge sur l'intérêt pour enrichir les collections de BAnQ avec de nouveaux documents en fonction de la demande des publics

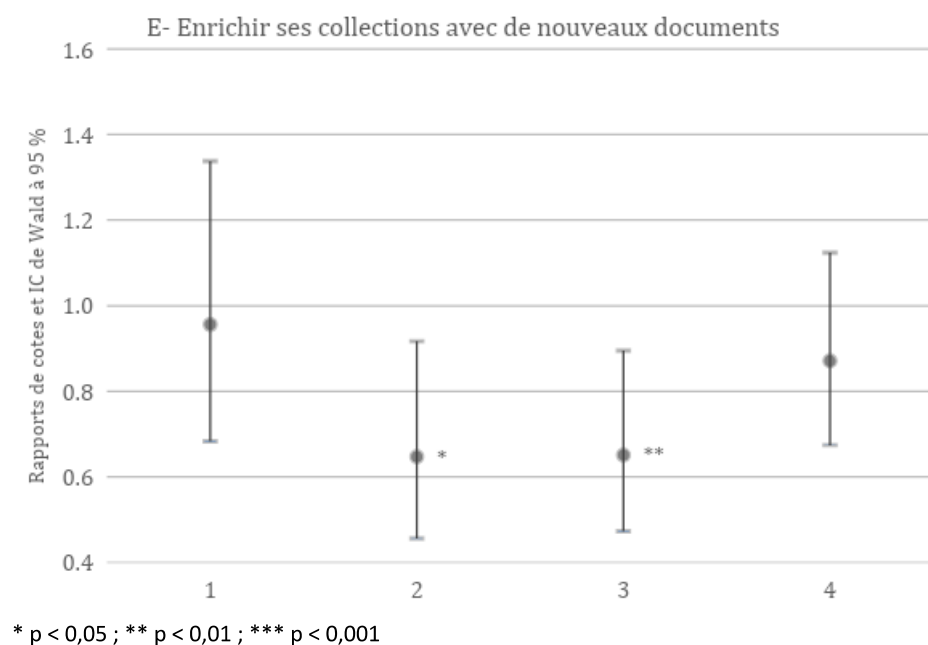


Figure 12 : Influence de l'âge et du score de littératie sur l'intérêt pour se faire recommander des documents en fonction de ses habitudes et préférences

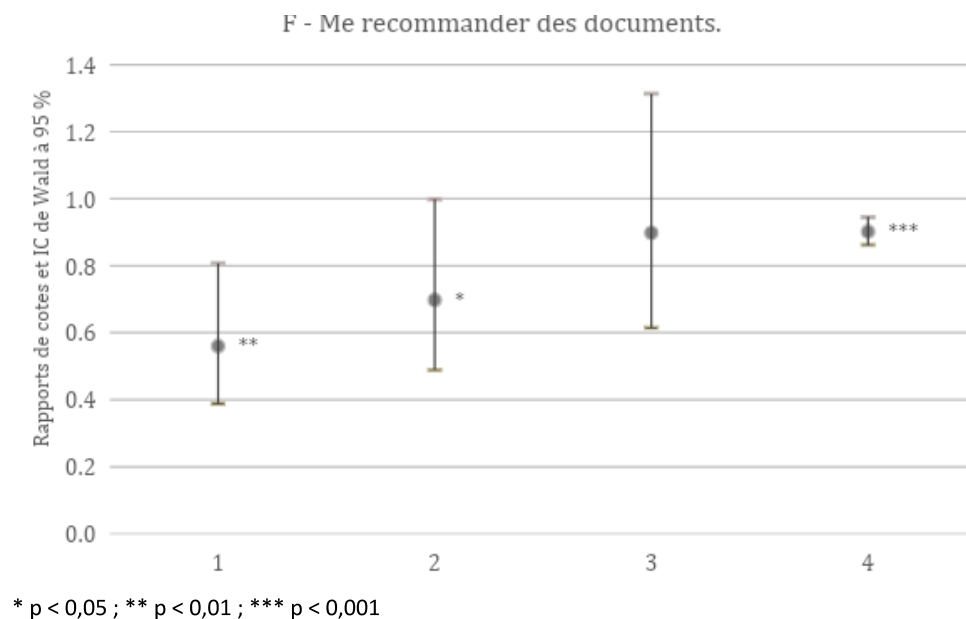
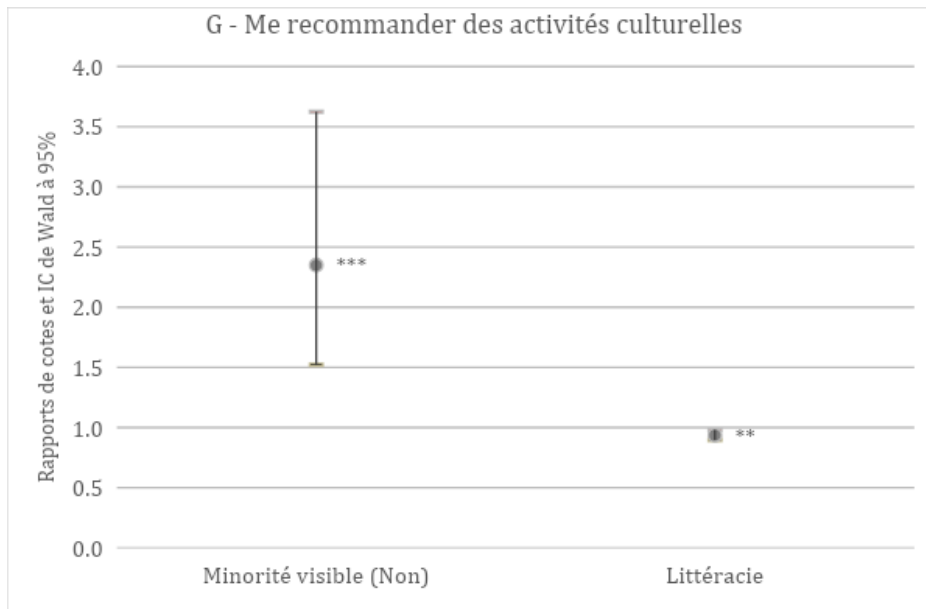
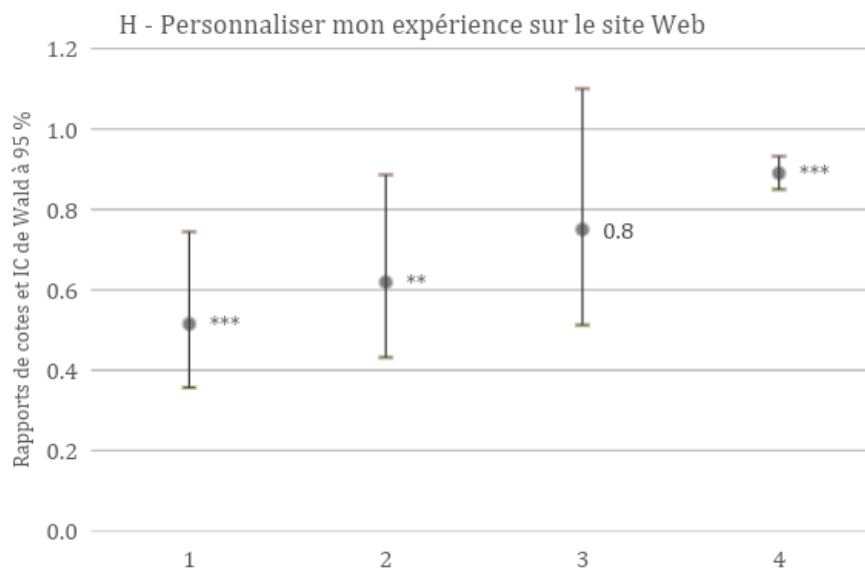


Figure 13 : Influence de l'âge et du score de littératie sur l'intérêt pour se faire recommander des activités culturelles en fonction des habitudes et préférences



* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Figure 14 : Influence de l'âge et du score de littératie sur l'intérêt pour que son expérience numérique sur le site Web de BAnQ soit personnalisée



* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

6.5. Perception de la gestion des données personnelles : mesure de score et typologie d'attitudes

Mode de collecte et d'analyse

La dimension de la position relative à la protection de la vie privée rassemble quatre sous-dimensions (Tableau 29) associées à différents cycles de vie des données : (I) la collecte des données, (J) leur exploitation (recoupement et minage pour personnaliser les services), (K) leur durée d'archivage et (L) leur partage avec des tiers. La sous-dimension L se déclinait en quatre propositions : le partage avec (L1) des ministères, (L2) maisons d'édition, (L3) des OBNL et (L4) des équipes de recherche universitaires. Trois des sous-dimensions (I, J, L) ont été mesurées par des questions sous la forme d'échelles de Likert à 5 points (Fortement d'accord, Plutôt d'accord, Ni d'accord ni en désaccord, Plutôt en désaccord, Fortement en désaccord). La sous-dimension K a été mesurée par une question à choix multiple sur la durée d'archivage des données (pas de conservation, un an, entre un et cinq ans, indéfiniment) pour laquelle était attribué un score de 0 à 3. L'option je ne sais pas était proposée à chaque question.

En rétrospective, la sous-dimension I a posé certains problèmes qui ont amené à l'exclure de la mesure du score final. La question comportait un biais de formulation : elle demandait de se prononcer sur la validation d'une pratique actuelle, ce qui tend à susciter plus d'avis favorables et rend également ambiguë l'interprétation des réponses défavorables. Une meilleure formulation aurait proposé une alternative : limiter la collecte de données personnelles (privilégier la protection de la vie privée) ou étendre la collecte de données (privilégier le développement de nouveaux services).

Un score a été attribué à chaque personne sondée en fonction de ses choix de réponse, pour un score total entre 5 (perception très protectrice) et 28 points (perception très ouverte). Voir en Annexe 5 le Tableau 49 pour le détail des fréquences de réponses sur le score total. Les scores totaux ont ensuite été répartis par niveau pour mieux distinguer différentes attitudes relatives à la gestion des données personnelles. Pour cela, nous avons organisé les cinq sous-dimensions concernant l'exploitation des données (J) et le partage des données avec des tiers (L1, L2, L3, L4) en une échelle de -5 à 5 (voir en annexe, Tableau 30). Par la suite, nous avons croisé cette échelle avec la question sur la conservation des données personnelles (K) pour valider la distribution des réponses par attitude et la pertinence de l'avoir incluse dans le score global (Tableau 31).

Tableau 29 : Perception relative à la protection des données personnelles : questions du formulaire et mesures de score

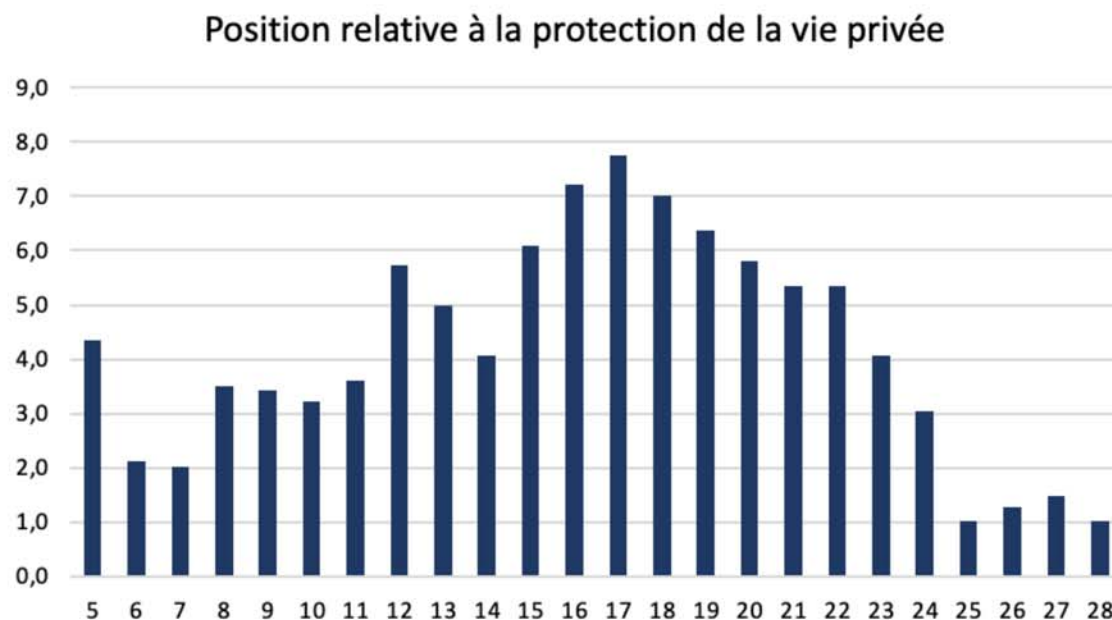
I Position concernant la collecte des données	
(Q9)	<i>Veillez indiquer dans quelle mesure vous êtes d'accord ou non avec chacune des affirmations suivantes. BAnQ devrait pouvoir :</i>
0 pts (non inclus)	a) Collecter seulement les données personnelles utiles aux services actuels <i>Fortement d'accord - Plutôt d'accord - Ni d'accord ni en désaccord - Plutôt en désaccord - Fortement en désaccord</i>
J Position concernant l'exploitation des données	
1-5 pt	b) Recouper différentes données à mon sujet pour personnaliser ses services (ex. : historique d'emprunt selon l'âge) <i>Fortement d'accord - Plutôt d'accord - Ni d'accord ni en désaccord - Plutôt en désaccord - Fortement en désaccord</i>
K Position concernant la conservation des données	
(Q10)	<i>BAnQ désire conserver mes données personnelles anonymisées en vue de mieux connaître les pratiques de ses abonnés et d'améliorer son offre de services. Je souhaiterais que mes données :</i>
0-3 pts	a) Ne soient pas conservées du tout b) Soient conservées moins d'un an b) Soient conservées entre 1 an et 5 ans c) Soient conservées indéfiniment
L Position concernant le partage des données	
(Q11)	<i>Veillez indiquer dans quelle mesure vous êtes d'accord ou non avec chacune des affirmations suivantes. BAnQ devrait pouvoir partager mes données personnelles, lorsqu'elles sont anonymisées, avec :</i>
1-5 pts	L1) un ministère du gouvernement québécois, pour concevoir des politiques publiques <i>Fortement d'accord - Plutôt d'accord - Ni d'accord ni en désaccord - Plutôt en désaccord - Fortement en désaccord</i>
1-5 pts	L2) une maison d'édition, pour lui permettre de mieux cibler ses clients <i>Fortement d'accord - Plutôt d'accord - Ni d'accord ni en désaccord - Plutôt en désaccord - Fortement en désaccord</i>
1-5 pts	L3) un OBNL, pour faire de la sensibilisation à la lecture <i>Fortement d'accord - Plutôt d'accord - Ni d'accord ni en désaccord - Plutôt en désaccord - Fortement en désaccord</i>
1-5 pts	L4) des équipes de recherche universitaires, afin d'étudier les pratiques de lecture de la population québécoise <i>Fortement d'accord - Plutôt d'accord - Ni d'accord ni en désaccord - Plutôt en désaccord - Fortement en désaccord</i>
Score total de 5 à 28 (5 = très protecteur ; 28 = très ouvert)	

L'échelle de niveaux a ensuite été découpée de manière à identifier **une typologie d'attitudes, qui s'étend des usagères et usagers les plus protecteurs de leurs données personnelles à une attitude plus ouverte à la valorisation de leurs données**. Les cinq types d'attitude identifiés sont les suivants : *très protecteur* (niveaux -5 à -4), *protecteur* (niveaux -3 à -2), ni l'un ni l'autre ou *neutre* (niveaux -1 à 1), *ouvert* (niveaux 2 à 3) et *très ouvert* (niveaux 4 à 5). Les très protecteurs sont ceux qui sont fortement en désaccord avec le fait de croiser différentes données personnelles et le fait de partager les données (même anonymisées) avec des tiers, et qui par ailleurs souhaitent limiter au maximum la durée d'archivage des données. Les très ouverts sont au contraire ceux qui sont fortement en accord avec ces pratiques.

Score global

La distribution des réponses présente globalement une forme en cloche qui culmine aux scores moyens (Figure 15). On observe toutefois un pic isolé de réponses au score le plus bas (5), et plus généralement une distribution plus grande aux scores inférieurs (soit les plus protecteurs) comparé aux scores les plus hauts (les plus ouverts) qui rassemblent les fréquences les plus faibles.

Figure 15 : Distribution globale du score de perception relative à la gestion des données personnelles (sans la sous-dimension I)



Attitudes

Dans l'ensemble, l'attitude des individus en regard de la protection des données personnelles est **partagée de manière presque égale** entre ceux qui sont « très protecteurs » ou « protecteurs » de leurs données (32,1%), ceux qui sont neutres par rapport à cette question (33,7%) et ceux qui sont « ouverts » ou « très ouverts » (34,2%) (Figure 16). On remarque par ailleurs une concentration des réponses au centre de l'échelle, ce qui indique que **les attitudes sont généralement neutres**. Toutefois, les résultats des sous-dimensions ne sont pas homogènes. Il faut donc aussi considérer la distribution des réponses par sous-dimension (Figures 17 et 18).

Figure 16 : Perception relative à la gestion des données personnelles (sous-dimensions J et L) : distribution par attitude (vue générale)

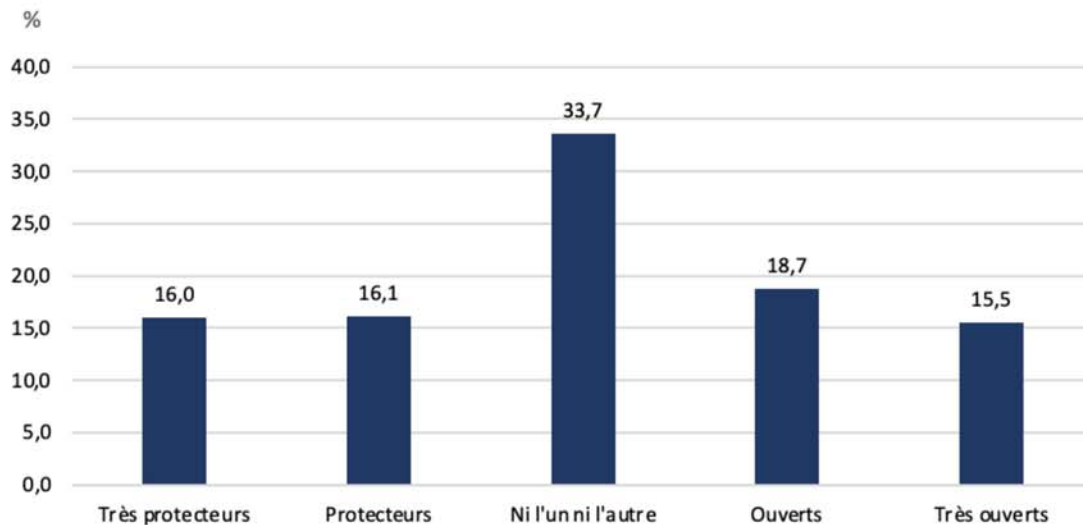


Tableau 30 : Perception relative à la gestion des données personnelles (sous-dimensions J et L) : distribution par attitude (détail des fréquences)

Attitude	Score 2	Fréquence	Pourcentage	Pourcentage groupé
Très protecteurs	-5	123	10,5	16,0
	-4	65	5,5	
Protecteurs	-3	124	10,6	16,1
	-2	65	5,5	
Ni l'un ni l'autre	-1	139	11,8	33,7
	0	113	9,6	
	1	143	12,2	
Ouverts	2	108	9,2	18,7
	3	112	9,5	
Très ouverts	4	81	6,9	15,5
	5	101	8,6	

Fréquence manquante = 256

Tableau 31 : Perception relative à la gestion des données personnelles : distribution par attitude (sous-dimensions J et L) et croisement avec sous-dimensions K

	Conservation des données				Total
	<i>Pas du tout</i>	<i>Moins d'un an</i>	<i>1 à 5 ans</i>	<i>Indéfiniment</i>	
Très protecteurs	102	50	23	6	181
	56,4%	27,6%	12,7%	3,3%	100%
Protecteurs	36	71	56	13	176
	20,5%	40,3%	31,8%	7,4%	100%
Ni l'un ni l'autre	29	129	167	38	209
	8,0%	35,5%	46,0%	10,5%	100%
Ouverts	9	46	110	44	209
	4,3%	22,0%	52,6%	21,1%	100%
Très ouverts	6	23	88	57	174
	3,5%	13,2%	50,6%	32,8%	100%
Total	182	319	444	158	1103
	16,5%	28,9%	40,3%	14,3%	100%

Fréquence manquante = 327

Score par sous-dimension

La très large majorité (près de 80%) des personnes abonnées participant au sondage se dit favorable (plutôt ou fortement d'accord) avec l'énoncé « *Collecter seulement les données personnelles utiles aux services actuels* » et seules 7% d'entre elles sont plutôt ou fortement en désaccord (Tableau 32). Mais la formulation de la question (demande de valider la pratique actuelle) pourrait avoir biaisé les résultats en direction des avis favorables.

Le recoupage des données suscite des avis très partagés (Tableau 33). Une proportion égale de personnes sondées se montre favorable (39,2%) et défavorable (37,8%) à ce que BAnQ puisse « *recouper différentes données à mon sujet pour personnaliser ses services* » et près du quart (23,1%) se disent neutres à ce sujet. La proportion des personnes sondées fortement en désaccord est plus grande que celle des personnes fortement d'accord (16,7% contre 11,2%).

Par ailleurs, 83,2% des personnes sondées acceptent que BAnQ conserve leurs données personnelles selon différents termes (Figure 19). Mais **69% souhaitent une limitation de la durée de conservation**. La plupart d'entre elles (40,1%) ont une préférence pour une durée de conservation entre 1 an et 5 ans (Tableau 34).

Le partage des données personnelles anonymisées est plus ou moins bien accepté selon les partenaires proposés (Figure 20). Le partage avec des **équipes de recherche universitaires** afin d'étudier les pratiques de lecture de la population est celui qui reçoit **l'approbation la plus forte**, 66,9% des répondants y étant favorables, alors que 18,3% ne le sont pas (Tableau 38). À l'opposé, le partage avec une **maison d'édition** afin de lui permettre de mieux cibler ses clients est celui qui reçoit le moins d'avis favorables, 20,5% contre 60,4 % qui sont **défavorables** avec un tel partage (Tableau 36). Entre les deux on retrouve les cas de l'OBNL pour faire de la sensibilisation à la lecture et d'un ministère du gouvernement québécois pour concevoir des politiques publiques, pour lesquels les avis favorables et défavorables sont **également partagés**. Respectivement, on observe 36,5% favorables contre 40,4% défavorables pour l'OBNL (Tableau 37) et 42,4% favorables contre 41,9% pour le ministère (Tableau 35). Ces deux propositions suscitent un nombre important d'avis fortement en désaccord (environ un tiers des personnes sondées dans les deux cas).

Figure 17 : Perception relative à la gestion des données personnelles : scores moyens par sous-dimension d'analyse

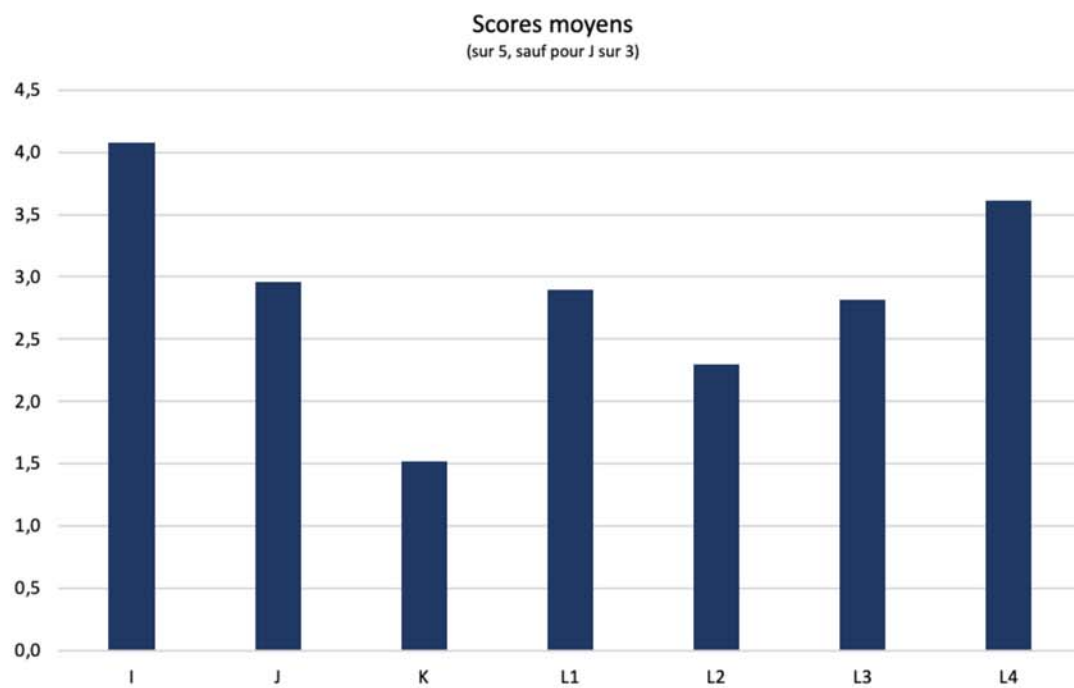


Figure 18 : Perception relative à la collecte et à l'exploitation des données personnelles (sous-dimensions I et J) : distribution des fréquences par sous-dimension d'analyse

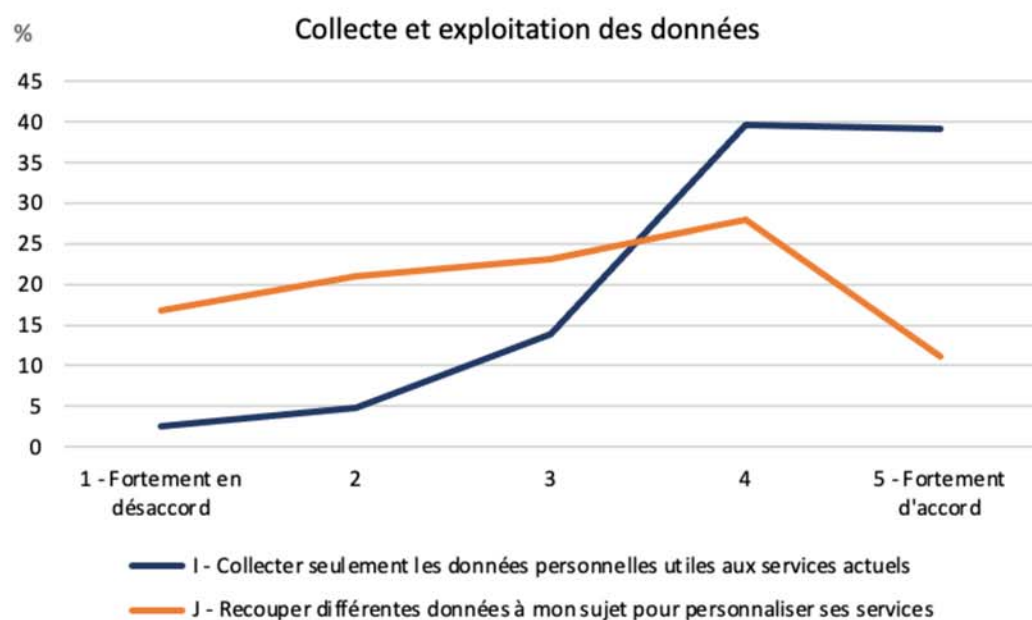


Tableau 32 : Collecter seulement les données personnelles utiles aux services actuels (sous-dimension I) : fréquence des réponses

	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
1 - Fortement en désaccord	34	2,5	34	2,5
Plutôt en désaccord	66	4,9	100	7,4
Ni d'accord ni en désaccord	187	13,8	287	21,2
Plutôt d'accord	535	39,6	822	60,8
5 - Fortement d'accord	530	39,2	1352	100,0

Fréquence manquante = 78

Tableau 33 : Recouper différentes données pour personnaliser les services (sous-dimension J) : fréquence des réponses

	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
1 - Fortement en désaccord	216	16,7	216	16,7
Plutôt en désaccord	272	21,1	488	37,8
Ni d'accord ni en désaccord	298	23,1	786	60,8
Plutôt d'accord	362	28,0	1148	88,9
5 - Fortement d'accord	144	11,2	1292	100,0

Fréquence manquante = 138

Figure 19 : Conservation des données personnelles (sous-dimension K) : distribution des fréquences

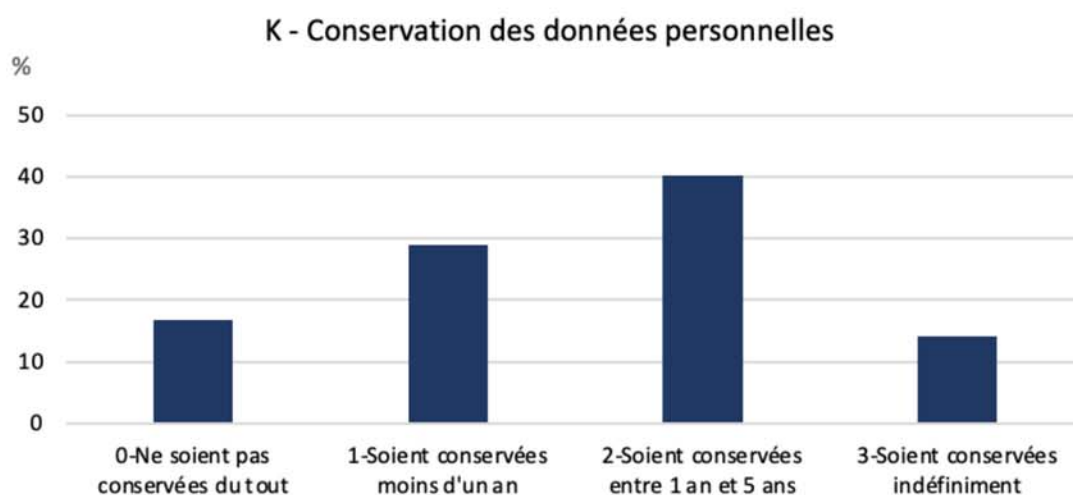


Tableau 34 : Conservation des données personnelles (sous-dimension K) : fréquence des réponses

	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
1 - Ne soient pas conservées du tout	220	16,8	220	16,8
Soient conservées moins d'un an	378	28,9	598	45,8
Soient conservées entre 1 an et 5 ans	524	40,1	1122	85,9
3 - Soient conservées indéfiniment	184	14,1	1306	100,0

Fréquence manquante = 124

Figure 20 : Perceptions relatives au partage des données personnelles anonymisées : distribution des fréquences par sous-dimension d'analyse (L1, L2, L3, L4)

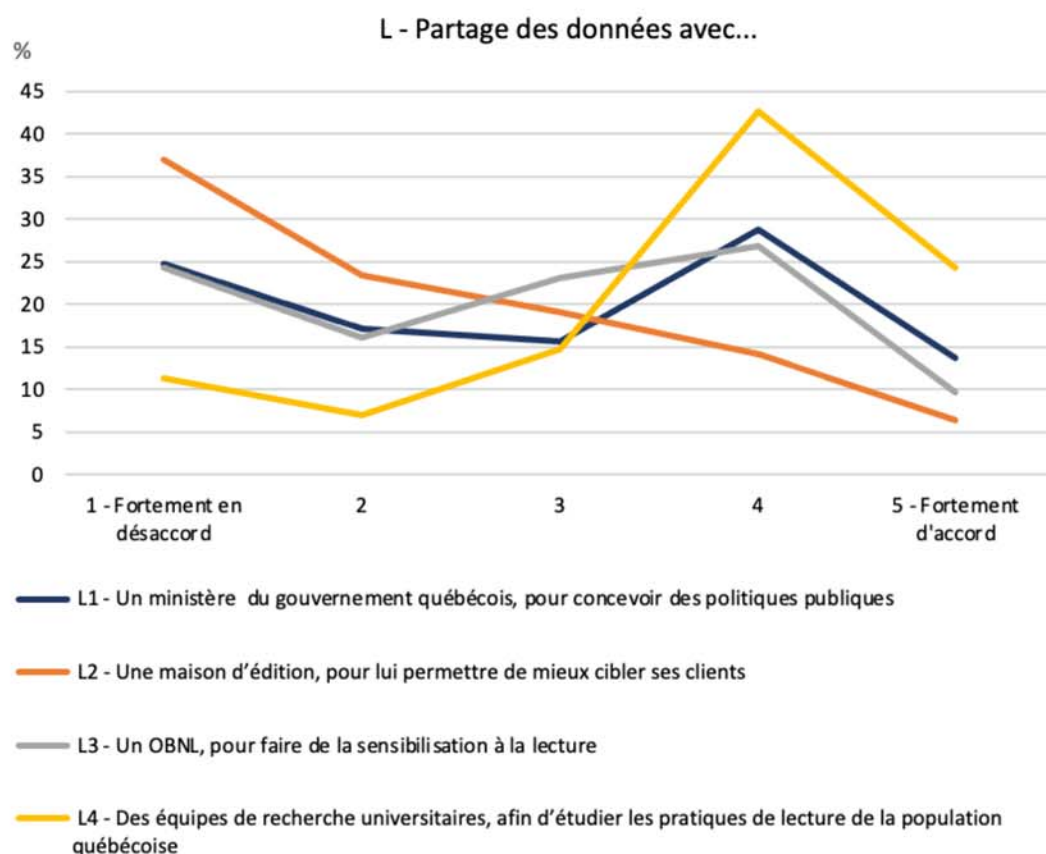


Tableau 35 : Partager ses données anonymisées avec un ministère (sous-dimension L1) : fréquence des réponses

	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
1 - Fortement en désaccord	336	24,8	336	24,8
Plutôt en désaccord	232	17,1	568	41,9
Ni d'accord ni en désaccord	213	15,7	781	57,6
Plutôt d'accord	391	28,8	1172	86,4
5 - Fortement d'accord	185	13,6	1357	100,0

Fréquence manquante = 73

Tableau 36 : Partager ses données anonymisées avec une maison d'édition (sous-dimension L2) : fréquence des réponses

	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
1 - Fortement en désaccord	478	37,0	478	37,0
Plutôt en désaccord	302	23,4	780	60,4
Ni d'accord ni en désaccord	247	19,1	1027	79,5
Plutôt d'accord	183	14,2	1210	93,7
5 - Fortement d'accord	82	6,4	1293	100,0

Fréquence manquante = 138

Tableau 37 : Partager ses données anonymisées avec une OBNL (sous-dimension L3) : fréquence des réponses

	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
1. Fortement en désaccord	307	24,3	307	24,3
Plutôt en désaccord	203	16,1	510	40,4
Ni d'accord ni en désaccord	291	23,1	801	63,5
Plutôt d'accord	339	26,9	1140	90,3
5. Fortement d'accord	122	9,7	1262	100,00

Fréquence manquante = 168

Tableau 38 : Partager ses données anonymisées avec des équipes de recherche universitaires (sous-dimension L4) : fréquence des réponses

	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
1. Fortement en désaccord	149	11,4	149	11,4
Plutôt en désaccord	91	7,0	240	18,3
Ni d'accord ni en désaccord	194	14,8	434	33,1
Plutôt d'accord	558	42,6	992	75,7
5. Fortement d'accord	318	24,3	1310	100,0

Fréquence manquante = 120

6.6. Perception au regard de la protection de la vie privée : déterminants sociodémographiques

Modèle d'analyse

Les cinq sous-dimensions étant mesurées de la même façon, sous forme d'échelle d'accord (de 1 à 5), nous avons encore une fois utilisé le modèle de régression logistique ordinale. Rappelons que ce modèle permet de mesurer l'effet net de certaines caractéristiques sur les probabilités d'avoir une position d'ouverture pour chacune des sous-dimensions, par rapport à un groupe de référence (entre parenthèses).

Le score global n'a pas été utilisé en régression logistique car les niveaux sont trop nombreux (entre 5 et 28) et parce que les résultats des sous-dimensions ne sont pas homogènes : l'influence des variables sociodémographiques n'a donc pas été calculée sur le score global, mais plutôt sur chaque sous-dimension.

Influence des variables sociodémographiques par sous-dimension

Les variables sociodémographiques qui ont une influence sur la perception relative à la gestion des données personnelles sont : le niveau de **diplomation** (influence sur deux sous-dimensions, quatre items), **l'âge** (influence sur une sous-dimension, deux items), le fait d'appartenir au **groupe LGBTQ+** (influence sur une sous-dimension, un item) et la **littératie des données personnelles** (influence sur une sous-dimension, un item).

Les autres variables mesurées (genre, occupation, secteur d'emploi, minorité visible) n'ont pas d'effet sur les sous-dimensions retenues du score de perception relative à la gestion des données personnelles.

Recouper les données (sous-dimension J)

De même, **des études universitaires diminuent d'environ de moitié les chances d'être favorable** à ce que BAnQ recoupe différentes données personnelles afin de personnaliser ses services, par rapport aux personnes sondées n'ayant aucun diplôme ou un diplôme d'études secondaires uniquement. Les personnes détenant un diplôme postsecondaire non-universitaire ne se distinguent pas des personnes sondées moins diplômées à ce sujet.

Durée de conservation de données (sous-dimension K)

Aucun facteur sociodémographique n'est en lien avec la durée de conservation des données personnelles.

Partage des données avec des tiers (sous-dimension L)

Détenir un ou des diplômes universitaires augmente les chances d'être favorable au partage des données personnelles avec des équipes de recherche universitaires (L4) comparativement aux personnes sondées n'ayant aucun diplôme ou un diplôme d'études secondaires uniquement. **Appartenir au groupe des plus jeunes (18 à 34 ans) augmente également les chances d'être favorable** au partage avec des équipes de recherche (L4) comparativement aux 65 ans et plus. Au contraire, appartenir au groupe âgé de 45 à 54 ans diminue les chances d'être favorable à un tel partage (voir Figure 25).

Détenir un diplôme universitaire supérieur au baccalauréat (comparativement à ne détenir aucun diplôme ou un diplôme d'études secondaires) est un facteur qui réduit les chances d'être en accord avec le partage des données personnelles avec une **maison d'édition** (L2 ; voir Figure 23).

Comme pour le partage avec des équipes de recherche universitaires, un diplôme d'études universitaires supérieures au baccalauréat augmente les chances d'être favorable avec un partage avec un **ministère** québécois afin de concevoir des politiques publiques (L1 ; voir Figure 22).

Enfin, **les plus jeunes**, âgés de 18 à 34 ans ont deux fois plus de chances comparativement aux 65 ans et plus d'être d'accord avec le partage des données personnelles avec une **OBNL** pour faire de la sensibilisation à la lecture (L3). De même, s'identifier à une **minorité sexuelle** augmente également les chances d'être en accord avec un tel partage (Figure 24).

Influence de la littératie des données personnelles par sous-dimension

Ici aussi, les personnes sondées ont exprimé leur perception après avoir été informées des bonnes réponses aux questions portant sur la littératie des données personnelles. Ainsi leur perception a pu être éclairée par une meilleure compréhension de la nature, de la portée et des enjeux des données personnelles.

Les résultats montrent que **le niveau de littératie des données personnelles a un impact** sur deux sous-dimensions (quatre items) de la perception relative à la gestion des données personnelles.

Un score élevé en littératie des données personnelles réduit les chances d'être favorable à ce que BANQ puisse recouper différentes données pour personnaliser ses services. Par exemple, une augmentation de 5 points dans le score de littératie diminue de 40% les chances d'être favorable (Figure 21).

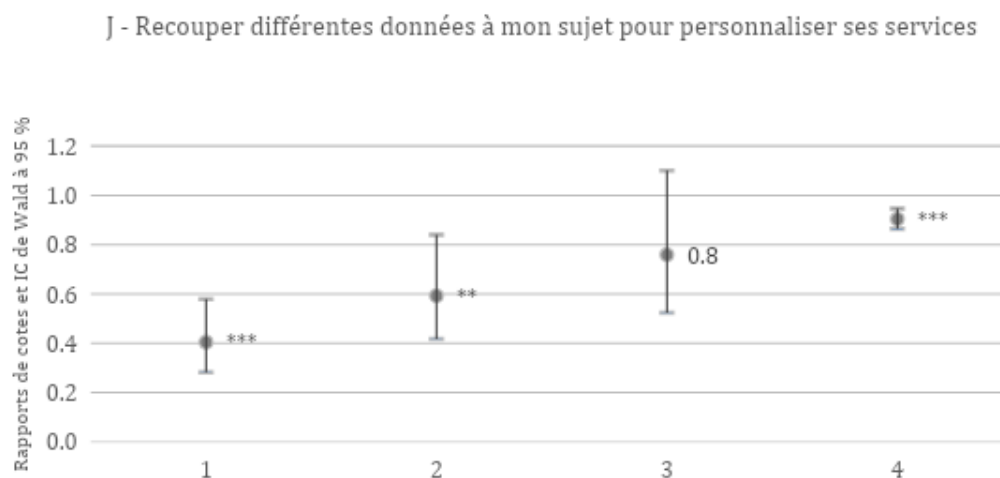
De plus, avoir obtenu un score élevé en littératie des données personnelles est un facteur qui **réduit les chances d'être en accord avec le partage des données personnelles avec une maison d'édition**. Par exemple, une augmentation de 5 points dans le score de littératie diminue de moitié les chances d'être favorable à ce type de partage.

Un score élevé en littératie des données personnelles **diminue légèrement les chances d’être en accord avec le partage des données avec un ministère québécois** afin de concevoir des politiques publiques. Ainsi, une augmentation de 5 points dans le score de littératie diminue de 25% les chances d’être favorable à ce type de partage.

Un score élevé en littératie des données personnelles **diminue les chances d’être favorable au partage des données avec une OBNL**. Dans ce cas, une augmentation de 5 points dans le score de littératie diminue de 30% les chances d’être favorable à ce type de partage.

En revanche, le niveau de littératie des données personnelles n’a pas d’effet sur le partage avec des équipes de recherche universitaires, ni sur la durée de conservation des données.

Figure 21 : Influence du niveau de diplomation et de la littératie sur la perception quant au recouplement des données personnelles (sous-dimension J)



* $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$

Figure 22 : Influence du niveau de diplomation et de la littératie des données personnelles sur la perception quant au partage des donnée personnelles avec un ministère québécois (sous-dimension L1)

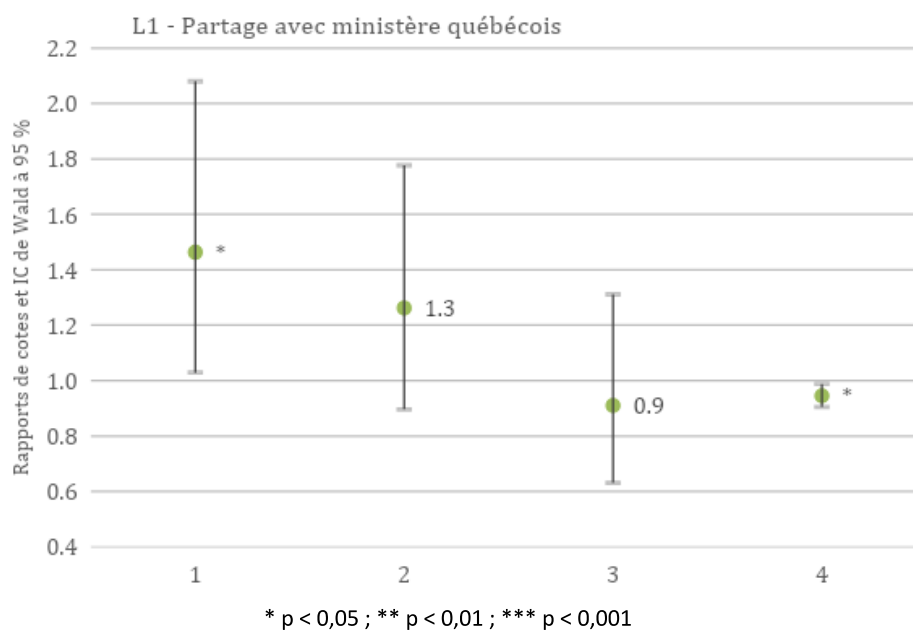


Figure 23 : Influence du niveau de diplomation et de la littératie des données personnelles sur la perception quant au partage des donnée personnelles avec une maison d'édition (sous-dimension L2)

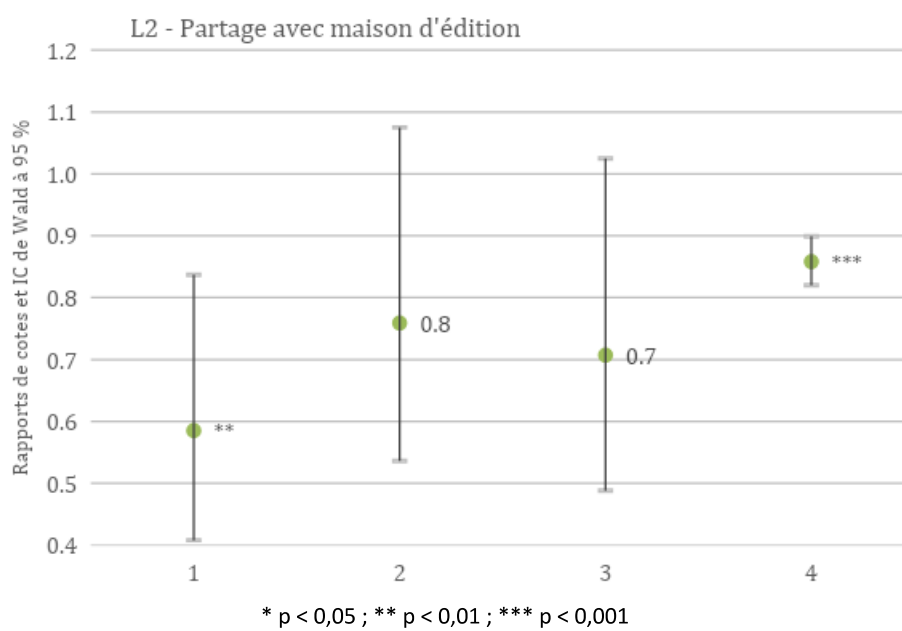


Figure 24 : Influence de l'âge, de l'affiliation au groupe LGBTQ+ et de la littératie des données personnelles sur la perception quant au partage des données personnelles avec une OBNL (sous-dimension L3)

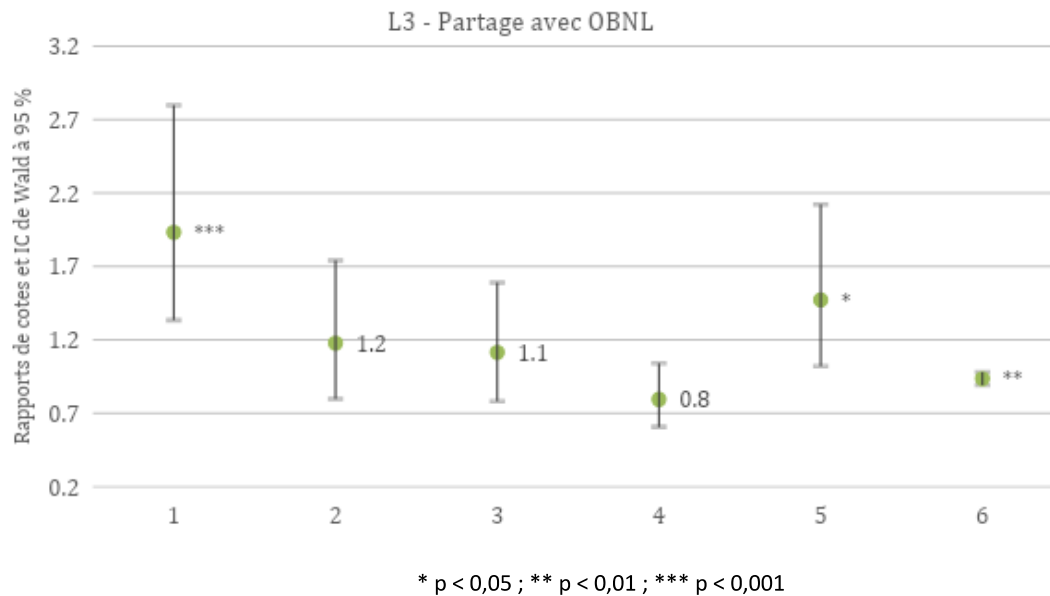
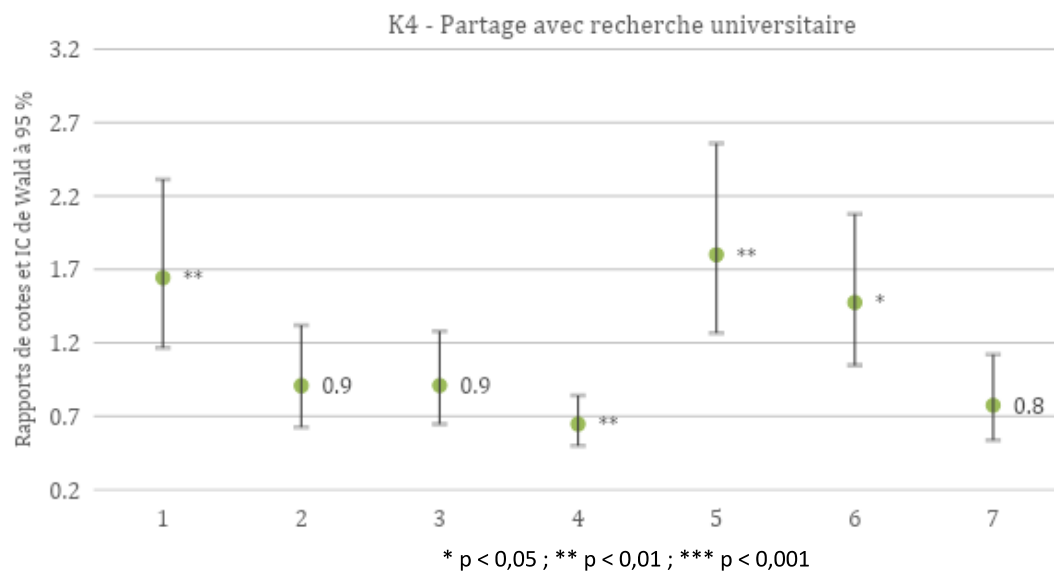


Figure 25 : Influence de l'âge et du niveau de diplôation sur la perception quant au partage des données personnelles avec des équipes de recherche universitaires (sous-dimension L4)



6.7. Relation entre littératie, intérêt et attitude

Modèle d'analyse

Afin de vérifier l'existence d'une relation entre la littératie des données personnelles et l'intérêt pour le développement de nouveaux services dans un premier temps, la littératie et l'attitude en regard de la protection des données personnelles dans un second temps et finalement entre l'intérêt et l'attitude, nous avons mené des tests d'indépendance du chi-deux. Pour ce faire nous avons croisé les variables de score définies en niveaux. Le test d'indépendance part de l'hypothèse qu'il n'existe pas de relation entre les variables. Donc que la répartition des proportions des lignes, par exemple, est identique d'une ligne à l'autre et que toute variation est due au hasard d'échantillonnage. La statistique du chi-deux mesure l'ampleur de la différence entre les résultats observés et ceux auxquels on aurait pu s'attendre, si l'hypothèse d'indépendance était avérée.

Littératie / Intérêt

On observe **une association entre le niveau de littératie des données personnelles et le niveau d'intérêt** pour le développement de nouveaux services fondés sur les données d'usage ($p < 0,0001$).

Un niveau de littératie très élevé est associé à un niveau d'intérêt très bas ou bas. Ainsi, près de la moitié (42,5%) des personnes sondées ayant un bas intérêt pour le développement de nouveaux services ont un niveau de littératie élevé. Et 33,8% de celles qui ont un niveau d'intérêt très bas ont un niveau de littératie très élevé. **En comparaison, les personnes sondées ayant un niveau de littératie bas à très bas expriment un intérêt élevé à très élevé pour de nouveaux services.** On observe notamment que parmi les personnes ayant un intérêt très élevé pour les nouveaux services, 20,9% ont un bas niveau de littératie (Tableau 39).

Littératie / Attitude

On observe le même type d'association, quoiqu'un peu moins marqué, entre le niveau de littératie des données personnelles et l'attitude des personnes sondées au sujet de la gestion de leurs données personnelles ($p < 0,01$).

Les résultats montrent une forme de polarisation entre d'un côté les personnes sondées ayant un niveau de littératie très élevé et, d'un autre côté, celles ayant un niveau bas à très bas. En effet, parmi les personnes sondées ayant une attitude protectrice ou très protectrice, 21% possèdent un niveau de littératie très élevé. Parmi celles qui ont une attitude plutôt neutre, plus du tiers (37,3%) ont un niveau de littératie élevé. À l'autre extrémité, le tiers (32,4%) des individus très ouverts au regard de la gestion de leurs données personnelles ont un niveau de littératie bas ou très bas (Tableau 40).

Attitude / Intérêt

Il existe aussi une association entre l'intérêt pour le développement de nouveaux services et l'ouverture à la valorisation des données par BAnQ.

Ainsi, les deux tiers des personnes sondées (67,1%) qui ont **une attitude très ouverte face à la valorisation de leurs données personnelles ont un niveau d'intérêt très élevé pour les nouveaux services**. À l'autre bout du spectre, près de la moitié (43,6%) de celles qui sont les plus protectrices de leurs données ont un intérêt bas ou très bas (Tableau 41).

Tableau 39 : Niveau d'intérêt pour le développement de nouveaux services, selon le niveau de littératie des données personnelles

Niveau d'intérêt	Niveau de littératie					Total
	Très bas	Bas	Moyen	Élevé	Très élevé	
Très bas	3	7	16	19	23	68
	4,4%	10,3%	23,5%	27,9%	33,8%	100,0%
Bas	2	13	32	57	30	134
	1,5%	9,7%	23,9%	42,5%	22,4%	100,0%
Moyen	4	41	61	59	40	205
	2,0%	20,0%	29,8%	28,8%	19,5%	100,0%
Élevé	14	79	110	125	54	382
	3,7%	20,7%	28,8%	32,7%	14,1%	100,0%
Très élevé	22	83	98	154	41	398
	5,5%	20,9%	24,6%	38,7%	10,3%	100,0%
Total	45	223	317	414	188	1187
	3,8%	18,8%	26,7%	34,9%	15,8%	100,0%

Fréquence manquante = 243

Tableau 40 : Niveau de littératie des données personnelles, selon l'attitude relative à la gestion des données personnelles (fréquence et pourcentage)

Niveau de littératie						
Attitude	Très bas	Bas	Moyen	Élevé	Très élevé	Total
Très protecteur	6	33	48	59	39	185
	3,2%	17,8%	26,0%	31,9%	21,1%	100,0%
Protecteur	5,0	34,0	50,0	58,0	40,0	187,0
	2,7%	18,2%	26,7%	31,0%	21,4%	100,0%
Neutre	12	69	97	147	69	394
	3,1%	17,5%	24,6%	37,3%	17,5%	100,0%
Ouvert	5	34	77	79	23	218
	2,3%	15,6%	35,3%	36,2%	10,6%	100,0%
Très ouvert	14	45	51	52	20	182
	7,7%	24,7%	28,0%	28,6%	11,0%	100,0%
Total	42	215	323	395	191	1166
	3,6%	18,4%	27,7%	33,9%	16,4%	100,0%

Fréquence manquante = 264

Tableau 41 : Niveau d'intérêt pour le développement de nouveaux services, selon l'attitude relative à la gestion des données personnelles (fréquence et pourcentage)

Attitude	Niveau d'intérêt					Total
	Très bas	Bas	Moyen	Élevé	Très élevé	
Très protecteur	34	41	44	33	20	172
	19,8%	23,8%	25,6%	19,2%	11,6%	100,0%
Protecteur	12	34	31	54	37	168
	7,1%	20,2%	18,5%	32,1%	22,0%	100,0%
Neutre	16	38	79	137	94	364
	4,4%	10,4%	21,7%	37,6%	25,8%	100,0%
Ouvert	4	11	26	75	93	209
	1,9%	5,3%	12,4%	35,9%	44,5%	100,0%
Très ouvert	1	2	14	38	112	167
	0,6%	1,2%	8,4%	22,8%	67,1%	100,0%
Total	67	126	194	337	356	1080
	6,2%	11,7%	18,0%	31,2%	33,0%	100,0%

Fréquence manquante = 350

7. SYNTHÈSE ET DISCUSSION DES RÉSULTATS

7.1. Littératie des données personnelles

(QR 1a) Un niveau plutôt élevé, mais des lacunes importantes

Le niveau de littératie des données personnelles des personnes sondées est plutôt bon avec une majorité de personnes sondées (près de 60%) situées entre les niveaux moyen et élevé. On trouve davantage de personnes sondées situées au niveau très élevé (14,9%) qu'au niveau très bas (5,8%). **Les deux tiers des personnes sondées (67%) identifient correctement que l'historique des emprunts à la bibliothèque est une donnée personnelle.**

Dans la composition de notre échantillon, les 55 ans et plus sont surreprésentés, or ils tendent à avoir un niveau plus bas. On peut donc faire l'hypothèse que **le portrait global du niveau de littératie des données personnelles de la population des personnes abonnées à BAnQ serait plus élevé que celui de notre échantillon.** Par ailleurs, les personnes diplômées universitaires tendent à avoir un score plus haut : étant surreprésentées dans la population des bibliothèques publiques, elles contribuent très probablement à la mesure d'un niveau de littératie des données personnelles élevé dans notre échantillon. Ainsi, on peut faire l'hypothèse que **la population du Québec aurait un niveau de littératie des données personnelles plus bas que celui des personnes abonnées à BAnQ.**

Nos résultats montrent un niveau de compétence légèrement plus élevé que ce que suggère l'étude de Zini (2020), bien qu'elle porte sur un objet différent (le profil d'utilisation des services numériques par les usagères et usagers de BAnQ) : elle concluait que 50% d'entre eux se déclarent à l'aise avec les interfaces numériques.

Si nos résultats font apparaître des scores généraux de littératie des données personnelles globalement bons, il faut cependant noter **certaines lacunes importantes. Ainsi, le quart des personnes sondées (26,1%) a un niveau bas ou très bas.** Globalement, le score tend à décroître au fur et à mesure que le degré de difficulté des questions augmente.

- **Les connaissances plus complexes sont moins bien maîtrisées** : seuls 14,1% des personnes sondées ont identifié toutes les bonnes réponses aux questions de *compréhension des implications des données personnelles* ; et elles sont seulement un tiers (35,1%) pour les questions de *réflexion critique sur la gouvernance* des données personnelles.
- Ainsi, **la compréhension des enjeux des données personnelles** est l'aspect de la littératie des données personnelles qui pose le plus de difficultés aux personnes sondées. **Les lacunes de connaissances les plus courantes concernent les données sensibles et l'anonymisation des données personnelles.** À ce titre, les deux tiers des

personnes sondées (67%) estiment de manière erronée qu'il n'y a aucun moyen de relier une usagère ou un usager à des informations anonymisées.

- En ce qui concerne les connaissances relatives à **l'historique d'emprunt**, qui est la principale source de données personnelles que BAnQ souhaite exploiter à l'avenir, un tiers des personnes sondées (33%) n'est pas conscient qu'il constitue un ensemble de données personnelles, et près de la moitié (51%) ne sont pas conscients qu'il peut potentiellement révéler des informations sensibles.

(QR 1b) Un niveau qui varie surtout en fonction de l'âge et la scolarisation

Nos résultats montrent que le niveau de littératie des données personnelles varie en fonction de certaines caractéristiques sociodémographiques des personnes abonnées (Figure 26).

- L'âge et la scolarisation (niveau de diplomation) influencent le niveau global de littératie des données personnelles. **Plus on est jeune ou diplômé, meilleures sont les chances d'obtenir un score élevé en littératie des données personnelles.**
- Le niveau de diplomation a un impact particulièrement fort sur la capacité à identifier les données personnelles, tandis que l'âge a une influence particulièrement importante sur la compréhension des enjeux des données personnelles.
- Par ailleurs, le fait de s'identifier au groupe LGBTQ+ influence la compréhension des enjeux des données personnelles : **les personnes abonnées qui s'identifient au groupe LGBTQ+ tendent à mieux comprendre les enjeux des données sensibles.**

En mettant en perspective nos résultats avec les travaux antérieurs, on peut souligner qu'ils concordent avec ceux de Park (2013a et 2013b) et Epstein et al. (2020) sur le fait que les facteurs déterminants de la littératie des données personnelles sont l'âge et le niveau d'éducation.

Âge :

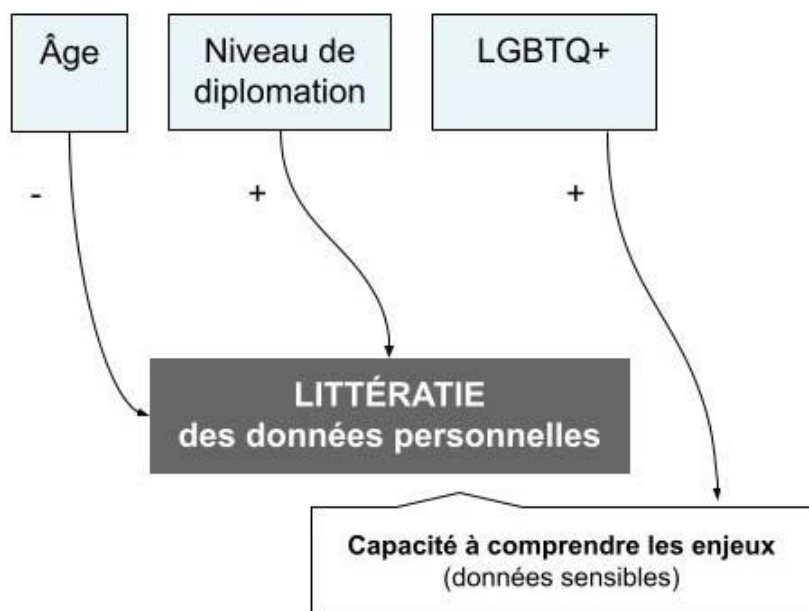
- Nos résultats rejoignent en particulier les travaux de Park (2013b) autour du fait que le niveau de littératie des données et l'âge sont négativement associés (plus l'âge augmente, plus le niveau de littératie baisse).
- Ils tendent aussi à soutenir l'hypothèse de Epstein et al. (2020) sur le fait que les connaissances liées au cadre légal des données personnelles pourraient être davantage présentes chez les groupes plus âgés. En effet, notre étude portait uniquement sur les deux premières dimensions du cadre OPL en excluant les connaissances du cadre légal et nous avons trouvé que l'âge est négativement associé à la littératie des données personnelles. À l'inverse, dans leur étude qui incluait les connaissances relatives au cadre légal de la protection des données personnelles, ils ont plutôt trouvé que l'âge est positivement associé à la littératie de la vie privée en ligne (les groupes plus âgés ont obtenu un meilleur score). Toutefois, l'étude de Park (2013b) intégrait elle aussi la connaissance du cadre légal dans sa mesure et elle a obtenu des résultats similaires aux nôtres, ce qui va à l'encontre de l'hypothèse. **Plus de travaux sont donc encore**

nécessaires pour comprendre si la connaissance du cadre légal de protection des données personnelles varie en fonction de l'âge.

Groupe LGBTQ+ :

- Nos résultats pourraient s'expliquer par le fait que les **risques particuliers d'atteinte à la vie privée des minorités sexuelles en ligne** (Robinson et al., 2020) les conduisent à être plus informées des enjeux des données personnelles, et notamment des données sensibles, dont l'orientation sexuelle fait partie. Cette interprétation résonne avec les résultats de l'enquête de Lutz et al. (2020) qui montrent que l'expérience d'une menace sur la vie privée tend à augmenter les préoccupations en matière de protection de la vie privée.

Figure 26 : Synthèse des hypothèses concernant l'influence des variables sociodémographiques sur le score global de littératie des données personnelles



7.2. Intérêt pour les nouveaux services fondés sur les données d'usage

(QR 2a) L'intérêt est fort pour la plupart des services projetés

Dans l'ensemble, le niveau d'intérêt pour le développement de nouveaux services fondés sur les données est fort : **les deux tiers (65,7%) des personnes sondées ont un niveau d'intérêt global élevé à très élevé.**

La proposition qui recueille le plus d'avis favorables est celle d'enrichir les collections de BAnQ grâce à l'historique d'emprunt des personnes abonnées (85,2% sont fortement ou plutôt d'accord avec cette utilisation). Cependant, la formulation positive de cette question dans le sondage a possiblement créé un biais favorable pour cette proposition. **Ce résultat n'est donc pas un indicateur clair.**

- **Il doit être interprété avec précaution, d'autant plus que certains risques liés à cette pratique ont été identifiés** par les travaux antérieurs (renforcement de la popularité des types de documents déjà les plus demandés au détriment de la découvrabilité des autres contenus).
- Dans une perspective inverse, un des bénéfices de **l'analyse des données d'usage pourrait être d'éclairer des « effets de niche »** : ainsi, selon Hanquinet et al. (2019), ces données permettent d'identifier des nuances et des spécificités des pratiques culturelles qui approfondissent leur compréhension au-delà des formes majoritaires.

En deuxième position (selon les scores moyens de réponse) se trouve la proposition de **recommander des documents** aux personnes abonnées en fonction de leurs habitudes et préférences.

- Ce résultat rejoint le constat de l'ABPQ (2021) : la proposition de se voir **offrir des suggestions de lecture** est un des services privilégiés par les personnes abonnées aux services des bibliothèques publiques qu'ils ont sondées.
- Comme pour l'enrichissement des collections, certains risques sont identifiés par la littérature académique : la recommandation algorithmique peut **enfermer les usagères et usagers dans des profils** qui déterminent l'offre qui leur est proposée et aller à **l'encontre de la diversification** des pratiques culturelles (Caillet, 2016 ; Donnat, 2016). Là encore, il s'agit de trouver un équilibre entre des tendances majoritaires qui se renforcent par effet de boucle (Fourcade, 2020) et des opportunités de nouvelles découvertes.

En troisième position se trouve la **personnalisation de l'expérience numérique sur le site Web** de BAnQ : elle recueille une majorité d'avis favorables, mais on note aussi un nombre important d'avis neutres (22,1% sont ni d'accord ni en désaccord). Ces résultats concordent également avec ceux de l'ABPQ (2021) qui indiquaient que l'amélioration de la navigation sur le site web recueille un certain intérêt de la part des personnes abonnées aux services des bibliothèques, mais dans une moindre mesure que les autres services. La navigation sur le site web faisait aussi

partie des préoccupations exprimées par les usagères et usagers de BAnQ lors de la consultation publique menée par l'institution en 2020 au sujet de sa nouvelle Stratégie numérique.

Enfin, en dernière position se trouve la proposition de **recommander d'autres activités culturelles**, qui récolte tout de même l'avis favorable de près de la moitié des personnes sondées, mais aussi un nombre important d'avis neutres (près d'un tiers des personnes sondées ne sont ni d'accord ni en désaccord). Ainsi, les services de recommandation croisée qui pourraient résulter de la mutualisation des données entre organismes culturels ne sont pas prioritaires par les personnes sondées, même si elles sont nombreuses à y être favorables.

(QR 2b) Certains aspects de l'intérêt sont influencés par la scolarisation et le fait d'appartenir à une minorité visible

Nos résultats montrent que le niveau d'intérêt des personnes abonnées pour le développement de nouveaux services fondés sur les données d'usage varie en fonction de deux caractéristiques sociodémographiques : **la scolarisation (niveau de diplomation) et le fait d'appartenir à une minorité visible**. Toutefois, l'influence de ces variables est limitée : elles n'affectent que certains aspects de l'intérêt (Figure 27).

Scolarisation :

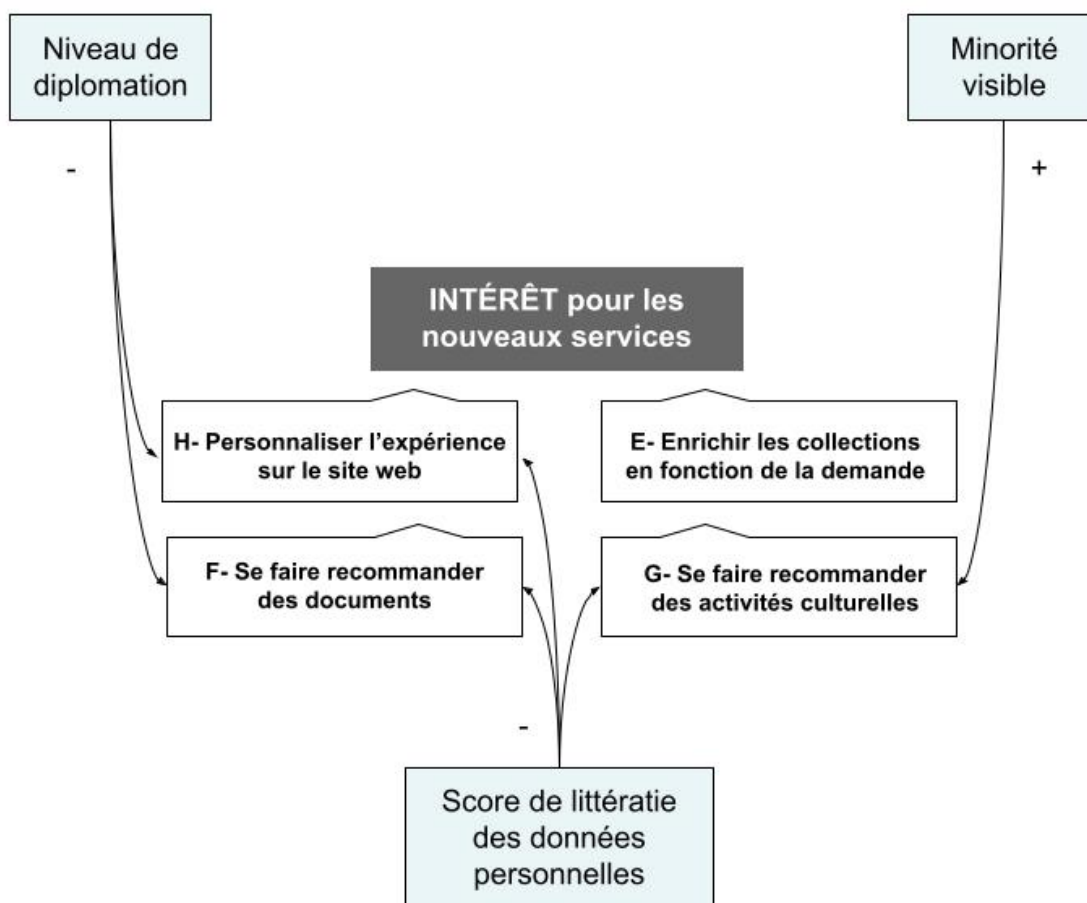
- Le score d'intérêt et le niveau de scolarisation sont négativement associés en ce qui concerne deux sous-dimensions : les personnes sondées **les plus scolarisées ont moins de chance d'être intéressées par la recommandation de documents ou la personnalisation de l'expérience sur le site Web de BAnQ**.
- Ainsi les personnes les plus diplômées, qui sont aussi les plus férues en littérature numérique, ont moins d'intérêt à se faire accompagner dans leurs choix de contenus et de personnalisation des services. Il est possible que leur compétences existantes, acquises durant la scolarisation, les outillent mieux pour guider leurs choix et leurs parcours de navigation en ligne.
- Le développement de nouveaux services de recommandation de documents et de personnalisation de l'expérience sur le site web de BAnQ pourrait donc surtout bénéficier aux personnes abonnées les moins diplômées.

Minorité visible :

- L'intérêt pour les nouveaux services et l'identification comme minorité visible sont positivement associés en ce qui concerne une sous-dimension : **faire partie d'une minorité visible double les chances d'être favorable à se faire recommander des activités culturelles**. On peut supposer que parmi les membres de ce groupe se trouve une part importante de personnes immigrantes et que leur moindre connaissance de l'offre culturelle au Québec les inciterait à chercher davantage de recommandations à ce sujet.
- Un parallèle pourrait par ailleurs être esquissé avec les résultats d'une étude récente sur les publics des arts de la scène, laquelle montre que les personnes immigrantes de

première génération tendent à avoir des pratiques culturelles plus intenses que celles de seconde génération ou que les personnes considérées plus strictement natives du Québec (Daigle et coll., 2019). Cette tendance était également observable dans la dernière *Enquête sur les pratiques culturelles du Québec* (2014) qui révèle que les personnes immigrantes de première génération fréquentent davantage la Grande Bibliothèque que celles de seconde génération ou que les personnes nées au Québec. Des travaux complémentaires sur les pratiques culturelles de la population immigrante et des groupes minoritaires seraient nécessaires pour étayer ces hypothèses. Nous pouvons également retenir de ces études la possible pertinence, sur le plan méthodologique, d'une distinction générationnelle plus fine dans la catégorisation des personnes immigrantes¹³.

Figure 27 : Synthèse des variables qui influencent l'intérêt pour les nouveaux services



¹³ Les auteurs remercient Rosaire Garon pour ses éclairages sur la question des pratiques culturelles de groupes issus de l'immigration.

(QR 2c) Une littératie des données personnelles élevée fait baisser l'intérêt pour la majorité des nouveaux services

Nos résultats montrent que pour une majorité des nouveaux services proposés, l'intérêt des personnes abonnées varie en fonction du niveau de littératie des données personnelles. Ainsi, le niveau d'intérêt et le score de littératie des données sont négativement associés en ce qui concerne trois sous-dimensions : **plus le score de littératie est élevé, plus le niveau d'intérêt baisse concernant la recommandation de documents, la recommandation d'activités culturelles et la personnalisation de l'expérience sur le site Web.**

On aurait pu supposer que le fait de connaître davantage les possibilités offertes par la valorisation des données d'usage développerait l'intérêt pour le développement de nouveaux services fondés sur ces données. C'est en tout cas ce que laissaient penser certains commentaires publiés par des citoyennes et citoyens lors de la consultation publique menée en 2020 par BAnQ au sujet de sa nouvelle Stratégie numérique. Mais c'est plutôt l'inverse qui est observé ici. Les résultats du sondage invitent plutôt à supposer qu'**un niveau élevé de littératie des données personnelles rendrait plus conscient des risques associés à ces pratiques, et par conséquent, il réduirait l'intérêt à se voir offrir de tels services.**

7.3. Perception de la gestion des données personnelles

(QR 3a) Des attitudes très partagées entre protection et ouverture

Au sujet de la gestion des données personnelles par BAnQ, les attitudes des individus participant au sondage sont très partagées : **un premier tiers (32,1%) sont très protecteurs ou protecteurs** de leurs données personnelles, et souhaitent limiter le recoupement, le partage et la durée de conservation des données. **Un second tiers (33,7%) se trouvent dans une attitude neutre**, autrement dit ni particulièrement protecteurs ni particulièrement ouverts. **Enfin un dernier tiers (34,2%) sont ouverts ou très ouverts** à ce que BAnQ valorise et conserve plus longuement leurs données. **En ce qui concerne plus spécifiquement l'exploitation des données personnelles, les avis sont aussi très partagés** : la proposition de recouper différentes données au sujet des personnes abonnées pour personnaliser les services de BAnQ suscite un nombre à peu près équivalent d'avis favorables (39,2%) et défavorables (37,8%), les autres se déclarant ni d'accord ni en désaccord.

Bien que notre analyse repose sur des catégories d'attitude différentes de celles utilisées dans les travaux précédents (voir section Enjeux pour la revue de la littérature) on peut tenter de les mettre en perspective. D'une part, **les études existantes montrent que l'attitude pragmatique est la plus répandue** (Westin, 2003 ; Motiwalla et Li, 2016 ; Elueze et Quan-Haase, 2018). Ces observations pourraient expliquer la part importante de l'attitude neutre (un tiers des personnes sondées) dans nos résultats.

D'autre part, **la fraction importante de personnes sondées exprimant une attitude protectrice ou très protectrice fait écho aux préoccupations des Canadiennes et Canadiens au sujet de la protection de la vie privée, mais les attitudes très négatives sont moins présentes chez les personnes que nous avons sondées.**

- Selon un sondage réalisé en 2020 pour le compte du Commissariat à la protection de la vie privée du Canada, « près de neuf Canadiens sur dix se disent préoccupés dans une certaine mesure par la protection de leur vie privée » (Phoenix SPI, 2021 : 10). Toutefois, mises en regard avec les résultats de notre enquête, les données de cette enquête suggèrent que nos répondantes et répondants sont moins nombreux à exprimer une attitude très négative dans le contexte de BAnQ : 16% sont « très protecteurs » dans notre enquête contre 39% de québécois se déclarant « énormément préoccupés » par la protection de leur vie privée (Phoenix SPI, 2021). Par contre, la part des personnes sondées exprimant une attitude « très ouverte » dans notre enquête (15,5%) est relativement proche de la part des citoyennes et citoyens « pas du tout préoccupés » (13%) dans l'enquête de Phoenix SPI (2021).
- Par ailleurs, alors que Kennedy et al. (2022) ont montré une forte présence de l'attitude critique face à la protection de la vie privée dans les services publics, nos résultats montrent plutôt que plus d'un tiers des personnes sondées sont ouvertes ou très ouvertes quant à l'utilisation de leurs données par BAnQ.

- **On peut donc faire l'hypothèse qu'il existe une moindre préoccupation concernant la gestion des données personnelles des utilisatrices et utilisateurs dans le contexte des bibliothèques publiques** comparé aux autres institutions et organismes privés. Cette hypothèse va dans le sens du rôle pressenti pour les bibliothèques publiques dans le défunt projet Quayside à Toronto. En effet, lors des débats sur le développement par Google de ce nouveau projet urbain, le Toronto Region Board of Trade (Ruttan et al., 2019) a plaidé pour que la gestion des données personnelles des citoyennes et citoyens captées dans les projets de ville intelligente soit confiée aux bibliothèques publiques sous la forme d'un « pôle de données civiques » (*civic data hub*) car **les bibliothèques auraient « les ressources, l'expertise et la confiance du public nécessaires » pour cela** (p. 1).
- **Une autre hypothèse serait que le type d'information personnelle détenue par BANQ est moins préoccupant comparé à d'autres informations personnelles** qui peuvent circuler sur Internet. En effet, les études existantes montrent que le degré de préoccupation varie en fonction des informations personnelles concernées : il est plus élevé pour des données de localisation (adresse, 75,4%) et pour les coordonnées (numéro de téléphone, 75,4%), et il est moins élevé pour les activités sociales (48,8%) et les opinions personnelles (35%) (Azzaria et Zimmermann, 2019). Les données collectées par BANQ concernent les loisirs culturels et les intérêts scientifiques, des catégories qui pourraient susciter un degré de préoccupation plus bas de la part des personnes sondées.

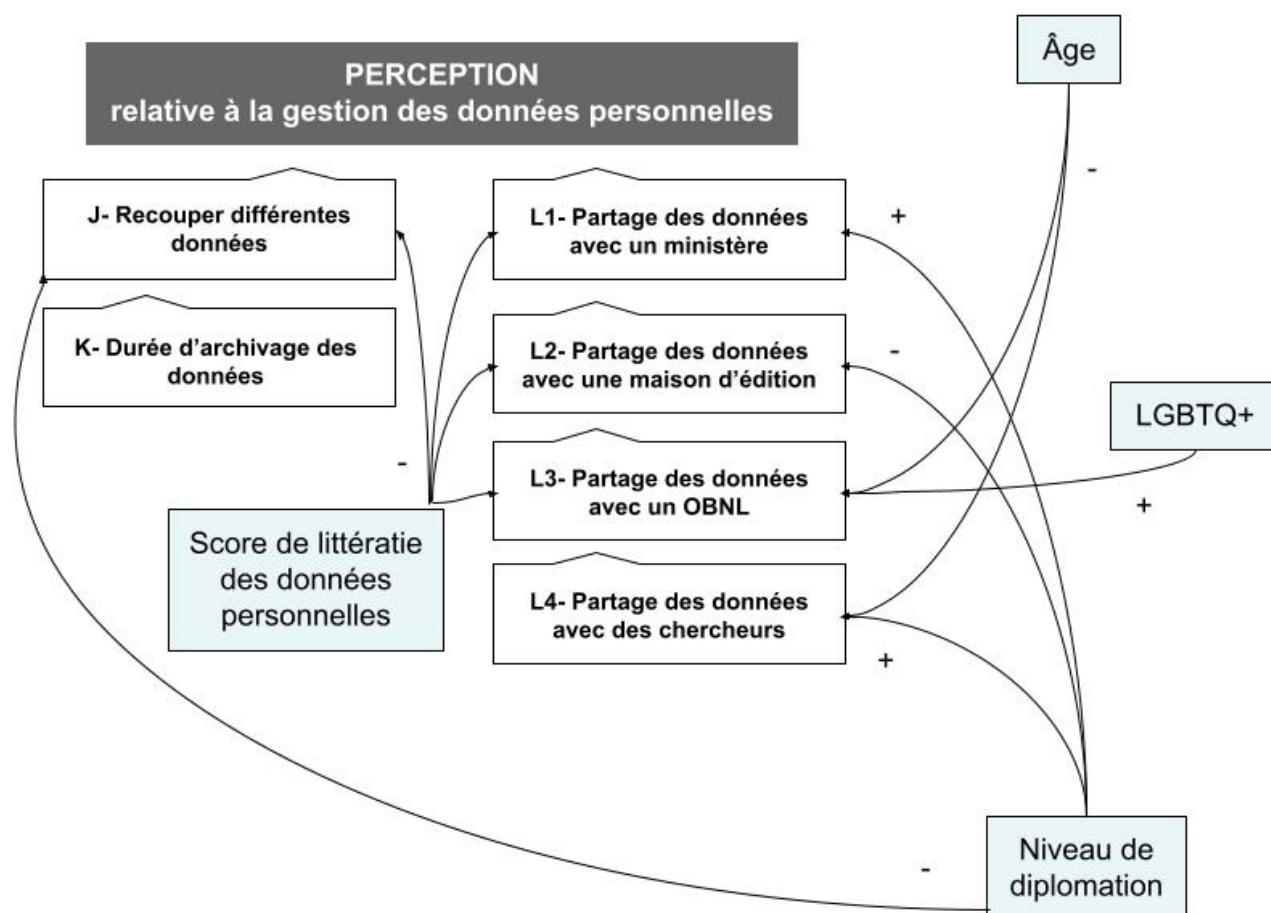
Sur le thème de l'archivage des données, une majorité se dégage plus nettement. **Plus des deux tiers des personnes sondées (69%) souhaitent limiter la durée de conservation de leurs données personnelles** : la période de conservation entre 1 an et 5 ans est privilégiée par un nombre important de personnes (40,1%).

Enfin, **concernant le partage des données personnelles anonymisées**, le partage avec des **équipes de recherche universitaires** afin d'étudier les pratiques de lecture de la population reçoit le plus fort taux d'approbation (66,9%). Ici, le fait que le sondage soit mené par une équipe universitaire peut avoir induit un biais favorable à cette proposition. Le partage avec une **maison d'édition** afin de lui permettre de mieux cibler ses clients reçoit le plus fort taux de désapprobation (60,4%). Les avis sont mitigés au sujet du partage avec un **ministère** du gouvernement québécois pour concevoir des politiques publiques (42,4% favorables) de même que pour le partage avec un **OBNL** pour faire de la sensibilisation à la lecture (36,6% favorables). Ces résultats rejoignent les constats faits par de précédentes études sur les différences de préoccupation en fonction de l'organisation qui gère les données. Ainsi, les individus participant à des enquêtes au Canada et au Québec déclarent être **moins préoccupés lorsque ce sont des organismes gouvernementaux et paragouvernementaux** que lorsque ce sont des entreprises privées ou commerciales qui utilisent leurs données personnelles (Phoenix SPI, 2021 ; Dilhac et al., 2018). En 2020, seuls 45% des Canadiennes et Canadiens jugeaient que les entreprises respectaient leur droit à la vie privée, contre 63% pour le gouvernement fédéral (Phoenix SPI, 2021).

(QR 3b) Certains aspects de la perception sont influencés par la scolarisation, l'âge et l'affiliation au groupe LGBTQ+

Nos résultats montrent que les perceptions des abonnés au sujet de protection de la vie privée varient en fonction de quatre caractéristiques sociodémographiques : **la scolarisation (niveau de diplomation), l'âge et le fait d'appartenir au groupe LGBTQ+**. Toutefois, ici aussi l'influence de ces variables n'est pas constante : elles n'ont d'effet que sur certains des aspects de l'intérêt, et parfois dans des sens contraires (Figure 28).

Figure 28 : Synthèse des variables qui influencent la perception relative à la gestion des données personnelles



Scolarisation

- Les résultats montrent que **le niveau de diplomation est le facteur qui influence le plus la perception** (quatre items sur sept), mais de façon **non homogène**. Ainsi, le niveau de diplomation a une influence dans des sens différents selon les éléments de perception mesurés. Ces résultats renforcent le constat que nous avons tiré de l'analyse de la littérature existante : **l'effet du niveau de diplomation sur les perceptions relatives à la vie privée n'est pas univoque et varie en fonction des contextes d'utilisation des données**.
- **Relativement à l'exploitation des données personnelles** par recoupement de différentes informations, le score de perception et le niveau de diplomation sont négativement associés : **plus on est diplômé, moins on a de chance d'être favorable à ces pratiques**. Ces résultats vont dans le même sens que les enquêtes de Phoenix SPI (2021), Blank et al. (2014) et Sheehan (2002) qui concluent à une association négative entre niveau de scolarisation et attitude ouverte ou confiance en ce qui concerne la protection de la vie privée.
- À l'inverse, le score de perception et **le partage des données avec des tiers qui sont des institutions à mission publique** (ministère, équipes de recherche universitaires) sont positivement associés : **plus on est diplômé, plus on a de chance d'être favorable à ce partage**. Dans ce cas, nos résultats rejoignent plutôt les conclusions de Phoenix SPI (2021) selon lesquelles **les groupes les plus diplômés** sont plus susceptibles d'accepter l'exploitation des renseignements personnels par le gouvernement (Phoenix SPI, 2021). Ces résultats rejoignent les études qui montrent que les groupes de population plus diplômés ont davantage confiance en les institutions publiques comparé aux groupes moins scolarisés.

Âge

- Selon nos résultats, l'âge a une influence sur certains aspects de la perception relative à la vie privée : en ce qui concerne le partage avec certains tiers, **l'âge et l'ouverture au partage des données sont négativement associés**. Ces résultats concordent avec les travaux antérieurs qui montrent que les préoccupations relatives à la protection de la vie privée tendent à augmenter en fonction de l'âge (Phoenix SPI, 2021 ; Azzaria et Zimmermann, 2019 ; McStay, 2020).
- **Ainsi, les groupes d'âge les plus jeunes (18 à 34 ans) tendent à être plus favorables au partage avec certains tiers**, mais ce n'est pas le cas pour toutes formes de partage. Les tiers pour lesquels les jeunes sont plus favorables au partage sont : une OBNL pour faire de la sensibilisation à la lecture (2 fois plus chances d'y être favorable comparé aux 65 ans et plus), et des équipes de recherche universitaires pour étudier les pratiques de lecture (1,6 fois plus de chance).
- Au contraire, appartenir au groupe âgé de 45 à 54 ans diminue les chances (0,6 fois) d'être favorable au partage avec les équipes de recherche. Quant aux groupes plus âgés (35 à 64 ans), ils ne sont ni plus ni moins favorables que le groupe des 65 ans et plus.

Affiliation au groupe LGBTQ+

- L'affiliation au groupe LGBTQ+ a une influence très localisée (sur un item d'une sous-dimension de la perception) : le fait d'appartenir à une minorité sexuelle et la perception relative au partage avec une OBNL pour des initiatives de sensibilisation à la lecture sont positivement associés. Ce résultat est difficilement interprétable. Il peut être mis en perspective avec une des conclusions de l'enquête de Kennedy et al. (2022) selon laquelle certains groupes minorisés peuvent être plus ouverts à l'exploitation de leurs données dans des contextes où ils perçoivent des bénéfices liés à l'offre de services adaptés à leurs besoins. Toutefois les initiatives de sensibilisation à la lecture ne sont pas directement associées aux besoins du groupe LGBTQ+, ce qui vient affaiblir l'hypothèse de cette interprétation dans le cas de notre enquête.

En conclusion, les résultats de notre enquête concordent avec les études antérieures sur le fait que **les principales variables sociodémographiques qui affectent la perception relative à la gestion des données personnelles sont l'âge et le niveau de diplomation**. Par ailleurs, nos résultats permettent de faire ressortir que l'affiliation au groupe LGBTQ+ a un effet sur ces perceptions mais qu'il est très limité. En revanche, notre enquête ne rejoint pas les conclusions de Tchéhouali (2021) au sujet de la variable du genre : nous n'avons pas observé d'effet de cette variable sur les perceptions et attitudes relatives à la gestion des données personnelles.

Enfin, **nos résultats sur l'effet de la variable du niveau de diplomation invitent à nuancer la compréhension de son influence** : son effet est différent selon les pratiques de données considérées (exploitation par croisement des données personnelles ou partage des données avec des tiers), et il varie aussi selon les contextes du partage (avec des institutions publiques ou avec une entreprise commerciale). Ces résultats nous amènent à faire ressortir que **l'étude des perceptions relatives à la protection de la vie privée en contexte ne peut supposer que ces perceptions sont homogènes pour toutes les pratiques d'une même organisation**. Au contraire, leur étude doit nécessairement prendre en compte la diversité des phases du cycle de vie des données (on peut être favorable à la collecte de données et défavorable à leur exploitation par recoupement) et la diversité des contextes de circulation de ces données (on peut accepter qu'elles soit partagées avec des tiers, mais seulement avec certaines catégories de tiers). Ces considérations résonnent avec la proposition de Nissenbaum (2009) qui invite à étudier les contextes spécifiques de circulation des données personnelles, et plus particulièrement les effets de continuité ou de rupture entre ces contextes. Ainsi, dans notre enquête, les personnes sondées les plus scolarisées tendent à être plus favorables à la circulation de leurs données entre des contextes voisins du point de vue de leur statut d'institution publique (d'une bibliothèque publique vers un ministère ou une université).

(QR 3c) Une littératie des données personnelles élevée rend plus protecteur quant à la gestion de ses données personnelles

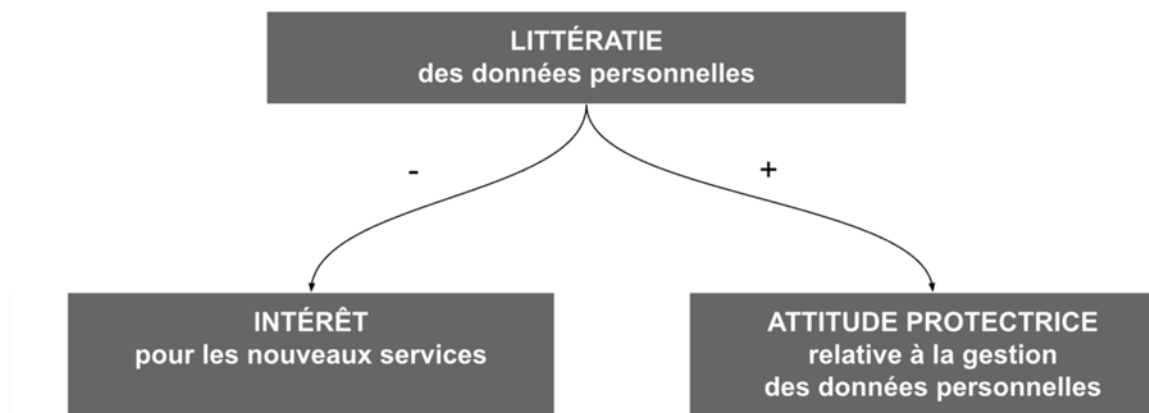
Les résultats montrent également que **le niveau de littératie des données personnelles a un impact important et univoque sur les perceptions relatives à la gestion des données personnelles**. Son effet est observable sur plus de la moitié des items relatifs à l'exploitation et au partage des données. Le sens de cette influence est constant : on constate une association négative entre le niveau de littératie des données personnelles et ces aspects de la perception. Autrement dit, **plus le score de littératie est élevé, moins on a de chances d'être favorable au fait que différentes données personnelles soient recoupées, et au fait de les partager avec un ministère, une maison d'édition et une OBNL**. Ces résultats concordent avec ceux de Kennedy et al. (2022) et de Hoffmann et al. (2016) qui portaient spécifiquement sur la mesure de la littératie des données personnelles.

7.4. Relation entre littératie, intérêt et attitude

(QR4) Les personnes abonnées qui ont plus de littératie sont plus protectrices et moins intéressées par les nouveaux services

Nos résultats montrent que **plus on a de littératie, plus on est protecteur de ses données personnelles et moins on est intéressé au développement de nouveaux services utilisant les données personnelles**. À l'inverse, plus on a d'intérêt pour ces nouveaux services, plus on est ouvert à la valorisation de ses données personnelles et moins on a de littératie des données personnelles (Figure 29).

Figure 29 : Synthèse des relations entre littératie des données personnelles, intérêt pour le développement de nouveaux services et attitude relative à la gestion des données personnelles



On peut faire **deux hypothèses d'explication**, potentiellement complémentaires.

- La première est que lorsque les personnes abonnées comprennent mieux la nature, les enjeux et les implications des données personnelles, elles seraient **plus conscientes des risques potentiels** que l'exploitation de ces données fait peser sur la vie privée, et donc elles seraient plus protectrices et moins intéressées à ce que leurs données soient exploitées pour développer de nouveaux services.
- La seconde est que lorsque les personnes abonnées comprennent mal la nature, les enjeux et l'incidence des données personnelles, c'est aussi qu'elles ont généralement **une plus faible littératie informationnelle et numérique** et donc moins de capacités pour trouver de l'information dans les ressources de BAnQ ; ces personnes seraient donc **plus intéressées à se faire offrir des services de médiation numérique**. Ces personnes se

soucieraient moins de potentiels impacts sur leur vie privée, soit parce qu'elles n'en seraient pas conscientes (faible littératie des données personnelles), soit parce qu'elles percevraient plus d'avantages que d'inconvénients (*privacy calculus*) à la collecte et à l'exploitation des données personnelles.

7.5. Modes d'encadrement des données d'usage identifiés par les personnes sondées

Les deux solutions les plus fréquemment identifiées par les personnes sondées pour **encadrer l'utilisation de leurs données personnelles** par BAnQ sont le fait de « donner la possibilité de refuser que [s]es données soient collectées » (81%) et le fait d'« informer précisément sur les données collectées et leur utilisation » (73%).

Premièrement, le fait de **donner la possibilité aux usagères et usagers de refuser que leurs données soient collectées** fait partie des bonnes pratiques déjà identifiées par les travaux sur le consentement à la collecte des données personnelles en contexte numérique (Laboratoire de Cyberjustice, 2019).

Deuxièmement, le fait d'**informer les usagères et usagers sur les données collectées et leur utilisation** est aussi une bonne pratique identifiée par les travaux académiques (Casemajor et al., 2021). Deux principaux défis ont été identifiés à cet égard dans les études existantes : d'une part l'accessibilité de cette information dans les interfaces des bibliothèques (Affonso et Sant'Ana, 2018) et d'autre part son intelligibilité étant donné la complexité des termes des contrats d'utilisation des données et l'asymétrie de l'information dont disposent les organisations et les usagères et usagers au sujet de la valorisation des données (Maceli, 2018 ; Bashir et al., 2015). Selon Trottier, la capacité des usagères et usagers à offrir un consentement éclairé en contexte numérique serait plus un privilège qu'un droit, car elle nécessite non seulement du temps mais également un niveau suffisant de littératie numérique (Trottier, 2018).

Enfin, l'option de « faire appel à un organisme de réglementation indépendant » (46%) pour contribuer à la gouvernance des données personnelles à BAnQ a été beaucoup moins choisie par les personnes sondées. Ce résultat diffère de l'étude de Hartman et al. (2020) menée au Royaume-Uni au sujet de la gestion des données personnelles, dans laquelle les personnes sondées ont choisi comme deuxième option la plus populaire le contrôle par un organisme de réglementation. Dans le contexte de notre étude, le moindre choix de cette option par les personnes sondées pourrait s'expliquer par la **tendance des citoyennes et citoyens canadiens à faire davantage confiance aux institutions publiques** (gouvernementales et paragouvernementales) qu'aux acteurs du privés dans la gouvernance des données personnelles. Ainsi, le fait de « faire appel à un organisme de réglementation indépendant » pourrait donc être perçu comme moins nécessaire dans le cas d'une institution publique comme BAnQ.

8. PISTES D'ACTION ET DE RÉFLEXION

8.1. Littératie des données personnelles

Piste d'action : contribuer à augmenter la littératie des données personnelles des citoyennes et citoyens

Étant donné que les données personnelles sont un des enjeux clés du débat actuel sur l'intelligence artificielle, et que BAnQ s'est donné comme objectif d'enrichir l'éducation numérique des publics (*Stratégie numérique 2021-2028*), des **actions éducatives pourraient être mises en place par BAnQ** pour contribuer à améliorer la littératie des données personnelles des citoyennes et citoyens (rappelons qu'un quart des personnes sondées ont un niveau bas à très bas). Plusieurs des personnes sondées ont d'ailleurs suggéré en commentaire que BAnQ offre des outils, des documents ou un atelier « d'éducation populaire » qui apporteraient « des éclaircissements aux nombreux usagers qui ne savent pas bien de quoi il s'agit ». Un meilleur partage de ces connaissances pourrait aussi contribuer à habiliter le dialogue démocratique sur ces sujets.

Piste d'action : améliorer en priorité les connaissances des citoyennes et citoyens sur les enjeux des données personnelles

Les plus grandes lacunes de connaissances des personnes sondées portent sur les *enjeux* des données personnelles. Ainsi, les actions éducatives à mettre en place pourraient prioriser la connaissance des éléments suivants : a) le fait que **l'historique d'emprunt** constitue une donnée personnelle ; b) le fait que des **données sensibles** peuvent y être présentes ; c) le fait que **l'anonymisation** des données personnelles ne peut jamais être garantie à cent pour cent.

Piste d'action : réduire le fossé de la vie privée en ciblant les groupes plus âgés et moins diplômés

Certains groupes d'usagers sont plus susceptibles d'avoir des lacunes en littératie des données personnelles. Pour réduire ce « fossé de la vie privée », des actions éducatives pourraient cibler les groupes qui en ont le plus besoin, soit **les personnes plus âgées (65 ans et plus) et les moins diplômées (en-dessous d'un diplôme postsecondaire)**.

8.2. Nouveaux services fondés sur les données d'usage

Piste d'action : cibler les nouveaux services qui contribueraient le plus à réduire la fracture numérique et à répondre à des besoins spécifiques

Le développement de nouveaux services fondés sur les données d'usage pourrait cibler en priorité les groupes les plus affectés par la fracture numérique pour **contribuer à réduire les inégalités sociales** liées au potentiel du numérique pour l'accès au savoir et à la culture (fracture numérique).

En particulier, la recommandation de documents et la personnalisation de l'expérience sur le site web de BAnQ pourraient surtout intéresser les personnes abonnées **les moins diplômées**. Les usagères et usagers **plus âgés** (65 ans et plus) pourraient également être ciblés car ils tendent à moins bien maîtriser les environnements numériques comparé aux plus jeunes.

Piste de réflexion : trouver un équilibre entre conforter la demande des usagers et leur faire découvrir de nouveaux horizons

Les projets qui s'appuient sur les données d'usage pour développer les collections (acquérir des documents en fonction de la demande) ou pour offrir des services aux usagères et usagers (recommander des documents, personnaliser la navigation sur le site web en fonction des habitudes et préférences) devraient viser un équilibre entre satisfaire les demandes des publics portées sur les types de documents les plus populaires et répondre à la mission de BAnQ de **valoriser l'enrichissement des connaissances dans toutes ses dimensions**. Plusieurs personnes sondées ont d'ailleurs souligné ce point en commentaire : « j'apprécie de me voir suggérer sur des tablettes à l'entrée des romans québécois, des ouvrages grand public, des manuels que je n'aurais jamais pensé consulter » ; les algorithmes tendent à « nous enfermer dans un cocon de confort [tandis que] la culture sert à dépasser ses limites ». Ainsi, les nouveaux services fondés sur les données d'usage devraient **favoriser la curiosité des usagères et usagers pour des documents moins connus ou plus nichés qui reflètent la diversité de la production québécoise** (et au-delà) sans se restreindre à l'offre déjà la plus populaire. Ce type d'action contribue à une plus grande **découvrabilité** de la collection.

Piste de réflexion : considérer que les données d'usage n'offrent qu'un portrait incomplet des pratiques des usagères et usagers

Les données d'usage offrent de nouvelles opportunités de connaissance des publics, mais il est important de garder en tête qu'elles ne fournissent pas un portrait complet ni totalement objectif des publics. La prise de décision dans le développement de nouveaux services devrait **aussi**

s'appuyer sur l'expertise des membres du personnel au contact avec les usagères et usagers qui ont une vision complémentaire de leurs pratiques et de leurs besoins.

Piste de réflexion : prendre en compte le coût écologique du développement de nouveaux services numériques et du stockage des données

Bien que notre enquête ne portât pas sur cet aspect, il est aujourd'hui incontournable de se poser la question du coût écologique du développement de nouveaux services numériques et du stockage des données d'usage lorsque celles-ci sont accumulées continûment. **Le principe de sobriété numérique invite à distinguer les besoins prioritaires** qui nécessitent impérativement la conception de nouveaux services numériques, et ceux pour lesquels le numérique offre un potentiel non essentiel qui ne nécessite pas un développement.

8.3. Protection de la vie privée

Le contexte législatif au Québec et au Canada a récemment changé pour renforcer les obligations relatives à la protection des données personnelles et au consentement des usagers. Au-delà des astreintes légales qui s'appliquent à toutes les organisations, les bibliothèques et archives publiques ont un rôle important à jouer dans la définition de bonnes pratiques pour la gouvernance des données personnelles étant donné la confiance du public envers ces institutions.

Piste d'action : mettre en œuvre des mécanismes de codesign des politiques et des services touchant aux données personnelles en amont pour impliquer davantage les usagères et usagers dans la gouvernance des données personnelles

L'implication des usagères et usagers en aval, dès la conception des nouvelles politiques ou nouveaux services, est un bon moyen de **prendre en compte leurs besoins et attentes en termes de protection de la vie privée**. En effet, lorsque la consultation a lieu en aval (une fois le projet déjà développé), les ajustements possibles sont généralement mineurs et permettent difficilement de corriger les lacunes. Un exemple de projet de codesign serait un atelier réunissant du personnel professionnel et des usagères et usagers des bibliothèques et/ou archives publiques pour imaginer collectivement un algorithme de recommandation qui favorise la découvrabilité des collections tout en introduisant de la sérendipité (le fait de découvrir au hasard).

Piste de réflexion : distinguer les informations personnelles nécessaires à collecter pour remplir la mission aujourd'hui vs celles qui sont porteuses pour les missions futures vs celles qui sont porteuses en termes d'innovation mais accessoires pour la mission

La mise à jour des cadres juridiques au Québec en 2023 établit que la collecte de données personnelles doit se limiter à ce qui est nécessaire pour remplir la mission de l'organisation. **Il s'agit donc de définir la part des données nécessaire et la part des données accessoire.** Dans le cas des bibliothèques et archives, certaines de ces données (ex. l'historique d'emprunt ou de consultation des documents) sont essentielles aux services fournis. Mais il y a-t-il des données annexes qui sont collectées pour des fins accessoires à la mission, c'est-à-dire qu'elles permettent de réaliser des actions secondaires sans lesquelles l'organisation peut tout de même remplir sa mission ? Une difficulté de déterminer la part nécessaire vient du fait qu'il faut non seulement interpréter la mission de l'organisme aujourd'hui, mais aussi se projeter dans l'avenir pour laisser ouvert le champ de l'innovation. Dans tous les cas, les principes de la collecte minimale, du consentement actif (*opt in* plutôt que *opt out*) et de la protection de la vie privée par défaut (*privacy by design*) devraient être considérés.

Piste de réflexion : prendre en compte la sensibilité plus aiguë de certains groupes d'usagers et usagers qui ont des enjeux spécifiques relatifs aux données personnelles

Certains groupes de population ont des enjeux particuliers liés à l'exposition de leur vie privée. L'enquête a par exemple fait ressortir que le fait de s'identifier au groupe LGBTQ+ est associé à une meilleure compréhension des enjeux des données sensibles. Il est donc important de **prendre en compte les besoins spécifiques des populations plus vulnérables** dans le développement de nouvelles politiques et de nouveaux services liés aux données personnelles.

Piste de réflexion : considérer les avantages et les inconvénients d'adopter une date limite pour la durée de conservation des données d'usage

Au Canada, il n'existe pas pour les bibliothèques et archives d'obligation de purger l'historique d'emprunt de documents, comme c'est le cas en Europe avec le Règlement général sur la protection des données (RGPD, 2016). Ce règlement stipule que les données d'emprunt doivent être détruites à la fin du 4^{ème} mois suivant la restitution de l'objet du prêt pour les informations concernant chaque prêt. Et en ce qui concerne les informations personnelles liées à l'identité de chaque personne qui emprunte, elles doivent être supprimées un an après le dernier prêt. Au Canada, par défaut, les données d'emprunt peuvent être conservées indéfiniment pour un abonné actif. L'enquête a fait ressortir que les personnes abonnées interrogées se prononcent majoritairement pour limiter cette durée de conservation, avec pour la plupart d'entre elles une préférence pour une durée de 1 an à 5 ans. **Si la conservation de ces données à long terme peut constituer une ressource potentielle d'innovation (analyse statistique, entraînement d'IA), il reste à considérer les bonnes pratiques établies ailleurs dans le monde ainsi que les attentes des personnes abonnées en termes de limite de durée de conservation.**

9. CONCLUSION ET NOUVELLES PISTES DE RECHERCHE

Cette enquête par questionnaire auprès des usagères et usagers de BAnQ aura permis de dresser un portrait, jusqu'alors inédit, des perceptions et opinions actuelles en matière de gestion des données personnelles dans l'une des plus importantes institutions culturelles du Québec. Ainsi, l'enquête offre un portrait détaillé des niveaux de littératie des données personnelles, une appréciation de l'intérêt des usagères et usagers pour différents types de services qui pourraient découler de l'exploitation de ces données personnelles et enfin une description du spectre d'attitudes des usagères et usagers par rapport aux questions de gouvernance des données personnelles à BAnQ. En croisant ces trois dimensions entre elles, ainsi qu'avec un ensemble de données sociodémographiques, notre analyse permet de dresser un portrait complet de la situation actuelle. Ce portrait mène à des constats sur la population desservie par l'institution, mais aussi à un ensemble de questions complexes qui devront continuer d'alimenter la réflexion du secteur des bibliothèques et archives en cette matière, tout comme la recherche scientifique sur ces enjeux.

À travers le cas de BAnQ, les résultats de cette enquête apportent un éclairage sur le contexte des institutions culturelles publiques, peu étudié à ce jour du point de vue de la protection des données personnelles, ainsi qu'une perspective sur le contexte québécois et canadien, encore sous-documenté par rapport à ceux des États-Unis ou de l'Europe. Il **serait éventuellement intéressant de procéder à des analyses comparatives entre différents contextes nationaux** afin d'en faire émerger les ressemblances et les différences. De manière similaire, il serait aussi souhaitable de s'interroger sur la montée en généralité que peut offrir ce cas, notamment par rapport à d'autres types d'institutions culturelles. Dans quelle mesure, les constats faits dans le cadre d'une institution du secteur des bibliothèques et archives **sont-ils comparables aux institutions culturelles dans d'autres secteurs** ? Y a-t-il des dynamiques qui marquent l'ensemble des secteurs culturels ? Qu'est-ce qui, au contraire, caractérise en propre le secteur des bibliothèques et archives par rapport à d'autres secteurs culturels ?

La contribution de l'enquête se situe sur deux plans : un plan de coconstruction des connaissances avec le partenaire de la recherche, BAnQ, et un plan scientifique. Sur le plan de **la coconstruction des savoirs**, l'enquête aura permis à l'institution d'approfondir ses connaissances sur ses usagères et usagers, celle-ci s'inscrivant dans la foulée de sondages et de consultations précédentes qui avaient permis d'établir de premiers constats. La présente enquête offre ainsi des pistes d'action, qui ont trait notamment au développement de la littératie des données personnelles et à la gouvernance participative en cette matière, mais aussi des pistes de réflexion qui devront alimenter le développement d'une expertise institutionnelle sur la gouvernance des données personnelles. Les apports des équipes de recherche pourront encore être utiles à ce chapitre, que ce soit par exemple pour la réalisation d'un atelier de codesign de politiques et de services ayant trait aux données personnelles, ou pour poursuivre la recherche sur certaines questions soulevées par l'enquête.

Sur le **plan scientifique**, l'enquête ouvre un certain nombre de pistes de recherche qui devraient être poursuivies. On note plus particulièrement trois principaux chantiers qui pourront être approfondis à partir de l'enquête actuelle. L'un des premiers éléments que l'enquête a permis de mettre au jour est *la relation à trois variables qui unit le niveau de littératie des données personnelles, l'intérêt plus ou moins grand pour de nouveaux services culturels envisageables à partir des données et les attitudes contrastées par rapport à la protection des données personnelles*. Une exploration plus en profondeur de cette relation complexe permettra de mieux comprendre comment ces trois variables interagissent entre elles.

Un second chantier concerne **la gouvernance des données personnelles dans les institutions culturelles**. On note entre autres dans l'enquête qu'un niveau élevé de littératie des données personnelles correspond généralement à un faible intérêt pour les nouveaux services et les recommandations culturelles en fonction de l'historique de consommation. Or, les deux éléments sont aussi liés à un niveau d'éducation élevé, caractéristique généralement associée aux populations qui fréquentent les institutions culturelles. Ainsi, il y a là une avenue à explorer qui pourrait ouvrir des possibilités quant à la démocratisation des arts et de la culture, mais qui pose néanmoins des questions importantes par rapport à la diversité des contenus et la libre circulation des œuvres.

Enfin, on peut envisager un dernier **chantier de recherche qui concerne le personnel des institutions culturelles**. En effet, dans le cadre de cette enquête nous avons recueilli un ensemble significatif de données sur les membres du personnel de BAnQ qui, pour le moment, demeure inexploité. Ainsi, il serait fort intéressant de mener sur ce deuxième jeu de données des analyses similaires à celles présentées ici, afin de voir dans quelle mesure les réponses des membres du personnel s'écartent ou se rapprochent de celles des usagères et usagers. Une mise en comparaison de ces deux groupes permettrait très certainement d'affiner notre compréhension des écarts qui peuvent exister entre eux. Un tel exercice semble prometteur du point de vue de la recherche, mais aussi du point de vue des politiques institutionnelles, puisqu'il mènerait naturellement à la mise en place de meilleures stratégies pour le développement d'une compréhension partagée des enjeux de gouvernance des données personnelles dans les institutions culturelles.

ANNEXES

1. Questionnaire

SONDAGE SUR LA GESTION ÉTHIQUE DE VOS DONNÉES PAR BAnQ

Le temps estimé pour remplir ce sondage est de 15 minutes. Ce sondage ne s'adresse qu'aux personnes de 18 ans ou plus qui sont abonnées à Bibliothèque et Archives nationales du Québec (BAnQ).

Cette étude est réalisée en collaboration avec l'Institut national de la recherche scientifique (INRS) et l'Université de Montréal. Une liseuse fera l'objet d'un tirage au sort parmi les personnes participantes (optionnel).

Objectifs :

- Le sondage permettra à BAnQ de mieux connaître le point de vue de ses abonnés relativement à la gestion éthique de leurs données. Plus précisément, les questions portent sur vos connaissances et votre opinion quant à l'utilisation de vos données par BAnQ, ainsi que sur votre profil.
- En participant, vous contribuez à la formulation de recommandations touchant la gestion des données pour le milieu des bibliothèques et des archives.
- Dans un but éducatif, les bonnes réponses à certaines questions vous seront données en s'affichant sur l'écran suivant.

Informations sur ma participation :

- Participer au sondage ne vous expose pas à des risques différents que ceux auxquels vous vous exposez dans votre vie de tous les jours. Le seul inconvénient est le temps passé à répondre au sondage.
- Vos réponses seront enregistrées et gardées confidentielles par l'équipe de BAnQ. Si vous vous inscrivez au tirage à la fin du sondage, ou pour une éventuelle étape de groupe de discussion (optionnel), votre adresse courriel sera collectée. Elle sera gardée de manière sécurisée et confidentielle par BAnQ.
- Les données récoltées seront la propriété de BAnQ. Elles seront anonymisées par BAnQ à l'aide d'un code, cryptées à l'aide d'un logiciel et protégées avec un mot de passe dans un ordinateur de BAnQ. Seules les données anonymisées seront transmises aux chercheurs de l'INRS pour être analysées.
- Les informations qui permettent de vous identifier seront détruites deux ans après la fin de l'étude. Seules les données anonymisées seront archivées, à l'INRS et à BAnQ, dans un ordinateur à accès sécurisé. Les données anonymes seront ainsi conservées indéfiniment pour servir à des recherches futures de même nature, par l'équipe de l'INRS. BAnQ se réserve le droit de partager les données anonymisées avec d'autres équipes de recherche dans le respect des règles d'éthique.
- Votre nom ne sera pas mentionné dans les publications qui découleront de l'étude.

Avertissements techniques :

- Une fois le sondage commencé, le bouton de retour en arrière de votre navigateur Web ne fonctionnera pas. L'utiliser pourrait vous faire quitter le sondage.
- Vous serez tout à fait libre de répondre ou non à une question posée, et vous pourrez quitter à tout moment le sondage. Dans ce cas, vos données ne seront pas utilisées.
- Si vous avez soumis votre adresse courriel dans le formulaire et que vous souhaitez retirer vos réponses, vous pouvez contacter la personne responsable à l'INRS (ci-dessous).

Cette recherche a obtenu un certificat en éthique de recherche. Pour toute question relative à votre consentement ou à votre participation, adressez-vous à :

- Nathalie Vachon, professionnelle de recherche à l'INRS : nathalie.vachon@inrs.ca, 514 499-4079.
- Pour contacter une personne de l'INRS extérieure au projet de recherche (le président ou la coordonnatrice du Comité d'éthique en recherche), vous pouvez appeler au 418-650-7434 ou écrire à cer@inrs.ca.

Je reconnais avoir lu et compris les informations ci-dessus. J'accepte de participer à la recherche selon les modalités décrites ci-dessus.

- J'accepte
- Je n'accepte pas (fermez alors la session)

A- PRÉAMBULE

Pour commencer, nous aimerions vous connaître un peu.

(Qi) Dans quel groupe d'âge vous situez-vous?

- Moins de 18 ans (**terminer la session si choisi**)
- 18-24 ans
- 25-34 ans
- 35-44 ans
- 45-54 ans
- 55-64 ans
- 65-74 ans
- 75-84 ans
- 85 ans et plus

(Qii) Êtes-vous abonné à BAnQ?

- Oui et j'utilise les services sur place à la Grande Bibliothèque ou les services en ligne.
- Oui et j'utilise seulement les services en ligne.
- Non, je ne suis pas abonné.

B- CONNAÎTRE VOS DONNÉES

Nous sommes prêts à commencer.

Saviez-vous que lorsque vous faites l'emprunt de documents auprès de BAnQ, l'institution enregistre vos emprunts, conserve votre historique d'emprunt et peut l'utiliser? En effet, l'information concernant vos emprunts est cumulée au fil du temps sous forme de données numériques.

Maintenant, explorons plus en détail le type d'informations que vous fournissez à BAnQ.

(Q1) À votre avis, quelles informations parmi les suivantes constituent des données personnelles? Plusieurs choix sont possibles.

- a) Mon nom
- b) Le nom de la PDG de BAnQ
- c) Mon adresse postale
- d) L'historique de mes emprunts à la bibliothèque
- e) Le mot de passe du réseau Internet sans fil du café où je me branche
- f) Je ne sais pas.

Message affiché à l'issue de la réponse :

Réponse : Les données personnelles sont des informations qui fournissent des renseignements sur une personne en permettant de l'identifier.

Votre nom, votre adresse postale, ainsi que l'historique de vos emprunts à BAnQ constituent des exemples de données personnelles puisqu'elles fournissent des renseignements sur vous et permettent de vous identifier. Poursuivons.

(Q2) Selon vous, lesquelles de ces informations vous concernant sont actuellement collectées par BAnQ? Plusieurs choix sont possibles.

- a) La municipalité où j'habite
- b) Le contenu des messages que j'envoie au service d'aide à distance de BAnQ
- c) Le contenu des messages que j'envoie à mes amis en étant connecté au réseau Internet sans fil de la Grande Bibliothèque
- d) La liste des documents favoris enregistrés dans mon dossier d'abonné de BAnQ
- e) Je ne sais pas.

Message affiché à l'issue de la réponse :

Réponse : BAnQ collecte actuellement les informations de tous ses abonnés concernant leur municipalité de résidence, le contenu des messages qu'ils envoient au service d'aide à distance de BAnQ, ainsi que la liste des documents favoris enregistrés à leur compte. Mais elle ne collecte pas le contenu des messages envoyés à vos amis quand vous êtes connecté au réseau Internet sans fil de la Grande Bibliothèque, qui reste privé.

(Q3) *Vrai ou faux? À partir de votre historique d'emprunt de documents, on peut déduire des renseignements sur votre état de santé, votre orientation politique ou votre orientation sexuelle.*

- a) Vrai
- b) Faux
- c) Je ne sais pas.

Message affiché à l'issue de la réponse :

Réponse : Vrai. *À partir de votre historique d'emprunt de documents, il serait théoriquement possible de déduire certains renseignements sur votre état de santé (ex. : emprunt de livres sur une maladie), sur vos orientations politiques (ex. : emprunt régulier d'une revue politiquement engagée) ou sur votre orientation sexuelle (ex. : emprunt régulier de magazines LGBTQ+).*

Bibliothèque et Archives nationales du Québec prend toutes les précautions nécessaires pour assurer la confidentialité des renseignements personnels. Pour plus d'information, voir la fiche pratique à la fin du sondage.

(Q4) *Vrai ou faux? Quand BAnQ rend anonymes vos données personnelles, il n'existe aucun moyen de vous relier à nouveau à l'information en question.*

- a) Vrai
- b) Faux
- c) Je ne sais pas.

Message affiché à l'issue de la réponse :

Réponse : Faux. *Même si cela est très difficile, il peut y avoir des moyens indirects de relier des informations à une personne même après l'anonymisation des données.*

Bibliothèque et Archives nationales du Québec prend toutes les précautions nécessaires pour assurer la confidentialité des renseignements personnels. Pour plus d'information, voir la fiche pratique à la fin du sondage.

(Q5) *BAnQ utilise divers services Web (ex. : Google Analytics, Facebook) qui collectent également des données à votre sujet. Est-il vrai ou faux que ces fournisseurs de services Web sont soumis aux mêmes règles que BAnQ concernant la protection de vos données?*

- a) Vrai
- b) Faux
- c) Je ne sais pas.

Message affiché à l'issue de la réponse :

Réponse : Faux. *Les fournisseurs de services (ex. : Google Analytics, Facebook) ont leurs propres règles de protection des données. BAnQ ne peut pas leur imposer ses règles.*

(Q6) À votre avis, quelles initiatives pourraient le mieux contribuer à encadrer l'utilisation de vos données personnelles par BAnQ? Plusieurs choix sont possibles.

- a) M'informer précisément sur les données collectées et leur utilisation
- b) Me donner la possibilité de refuser que mes données soient collectées
- c) Publier un guide d'utilisation du catalogue en ligne de BAnQ
- d) Faire appel à un organisme de réglementation indépendant pour surveiller ce qui est fait avec mes données
- e) Je ne sais pas.

Message affiché à l'issue de la réponse :

Réponse : La transparence à l'égard des données collectées et de leurs usages, le libre choix d'accepter ou de refuser la collecte de ces données ainsi que le recours à un organisme de surveillance indépendant pourraient contribuer à un encadrement plus éthique des données personnelles recueillies par BAnQ.

(Q7) Selon vous, les données ouvertes sont :

- a) Les messages courriel déjà consultés
- b) Des informations rendues publiques sur le Web
- c) Des données numériques dont l'accès et l'utilisation sont laissés libres aux usagers
- d) Des données personnelles ayant été volées par des pirates informatiques
- e) Je ne sais pas.

Message affiché à l'issue de la réponse :

Réponse : Les données ouvertes sont des données numériques dont l'accès et l'utilisation sont laissés libres aux usagers. Ces données sont par exemple produites par des municipalités ou des établissements publics.

BAnQ rend disponibles beaucoup de données ouvertes via la plateforme « Données Québec ». Pour en savoir plus, voir la fiche pratique à la fin du sondage.

C- VOTRE POINT DE VUE

À présent, nous aimerions avoir votre point de vue sur les bonnes pratiques à mettre en place pour gérer et utiliser vos données personnelles. En effet, ces données peuvent être utilisées pour mettre sur pied de nouveaux services de BAnQ.

(Q8) Veuillez indiquer dans quelle mesure vous êtes d'accord ou non avec chacune des affirmations suivantes.

BAnQ devrait pouvoir utiliser mon historique d'emprunt pour :

- a) Enrichir ses collections avec de nouveaux documents en fonction de la demande du public
- b) Me recommander des documents
- c) Me recommander des activités culturelles
- d) Personnaliser mon expérience numérique sur le site Web de BAnQ

- Fortement d'accord
- Plutôt d'accord
- Ni d'accord ni en désaccord
- Plutôt en désaccord
- Fortement en désaccord
- Je ne sais pas.

(Q9) Veuillez indiquer dans quelle mesure vous êtes d'accord ou non avec chacune des affirmations suivantes.

BAnQ devrait pouvoir :

- a) Collecter seulement les données personnelles utiles aux services actuels
- b) Recouper différentes données à mon sujet pour personnaliser ses services (ex. : historique d'emprunt selon l'âge)

- Fortement d'accord
- Plutôt d'accord
- Ni d'accord ni en désaccord
- Plutôt en désaccord
- Fortement en désaccord
- Je ne sais pas.

(Q10) BAnQ désire conserver mes données personnelles anonymisées en vue de mieux connaître les pratiques de ses abonnés et d'améliorer son offre de services. Je souhaiterais que mes données :

- a) Ne soient pas conservées du tout
- b) Soient conservées moins d'un an
- b) Soient conservées entre 1 an et 5 ans
- c) Soient conservées indéfiniment
- d) Je ne sais pas.

(Q11) Veuillez indiquer dans quelle mesure vous êtes d'accord ou non avec chacune des affirmations suivantes.

BAnQ devrait pouvoir partager mes données personnelles, lorsqu'elles sont anonymisées, avec :

- a) Un ministère du gouvernement québécois, pour concevoir des politiques publiques
- b) Une maison d'édition, pour lui permettre de mieux cibler ses clients
- c) Un OBNL, pour faire de la sensibilisation à la lecture
- d) Des équipes de recherche universitaires, afin d'étudier les pratiques de lecture de la population québécoise

- Fortement d'accord
- Plutôt d'accord
- Ni d'accord ni en désaccord

- Plutôt en désaccord
- Fortement en désaccord
- Je ne sais pas.

Message affiché à l'issue de la réponse :

Nous y sommes presque! Pour mieux comprendre vos réponses, nous aimerions en savoir plus à votre sujet. Cela nous permettra de faire des recommandations plus précises à BAnQ sur l'utilisation éthique de vos données personnelles.

D - VOTRE PROFIL

(Q12) *Quel est votre genre?*

Par genre, on entend le genre actuel, qui peut différer du sexe assigné à la naissance ou de celui inscrit dans les documents légaux.

- Masculin
- Féminin
- Autre
- Je préfère ne pas répondre.

(Q13) *Dans quelle catégorie occupez-vous votre emploi actuel (ou avez-vous occupé votre dernier emploi)?*

- Membre du personnel de BAnQ
- Membres des corps législatifs et cadres supérieurs
- Affaires, finance et administration
- Sciences naturelles et appliquées et domaines apparentés
- Santé
- Enseignement, droit et services sociaux, communautaires et gouvernementaux
- Arts, culture, sports et loisirs
- Vente et services
- Métiers, transport, machinerie et domaines apparentés
- Ressources naturelles, agriculture et production connexe
- Fabrication et services d'utilité publique
- Ne s'applique pas
- Je préfère ne pas répondre.

(Q14) *Quel est le plus haut certificat, diplôme ou grade que vous avez obtenu?*

- Aucun diplôme, certificat ou grade
- Diplôme d'études secondaires ou attestation d'équivalence
- Certificat ou diplôme d'apprenti, d'une école de métiers, d'un collège, d'un cégep ou d'un autre établissement non universitaire
- Diplôme ou grade universitaire inférieur au baccalauréat (certificat, mineure, majeure)
- Baccalauréat
- Diplôme ou grade universitaire supérieur au baccalauréat

- Je préfère ne pas répondre.

(Q15) *Au cours de la dernière année, quelle était votre occupation principale?*

- Travailleur rémunéré (y compris en congé de maladie, parental, etc.)
- Travailleur autonome
- Aux études
- Sans emploi ou à la recherche d'emploi
- Au foyer (s'occuper des enfants, mais pas retraité ni en congé)
- Retraité
- Je préfère ne pas répondre.

(Q16) *Dans quelle catégorie de revenu avant impôt votre ménage se situe-t-il?*

- Moins de 30 000 \$
- 30 000 \$ - 49 999 \$
- 50 000 \$ - 69 999 \$
- 70 000 \$ - 89 999 \$
- 90 000 \$ - 124 999 \$
- 125 000 \$ - 199 999 \$
- 200 000 \$ ou plus
- Je préfère ne pas répondre.

(Q17) *Quel est votre statut de citoyenneté au Canada?*

- Citoyen canadien
- Résident permanent
- Permis de séjour temporaire (travail, études)
- Réfugié
- Je préfère ne pas répondre.

(Q18) *Êtes-vous Autochtone, c'est-à-dire membre d'une Première Nation, Métis ou Inuk (Inuit)?*

- Oui
- Non
- Je préfère ne pas répondre.

(Q19) *Êtes-vous membre d'un groupe minoritaire visible (c'est-à-dire ni blanc ni autochtone)?*

- Oui
- Non
- Je préfère ne pas répondre.

(Q20) Vous identifiez-vous à un groupe LGBTQ2+?

Le terme LGBTQ2+ inclut les personnes lesbiennes, gaies, queer, pansexuelles, asexuelles, transgenres, non binaires et autres minorités sexuelles.

- Oui
- Non
- Je préfère ne pas répondre.

Nous avons terminé !

« Si vous voulez vous inscrire au tirage au sort d'une liseuse électronique, vous porter volontaire pour d'éventuels groupes de discussion en lien avec ce sondage ou nous transmettre un commentaire, les questions suivantes vous permettent de le faire. Sinon, cliquez sur terminer. »

(Q21) Je désire (plusieurs choix sont possibles) :

- M'inscrire au tirage au sort d'une liseuse
- M'inscrire pour un groupe de discussion
- Laisser un commentaire
- Terminer ici

(Q22) Voici mon adresse courriel :

(Case ouverte)

Je désire :

- Laisser un commentaire
- Terminer ici

(Q23) Avez-vous des suggestions ou commentaires à faire au sujet de ce sondage?

(Case ouverte)

Message de fin de sondage

Bibliothèque et Archives nationales du Québec vous remercie pour votre temps et votre participation.

Pour toute question relative aux pratiques et politiques de BAnQ, vous pouvez vous adresser à : (COURRIEL, responsable BAnQ)

Pour en savoir plus sur les données personnelles, cliquez ici pour accéder à une fiche pratique :

https://docs.google.com/document/d/1moQvrfZH4IAxyhk5NMMvO4n_xvyglnWus8j6Khl8Qjw/edit?usp=sharing
[Lien vers le document d'accompagnement hébergé en ligne]

2. Test de redressement de l'échantillon

Bilan des variables exploitables

L'échantillon présente une distribution sociodémographique satisfaisante sur les variables suivantes :

- *Type d'abonnement* ;
- *Localisation géographique*.

Des biais de l'échantillon ont été identifiés sur les variables suivantes :

- *Genre* : surreprésentation des hommes et une sous-représentation des femmes ;
- *Âge* : surreprésentation des groupes plus âgés et sous-représentation des jeunes.

Des lacunes dans les études de référence rendent incertaine l'évaluation des variables suivantes :

- *Occupation* : hypothèse forte d'un biais de l'échantillon vers la surreprésentation des personnes retraitées et la sous-représentation des personnes aux études ;
- *Niveau de diplomation* : hypothèse forte que la surreprésentation des personnes diplômées universitaires reflète la population des personnes abonnées à BAnQ ;
- *Secteur d'emploi* : hypothèse que la surreprésentation des secteurs « Enseignement, droit et services sociaux, communautaires et gouvernementaux » et « Arts, culture, sports et loisirs » reflète la population personnes abonnées aux services de BAnQ ;
- *Minorité visible* : hypothèse que la sous-représentation des minorités visibles reflète la population des personnes abonnées à BAnQ ;
- *Groupe LBGTQ+* : hypothèse que la légère surreprésentation de ce groupe dans l'échantillon reste satisfaisante pour l'analyse.

Tableau 42 : Composition de l'échantillon comparé à la population de référence

Distribution satisfaisante	Biais de l'échantillon	
	Surreprésentation	Sous-représentation
[Confirmé par les études de référence]		
<i>Type d'abonnement</i> (moitié-moitié en ligne et tous services)	<i>Genre : hommes</i> (fort ; +12,5%)	<i>Genre : femmes</i> (modéré ; - 7,8%)
<i>Localisation géographique</i> (40% Montréal)	<i>Âge : 65+ ans</i> (très fort ; + 37,6%)	<i>Âge : 18-44 ans</i> (fort ; - 18,5%)
[Hypothèses]		
<i>Niveau de diplomation</i> (forte surreprésentation des diplômé.e.s universitaires)	<i>Occupation : retraité.e.s</i> (très fort ; 47,8% total)	<i>Occupation : aux études</i> (très fort ; 5,5% total)
<i>Secteur d'emploi</i> (surreprésentation des métiers en enseignement, art et culture)		
<i>Minorité visible</i> (sous-représentation modérée)		
<i>Groupe LBGTQ+</i> (faible sur-représentation)		

Contraintes lors de la diffusion du sondage

Initialement, la méthodologie d'administration du sondage prévoyait de vérifier périodiquement lors de la diffusion du sondage la composition de l'échantillon, afin de déceler de possibles biais dans la distribution des personnes sondées (variable d'âge, de genre, de région de résidence et de type d'abonnement selon les données d'enquête interne fournies par BAnQ). L'objectif était de procéder à une post-stratification de l'échantillon en envoyant des relances ciblées à certains groupes d'usagers et usagers via l'infolettre de BAnQ. Toutefois, il n'a pas été possible (de par les contraintes internes à BAnQ) d'utiliser l'infolettre pour cibler des groupes de personnes abonnées en fonction de leurs caractéristiques sociodémographiques. De plus, une cyberattaque majeure a affecté l'INRS durant la période de diffusion du sondage, rendant impossible l'accès au poste informatique de la membre de l'équipe affectée à la tâche de suivi de la collecte des réponses. Pour ces raisons, **il n'a pas été possible de faire un redressement au moment de la collecte des données.**

Données de référence sur la population cible : lacunes

Le redressement a posteriori de l'échantillon implique une pondération des résultats en fonction des caractéristiques sociodémographiques de la population cible. **Certaines lacunes dans les statistiques existantes sur la population des personnes abonnées compliquent cette opération** (Tableau 9). Ces lacunes sont différentes en fonction du type d'abonnement (tous services et à distance).

Premièrement, la variable de la localisation géographique n'a pas été collectée de la même manière dans l'étude de référence (code RTA) et dans le sondage (localisation de l'adresse IP). Il est donc possible que des écarts existent entre ces deux mesures. Cette variable n'a donc pas été considérée pour le test de redressement.

Deuxièmement, chez les personnes abonnées à distance, 27% n'ont pas d'information sur l'âge (n=111 571), dont 14% n'ont pas non plus d'information sur le sexe (n=57 989). Chez les personnes abonnées tous services seulement 2% n'ont pas d'indication pour l'âge (n=3 635), un nombre équivalent (n=3409) n'ont pas d'information sur le sexe, dont 0,7% (n=1 184) n'ont pas d'information sur aucune des deux caractéristiques.

Troisièmement, **les deux types d'abonnements ont des profils différents quant à la sous-déclaration de l'âge ou du sexe.** Chez les personnes abonnées tous services, ce sont plutôt les femmes qui ne fournissent pas leur âge, alors que c'est l'inverse chez les abonnés à distance, ce sont plutôt les hommes qui ne répondent pas à la question sur l'âge. Pour la non-déclaration du sexe, chez les personnes abonnées tous services, ce sont plutôt les jeunes de moins de 35 ans qui ne le fournissent pas. Alors que chez les personnes abonnées à distance ce sont plutôt les 65 ans et plus qui ne dévoilent pas leur âge.

Ainsi **une grande partie de l'information manque, surtout pour les personnes abonnées à distance.** Redresser l'échantillon en fonction d'une information lacunaire pourrait introduire de

nouveaux biais. De plus, comme les caractéristiques sociodémographiques des personnes abonnées tous services sont différentes de celles des personnes abonnées à distance, il n'est pas possible de se servir du sous-groupe des personnes abonnées tous services (30% des personnes abonnées) pour redresser l'échantillon de toutes les personnes abonnées (qui se partagent à 50-50 pour le type d'abonnement).

Si on se fie aux distributions des personnes abonnées pour qui on a les informations sur l'âge et le sexe, on constate que **pour l'abonnement tous services, les femmes sont surreprésentées chez les moins de 45 ans, et chez les abonnés à distance les femmes sont largement surreprésentées chez les moins de 75 ans** alors que **les hommes sont surreprésentés chez les 75 à 84 ans et particulièrement chez les 85 ans et plus**. Mais on peut s'interroger sur l'exactitude de ce portrait s'il n'inclut pas les 112 000 personnes abonnées qui n'ont pas fourni leurs informations. Supposer que les non-répondantes et non-répondants se répartissent de la même façon que les répondantes et répondants pour l'âge et le sexe est une hypothèse forte.

Tableau 43 : Données de référence sur la population des abonnés (lacunes)

	CAT_AGE	Tous services				À distance			
		SEXE				SEXE			
		F	M	n.a	Total	F	M	n.a	Total
Fréquence	18 à 24 ans	13 816	9 402	474	23 692	23 574	14 989	45	38 608
Pourcentage		7,79	5,30	0,27	13,35	5,68	3,61	0,01	9,30
Pctage en ligne		58,32	39,68	2,00		61,06	38,82	0,12	
Pctage en col.		14,68	11,76	13,90		10,81	10,81	0,08	
	25 à 34 ans	23 787	17 436	723	41 946	31 669	19 670	20	51 359
		13,40	9,83	0,41	23,64	7,63	4,74	0,00	12,38
		56,71	41,57	1,72		61,66	38,30	0,04	
		25,28	21,81	21,21		14,53	14,19	0,03	
	35 à 44 ans	18 241	15 894	412	34 547	28 717	16 082	39	44 838
		10,28	8,96	0,23	19,47	6,92	3,88	0,01	10,81
		52,80	46,01	1,19		64,05	35,87	0,09	
		19,38	19,88	12,09		13,17	11,60	0,07	
	45 à 54 ans	12 172	12 304	247	24 723	27 842	15 890	52	43 784
		6,86	6,93	0,14	13,93	6,71	3,83	0,01	10,55
		49,23	49,77	1,00		63,59	36,29	0,12	
		12,93	15,39	7,25		12,77	11,46	0,09	
	55 à 64 ans	9 293	9 686	175	19 154	30 894	16 906	56	47 856
		5,24	5,46	0,10	10,79	7,44	4,07	0,01	11,53
		48,52	50,57	0,91		64,56	35,33	0,12	
		9,87	12,12	5,13		14,17	12,20	0,10	
	65 à 74 ans	9 105	9 037	122	18 264	30 156	19 855	70	50 081
		5,13	5,09	0,07	10,29	7,27	4,78	0,02	12,07
		49,85	49,48	0,67		60,21	39,65	0,14	
		9,67	11,31	3,58		13,83	14,32	0,12	
	75 à 85 ans	4 078	4 039	52	8 169	11 274	10 700	39	22 013
		2,30	2,28	0,03	4,60	2,72	2,58	0,01	5,30
		49,92	49,44	0,64		51,22	48,61	0,18	
		4,33	5,05	1,53		5,17	7,72	0,07	
	85 ans et plus	2 240	1 066	20	3 326	2 200	2 645	19	4 864
		1,26	0,60	0,01	1,87	0,53	0,64	0,00	1,17
		67,35	32,05	0,60		45,23	54,38	0,39	
		2,38	1,33	0,59		1,01	1,91	0,03	
	n.a	1 377	1 074	1 184	3 635	31 693	21 889	57 989	111 571
		0,78	0,61	0,67	2,05	7,64	5,27	13,97	26,89
		37,88	29,55	32,57		28,41	19,62	51,97	
		1,46	1,34	34,73		14,54	15,79	99,42	
	Total	94 109	79 938	3 409	177 456	218 019	138 626	58 329	414 974
		53,03	45,05	1,92	100,00	52,54	33,41	14,06	100,00

3. Littératie des données personnelles

Tableau 44 - Identifier les données personnelles : score de la sous-dimension

A	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
0	46	3,2	46	3,2
1	152	10,7	198	13,9
2	399	28,0	597	41,9
3	827	58,1	1424	100,0

Fréquence manquante = 6

Tableau 45 - Connaître la nature et l'étendue des données collectées : score de la sous-dimension

B	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
0	117	8,2	117	8,2
1	234	16,5	351	24,7
2	427	30,1	778	54,8
3	643	45,3	1421	100,0

Fréquence manquante = 9

Tableau 46 - Comprendre les enjeux : score de la sous-dimension

C	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
0	273	19,1	273	19,1
1	479	33,5	752	52,6
2	477	33,4	1229	85,9
3	201	14,1	1430	100,0

Tableau 47 - Réfléchir à la gouvernance : score de la sous-dimension

D	Fréquence	Pourcentage	Fréquence cumulée	Pctage cumulé
0	82	5,8	82	5,8
1	316	22,3	398	28,0
2	524	36,9	922	64,9
3	498	35,1	1420	100,0

Fréquence manquante = 10

4. Intérêt pour le développement de nouveaux services

Tableau 48 : Fréquences et pourcentage des réponses pour le score total de d'intérêt pour les nouveaux services

Score d'intérêt	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
4 (très bas)	33	2,8	33	2,8
5	2	0,2	35	2,9
6	9	0,8	44	3,7
7	25	2,1	69	5,8
8	39	3,3	108	9,0
9	24	2,0	132	11,0
10	49	4,1	181	15,1
11	34	2,8	215	18,0
12	75	6,3	290	24,2
13	75	6,3	365	30,5
14	136	11,4	501	41,9
15	139	11,6	640	53,5
16	204	17,0	844	70,5
17	110	9,2	954	79,7
18	69	5,8	1023	85,5
19	42	3,5	1065	89,0
20 (très élevé)	132	11,0	1197	100,0

Fréquence manquante = 233

5. Perception relative à la gestion des données personnelles

Tableau 49 : Fréquences et pourcentage des réponses pour le score total de perception relative à la gestion des données personnelles

Score (sans I)	Fréquence	Pourcentage	Fréquence cumulée	Pourcentage cumulé
5	47	4,3	47	4,3
6	23	2,1	70	6,5
7	22	2,0	92	8,5
8	38	3,5	130	12,0
9	37	3,4	167	15,4
10	35	3,2	202	18,7
11	39	3,6	241	22,3
12	62	5,7	303	28,0
13	54	5,0	357	33,0
14	44	4,1	401	37,1
15	66	6,1	467	43,2
16	78	7,2	545	50,4
17	84	7,8	629	58,1
18	76	7,0	705	65,2
19	69	6,4	774	71,5
20	63	5,8	837	77,4
21	58	5,4	895	82,7
22	58	5,4	953	88,1
23	44	4,1	997	92,1
24	33	3,1	1030	95,2
25	11	1,0	1041	96,2
26	14	1,3	1055	97,5
27	16	1,5	1071	99,0
28	11	1,0	1082	100,0

Fréquence manquante = 348

BIBLIOGRAPHIE

ABPQ - Association des bibliothèques publiques du Québec. (2021). *Stratégie numérique Résultats de la consultation auprès des usagers*. Montréal : ABPQ.

Affonso, Elaine Parra et Ricardo C. Sant'Ana. (2018). « Privacy awareness issues in user data collection by digital libraries ». *IFLA Journal*, 44(3) : 170-182.

Altman, Irwin. (1975). *The Environment and Social Behavior: Privacy, Personal Space, Territory, Crowding*. Monterey : Brooks/Cole, 256 p.

Altman, Irwin. (1977). « Privacy Regulation: Culturally Universal or Culturally Specific? ». *Journal of Social Issues*, 33 (3) : 66-84.

Apprich, Clemens, Wendy Hui K. Chun, Florian Cramer et Hito Steyerl. (2018). *Pattern Discrimination*. Lüneberg, Germany : Meson Press, 123 p.

Autorité canadienne pour les enregistrements Internet (ACEI). (2020). *Attitudes des Canadiens sur les enjeux relatifs à l'Internet*. En ligne. <https://www.cira.ca/fr/resources/letat-de-linternet/rapport/attitudes-des-canadiens-%20sur-les-enjeux-relatifs-a-linternet> (consulté le 21 juin 2023).

Azzaria, Georges et Hélène Zimmermann. (2019). « Pratiques et représentations des internautes québécois en matière de consentement, de droit à la vie privée et de droit d'auteur ». *Lex-electronica* 24, hors-série, 48 p.

Badouard, Romain, Clément Mabi et Guillaume Sire. (2016). « Inciter, contraindre, encadrer. Trois logiques de gouvernamentalité numérique ». *French Journal for Media Research*, (6) : 2-17.

Barnes, Susan B. (2006). « A privacy paradox: Social networking in the United States ». *First Monday*, 11(9).

Barth, Susanne, Menno D. de Jong, Marianne Junger, Pieter H. Hartel et Janina C. Roppelt. (2019). « Putting the privacy paradox to the test: Online privacy and security behaviors among users with technical knowledge, privacy awareness, and financial resources ». *Telematics and Informatics*, 41 : 55-69.

Bartsch, Miriam et Tobias Dienlin. (2016). « Control your Facebook: An analysis of online privacy literacy ». *Computers in Human Behavior*, 56 : 147-154.

Bashir, Masooda, Carol Hayes, April D. Lambert et Jay P. Kesan. (2015). « Online privacy and informed consent: The dilemma of information asymmetry ». *Proceedings of the Association for Information Science and Technology*, 52(1) : 1-10.

Beauchemin, William-J., Noémie Maignien et Nadia Duguay. (2020). *Portraits d'institutions culturelles montréalaises. Quels modes d'action pour l'accessibilité, l'inclusion et l'équité ?* Sainte-Foy : PUL, 410 p.

Beaudouin, Valérie, Isabelle Garron et Nicolas Rollet. (2016). « 'Je pars d'un sujet, je rebondis sur un autre' : pratiques et usages des publics de Gallica ». Rapport de recherche. Paris : Télécom Paris Tech / Bibliothèque nationale de France / Labex Obvil, 29 p.

Beck, Estelle. (2018) « Les données personnelles des usagers en bibliothèque : de beaux défis de gestion en perspective » *Documentation et bibliothèques*, 64(1) : 16-27.

Beer, David. (2015). « Productive measures: Culture and measurement in the context of everyday neoliberalism ». *Big Data & Society*, 2(1) : 1-12.

Benghozi, Pierre-Jean et Françoise Benhamou. (2008). « Longue traîne : levier numérique de la diversité culturelle ? ». *Culture prospective*, 1(1) : 1-11.

Bennett, Colin J. et Christopher Parsons. (2013). « Privacy and surveillance: the multidisciplinary literature on the capture, use and disclosure of personal information in cyberspace ». In *The Oxford Handbook of Internet Studies*, sous la dir. de William H. Dutton. Oxford : Oxford University Press, 586-508.

Bergström, Annika. (2015). « Online privacy concerns: A broad approach to understanding the concerns of different groups for different uses ». *Computers in Human Behavior*, 53 : 419-426.

Blank, Grant, Gillian Bolsover et Elizabeth Dubois. (2014). *A New Privacy Paradox: Young people and privacy on Social Network Sites*. Article présenté à la conférence annuelle de l'American Sociological Association.

Boisselet, Romain. (2018). « (Auto)-censure et surveillance de masse, quels impacts pour les bibliothèques ? ». *Bulletin des bibliothèques de France (BBF)*, 15. En ligne. https://bbf.enssib.fr/tour-d-horizon/auto-censure-et-surveillance-de-masse-quels-impacts-pour-les-bibliotheques_68537 (consulté le 21 juin 2023).

Boisson, Alexis. (2021). « OPT IN : du "slogan" au design du site Internet. L'œuvre pédagogique de la CNIL dans ses lignes directrices sur les cookies et autres traceurs ». *Revue Lamy Droit civil*, 2021(188) : 37-41.

Bonikowska, Aneta, Claudia Sanmartin et Marc Frenette. (2019). *Littératie des données et comment la mesurer dans la fonction publique*. Ottawa : Statistiques Canada, 25 p.

Bowers, Stacey L. (2008). « Self-Service Holds: A Violation of Library Patrons' Privacy ». *Libraries*, 47(4) : 54-57.

boyd, dana et Eszter Hargittai. (2010). « Facebook privacy settings: Who cares? » *First Monday*, 15(8).

Caillet, Bruno. (2016-2017) « Quelles réponses au procès d'élitisme fait aux structures culturelles ? » *Culture et Recherche, Les publics in situ et en ligne*, Hiver (134) : 36-39.

Calzada Prado, Javier et Miguel Ángel Marzal. (2013). « Incorporating data literacy into information literacy programs: Core competencies and contents ». *Libri*, 63(2) : 123-134.

Cardon, Dominique. (2015). *À quoi rêvent les algorithmes? Nos vies à l'heure des big data*. Paris : Seuil, 112 p.

Casemajor, Nathalie, Guillaume Sirois, Lily-Cannelle Mathieu, Alexis Morin et Louis Melançon. (2021). *La gouvernance des données d'usage : enjeux éthiques et perceptions des publics dans les bibliothèques et archives. Synthèse de la littérature*. Montréal : INRS, 108 p.

Cavoukian, Ann. (2006). *The 7 Foundational Principles Implementation and Mapping of Fair Information Practices*. Toronto : Commissariat à l'information et à la vie privée de l'Ontario. 12 p.

Choi, Hanbyul, Jonghwa Park et Yoonhyuk Jung. (2018). « The role of privacy fatigue in online privacy behavior ». *Computers in Human Behavior*, 81 : 42-51.

Choi, SoeYoon. (2022). « Privacy Literacy on Social Media: Its Predictors and Outcomes ». *International Journal of Human-Computer Interaction*, 39(1) : 217-232.

Chun, Wendy Hui Kyong. (2021). *Discriminating Data: Correlation, Neighborhoods, and the New Politics of Recognition*. Cambridge (É-U) : MIT Press.

Commissariat à la protection de la vie privée du Canada (CPVPC). (2019). *Réforme des droits sur la vie privée : Pour faire respecter les droits et rétablir la confiance envers le gouvernement et l'économie numérique. Rapport annuel au Parlement 2018-2019 concernant la Loi sur la protection des renseignements personnels et la Loi sur la protection des renseignements personnels et les documents électroniques*. Ottawa : CPVPC, 1913-3375.

Commission d'accès à l'information du Québec (CAIQ). (2021). *Intelligence artificielle et protection des renseignements personnels. Retour sur la consultation sur les principes en intelligence artificielle*. En ligne. https://www.cai.gouv.qc.ca/documents/CAI_M_consultation_AI_PRP.pdf (consulté le 1er juillet 2023).

Commission d'accès à l'information du Québec (CAIQ). (2023). *Protection des renseignements personnels*. En ligne. <https://www.cai.gouv.qc.ca/organismes/protection-des-renseignements-personnels/> (consulté le 15 juillet 2023).

Couldry, Nick et Alison Powell. (2014). « Big Data from the Bottom Up ». *Big Data & Society*. 1(2) : 1-5.

Curry, Ann. (2005). « If I Ask, Will They Answer? Evaluating Public Library Reference Service to Gay and Lesbian Youth ». *Reference & User Services Quarterly*, 45(1) : 65-75.

DAIGLE/SAIRE. (2020). *Étude des publics des arts de la scène au Québec*. En ligne. http://www.daiglesaire.ca/upload/pdf/DS_GTFAS_Etude_des_publics.pdf (consulté le 1er juillet 2023).

D'Alché-Buc, Florence, Valérie Beaudouin, Emmanuelle Bernès, Philippe Chevalier, Aude Le Moullec-Rieux, Adrien Nouvellet, Christophe Prieur et François Roueff. (2017). *Analysis of Gallica and Data BnF logs and Modelling of Behaviour Patterns: Presentation of the Main Results*. Paris : Bibliothèque nationale de France (BnF), Télécom ParisTech, 6 p.

Dencik, Lina et Jonathan Cable. (2017). « Digital Citizenship and Surveillance | The Advent of Surveillance Realism: Public Opinion and Activist Responses to the Snowden Leaks ». *International Journal of Communication*, 11(2017) : 763-781.

Denis, Juliette et Joëlle Farchy. (2020). *La culture des données: intelligence artificielle et algorithmes dans les industries culturelles*. Paris : Presses des Mines, 266 p.

Denouël, Julie et Fabien Granjon. (2011). *Communiquer à l'ère numérique*. Paris : Presses des Mines.

Dilhac, Marc-Antoine, Christophe Abrassart et Nathalie Voarino. (2018). « Rapport de la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle. » Montréal : Université de Montréal. En ligne.

<https://www.google.com/url?q=https://papyrus.bib.umontreal.ca/xmlui/handle/1866/22498>
(consulté le 1er août 2023).

Donnat, Olivier. (2017). « La question du public, d'un siècle à l'autre. » *Culture et Recherche, Les publics in situ et en ligne*. Hiver (134) : 6-8.

Draper, Nora A. et Joseph Turow. (2019). « The corporate cultivation of digital resignation ». *New Media & Society*, 21(8) : 1824-1839.

Duguay, Stefanie. (2016). « 'He has a way gay Facebook than I do': Investigating sexual identity disclosure and context collapse on a social networking site ». *New Media & Society*, 18(6) : 891-907.

Elmer, Greg. (2003). *Profiling Machines: Mapping the Personal Information Economy*. Cambridge (É-U) : MIT Press, 176 p.

Elueze, Isioma et Anabel Quan-Haase. (2018). « Privacy attitudes and concerns in the digital lives of older adults: Westin's privacy attitude typology revisited ». *American Behavioral Scientist*, 62(10) : 1372-1391.

Epstein, Dmitry et Kelly Quinn. (2020). « Markers of Online Privacy Marginalization: Empirical Examination of Socioeconomic Disparities in Social Media Privacy Attitudes, Literacy, and Behavior ». *Social Media + Society*, 6(2).

Eshet-Alkalai, Yoram et Eran Chajut. (2009). « Changes over time in digital literacy ». *CyberPsychology & Behavior*, 12(6) : 713-715.

Eubanks, Virginia. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York : St Martin's Press, 272 p.

Ferrari, Anusca. (2013). *DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe*. Luxembourg : Publications Office of the European Union.

- Fogel, Joshua et Elham Nehmad. (2009). « Internet social network communities: Risk taking, trust, and privacy concerns ». *Computers in Human Behavior*, 25(1) : 153-160.
- Fotopoulou, Aristeia. (2021). « Conceptualising critical data literacies for civil society organisations: agency, care, and social responsibility ». *Information, Communication & Society*, 24(11) : 1640-1657.
- Fourcade, Marion et Fleur Johns. (2020). « Loops, ladders and links: the recursivity of social and machine learning ». *Theory and Society*, 49(5-6) : 803-832.
- Frener, Regine et Sabine Trepte. (2022). « Theorizing Gender in Online Privacy Research ». *Journal of Media Psychology*, 34(2) : 77-88.
- Gambs, Sébastien. (2020). « Respect de la vie privée et problématiques éthiques à l'ère des données massives ». Présentation. Colloque *Éthique, justice et connaissance* organisé par le Centre de recherche en éthique de Montréal. En ligne. https://www.ciqss.org/sites/default/files/documents/S%C3%A9bastien%20Gambs_%C3%A9thiques.pdf (consulté le 21 juin 2022).
- Gauthier, Sabine. (2023). « L'intelligence artificielle, le partage des données et les droits humains. » *Université de Sherbrooke*. En ligne. <https://www.usherbrooke.ca/actualites/series-udes/videos/videos-details/50160> (consulté le 1er août 2023).
- Gebre, Engida H. (2018). « Young Adults' Understanding and Use of Data: Insights for Fostering Secondary School Students' Data Literacy ». *Canadian Journal of Science, Mathematics and Technology Education*, 18 : 330-341.
- Gerbet, Thomas. (2021). « La Grande Bibliothèque ferme à cause de la fuite de données de La Place 0-5 ». *Radio-Canada*. En ligne. <https://ici.radio-canada.ca/nouvelle/1793356/fuite-donnees-fermeture-grande-bibliotheque-inmedia> (consulté le 20 juillet 2023).
- Gillespie, Tarleton et Kristen A. Foot. (2014). « The Relevance of Algorithms » In *Media Technologies: Essays on Communication, Materiality and Society*, dir. de Tarleton Gillespie et Pablo J. Boczkowski. Cambridge (É-U) : MIT Press, 167-193.
- Gingras, Marie-Ève et Hélène Belleau. (2015). *Avantages et désavantages du sondage en ligne comme méthode de collecte de données : une revue de la littérature*. Montréal : INRS - Centre Urbanisation Culture Société, 15 p.
- Gitelman, Lisa. (2013). *Raw Data Is An Oxymoron*. Cambridge (É-U) : MIT Press, 192 p.
- Gliga, Adriana. (2023). « Loi 25 au Québec : votre organisation est-elle bien préparée ? » *MNP*. En ligne. <https://www.mnp.ca/fr/points-de-vue/bibliotheque/loi-25-au-quebec-votre-organisation-est-elle-bien-prepree> (consulté le 21 juillet 2023).
- Gray, Jonathan, Carolin Gerlitz et Liliana Bounegru. (2018). « Data infrastructure literacy ». *Big Data & Society*, 5(2).

- Gui, Marco. (2007). « Formal and substantial Internet information skills: The role of socio-demographic differences on the possession of different components of digital literacy. » *First Monday*, 12(9) : 1-16.
- Guyan, Kevin. (2022). « Fixing the Wrong Problems: Queer Communities and the False Promise of Unbiased and Equal Data Systems ». *European Data Protection Law Review*, 8(4) : 455-461.
- Haight, Michael, Anabel Quan-Haase et Bradley A. Corbett. (2014). « Revisiting the Digital Divide in Canada: The Impact of Demographic Factors on Access to the Internet, Level of Online Activity, and Social Networking Site Usage ». *Information, Communication & Society*, 17(4) : 503-519.
- Hanquinet, Laurie, Dave O'Brien et Mark Taylor. (2019). « The coming crisis of cultural engagement? Measurement, methods, and the nuances of niche activities ». *Cultural Trends*, 28(2/3) : 198-219.
- Hargittai, Eszter. (2002). « Second-Level Digital Divide: Differences in People's Online Skills ». *First Monday*, 7(4).
- Hargittai, Eszter. (2005). « Survey Measures of Web-Oriented Digital Literacy ». *Social Science Computer Review*, 23(3) : 371-379.
- Hargittai, Eszter et Kerry Dobransky. (2017). « Old dogs, new clicks: Digital inequality in skills and uses among older adults ». *Canadian Journal of Communication*, 42(2) : 195-212.
- Hargittai, Eszter et Eden Litt. (2013). « New strategies for employment? Internet skills and online privacy practices during people's job search ». *IEEE Security & Privacy*, 11(3) : 38-45.
- Hargittai, Eszter et Alice Marwick. (2016). « 'What Can I Really Do?' Explaining the Privacy Paradox with Online Apathy ». *International Journal of Communication* 10(21) : 3737-3758.
- Hartman, Todd, Helen Kennedy, Robin Steedman et Rhianne Jones. (2020). « Public perceptions of good data management: Findings from a UK-based survey ». *Big Data & Society*, 7(1).
- HCDH - Haut-Commissariat des Nations Unies aux droits de l'homme. (2021). « Michelle Bachelet : les risques d'atteinte à la vie privée liés à l'intelligence artificielle nécessitent une action urgente. » En ligne. <https://www.ohchr.org/fr/press-releases/2021/09/artificial-intelligence-risks-privacy-demand-urgent-action-bachelet> (consulté le 1er août 2023).
- Hobbs, Renee. (2010). « Digital and Media Literacy: A Plan of Action ». *The Aspen Institute Communications and Society Program*. En ligne. https://www.aspeninstitute.org/wp-content/uploads/2010/11/Digital_and_Media_Literacy.pdf (consulté le 24 juin 2023).
- Hoffman, Christian Pieter, Christopher Lutz et Giulia Ranzini. (2016). « Privacy cynicism: A new approach to the privacy paradox ». *Journal of Psychosocial Research on Cyberspace*, 10(4) : article 7.
- Hoy, Mariea G. et George Milne. (2010). « Gender Differences in Privacy-Related Measures for Young Adult Facebook Users ». *Journal of Interactive Advertising*, 10(2) : 28-45.

Iacovino, Livia et Malcolm Todd. (2007). « The long-term preservation of identifiable personal data: a comparative archival perspective on privacy regulatory models in the European Union, Australia, Canada and the United States ». *Archival Science*, 7(1) : 107-127.

Institut Fasken. (2022). *La réglementation sur l'intelligence artificielle au Canada et à l'étranger : comparaison entre le projet de loi C-27 édictant la LIAD et la législation sur l'IA de l'Union européenne*. En ligne. <https://www.fasken.com/fr/knowledge/2022/10/18-the-regulation-of-artificial-intelligence-in-canada-and-abroad> (consulté le 1er août 2023).

Jarvie, Max, George R. Wray, Catherine Labasi et Daniel E. Khoury. (2022). « Projet de loi C-27 : incidence des nouvelles lois canadiennes en matière de protection de la vie privée sur les entreprises » *Borden Ladner Gervais*. En ligne. <https://www.blg.com/fr/insights/2022/06/canadas-artificial-intelligence-and-data-act-impact-for-businesses> (consulté le 1er août 2023).

Kennedy, Helen et Giles Moss. (2015). « Known or knowing publics? Social media data mining and the question of public agency ». *Big Data & Society*, 2(2) : 1-11.

Kennedy, Helen, Susan Oman, Hannah Ditchfield, Mark Taylor, Jo Bates, Itzelle Medina-Perea, Lulu Pinney et Monika Fratzczak. (2022). *Data Matters Are Human Matters: final Living With Data report on public perceptions of public sector data uses*. Sheffield, Angleterre : Nuffield Foundation.

Kezer, Murat, Barış Sevi, Zeynep Cemalcilar et Lemi Baruh. (2016). « Age differences in privacy attitudes, literacy and privacy management on Facebook ». *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 10 (1).

Kirsch, Irwin et Marilou Lennon. (2005). « Cadre conceptuel de la littératie en matière de technologie de l'information et des communications (TIC). » In *Mesurer la littératie et les compétences des adultes : Des nouveaux cadres d'évaluation*, sous la dir. de Scott T. Murray, Yvan Clermont et Marilyn Brinkley. Ottawa : Statistiques Canada.

Laboratoire de Cyberjustice. (2019). « Guide de bonnes pratiques concernant la création de politiques de confidentialité dans le secteur de la culture. » sous la dir. de IVADO et Synapse C. Montréal : Laboratoire de Cyberjustice, 29 p.

Lapointe, Marie-Claude, Jason Luckerhoff et Lucile Niort. (2022) *Portrait des emprunteurs de livres numériques à BANQ*, Montréal : BANQ, 66 p.

Larson, Ronald. (2019). « Library stakeholder attitudes and new technology ». *Library Management* 40(1/2) : 28-140.

Legoux, Renaud et Simon Lord. (2019). « La culture des données, un élément qui change la donne ». *Gestion*, 44(1) : 96-99.

Li, Xiaoqian, Wenhong Chen et Joseph D. Straubhaar. (2018). « Privacy at the Margins| Concerns, Skills, and Activities: Multilayered Privacy Issues in Disadvantaged Urban Communities ». *International Journal of Communication*, 12(22) : 1269-1290.

Litt, Eden. (2013). « Understanding social network site users' privacy tool use ». *Computers in Human Behavior*, 29(4) : 1649-1656.

Lutz, Christoph, Christian P. Hoffman et Giulia Ranzini. (2020). « Data capitalism and the user: An exploration of privacy cynicism in Germany ». *New Media & Society*, 22(7) : 1168-1187.

Maceli, Monica G. (2018). « Encouraging patron adoption of privacy-protection technologies: Challenges for public libraries ». *IFLA Journal*, 44(3) : 195-202.

Malik, Aqdas, Kari Hiekkanen et Marko Nieminen. (2016). « Privacy and trust in Facebook photo sharing: age and gender differences ». *Program: electronic library and information systems*, 50(4) : 462-480.

Margulis, Stephen T. (2003). « Privacy as a Social Issue and Behavioral Concept ». *Journal of Social Issues*, 59(2) : 243-261.

Markham, Annette N. (2019). « Critical pedagogy as a response to datafication ». *Qualitative Inquiry*, 25(8) : 754-760.

Marwick, Alice, Claire Fontaine et danah boyd. (2017). « 'Nobody Sees It, Nobody Gets Mad': Social Media, Privacy, and Personal Responsibility Among Low-SES Youth ». *Social Media + Society*, 3(2) : 1-14.

Marwick, Alice E. (2023). *The Private Is Political: Networked Privacy and Social Media*. New Haven: Yale University Press, 384 p.

Marwick, Alice E. et danah boyd. (2014). « Networked privacy: How teenagers negotiate context in social media ». *New Media & Society*, 16(7) : 1051-1067.

Mathson, Stephanie et Jeffrey Hancks. (2007). « Privacy Please? A Comparison Between Self-Checkout and Book Checkout Desk Circulation Rates for LGBT and Other Books ». *Journal of Access Services*, 4(3-4) : 27-37.

McConnell, Elizabeth, Bálint Néray, Bernie Hogan, Aaron Korpak, Antonia Clifford et Michelle Birkett. (2018). « 'Everybody Puts Their Whole Life on Facebook': Identity Management and the Online Social Networks of LGBTQ Youth ». *International Journal of Environmental Research and Public Health*, 15(6) : 1078.

Ministère de la Culture et des Communications du Québec (MCCQ) - (2013) « Chiffres à l'appui. Culture, communications et technologies numériques au Québec. » En ligne.
<https://mcc.gouv.qc.ca/fileadmin/documents/publications/chiffres-appui-techno-numeriques-mcc-mars2013.pdf> (consulté le 1er juillet 2023).

Ministère de la Culture et des Communications du Québec (MCCQ) (2014) *Chiffres à l'appui du ministère de la Culture et des Communications du Québec*. En ligne.
https://www.mcc.gouv.qc.ca/fileadmin/documents/publications/ChiffresAlAppui-Numerique-MCC-MAI_2014.pdf (consulté le 1er juin 2023).

McStay, Andrew. (2020). « Emotional AI, soft biometrics and the surveillance of emotional life: An unusual consensus on privacy ». *Big Data & Society*, 7(1) : 1-12.

Motiwalla, Luvai F. et Xiao-Bai Li. (2016). « Unveiling consumer's privacy paradox behaviour in an economic exchange ». *International Journal of Business Information Systems*, 23(3) : 307-329.

Nissenbaum, Helen. (2009). *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford (É-U) : Stanford University Press.

Nouvellet, Adrien, Valérie Beaudouin, Florence D'Alché-Buc, Christophe Prieur et François Roueff. (2017). « Analyse des traces d'usage de Gallica ». *Télécom ParisTech*. En ligne. <https://hal.archives-ouvertes.fr/hal-01709264> (consulté le 1er juillet 2023).

Oh, Sarah S, Kyoung-A Kim, Minsu Kim, Jaeuk Oh, Sang H. Chu et JiYeon Choi. (2021). « Measurement of digital literacy among older adults: Systematic review [Review] ». *Journal of Medical Internet Research*, 23(2:e26145).

Pangrazio, Luci, Anna-Lena Godhe et Alejo González López Ledesma. (2020). « What is digital literacy? A comparative review of publications across three language contexts ». *E-Learning and Digital Media*, 17(6) : 442-459.

Pangrazio, Luci et Julian Sefton-Green. (2020). « The social utility of 'data literacy' ». *Learning, Media and Technology*, 45(2) : 208-220.

Pangrazio, Luci et Neil Selwyn. (2018). « 'It's Not Like It's Life or Death or Whatever': Young People's Understandings of Social Media Data ». *Social Media + Society*, 4(3) : 1-9.

Pangrazio, Luci et Neil Selwyn. (2019). « 'Personal data literacies': A critical literacies approach to enhancing understandings of personal digital data ». *New Media & Society*, 21(2) : 419-437.

Papacharissi, Zizi. (2010). « Privacy as a luxury commodity. » *First Monday*, 15(8).

Park, Yong Jin. (2013a). « Digital literacy and privacy behavior online ». *Communication Research*, 40(2) : 215-236.

Park, Yong Jin. (2013b). « Offline status, online status: Reproduction of social categories in personal information skill and knowledge ». *Social Science Computer Review*, 31(6) : 680-702.

Park, Yong Jin et S. Mo Jang. (2014). « Understanding privacy knowledge and skill in mobile communication ». *Computers in Human Behavior*, 38 : 296-303.

Pasek, Anne, Rena Bivens et Mél Hogan. (2019). « Data segregation and algorithmic amplification: A conversation with Wendy Hui Kyong Chun ». *Canadian Journal of Communication*, 44(3) : 455-469.

Petronio, Sandra. (2002). *Boundaries of Privacy: Dialectics of Disclosure*. New York : SUNY Press.

Phoenix SPI. (2019). *Rapport final : Sondage auprès des Canadiens sur la protection de la vie privée de 2018-2019*. Réalisé pour le Commissariat à la protection de la vie privée du Canada. En ligne.

https://www.priv.gc.ca/fr/mesures-et-decisions-prises-par-le-commissariat/recherche/consulter-les-travaux-de-recherche-sur-la-protection-de-la-vie-privee/2019/por_2019_ca/ (consulté le 1er juillet 2023).

Quan-Haase, Anabel, Carly Williams, Maria Kicevski, Isioma Elueze et Barry Wellman. (2018). « Dividing the grey divide: Deconstructing myths about older adults' online activities, skills, and attitudes ». *American Behavioral Scientist*, 62(9) : 1207-1228.

Quan-Haase, Annabel et Dennis Ho. (2020). « Online privacy concerns and privacy protection strategies among older adults in East York, Canada ». *Journal of the Association for Information Science and Technology*, 71(9) : 1089-1102.

Raynes-Goldie, Kate. (2010). « Aliases, creeping, and wall cleaning: Understanding privacy in the age of Facebook ». *First Monday*, 15(1).

Risdale, Chantel, James Rothwell, Mike Smit, Hossam Ali-Hassan, Michael Bliemel, Dean Irvine, Daniel Kelley, Stan Matwin et Brad Wuetherick. (2015) *Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report*. Halifax : Dalhousie University.

Ritzhaupt, Albert D., Feng Liu, Kara Dawson et Ann E. Barron. (2013). « Differences in student information and communication technology literacy based on socio-economic status, ethnicity, and gender: Evidence of a digital divide in Florida schools ». *Journal of Research on Technology in Education*, 45(4) : 291-307.

Roberts, Lynne D., Joel A. Howell, Kristen Seaman et David C. Gibson. (2016). « Student Attitudes toward Learning Analytics in Higher Education: 'The Fitbit Version of the Learning World' ». *Frontiers in Psychology*, 7 : 1-11.

Robinson, Laura, Jeremy Schulz, Grant Blank, Massimo Ragnedda, Hiroshi Ono et al. (2020). « Digital inequalities 2.0: Legacy inequalities in the information age ». *First Monday*, 25 (7).

Rogers, Everett M. (2010). *Diffusion of Innovations, 4th Edition*. New York : Free Press.

Rollet, Nicolas, Valérie Beaudouin et Isabelle Garron. (2017). « Vidéo-ethnographie des usages de Gallica : une exploration au plus près de l'activité ». *Document Numérique* 20(2/3) : 97-114.

Rouvroy, Antoinette et Thomas Berns. (2013). « Gouvernamentalité algorithmique et perspectives d'émancipation ». *Réseaux*, 1(177) : 163-196.

Ruppert, Evelyn. (2018). *Sociotechnical imaginaries of Different Data Futures*. Rotterdam : Erasmus Universiteit Rotterdam.

Ruttan, Craig, Raly Chakarova, Natasha Apollonova, Patrick Gill et Brian Kelcey. (2019). *BiblioTech: A City-Building Proposal for the Toronto Public Library to Establish a Civic Data Hub*. Réalisé pour le Toronto Board of Trade. En ligne. <https://bot.com/Resources/Resource-Library/BiblioTech> (consulté le 1er juillet 2023).

- Samarati, Pierangela. (2001). « Protecting respondents identities in microdata release ». *IEEE Transactions on Knowledge and Data Engineering*, 13(6) : 1010-1027.
- Schallum, Pierre. (2023). « Le principe de nécessité ou de minimisation des données comme condition à l'innovation responsable : données de qualité, vie privée, cybersécurité et environnement ». *Technologie et innovation*, 8 : 1-9.
- Sheehan, Kim B. (1999). « An investigation of gender differences in on-line privacy concerns and resultant behaviors ». *Journal of Interactive Marketing*, 13(4) : 24-38.
- Sheehan, Kim B. (2002). « Toward a Typology of Internet Users and Online Privacy Concerns ». *The Information Society*, 18(1) : 21-32.
- Sillitoe, Paul J. (1998). « Privacy in a public place: Managing public access to personal information controlled by archives services ». *Journal of the Society of Archivists* 19(1) : 5-15.
- Stoilova, Mariya, Sonia Livingstone et Rishita Nandagiri. (2020). « Digital by Default: Children's Capacity to Understand and Manage Online Data and Privacy ». *Media and Communication*, 8(4) : 197-207.
- Taylor, Yvette, Emily Falconer et Ria Snowden. (2014). « Queer youth, Facebook and faith: Facebook methodologies and online identities ». *New Media & Society*, 16(7) : 1138-1153.
- Tchéhouali, Destiny et Josée Plamondon. (2019). *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 ; Coalition pour la Culture et les Médias, 30 p.
- Tchéhouali, Destiny. (2021). *ALTER ALGO. Les algorithmes sont-ils vraiment nos alter ego numériques ?* Montréal : ISOC Québec.
- Travis, Tiffini A. et Christian Ramirez. (2020). « Big Data and Academic Libraries: The Quest for Informed Decision-Making ». *Libraries and The Academy*, 20(1): 33-47.
- Tremblay, Alain. (2023). « Comprendre les grands principes de la nouvelle Loi 25 au Québec ». *Tink*. En ligne. <https://www.tink.ca/perspectives/comprendre-les-grands-principes-de-la-nouvelle-loi-25-au-quebec> (consulté le 1er juillet 2023).
- Trepte, Sabine, Doris Teutsch, Philipp K. Masur, Carolin Eicher, Mona Fischer et Alisa Hennhöfer. (2014). « Do People Know About Privacy and Data Protection Strategies? Towards the 'Online Privacy Literacy Scale' (OPLIS) ». In *Reforming European Data Protection Law*, sous la dir. de Serge Gutwirth, Ronald Leenes et Paul de Hert. Dordrecht : Springer Netherlands, 333-365.
- Trottier, Daniel. (2018). « Privacy and Surveillance ». In *The Sage Handbook of Social Media*, sous la dir. de Jean Burgess, Alice Marwick et Thomas Poell. Londres : SAGE, 463-480.
- Tummon, Nikki et Dawn McKinnon. (2018). « Attitudes and practices of Canadian academic librarians regarding library and online privacy: A national study ». *Library & Information Science Research*, 40(2) : 86-97.

van Deursen, Alexander et Jan van Dijk. (2011). « Internet skills and the digital divide ». *New Media & Society*, 13(6) : 893-911.

van Dijck, José. (2014). « Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology ». *Surveillance & Society*, 12(2) : 197-208.

Véliz, Carissa. (2021). *Privacy is Power*. New York : Melville House.

Villani, Cédric. (2018). In *Donner un sens à l'intelligence artificielle : pour une stratégie nationale et européenne*. Paris : Mission parlementaire du gouvernement français, 233 p.

Vitak, Jessica, Yuting Liao, Priya Kumar et Mega Subramaniam. (2018). « Librarians as Information Intermediaries: Navigating Tensions Between Being Helpful and Being Liable ». In *Transforming Digital Worlds*, sous la dir. de Gobinda Chowdhury, Julie McLeod, Val Gillet et Peter Willet. New York : Springer, 693-702.

West, Sarah M. (2019). « Data Capitalism: Redefining the Logics of Surveillance and Privacy ». *Business & Society*, 58(1) : 20-41.

Westin, Alan F. (1967). *Privacy and Freedom*. New York : Athenum.

Westin, Alan F. (2003). « Social and Political Dimensions of Privacy ». *Journal of Social Issues*, 59(2) : 431-453.

Wissinger, Christina L. (2017). « Privacy literacy: From Theory to Practice ». *Communications in Information Literacy*, 11(2) : 378-389.

Young, Alyson L. et Anabel Quan-Haase. (2013). « Privacy protection strategies on Facebook: The Internet privacy paradox revisited ». *Information, Communication & Society* 16(4) : 479-500.

Ytre-Arne, Brita et Ranjana Das. (2019). « An Agenda in the Interest of Audiences: facing the Challenges of Intrusive Media Technologies ». *Television & New Media*, 20(2) : 184-219.

Zini, Élodie. (2020). « L'utilisation des personas à Bibliothèque et Archives nationales du Québec : une rencontre avec notre public ». *Documentation et bibliothèques*, 66(3) : 16-28.

Zuboff, Shoshana. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York : PublicAffairs, 704 p.

Zukowski, Tomasz et Irwin Brown. (2007). « Examining the influence of demographic factors on internet users' information privacy concerns ». *ACM International Conference Proceeding Series*, 226 : 197-204.