

Les métiers des données : à l'intérieur, à l'extérieur et entre les mondes de la biblio- théconomie et de l'archivistique

Julien A. Raemy, assistant HES en sciences de l'information à la HEG-Genève // HES-SO

René Schneider, professeur ordinaire au sein de la filière ID à la HEG-Genève // HES-SO

Introduction

Le terme «métiers des données» désigne un nombre de plus en plus rapide et croissant de profils professionnels classiques qui sont complétés par le préfixe «Data». Il s'agit d'une réaction aux changements technologiques et sociaux qui vont de pair avec le flux toujours croissant d'informations ou de données de meilleure qualité ainsi que les changements technologiques et sociaux qui en découlent. Bien entendu, ces changements ne s'arrêtent pas aux professions des sciences de l'information.

Dans cet article, après une brève introduction, le lien entre l'information et les données sera abordé afin de dresser un panorama des professions actuelles et émergentes dans le domaine des données. Enfin, les trois ou quatre profils professionnels qui ont un lien particulièrement étroit avec les sciences de l'information seront examinés plus en détail : celui du bibliothécaire de données, de l'archiviste de données et du curateur de données, ainsi que la position particulière du scientifique de données.

Information et données

La science de l'information, qui - d'après son nom - se définit par le concept d'information et le désigne logiquement comme son objet principal et dont la recherche théorique et pratique constitue son intérêt pour le savoir, a largement déterminé dans le passé son rapport à ce dernier. Dans le monde anglo-saxon, l'information est souvent comprise comme une « connaissance communiquée » et dans le monde germanophone, elle est également appelée de manière pragmatique mais en moins élégante en tant que « connaissance en action ». Schématiquement, l'information est aussi généralement comprise comme sur-ordonnée ou post-ordonnée à la notion de « données » et sous-ordonnée ou pré-ordonnée à la notion de « savoir ».

Dans ce contexte, nous comprenons également la relation plutôt peu soignée entre l'information et les données, qui se justifie généralement par le fait que les données en tant que telles n'ont pas de sens et que leur étude en tant que telle ou leur analyse approfondie sont négligeables au point de ne pouvoir remplir leur rôle que dans un contexte particulier.

Cette approche peut être qualifiée d'historique, sinon d'obsolète, depuis la création du Big Data et l'établissement de diplômes en science des données qui y sont associés.

Selon la compréhension des scientifiques des données, les données ont déjà un sens (par exemple uniquement par la valeur qu'elles quantifient et par laquelle elles diffèrent des autres données, ou par la disposition dans une matrice, c'est-à-dire la relation aux autres données), implicitement (ou même intrinsèque) en plus d'être détectable ou explicite (ou même extrinsèque) par les algorithmes des scientifiques des données qui se fondent sur les métadonnées qui les accompagnent.

Cette appréhension aura inévitablement un impact sur les sciences de l'information et les professions connexes en raison de l'importance croissante de la science des données, dont l'impact sur les sciences au sens large et la société dans son ensemble n'est ni prévisible ni délimitable. La nouvelle discipline scientifique de la science des données (en définitive une fusion de l'informatique et des mathématiques avec les points focaux respectifs du *Machine Learning* et de la statistique) s'est établie pratiquement à la hâte et se caractérise par un rôle particulier dans la construction de ces disciplines scientifiques.

Ce rôle se caractérise également par le fait qu'au travers de l'attention et de l'importance des données (en tant que données ouvertes, liées ou de recherche), la question se pose de savoir dans quelle mesure les résultats, c'est-à-dire la production de la recherche scientifique dans les différentes disciplines, peuvent servir de contribution à la discipline des données scientifiques. Si l'on compare dans ce contexte le rôle particulier des sciences de l'information, qu'elles ont toujours joué en tant que successeur de la bibliothéconomie, à savoir en tant que prestataire de services dans toutes les disciplines, on obtient une image globale dans laquelle les différentes disciplines sont flanquées des sciences de l'information et des données.

Data Science : chance ou menace

Pour toutes les sciences - y compris les sciences de l'information - cela soulève la question d'un positionnement par rapport à la science des données et la question de savoir si la relation qui en résulte est une relation de concurrence ou de collaboration et si le changement résultant doit être considéré comme une opportunité ou une menace.

Ce qui est clair en tout cas, indépendamment des noms respectifs, c'est que l'influence de l'informatique et des technologies de l'information sur ces professions ne diminue pas, mais qu'elle s'accroît de façon incontrôlée.

En ce qui concerne la question de l'opportunité ou de la menace, on peut dire que ce changement peut être considéré comme une opportunité rentable, car les scientifiques se plaignent constamment qu'ils consacrent encore la majeure partie de leur travail à nettoyer et à traiter les données essentielles à leur travail, et ce dans l'incertitude quant à la possibilité de trouver toutes les données pertinentes ou de les rendre suffisamment disponibles. Les bibliothécaires, archivistes et documentalistes peuvent répondre à ce besoin en traitant les données du début de leur création jusqu'à leur archivage ou désherbage de la même manière qu'ils le font depuis longtemps avec les livres, documents et archives. En gros, en leur fournissant des métadonnées pour que les données puissent être trouvées et (ré)utilisées.

Le défi consiste certainement à normaliser les formats et les métadonnées, à convertir ou à migrer les médias et, enfin et surtout, au volume ou à la masse des données et

à leur complexité (en tant que base de données, application multimédia ou web). Sans aucun doute, il manque encore des modèles généraux, des infrastructures complètes et, enfin et surtout, des emplois pour le personnel qualifié.

Une menace découle de la question qui se pose pour toutes les professions en général, à savoir si une grande partie du travail ne peut pas être effectuée dans un avenir prévisible par les machines qui exécutent elles-mêmes les processus des scientifiques de données.

Cependant, l'intelligence artificielle est également connue depuis des décennies de ses promesses qui souvent ne se sont jamais converties en réalité tangible. C'est pour cette raison que nous avons actuellement - et probablement pour les décennies à venir - besoin de spécialistes en Information documentaire qui savent traiter et manipuler les données. Ces profils seront décrits dans le prochain chapitre.

Nouveaux profils professionnels

À l'heure actuelle on peut dénombrer trois nouveaux profils professionnels à l'intérieur du monde des sciences de

l'information : les bibliothécaires de données, les archivistes de données ainsi que les curateurs de données. À l'extérieur de ce monde se trouvent, entre autres, les *Data Journalist*, les *Data Analyst*, les *Data Engineer*, les *Personal Data Broker*, etc. Il existe également des profils qui se trouvent entre ces deux mondes tels que les *Data Steward* et les *Data Protection Officer*, rôle régi au sein du règlement général sur la protection des données (RGPD).

Sur la base des explications données ci-dessus, il est presque évident qu'au sein des sciences de l'information avec les professions classiques de bibliothécaire, archiviste et documentaliste, il y aura de plus en plus de bibliothécaires de données et d'archivistes de données. Outre les professions classiques, auxquelles seul le préfixe «Data» a été ajouté en anglais, il existe également de nouvelles professions qui doivent fonctionner dans l'environnement des bibliothèques et des archives, en particulier celle du curateur de données, mais aussi celle de *Data Steward*, de *Data Supporter* ou de façon plus générale de *Data Manager*.

Afin d'appréhender au mieux cette problématique, les profils professionnels importants gravitant autour des sciences de l'information sont décrits ci-dessous :

Bibliothécaire de données (Data Librarian)

Les bibliothécaires de données sont particulièrement actifs dans la gestion des données de recherche. Contrairement à l'archivage des données, la gestion (des données de recherche) commence beaucoup plus tôt et comporte au moins deux phases (la planification et la mise en œuvre) avant l'archivage à long terme. Les bibliothécaires de données, qui travaillent principalement dans des bibliothèques scientifiques, sont particulièrement actifs dans ces phases.

Le concept artistique de la «datathèque», qui existe peu ou prou, peut cependant être utile pour tracer une ligne avec les archives à long terme. Les bibliothécaires de données sont généralement plus proches des activités de ceux qui produisent les données et les aident, par exemple, à anticiper la création des données dès la phase de planification des projets et à préparer le terrain pour une publication réussie à une date ultérieure (généralement après la fin du projet).

En règle générale, les bibliothécaires de données ont également une grande affinité pour le libre accès ou travaillent déjà dans ce domaine et appliquent leur pratique professionnelle avant tout à la mise à disposition et à la gestion de dépôts de données.

Data Steward

Le Data Steward est soit un bibliothécaire soit un scientifique qui a de très bonnes connaissances sur la gestion des données de la recherche. Cette personne est intégrée au sein d'une équipe, en quelque sorte un bibliothécaire embarqué, qui va délivrer les données au curateur de données.

Curateur de données (Data Curator)

Le seul titre professionnel apparemment nouveau est celui de curateur de données. Les curateurs travaillent habituellement dans des musées et s'occupent principalement de la préparation et de l'organisation d'expositions. Ils sélectionnent les objets appropriés dans la collection du musée et s'occupent également du traitement et de la conservation des objets. Toutes ces activités peuvent être transférées de manière transparente aux données (d'une archive ou d'un référentiel). Les curateurs de données revendiquent donc aussi à tout moment - de la création à la réutilisation, en passant par la suppression ou la destruction des ensembles de données.

Seul le profil professionnel du conservateur de données semble avoir le po-

tentiel de reprendre des tâches d'autres domaines, de fusionner avec de nouveaux domaines de responsabilité et de devenir actif dans un large éventail d'institutions.

Archiviste de données (Data Archivist)

Dans le cas de l'archiviste de données, une description de poste classique avec ses domaines d'activité traditionnels, en particulier le stockage à long terme de documents, est transférée dans le domaine des données. Comme les archivistes ont déjà commencé il y a quelque temps à s'occuper des documents numérisés, de documents dits nés numériques ou encore de l'archivage du Web, l'extension au domaine des données de recherche est parfaitement intégrée dans un développement déjà existant

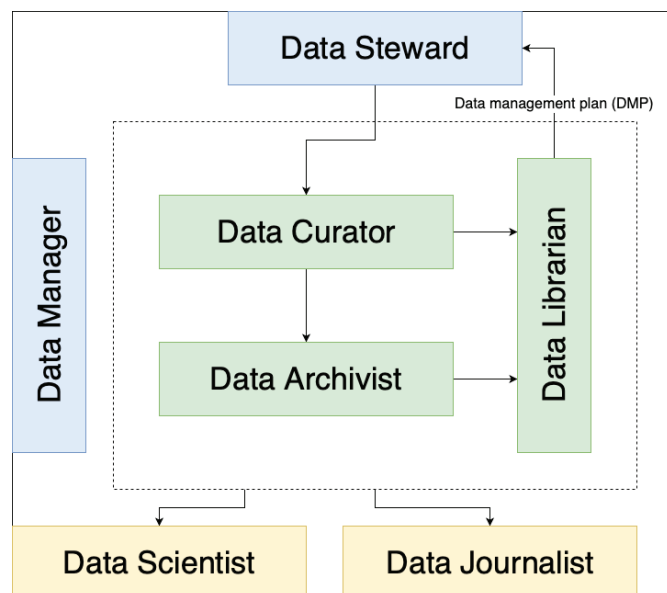
Scientifique de données (Data Scientist)

Le scientifique de données est une personne chargée de recueillir, d'analyser et d'interpréter de grandes quantités de données. Il va s'intéresser aux données déjà présentes, généralement celles reçues, nettoyées, indexées et valorisées par les archivistes de données et les curateurs de données.

Journaliste de données (Data Journalist)

Les journalistes de données puisent et analysent de grandes quantités de données statistiques, et par le biais de la programmation informatique et/ou d'outils de visualisation, mettent en lumière des faits avec leurs compétences de journaliste.





Conclusion

En conclusion, on peut dire que le développement technique continu et l'importance croissante des données et leur importance croissante dans les sciences conduiront les professions classiques de l'information à se tourner de plus en plus vers les données et à y appliquer leurs compétences fondamentales dans le domaine des métadonnées.

Les clients de ces services comprendront à leur tour des professions qui ont la désignation «Data» devant leurs dési-

gnations traditionnelles, mais il se peut aussi qu'un jour des machines accèdent aux données traitées et aux enregistrements de métadonnées via les interfaces correspondantes.

Cela implique également que les institutions de formation doivent se préparer à ce changement de paradigme et développer ainsi les compétences des futurs professionnels en sciences de l'information dans le traitement de données à petite et large échelle. ◆