

Moissonnage¹ des sites Web gouvernementaux

Cette fiche informative s'adresse à tous les organismes publics visés par la [Loi sur les archives](#).

Présentation

Internet est aujourd'hui une source d'information incontournable tant par la richesse de son contenu que par l'éventail des sujets abordés. Il offre une image de la société à un moment de son histoire et présente souvent des contenus exclusifs (non disponibles sous une autre forme). La conservation de sites Web devient donc essentielle pour une institution vouée à la constitution, à la préservation et à la diffusion du patrimoine documentaire comme Bibliothèque et Archives nationales du Québec (BANQ).

À BANQ, le moissonnage des sites Web est effectué par une équipe relevant de la Direction du dépôt légal et des acquisitions. Il s'inscrit dans les activités de constitution et de diffusion du patrimoine documentaire québécois publié. Les sites moissonnés sont destinés à une conservation permanente par BANQ.

Programme de collecte

Alors que le dépôt légal a pour objectif de rassembler tout le patrimoine documentaire publié, la collecte exhaustive des sites Web québécois est impossible. Parmi les nombreuses contraintes, notons : l'ampleur du corpus (en 2016, il y avait plus de 2,5 millions de sites enregistrés dont le nom de domaine se terminait par « .ca »), la mise à jour continue des sites Web, les ressources disponibles à BANQ et des obligations juridiques (nécessité d'obtenir une licence pour la diffusion).

C'est pourquoi BANQ procède à un moissonnage sélectif de sites Web québécois. Le programme distingue trois types de sites :

1. sites des organismes assujettis à la Loi sur les archives;
2. sites thématiques (ex. : entreprises privées);
3. sites événementiels (ex. : élections).

¹ Technique d'extraction automatique de données à partir d'un ou de plusieurs sites Web dans le but d'utiliser celles-ci, après traitement, dans un autre contexte. Office québécois de la langue française, Le grand dictionnaire terminologique, http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?id_Fiche=26507119, consulté le 23 novembre 2018.

En complément des fonds d'archives qu'elle détient, BANQ collecte tous les sites Web des ministères et organismes publics visés par les paragraphes 1 à 3 de l'annexe de la Loi sur les archives. BANQ vise aussi la collecte des sites Web des organismes inclus dans le Cadre de classement des publications gouvernementales du Québec.

Le moissonnage des sites Web des organismes visés par les paragraphes 4 à 7 de l'annexe de la Loi sur les archives sera effectué dans la mesure où l'organisme aura accordé à BANQ son autorisation par la signature d'une [licence](#)². BANQ ne diffuse aucun site sans obtenir une licence au préalable. Lorsque BANQ collecte le site d'un organisme, celui-ci n'a plus à le conserver en vertu de son calendrier de conservation.

Diffusion

La licence accordée à BANQ détermine si la diffusion des sites Web collectés s'effectue sur le portail de BANQ ou uniquement sur des postes informatiques à l'intérieur de nos locaux. Les sites sont consultés via une interface spécifique, [Archivage Web](#), disponible sur le portail de BANQ³. L'interface permet la recherche par regroupement thématique, par mots-clés ou par URL. BANQ ne peut garantir un affichage parfait des sites collectés.

Contact

Pour plus de détails, n'hésitez pas à consulter le [programme de collecte de sites Web](#)

Si vous avez des questions, vous pouvez contacter l'archiviste responsable de conseiller votre organisme.

Si votre site Web fait l'objet d'une refonte ou qu'un nouveau site est mis en ligne, si votre calendrier de conservation contient des règles concernant la fréquence d'archivage, veuillez aviser ou communiquer cette fréquence directement avec l'équipe de l'archivage Web par courriel : archivageweb@banq.qc.ca afin d'assurer que le site modifié soit archivé.

² La licence accordée à BANQ autorise le moissonnage (reproduction et archivage), la conservation à long terme et la diffusion du site collecté. Le producteur du site demeure titulaire des droits d'auteur.

³ Afin d'éviter toute confusion, un bandeau dans le haut des pages consultées indique qu'il s'agit d'un site archivé.