Catherine Laplante

Les données de recherche : comment évaluer pour mieux conserver?

Table des matières

Introduction

Ce travail s'intéresse aux données de recherche, selon la perspective archivistique de l'évaluation. La forte croissance de la production de données de recherche, qu'on pourrait qualifier de « déluge de données » (Guindon, 2013, p. 189), a rendu d'autant plus nécessaire d'assurer leur saine gestion. Associé à la problématique de justifier le stockage de cette quantité exponentielle se trouve le besoin de sélectionner quelles données doivent être conservées, dans l'optique de leur potentiel de réutilisation et, par extension, du partage des connaissances. À cette fin, les organismes subventionnaires exigent de plus en plus des chercheurs de fournir un plan de gestion des données qui doit être pensé en amont du projet de recherche (Portage, s. d.-a). Ce travail de planification dès la conception n'est pas sans rappeler qu'il s'agit d'un principe inhérent à l'archivistique (Da Sylva, 2017). Toutefois, les interventions des archivistes en matière de gestion des données de recherche représentent actuellement plus l'exception que la règle. Leur expertise en évaluation et en préservation

justifierait pourtant leur contribution auprès des chercheurs. De surcroît, les archivistes sont déjà reconnus pour leur rôle dans la gestion du matériel né numérique (Dooley, 2015).

L'objectif de ce travail est de faire une synthèse des lignes directrices en matière d'évaluation des données de recherche afin de déterminer les pratiques adéquates à mettre en place pour leur conservation. Conjointement à cette démarche, nous aborderons le point de vue archivistique sur cette question. À partir d'une revue de la littérature touchant aux aspects de l'évaluation et de la conservation, des lieux qui s'affairent à ce qui est appelé la « curation » des données ont été identifiés, tant au Canada qu'à l'international. Les mesures d'évaluation mises en place par ces organisations seront examinées et fourniront des pistes de réponses intéressantes à analyser.

Dans un premier temps, nous définirons le concept de données de recherche ainsi que le cycle de vie auquel elles sont soumises. Dans un deuxième temps, le but et les enjeux de la conservation des données seront exposés, puisque ceux-ci ont nécessairement une incidence sur les décisions prises au moment de l'évaluation. Dans un troisième temps, nous présenterons des exemples d'application de la fonction d'évaluation aux données de recherche, plus exactement des critères issus de notre recension des écrits. Ce cheminement aboutira finalement à une synthèse des pratiques d'évaluation à appliquer pour la conservation adéquate des données de recherche.

Les données de recherche : quelques définitions

Pour instaurer des mesures efficaces en matière d'évaluation et de préservation des données de recherche, il importe d'abord de bien connaître les caractéristiques de cet objet numérique. Afin de circonscrire le sujet des données de recherche dont il est ici question, des définitions puisées dans différentes sources seront présentées.

La définition qui semble faire consensus, peut-être en raison de son apparente complétude, est celle du *CASRAI Dictionary*, reprise par le réseau Portage (s. d.-b) de même que par Johnston (2017). Les données de recherche y sont définies selon quatre aspects :

- Des sources primaires qui soutiennent des projets de recherche, des études académiques ou des travaux artistiques.
- Elles peuvent être utilisées comme preuve pour valider des résultats.
- Elles peuvent prendre la forme de données expérimentales, de données d'observation, de données opérationnelles, de données de tiers, de données du secteur public, de données de suivi, de données traitées ou de données réutilisées.
- Tout autre contenu numérique et non numérique a le potentiel de devenir des données de recherche. (Portage, s. d.-b)

Da Sylva (2017) définit quant à elle les données de recherche comme :

> [des] données générées à l'intérieur d'un projet de recherche, en milieu académique, gouvernemental ou industriel : par exemple, des observations sur le terrain, des réponses à des sondages ou questionnaires, des données créées par des processus de simulation par ordinateur, etc. (Da Sylva, 2017, p. 7)

La définition proposée par Guindon provient du Conseil de recherches en sciences humaines (CRSH) :

> Ces données comprennent des ensembles quantitatifs de données sociales, politiques et économiques, des renseignements qualitatifs sous forme numérique, des données de recherche expérimentale, des bases de données d'images et de sons fixes et mobiles, ainsi que d'autres objets numériques utilisés à des fins d'examen analytique. (Conseil de recherches en sciences humaines, cité dans Guindon, 2013, p. 190)

En complément de cette définition, Guindon souligne à juste titre que les données de recherche doivent être distinguées du matériel dit de publication, c'est-à-dire « les produits de la communication scientifique – articles,

monographies, thèses ou comptes rendus de conférence » (Guindon, 2013, p. 190).

Dès lors, des points communs entre ces quelques définitions ressortent. On peut résumer que les données de recherche, générées sous une variété de formes, sont souvent définies par les nombreux exemples qui les composent. De plus, elles sont issues de sources diverses et couvrent des disciplines de toutes sortes. En plus de cette hétérogénéité marquée, les données se caractérisent souvent par leur abondance. Enfin, soulignons le côté brut des données, à savoir que certaines n'auront pas encore été traitées, suivant, entre autres, l'étape de leur cycle de vie.

Modèles conceptuels du cycle de vie des données de recherche

Les données de recherche désormais mieux définies, il convient à présent d'exposer quelques modèles représentant leur cycle de vie. S'attarder à l'évolution et aux multiples transformations subies par les données au cours de ce cycle de vie permettra de cibler les moments où les archivistes sont susceptibles d'intervenir pour mettre en place des outils de gestion ou, plus spécifiquement, d'évaluation.

Le cycle de vie des données de recherche diffusé par Portage (s. d.-b) rappelle certaines fonctions archivistiques, telles que la création, la diffusion et la conservation (Figure 1). Nous pouvons supposer que la fonction d'évaluation pourrait s'insérer dans l'étape du traitement, ou même déjà initialement au moment de la planification :

Figure 1 : Le modèle diffusé par Portage (s. d.-b)

Le modèle proposé par le Digital Curation Centre (DDC) accorde une place plus évidente à l'étape de l'évaluation et de la sélection, en soulignant du même coup la disposition de certaines données (Figure 2). Cette dernière peut se traduire par un transfert des données dans un autre centre mieux adapté à leur conservation, ou, au contraire, par leur destruction.

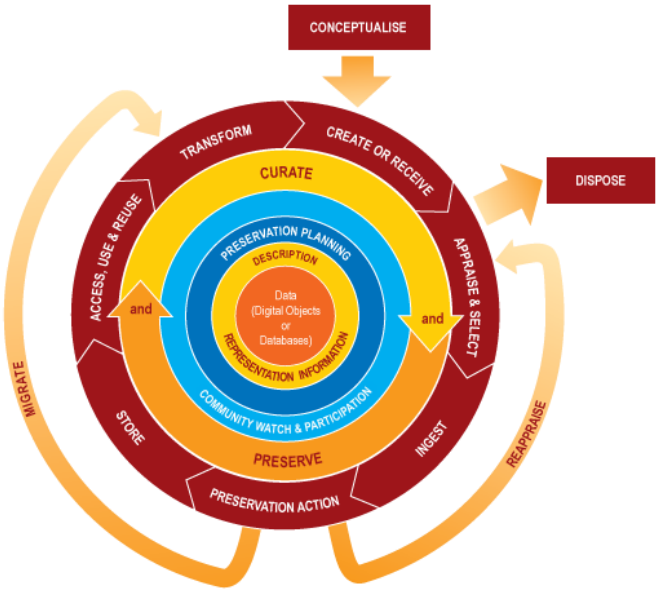

Figure 2 : DCC Curation Lifecycle Model (s. d.)

Dans le modèle du DDC, il est intéressant de mentionner la présence de la « réévaluation », une action occasionnelle qui s'exécute lorsque les données échouent les procédures de validation et doivent être évaluées de nouveau (DCC, s. d.).

Au demeurant, l'ère numérique a fait éclore de nouvelles visions en archivistique qui remettent en question les modèles plus linéaires. Les frontières devenant floues entre les étapes du cycle de vie traditionnel, une approche dite intégrée des archives a émergé : le records continuum. Par rapport au cycle de vie des données de recherche, Rombouts et Tjalsma (2010) font référence au records continuum pour poser la question : « À quel moment doivent être prises les décisions de sélection dans ce processus en continu? » À cela, les auteurs répondent que l'option préférable est de prendre les décisions de sélection à la source, c'est-à-dire au moment de la création.

Enfin, nous aurions pu aussi aborder plus en profondeur le modèle OAIS, bien connu du milieu de la gestion documentaire numérique. Ce modèle de haut niveau est en effet couramment appliqué pour la préservation à long terme des données de recherche, notamment en ce qui a trait aux dépôts numériques (Guindon, 2013; Carlson, 2014).

Pourquoi conserver les données de recherche?

La préservation des données de recherche ne saurait être abordée sans définir le concept de « curation » qui lui est inhérent, à savoir : « maintaining and adding value to a trusted body of digital information for current and future use » (Digital Curation Centre, cité dans Poole, 2016, p. 962). En d'autres mots, la curation a pour objectif de mettre à la disposition des chercheurs, autres que le créateur, les données de recherche; les actions posées pour en accroître la valeur consistent ce faisant à les rendre plus accessibles en leur appliquant, par exemple, des métadonnées (Carlson, 2014). La réutilisation constitue donc le but ultime de la préservation à long terme des données (Johnston, 2017). D'un point de vue archivistique, nous pouvons considérer que la curation met ainsi l'accent sur la valeur secondaire des données de recherche, en raison d'une utilité scientifique future.

Il est important de comprendre pourquoi les données de recherche doivent être conservées, puisque cela a une incidence sur la fonction d'évaluation. La

mission de l'organisation chargée de la conservation des données y est aussi intimement liée. Cet aspect sera abordé plus en profondeur ultérieurement.

Les définitions et modèles précédents ont permis de mettre en relief la finalité de l'accès et de la réutilisation des données. À cela s'ajoute l'importance de la reproductibilité, ainsi que le renchérit le rapport de Beagrie qui fait état de deux raisons prédominantes pour conserver les données de recherche :
« Research Integrity and Reproducibility (availability of the data supporting the findings in research); and the Potential for Reuse (availability of data for sharing with other users) » (Beagrie, 2019, p. 4).

Essentiellement, la conservation des données se fait donc en vue du partage (Childs, McLeod, Lomas et Cook, 2014). Cela s'insère dans une vision globale d'enrichissement de la recherche à plus large échelle, au bénéfice de toute la communauté scientifique. Les principales raisons de conservation invoquées par Portage abondent dans ce sens :

- Accélère le progrès scientifique
- Améliore l'efficacité de la recherche
- Accroît la visibilité et les retombées de la recherche
- Permet de respecter les politiques des organismes de financement
- Permet d'éviter la répétition des travaux de recherche
- Garantit l'intégrité de la recherche et la validation des résultats (Portage, s. d.-a)

Malgré la légitimité de ces motifs, les chercheurs semblent démontrer une certaine réticence à rendre disponibles leurs données de recherche (Poole, 2016). La concurrence entre les chercheurs, induite par la quête du financement et de la publication, explique en partie cette prise de position. D'autres enjeux, qui seront examinés dans une prochaine section, ont aussi pour effet de dissuader les chercheurs à conserver et à partager leurs données de recherche. Cela dit, le manque d'intérêt de la part des chercheurs pourrait s'expliquer par le fait qu'ils ne se sentent tout simplement pas concernés par la préservation des données qu'ils ont créées, puisqu'ils n'en ont plus besoin une fois leur projet de recherche terminé (Doorn et Tjalsma, 2007). En contrepartie, les organisations responsables de la préservation à long terme, elles, exercent rarement une influence au moment de la création des données. Cette dichotomie explique probablement les nombreuses initiatives émergentes pour accompagner les chercheurs dans la gestion de leurs données de recherche.

Combien de temps conserver?

La réflexion précédente conduit tout naturellement à s'interroger sur la durée de cette conservation dite « à long terme ». Est-ce que ce qualificatif signifie d'emblée « à perpétuité »? Selon le Digital Curation Center (2014), le terme veut plutôt dire « au-delà du projet de recherche ». L'organisme propose ainsi de réévaluer l'intérêt de conserver des collections de données de manière périodique.

Rappelons que fondamentalement en archivistique, l'évaluation consiste à fixer des périodes de temps pendant lesquelles les valeurs primaires et secondaires s'appliquent (Couture, 1996-1997). Cette composante est toutefois à peine effleurée lorsqu'il est question des données de recherche.

Dennie et Guindon (2017) ont demandé directement l'avis des principaux intéressés, à savoir les créateurs des données. Les résultats de leur enquête sur les attitudes des chercheurs en matière de gestion des données de recherche ont dévoilé leur opinion sur la question : « Combien de temps vos données devraient-elles être préservées? ». Selon les réponses, 41% des chercheurs estiment que les données devraient être conservées indéfiniment (Dennie et Guindon, 2017, p. 67-68).

Néanmoins, selon les experts des données, fixer une durée de conservation n'est pas chose aisée et cette durée variera selon les domaines de recherche (Rombouts et Tjalsma, 2010). Certaines bibliothèques de recherche proposent de manière un peu arbitraire une conservation de plus de 100 ans (Kimpton et Minton Morris, 2014). Parfois, le cas de figure s'avère simple à régler lorsque des obligations légales dictent la durée de conservation. Chose certaine, les coûts de conservation ont une incidence sur les décisions prises à ce niveau. En effet, la conservation permanente des données implique des investissements importants, tant dans le stockage que dans les mesures déployées pour assurer la lisibilité des formats dans le futur pour compenser l'évolution technologique (Cox et Verbaan, 2018). Or il est très difficile d'évaluer ces coûts à long terme. L'amenuisement des ressources financières au fil du temps conjointement à la nécessité de faire de la place aux nouveaux jeux de données viendront compromettre les promesses d'une conservation permanente (Rombouts et Tjalsma, 2010).

Compte tenu de la difficulté à établir la durée de conservation des données de recherche, le savoir-faire des archivistes en matière d'évaluation s'avérerait fort utile. Le groupe de travail Aurore sur les archives de la recherche de l'Association des archivistes français (AAF) s'est d'ailleurs penché sur la question en produisant un référentiel de gestion des archives scientifiques, dont une part est consacrée aux données (AAF, section Aurore, 2019). À l'image d'un calendrier de conservation, ce référentiel propose à titre indicatif une durée d'utilité administrative (DUA) débutant à la clôture du projet et au terme de laquelle un sort final est appliqué (conservation définitive et intégrale, tri ou destruction). Cette DUA, qui se rapporte davantage à un usage scientifique qu'à une réelle utilité administrative, a été fixée en tenant compte de différentes contraintes, notamment légales ou normatives. En ce qui concerne les données de recherche, la DUA et le sort final varient selon une typologie bien circonscrite (données brutes, travaillées, techniques, d'enregistrement, etc.). Suivant le cas, la DUA suggérée oscille donc entre 2 et 20 ans, ou est établie de façon non chiffrée (« durée de conservation autorisée par la CNIL » ou « selon les besoins de l'organisme »). Somme toute, cette ressource permet d'alimenter la réflexion au sujet de l'application d'outils archivistiques « traditionnels » dans un contexte numérique et, plus spécialement, à des fins de gestion des données. Surtout, il s'agit d'un exemple concret de l'expertise des archivistes pour établir des règles de conservation définies en fonction du cycle de vie des données.

Quelques enjeux de la conservation des données de recherche

Cette section a pour but de présenter quelques enjeux et possibles limitations à la conservation des données de recherche. Comme vu précédemment, la conservation des données est mise en place dans le but de pouvoir y accéder dans le futur. Les obstacles à cet accès ou à la conservation elle-même peuvent donc avoir des répercussions sur les critères d'évaluation. Quatre enjeux seront abordés : la qualité des données, la confidentialité, la propriété intellectuelle et les coûts.

La qualité des données

La qualité des données touche aux principes d'authenticité et d'intégrité, fondamentaux en archivistique. Il s'agit effectivement d'un enjeu important pour les données puisqu'elles auront subi un traitement au cours de leur migration (Ray, 2014). Quel est le degré acceptable de changement? Pour remédier aux altérations pouvant invalider l'utilisation des données, une documentation permettant de « rétablir, au besoin, l'information correcte » s'avère essentielle

(Da Sylva, 2017, p. 12). En d'autres termes, il est primordial de documenter la chaîne de traçabilité (Dooley, 2015).

La confidentialité

Le respect de la confidentialité s'insère dans la catégorie des aspects éthiques et légaux. Le partage des données est en effet compromis par l'obligation d'assurer la confidentialité des informations sensibles (Council on Library and Information Resources, 2013). Les données générées par les projets de recherche impliquant des sujets humains, comme en sciences sociales ou médicales, sont particulièrement concernées par cet enjeu. Lorsqu'elle est possible, l'anonymisation des données fait partie des solutions pouvant être appliquées (Da Sylva, 2017).

La propriété intellectuelle

Le partage des données de recherche vient toucher à une autre problématique légale et éthique : la propriété intellectuelle. En effet, « le chercheur est en droit de se voir attribuer la paternité des données » (Da Sylva, 2017, p. 8). Est-ce que le créateur des données acceptera d'offrir le fruit de son labeur pour que d'autres puissent l'exploiter? Certaines licences pourraient être accordées pour encadrer la réutilisation (Da Sylva, 2017). Il est à mentionner que la propriété intellectuelle diffère de la propriété physique des données (Dooley, 2015). Par exemple, les institutions qui acquièrent les données de chercheurs devront tout de même mettre en place des mesures pour que les accès respectent les droits d'auteurs.

Les coûts

Les coûts ont été évoqués auparavant par rapport à leur impact sur la durée de conservation. Ils sont ainsi à considérer au moment même de choisir quelles données conserver. Par extension, les organismes subventionnaires, qui financent en quelque sorte la création des données, exercent un certain ascendant sur les décisions qui seront prises : « Funding agencies play a very important role, as they provide the investments necessary for creating the data and thus are in a position to influence the long-term life of the data » (Doorn et Tjalsma, 2007, p. 9). Étant donné l'incidence des coûts sur la curation des données, des modèles conceptuels sont même utilisés pour estimer les coûts de préservation en contrepartie des « clients » potentiels (Burgi, Blumer et Makhlouf Shabou, 2017).

Exemples de critères d'évaluation

La mise en contexte précédente nous amène désormais à présenter des exemples de critères d'évaluation relevés dans la littérature. Pour un rappel, l'évaluation est l'acte de juger de la valeur d'un objet (Couture, 1996-1997). Ces critères constituent donc un guide pour attribuer une valeur aux données de recherche afin de déterminer leur conservation ou leur élimination, selon l'angle choisi.

Évaluer : un processus décisionnel en cinq étapes

Afin de faciliter la sélection des données à conserver, le Digital Curation Center (2014) propose un guide pour que les chercheurs abordent l'évaluation de manière séquentielle. La première étape vise à établir les buts que la réutilisation des données permettrait d'atteindre. Par le fait même, cette démarche fait prendre conscience aux chercheurs de la valeur que représentent leurs données pour l'avancement des connaissances à plus grande échelle. L'importance de l'intégrité de la recherche et le potentiel de publication d'articles sur les données elles-mêmes sont aussi soulignés. La deuxième étape

met l'accent sur les obligations de garder certaines données en considérant les deux extrêmes du spectre légal : les données qui doivent être rendues publiques ou celles qui, à l'opposé, présentent des conditions d'accès restreintes. C'est à cette étape que les politiques de diffusion et les enjeux liés au respect des renseignements personnels doivent être examinés. Troisièmement, la valeur à long terme des données doit être soupesée selon différentes perspectives. Est-ce que les données sont suffisamment documentées et de qualité? Peut-on entrevoir une demande potentielle pour les données? S'agit-il de l'unique exemplaire du jeu de données? Évaluer la valeur à long terme représente un grand défi puisque, tout comme pour les archives, il est difficile de prévoir quels seront les besoins de la recherche dans le futur. La quatrième étape consiste à évaluer les coûts de conservation de manière réaliste. Enfin, la cinquième et dernière étape fait état des constats et décisions résultant du cheminement suivi jusque-là, en concrétisant le tout dans un tableau synthèse.

Sept critères généraux à inclure dans une politique d'évaluation

Un autre guide, cette fois élaboré par Whyte et Wilson (2010), met davantage en lumière l'importance de la politique d'évaluation pour gérer adéquatement les données de recherche. Selon les auteurs, sept critères généraux doivent y être intégrés. Le premier mentionné est la pertinence du lien entre l'acquisition des données de recherche et la mission du centre d'archivage ou du bailleur de fonds. À cette fin, la mission de l'organisme doit être examinée en profondeur, notamment ses priorités stratégiques, la cohérence avec le développement des collections (mentionnée aussi par Johnston, 2017), ses codes de conduite par rapport à la recherche, etc. Le deuxième critère se rapporte à la valeur scientifique et historique des données, qui rappelle le concept de valeur secondaire propre à l'archivistique. Le troisième concerne le caractère unique des données et, par extension, le risque de perte des copies existantes. Le quatrième renvoie au potentiel de réutilisation, en y incluant notamment les enjeux légaux et technologiques. Le cinquième critère porte sur le caractère de non-reproductibilité des données, c'est-à-dire la difficulté de recréer les données en raison de conditions naturelles non répétables ou de coûts trop importants. Le sixième fait référence aux considérations financières. Enfin, le septième critère met l'accent sur l'importance d'une documentation complète pour accompagner les données conservées.

Les données optimales à conserver

Dans son enquête sur le terrain, Beagrie (2019) brosse un portrait exhaustif de la conservation des données de recherche. Nous mettrons donc seulement en exergue certains points, tirés essentiellement de son tableau synthèse « Optimal data to keep » (Beagrie, 2019, p. 36). Le premier point relève l'importance d'une adéquation avec la mission. Néanmoins, dans la compilation des résultats recueillis, il semble s'agir d'une catégorie un peu fourre-tout. On y trouve les exigences des organismes subventionnaires et des éditeurs, les obligations légales, le potentiel de réutilisation, les qualités d'unicité, etc. Les deuxième et troisième points sont plus inédits par rapport à ce qui a été mentionné dans notre analyse jusqu'à maintenant. Ils se rapportent à l'essence même des données, soit leur niveau (données brutes, données traitées, etc.) ainsi que le type de source dont elles sont issues (données d'observation, expérimentales, secondaires, négatives). L'importance de la documentation des données est aussi soulignée par le rapport, de même que les normes à respecter propres à chaque discipline.

Évaluer selon un ordre de priorités

La politique d'évaluation de l'Inter-university Consortium for Political and Social Research (ICPSR, s. d.-a-b) est basée sur des critères aidant à déterminer

l'ordre de priorité d'acquisition des données de recherche. Dans un premier temps, l'organisation repère les données intéressantes à conserver selon des critères généraux : les données doivent notamment soutenir sa mission et présenter une valeur fondamentale pour l'avancement de la recherche ou de l'enseignement dans le domaine des sciences sociales. Une attention est aussi accordée au type de données (données sur la diversité, données non standards, données interdisciplinaires ou internationales, etc.). Une fois que les jeux de données présentant un intérêt ont été identifiés, l'organisation évalue la priorité d'acquisition en fonction d'une approche coût-avantage : « If there are one or more concerns reducing the priority-level of a data resource, ICPSR considers the potential benefits and costs associated with acquiring the data and acquires, in the short-term, only what it has the capacity to accept. » (ICPSR, s. d.-a). Les critères de priorité semblent converger vers une certaine promotion des données ouvertes : données appartenant au domaine public, possibilité de réutilisation, considération des droits d'auteurs, lisibilité du format, justification des restrictions d'usage, etc. Enfin, la dimension des coûts est aussi largement considérée au cours de l'analyse.

Les données à ne pas conserver

L'optique d'évaluation adoptée par l'enquête de Kung et Campbell (2016) est axée sur l'élimination plutôt que la conservation. La formulation des critères sera donc nécessairement caractérisée par une tournure plus négative. Voici les huit critères selon lesquels les données ne devraient pas être conservées :

> […] sensitive or confidential; proprietary; easily replicable; do not have good metadata; test, pilot, or intermediate data; bad or junk data; data that cannot be used by others for a variety of reasons; and older data that are not used and have no obvious cultural or historical value. (Kung et Campbell, 2016, p. 53)

Mentionnons aussi que l'enquête est née du besoin de définir des critères plus spécifiques au domaine de la santé. Est-ce parce qu'il est reconnu que ce domaine est plus sujet à l'élimination des données, en raison d'une production abondante de données et de restrictions légales sur la confidentialité? Somme toute, un recoupement avec les critères déjà identifiés auparavant demeure possible. Ce constat nous laisse croire que peu importe l'angle de l'évaluation (élimination ou conservation), celui-ci a peu d'incidence au bout du compte sur les données sélectionnées pour une préservation à long terme.

Synthèse des bonnes pratiques à mettre en place

Il s'agit de l'aboutissement des sections précédentes : cette récapitulation sera essentiellement composée d'un tableau synthèse illustrant les critères d'évaluation présentés précédemment. Des pratiques complémentaires seront aussi mises de l'avant pour garantir une curation efficace des données de recherche.

Tableau 1. Synthèse des critères d'évaluation selon certains auteurs

<table>
<tr>
<td>DDC (2014) :
- Reuse purposes
- Legal or policy compliance risks
- Long-term value
- Weigh up the costs
- Complete your data appraisal</td>
<td>Whyte et Wilson (2010) :
- Relevance to Mission
- Scientific or Historical Value
- Uniqueness
- Potential for Redistribution
- Non-Replicability
- Economic Case
- Full Documentation</td>
</tr>
<tr>
<td>Beagrie (2019) :
- Relevant to mission
- Data Level
- Data Source Type
- Supplementary materials
- Cost
- Curation Level
- Disciplinary norms</td>
<td>ICPSR (s. d.) :
- Substantive value for research and/or instruction
- Support its mission
- Data Area
- Data Availability
- Security, Privacy, and Confidentiality Considerations
- Copyright and other Legal Issues
- Data Quality
- Data Format
- Financial Considerations</td>
</tr>
<tr>
<td colspan="2">Kung et Campbell (2016) :
- Bad or junk data
- Data that cannot be used by others
- Data that are easily replicable
- Data without good metadata
- Older data that are not used and have no obvious cultural or historical value
- Pilot, test, or intermediate data
- Proprietary data
- Sensitive or confidential data</td>
</tr>
</table>

Formuler les critères d'évaluation en fonction de la mission

La démarche précédente a mis en évidence l'importance qu'exerce la mission dans l'établissement des critères d'évaluation. Rappelons que ce principe est capital en archivistique. Voici une mise en relief de la diversité des missions selon des intervenants impliqués dans la gestion des données de recherche :

- Le chercheur lui-même sélectionne ses données en fonction des besoins et de l'évolution de son projet de recherche.
- Les organismes subventionnaires exigent des plans de gestion de données pour que le financement qu'ils accordent soit utilisé à bon escient, ainsi que pour maximiser le retour sur l'investissement (Digital Curation Center, 2014; Portage, s. d.-a).
- Les bibliothèques des institutions académiques, conformément à leur mission d'enseignement et de recherche, ont pour mission d'« appuyer la diffusion et le rayonnement des savoirs produits par les membres de la communauté » et de leur offrir des services spécialisés pour soutenir leurs activités (Les Bibliothèques de l'Université de Montréal, s. d.).
- Les réseaux de soutien accompagnent les chercheurs dès le début de leur projet de recherche (ex. : Portage) dans le but d'assurer la bonne gestion de leurs données et d'en garantir le partage par la suite.

- Les institutions recueillent dans leurs dépôts numériques les données de recherche à la fin du projet de recherche pour le développement des connaissances dans leur domaine (ex. : ICPSR).
- Les communautés de recherche nationales et internationales considèrent les données de recherche comme un bien public et prônent un accès ouvert aux données de recherche (ex. : Données de recherche Canada, 2019).

Visiblement, la variabilité de la mission des organismes au sein desquels les archivistes peuvent être amenés à gérer des données de recherche rend impossible l'établissement de critères d'évaluation universels. Par conséquent, il est essentiel que les archivistes adaptent leur pratique de manière à s'arrimer à la mission de l'organisme qu'ils servent. De surcroît, le choix de conserver ou non des données de recherche doit être cohérent avec cette mission et assurer une complémentarité avec les collections déjà archivées.

Les points précédents suggèrent également que les interventions en matière d'évaluation doivent s'effectuer à différents moments du projet de recherche, en fonction de la mission de tout un chacun. Enfin, soulignons l'importance de définir clairement les rôles de chaque intervenant lorsque la collaboration est requise pour tendre vers un même but et, ainsi, éviter les silos.

Collaborer avec le chercheur dès la création des données

La totalité de nos lectures souligne l'importance de la collaboration entre tous les individus et organismes concernés par la gestion des données de recherche. McGovern (2018) y consacre d'ailleurs l'entièreté de son article, en mettant l'accent sur la contribution des archivistes dans les bibliothèques universitaires.

En outre, établir une bonne communication avec le créateur des données est primordial pour garantir le succès de la préservation (Grant, 2017). Les chercheurs doivent en effet être impliqués dans le processus décisionnel en lien avec l'évaluation des données (Kung et Campbell, 2016).

Adapter les critères en fonction du type de données de recherche et des disciplines

Force est de constater que les pratiques à mettre en place doivent être adaptées selon le milieu, les politiques à respecter et les particularités des données. En effet, selon Beagrie (2019, p. 4) : « Not all research data is the same: it is highly varied in terms of data level; data type; and origin. In addition, not all disciplines are in the same place or have identical needs. » Ce constat est partagé par Poole (2015, p. 115) : « Appraisal and selection criteria should map to specific research communities ». Rombouts et Tjalsma (2010) soulignent eux aussi qu'il existe de trop grandes disparités entre les disciplines pour déterminer des critères de sélection spécifiques. Les auteurs suggèrent de dresser des lignes directrices générales selon trois principales raisons de conserver les données : la réutilisation, la vérification et le patrimoine (« reuse, verification and heritage ») (Rombouts et Tjalsma, 2010, p. 13).

Tenir compte des coûts de conservation à long terme

La problématique des coûts de conservation a été récurrente lors de notre analyse des critères d'évaluation. Cela peut notamment s'expliquer par la grande quantité d'acteurs impliqués dans la préservation à long terme des données de recherche : les chercheurs, les organismes subventionnaires, les institutions académiques, les organismes chargés de stocker les données dans leurs dépôts numériques… tout autant de ressources humaines demandant d'être rémunérées. C'est sans compter les ressources matérielles exigeant elles aussi un financement. De plus, l'environnement numérique impose un traitement et une maintenance périodique des données, ce qui fait grimper les coûts (Whyte et

Wilson, 2010). À cela s'ajoute le contexte de la recherche qui doit composer de façon générale avec des restrictions budgétaires de plus en plus importantes.

Documenter exhaustivement

La nécessité de la documentation des données de recherche touche deux dimensions liées au potentiel de réutilisation : les métadonnées et le contexte de création. Les premières sont essentielles pour un repérage efficace des données et l'encadrement de leur utilisation (Da Sylva, 2017). En ce qui concerne le contexte de création, la documentation est requise pour garantir l'authenticité les données, mais aussi pour retracer tout changement qui aurait pu les altérer et compromettre leur intégrité (Ray, 2014).

Bilan

Cette synthèse peut être résumée en trois principales lignes directrices à appliquer lors de l'évaluation des données de recherche :

1. Adapter les critères d'évaluation en fonction de la mission de l'organisme et du type de données à conserver.
2. Collaborer avec le chercheur dès la création des données afin d'assurer une gestion efficace des données tout au long du processus et une documentation complète et conforme.
3. Élaborer les critères d'évaluation en tenant compte des contraintes de la conservation à long terme, particulièrement celles d'ordre pécuniaire.

Conclusion

Ce travail a mis en lumière la complexité de l'évaluation des données de recherche. Cela s'explique notamment par la difficulté de définir les données de recherche elles-mêmes, puisqu'elles sont multiformes et hétérogènes. Les modèles du cycle de vie des données de recherche ont néanmoins permis de cibler le moment optimal pour effectuer l'évaluation, soit lors de la création. Porter attention aux raisons de la conservation a contribué à déterminer que les critères doivent s'articuler dans l'objectif de partager et de réutiliser les données de recherche. Certaines contraintes peuvent néanmoins venir teinter la formulation de ces critères d'évaluation. Cette mise en contexte établie, différents critères d'évaluation issus de la recension des écrits ont été présentés. Tout cela a abouti à une synthèse des bonnes pratiques en matière d'évaluation.

Cette ultime étape a concrétisé l'atteinte de notre objectif : dresser des lignes directrices en matière d'évaluation des données de recherche afin d'en assurer la conservation. Bien qu'il soit possible de mettre en place de telles mesures en les adaptant aux particularités de chaque type de données de recherche, le succès de la démarche ne pourrait être garanti sans une collaboration étroite avec le créateur des données. La contribution des archivistes, du fait de leur expertise en matière d'évaluation, est en ce sens souhaitable.

Dans une étude ultérieure, il serait intéressant de se pencher sur un examen plus approfondi des politiques d'acquisition des données de recherche selon le point de vue archivistique, afin de vérifier si l'essence de la mission du centre conservateur des données y est bien intégrée. Par analogie avec l'acquisition de fonds d'archives, cette politique devrait normalement aussi être en cohérence avec les collections de données déjà présentes dans les dépôts numériques, ou rechercher une complémentarité dans l'acquisition de nouveaux jeux de données. Il serait profitable de pousser la démarche encore plus loin et d'examiner dans quelle mesure les critères d'évaluation utilisés pour les documents d'archives peuvent être transposés aux données. La pertinence de l'implication des archivistes dans la gestion des données de recherche en serait d'autant plus démontrée.

Bibliographie

AAF (Association des archivistes français), section Aurore. (2019). *Référentiel de gestion des archives de la recherche*. Repéré à https://www.archivistes.org/IMG/pdf/referentiel_recherche_intro_septembre2012_corrige_.pdf

Beagrie, N. (2019). *What to Keep: A Jisc research data study*. Repéré à http://repository.jisc.ac.uk/id/eprint/7262

Burgi, P.-Y., Blumer, E. et Makhlouf Shabou, B. (2017). Research data management in Switzerland: National efforts to guarantee the sustainability of research outputs. *IFLA Journal, 43*(1), 5–21. https://doi.org/10.1177/0340035216678238

Carlson, J. (2014). The Use of Life Cycle Models in Developing and Supporting Data Services. Dans J. M. Ray (dir.), *Research data management: practical strategies for information professionals* (p. 63-86). West Lafayette, IN: Purdue University Press.

Childs, S., McLeod J., Lomas E. et Cook G. (2014). Opening research data: Issues and opportunities. *Records Management Journal*, *24*(2), 142-162. https://doi.org/10.1108/RMJ-01-2014-0005

Council on Library and Information Resources. (2013). *Research data management: principles, practices, and prospects*. Repéré à https://apo.org.au/node/39799

Couture, C. (1996-1997). L'évaluation des archives : état de la question. *Archives, 28*(1), 3-21. Repéré à http://www.archivistes.qc.ca/revuearchives/vol28_1/28-1-couture.pdf

Cox, A. et Verbaan, E. (2018). *Exploring research data management*. London, United Kingdom: Facet Publishing.

Da Sylva, L. (2017). Les données et leurs impacts théoriques et pratiques sur les professionnels de l'information. *Documentation et bibliothèques, 63*(4), 5–34. https://doi.org/10.7202/1042308ar

Dennie, D. et Guindon, A. (2017). Résultats d'une enquête sur les pratiques et attitudes des chercheurs de l'Université Concordia en matière de gestion des données de recherche. *Documentation et bibliothèques*, *63*(4), 59–72. https://doi.org/10.7202/1042311ar

DCC (Digital Curation Centre). (s. d.). DCC Curation Lifecycle Model. Repéré à http://www.dcc.ac.uk/resources/curation-lifecycle-model

DCC. (2014). *Five steps to decide what data to keep: DCC Checklist for Appraising Research Data*. Repéré à http://www.dcc.ac.uk/sites/default/files/documents/publications/Five%20Steps%20to%20decide%20what%20data%20to%20keep.pdf

Données de recherche Canada. (2019). Qui nous sommes. Repéré à https://www.rdc-drc.ca/fr/qui-nous-sommes/

Dooley, J. (2015). *The Archival Advantage: Integrating Archival Expertise into Management of Born-digital Library Materials*. Repéré à http://www.oclc.org/content/dam/research/publications/2015/oclcresearch-archivaladvantage-2015.pdf

Doorn, P. et Tjalsma H. (2007). Introduction: Archiving research data. *Archival Science*, *7*(1), 1-20. doi: 10.1007/s10502-007-9054-6

Grant, R. (2017). Recordkeeping and research data management: A review of perspectives. *Records Management Journal*, *27*(2), 159-174. https://doi.org/10.1108/RMJ-10-2016-0036

Guindon, A. (2013). La gestion des données de recherche en bibliothèque universitaire. *Documentation et bibliothèques*, *59*(4), 189–200. https://doi.org/10.7202/1019216ar

ICPSR (Inter-university Consortium for Political and Social Research). (s. d.-a). Details on Appraisal Critera. Repéré à https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/lifecycle/details.html

ICPSR. (s. d.-b). Selection and Appraisal. Repéré à https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/lifecycle/selection.html#criteria

Johnston, L. R. (2017). *Curating research data* (vol. 1 et 2). Chicago, IL: Association of College and Research Libraries.

Kimpton, M. et Minton Morris, C. (2014). Managing and Archiving Research Data: Local Repository and Cloud-Based Practices. Dans J. M. Ray (dir.), *Research data management: practical strategies for information professionals* (p. 223-238). West Lafayette, IN: Purdue University Press.

Kung, J. Y. C. et Campbell, S. (2016). What Not to Keep: Not All Data Have Future Research Value. *Journal of the Canadian Health Libraries Association (JCHLA), 37*(2), 53–57. https://doi.org/10.5596/c16-013

Les Bibliothèques de l'Université de Montréal. (s. d.). Mission, vision et engagements. Repéré à https://bib.umontreal.ca/les-bibliotheques-udem/mission-vision

McGovern, N. Y. (2018). Radical Collaboration: An Archival View. *Research Library Issues*, (296), 53–61. https://doi.org/10.29242/rli.296.8

Poole, A. H. (2015). How has your science data grown? Digital curation and the human factor: a critical literature review. *Archival Science, 15*(2), 101-139. https://doi.org/10.1007/s10502-014-9236-y

Poole, A. H. (2016). The conceptual landscape of digital curation. *Journal of Documentation, 72*(5), 961-986. https://doi.org/10.1108/JD-10-2015-0123

Portage. (s. d.-a). Foire aux questions. Repéré à https://portagenetwork.ca/fr/comment-gerer-vos-donnees/foire-aux-questions/

Portage. (s. d.-b). *Gestion des données de recherche : informations de base*. Repéré à https://portagenetwork.ca/wp-content/uploads/2017/06/Portage-Info-de-base-GDR.pdf

Ray, J. M. (dir.). (2014). *Research data management: practical strategies for information professionals*. West Lafayette, IN: Purdue University Press.

Rombouts, J. et Tjalsma, H. (2010). *Selection of Research Data. Guidelines for appraising and selecting research data. A report by DANS and 3TU.Datacentrum*. Repéré à http://resolver.tudelft.nl/uuid:dbab8a19-542a-4c4d-96b4-df8cc39333db

Whyte, A. et Wilson, A. (2010). *How to Appraise & Select Research Data for Curation*. Repéré à http://www.dcc.ac.uk/sites/default/files/documents/How%20to%20Appraise%20and%20Select%20Research%20Data.pdf