



# Bibliothèque, Open Science, Open Data et données de la recherche au Canada : quels enjeux ?

Alexandre Tur

---

Les billets d'EnssibLab  
21 juillet 2015



Alexandre Tur, chartiste de formation, a obtenu le diplôme de conservateur des bibliothèques de l'Enssib en 2015. Il a effectué son stage professionnel au Canada dans le département des services aux chercheurs de la Queen's University de Kingston (Ontario) où il a travaillé en priorité sur l'aide à la gestion des données de la recherche, notamment pour les chercheurs en Humanités, et à la promotion de l'Open Access au sein de l'université. Il est également l'auteur du mémoire d'étude "Accompagner les citoyens dans l'acquisition d'une culture numérique : le rôle des bibliothèques de lecture publique dans la formation au numérique". Il revient pour EnssibLab sur son expérience et nous livre avec beaucoup d'acuité ses réflexions sur la

tension, voire la confusion, qu'il peut y avoir entre ces 3 notions essentielles au paysage actuel de la recherche et des bibliothèques : gestion des données de la recherche, ouverture des résultats de la recherche (Open Science) et ouverture des données publiques (Open Data).



Dans le cadre de son stage professionnel à [la bibliothèque de la Queen's University de Kingston](#) (Ontario, Canada), Alexandre Tur a eu l'occasion de proposer plusieurs formations internes sur **les enjeux de la gestion des données de la recherche** en Humanités numériques. Sans grande surprise, à l'instar de la situation des chercheurs en France, il observe les dérives du débat soit vers la réticence des chercheurs à **partager leurs données** soit vers leur difficulté à **trouver des données utilisables comme sources** pour leurs recherches.

Cet état de fait pourrait entre autres s'expliquer par une tension entre **gestion des données de la recherche, ouverture des résultats de la recherche (Open Science) et ouverture des données publiques (Open Data)** liée à une certaine confusion des vocabulaires. Peut-être le cas des Humanités est-il particulièrement propice à déceler cette tension ; dans la mesure où l'importance d'une rigoureuse gestion des données de la recherche n'est pas (encore) aussi établie que dans certaines disciplines scientifiques, les enjeux dérivés de la mise à disposition de ces données sont d'autant plus visibles. Pour autant, cette tension est indubitablement le résultat de **politiques publiques**, peut-être trop bien intériorisées par les bibliothécaires.

En effet, *Open Data* et *Open Science* sont les lames de fond des politiques mondiales liées à la recherche scientifique publique et universitaire. De fait, si l'on parle de données de la recherche dans les bibliothèques canadiennes, c'est principalement pour anticiper **les mandats open access** que devraient prochainement mettre en place les agences de financement de la recherche pour les bénéficiaires de leurs bourses, à la manière de ce qui existe déjà pour les publications scientifiques. Encore faut-il, pour mettre à disposition ses données de recherche, savoir ce qu'elles sont et si on en dispose, ce qui est loin d'être acquis pour tous les chercheurs en sciences humaines.

Cependant, cette situation ne relève pas d'une quelconque spécificité canadienne. Depuis [l'Initiative de Budapest](#), les politiques en faveur de l'*Open Science* sont globales. Lorsqu'on veut ne serait-ce que **définir les données de la recherche**, on se réfère généralement aux [Principes et lignes directrices pour l'accès aux données de la recherche financée sur fonds publics](#) de l'OCDE, dont le titre témoigne bien de l'intention dans laquelle elles ont été rédigées, et qui sont reprises assez littéralement par la plupart des états membres, Canada compris.

Certes, l'argument du partage des résultats la recherche publique est parfaitement recevable ; il n'empêche nullement de signaler par ailleurs d'autres avantages d'une bonne gestion des données de la recherche, comme une meilleure **efficacité dans la production de données**, une

garantie de **conservation pérenne**, une **facilité de citation, de partage et de réutilisation**, etc. Pour beaucoup de chercheurs en Humanités, rationaliser la production et l'usage de données dès le stade du projet de recherche peut permettre de **prendre conscience de l'importance de telles données** (encore largement absentes des méthodologies) au côté des « sources », et d'améliorer grandement l'efficacité dans la création et le traitement de ces données. C'est-à-dire tout au long de leur recherche et non seulement en amont et/ou en aval.

Dans les faits d'ailleurs, les *data librarians* rendent bien (et consciemment !) des services de **gestion des données de la recherche** : aide à la rédaction préliminaire d'un **plan de gestion des données**, à l'optimisation de la collecte de celle-ci en fonction des outils d'analyse, au traitement des données, à leur sécurisation et anonymisation, à la mise en place d'une préservation pérenne et de la citabilité qui l'accompagne, etc. C'est souvent le cas dès que la bibliothèque commence à endosser la gestion des données de la recherche – y compris en Humanités, même s'il s'agit souvent de la dernière discipline prise en charge (la plus complexe peut-être). A la [Bodleian Library d'Oxford](#), le [projet SUDAMIH](#) (SUpporting DAta MAagement Infrastructure for the Humanities, 2009-2011) a permis de mettre en place un service complet d'assistance à l'organisation de bases de données (*database as a service*), qui dépasse largement les questions de partage et de réparabilité des données.

A l'opposé – mais tout près du point de vue des politiques publiques – les données de la recherche sont (trop) volontiers associées à l'ouverture des données publiques. Il suffit pour s'en rendre compte de se rapporter au mémoire de conservateur de Rémi Gaillard soutenu à l'Enssib en 2014, « De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ? », au titre là encore évocateur, et qui constitue de fait la principale synthèse francophone de référence sur les données de la recherche. Or, si l'on peut ici encore souligner la responsabilité pour un chercheur de publier ses données financées sur fonds publics, et se réjouir des opportunités de réutilisation, un monde entier sépare **les données gouvernementales publiques**, dont l'ouverture fait l'enjeu des politiques « *Open Data* », **des données de la recherche proprement dites**, au public essentiellement universitaire, et dont le potentiel économique reste, dans la plupart des cas, à démontrer.

C'est peut-être particulièrement le cas dans les bibliothèques universitaires canadiennes pour des raisons historiques. Elles sont en effet, depuis le milieu du XXe siècle, des vecteurs identifiés et privilégiés de « **l'information gouvernementale** » : il s'agit d'un secteur complet de la politique documentaire, comprenant publications administratives et législatives diverses, émanant des différentes institutions canadiennes et américaines, mais également toute une production statistique et de données publiques. Pour une évaluation de l'importance de ces collections, on dira qu'à Queen's, la seule partie imprimée couvre un étage presque entier de la bibliothèque et qu'un bibliothécaire à plein temps en est responsable. Sans compter, bien sûr, les classeurs de disquettes et de CD-ROM camouflés dans les bureaux ni les données dématérialisées, aujourd'hui très majoritaires. Naguère encore vendues au prix fort sur support magnétique, ces données sont de plus en plus fréquemment **mises en ligne en open data**, mais la bibliothèque est restée dans la continuité de sa mission d'origine en repérant les diverses sources de données accessibles pour en faciliter l'accès aux chercheurs.

Outre les données statistiques, prêtes à l'emploi (graphiques, pourcentages, tableaux résultant de calculs déjà effectués sur de larges échantillons statistiques), la bibliothèque fournit également les résultats bruts (rendus anonymes) d'enquêtes et de sondages ; enfin, comme 25 autres bibliothèques universitaires, Queen's héberge par ailleurs un entrepôt sécurisé contenant des « micro-données » sensibles (comprendre suffisamment détaillées pour présenter un risque pour la vie privée des personnes), dont l'accès est restreint par accréditation, sur le modèle de ce qu'on peut connaître en France pour le dépôt légal du web ou de l'audiovisuel.

Or cet entrepôt sécurisé est intitulé... « Queen's Research Data Centre », car toutes ces données sont évidemment **mises à disposition dans le seul but d'alimenter la recherche**. Du point de vue d'un chercheur en sciences sociales, elles sont indubitablement de précieuses sources pour un projet de recherche... mais en aucune manière ce que nous appelons « données de la recherche »

et qui regroupe plutôt les données issues de la recherche universitaire. La proximité du nom alimente toutefois les confusions, d'autant que la même équipe de *data librarians* est responsable des deux domaines et que les [ressources sont très peu différenciées](#) sur le site internet et dans les formations.

**Les bibliothécaires sont eux-mêmes au cœur de cette ambiguïté** : quand l'activité de *data librarian* s'est étendue aux données de la recherche (soit assez récemment), ces dernières ont été vues comme une nouvelle source de données, à côté des statistiques et données gouvernementales, ouvertes ou non. Ce n'est pas faux, bien sûr, et cela ouvre des perspectives très intéressantes : la **plateforme d'archivage des données de la recherche de l'Ontario, <Odesi>**, essentiellement consacrée aux données issues de sondages, offre un impressionnant service de visualisation et de réutilisation des données en ligne, permettant notamment de combiner données publiques, données issues d'un projet de recherche précédent et données produites par le chercheur lui-même. Cependant, si l'on se place dans la perspective du chercheur (et de son accompagnement), il n'est pas certain qu'il s'agisse là de l'enjeu prioritaire des données de la recherche.

Par ailleurs, il ne s'agit d'ailleurs pas uniquement d'un défaut des bibliothèques canadiennes : en cherchant sur des annuaires comme [databib.org](#) ou [re3data.org](#) des entrepôts disciplinaires de données de la recherche en Humanités, on peut se rendre compte qu'un certain nombre des sites référencés ne sont **pas vraiment des plateformes de dépôt à destination des chercheurs**, mais, assez fréquemment, des bases de données permettant de stocker, de mettre à disposition, voire (parfois) de valoriser les données produites par quelques projets d'envergure, soit uniquement de collecte (sondages, mais aussi corpus linguistiques ou textuels, etc.) soit de recherche achevée.

Il est vrai que certains projets « de recherche », en SHS du moins, ont **davantage pour objectif d'extraire et de mettre à disposition** de la communauté une large quantité de données susceptibles d'alimenter les recherches individuelles **que de traiter eux-mêmes ces résultats** d'une manière plus « classique ». En ce qui concerne les bibliothèques, c'est notamment le cas des nombreux projets de numérisation et d'exploitation des collections spéciales (cartes, livres anciens, etc.). Si l'institution qui produit ces données a tendance à les considérer comme « données de la recherche », elles ne le sont plus à partir du moment où un autre chercheur les utilise comme sources. Surtout, la plupart des données gouvernementales qui font l'objet des politiques d'*OpenData* ne sont pas réellement rendues accessibles pour faciliter (ou non) la recherche universitaire : **les motivations sont bien plus largement économiques** sinon éthiques. Au final, ce qui est peut-être le plus regrettable, c'est la confusion entretenue dans le vocabulaire. Pourquoi ne pas reconnaître les données de la recherche comme un concept spécifique, d'abord lié à la recherche universitaire, et dont les relations avec les données publiques en amont, et la publication en libre accès en aval, ne sont que ponctuelles ? Peut-être est-ce évident dans certaines disciplines. Ça ne l'est en aucune manière en Humanités, et si les bibliothèques veulent accompagner l'éveil de ce champ de recherche à la problématique de la gestion avisée et utile de ses données, elles doivent être elles-mêmes bien au fait des définitions et des distinctions à opérer.

---

## Notes

[1] A signaler également l'intervention de Pierre Naegelen, [Données de la recherche : quel positionnement et quels rôles pour les bibliothèques ?](#), produite à l'occasion de la journée d'étude toulousaine du 15 juin 2015: "[Données en partage : enjeux et acteurs des données de la recherche aujourd'hui](#)".