Document generated on 07/03/2025 9:44 p.m.

Documentation et bibliothèques

Résultats d'une enquête sur les pratiques et attitudes des chercheurs de l'Université Concordia en matière de gestion des données de recherche

The Practices and Attitudes of Researchers at Concordia University with Regard to Research Data Management: Results of an Investigation

Danielle Dennie and Alex Guindon

Volume 63, Number 4, October–December 2017

Les données et les sciences de l'information

URI: https://id.erudit.org/iderudit/1042311ar
DOI: https://doi.org/10.7202/1042311ar

See table of contents

Publisher(s)

Association pour l'avancement des sciences et des techniques de la documentation (ASTED)

ISSN

0315-2340 (print)
2291-8949 (digital)

Explore this journal

Cite this article

Dennie, D. & 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

Article abstract

A survey and a series of interviews of Concordia University (Montréal) professors were undertaken in 2015-2016 to better understand the needs and attitudes of Canadian researchers in the area of data management. The majority of researchers want to better manage, preserve and share their research data. They also recognise the merits of doing so. However, certain barriers prevent them from better managing their data, namely the lack of incentives, the shortage of human and technological resources, confidentiality issues and the ability to exercise a certain level of control on the use of the data by others.

Tous droits réservés © Association pour l'avancement des sciences et des techniques de la documentation (ASTED), 2017

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

RÉSULTATS D'UNE ENQUÊTE SUR LES PRATIQUES ET ATTITUDES DES CHERCHEURS DE L'UNIVERSITÉ CONCORDIA EN MATIÈRE DE GESTION DES DONNÉES DE RECHERCHE

Danielle Dennie

Bibliothécaire responsable des communications savantes et de la réserve électronique
Bibliothèque de l'Université Concordia

danielle.dennie@concordia.ca

Alex Guindon

Bibliothécaire responsable des services de données géospatiales et numériques
Bibliothèque de l'Université Concordia

alex.guindon@concordia.ca

RÉSUMÉ | ABSTRACT

Selon les résultats d'un sondage et d'une série d'entrevues réalisés auprès de professeurs à l'Université Concordia à Montréal en 2015-2016 pour comprendre les besoins et les attitudes des chercheurs canadiens en matière de gestion des données, la majorité des chercheurs souhaitent mieux gérer, préserver et partager leurs données de recherche et reconnaissent les avantages de ce faire. En revanche, certaines contraintes les en empêchent, tels le manque de mesures incitatives ou de ressources humaines et technologiques, les problèmes associés à la confidentialité ou encore la volonté de garder un certain niveau de contrôle sur l'utilisation des données par autrui.

The Practices and Attitudes of Researchers at Concordia University with Regard to Research Data Management : Results of an Investigation

A survey and a series of interviews of Concordia University (Montréal) professors were undertaken in 2015-2016 to better understand the needs and attitudes of Canadian researchers in the area of data management. The majority of researchers want to better manage, preserve and share their research data. They also recognise the merits of doing so. However, certain barriers prevent them from better managing their data, namely the lack of incentives, the shortage of human and technological resources, confidentiality issues and the ability to exercise a certain level of control on the use of the data by others.

Introduction

Depuis maintenant plusieurs années, la question de la gestion des données de recherche (GDR) a été l'objet de beaucoup de discussions dans le milieu universitaire. On a d'abord déploré le peu d'efforts consacrés à la description, à l'archivage et au partage de ces données. La célèbre revue *Nature* a même publié un éditorial à ce sujet en 2009 (Data 2009). Puis, les choses se sont mises à évoluer peu à peu. Des états de la question ont été publiés et des groupes et des associations ont été formés, par exemple le Research Data Alliance, CODATA ou Datacite. Par la suite, des centres d'expertise ont émergé tels que le Digital Curation Centre en Grande-Bretagne et le Australian National Data Service (ANDS) (aux États-Unis, l'Inter-University Consortium for Political and Social Research [ICPSR] existait déjà depuis plusieurs années). Finalement, les organismes de financement de la recherche ont commencé à exiger des actions de la part des chercheurs et des institutions universitaires afin d'assurer la pérennité et le partage des données issues de la recherche scientifique. En 2011, le Research Councils UK (RCUK) qui regroupe les principaux organismes subventionnaires du Royaume-Uni a publié une série de principes

relatifs à la GDR. La même année, la National Science Foundation (NSF) aux États-Unis a commencé à exiger, comme condition de remise des subventions de recherche, que les chercheurs soumettent des plans de gestion des données. De façon générale, dans la plupart des pays anglo-saxons (États-Unis, Royaume-Uni et Australie) et, dans une certaine mesure, en Europe, on a donc assisté à un véritable démarrage des initiatives visant l'adoption de meilleures pratiques de GDR.

Pendant ce temps au Canada les développements se sont fait attendre. Ce n'est qu'en 2016 que les grandes agences subventionnaires fédérales (*tri-agency*) ont émis une *Déclaration de principes des trois organismes sur la gestion des données numériques*[1]. Nous attendons d'ailleurs, au cours de l'année qui vient, le dépôt par ces agences de politiques officielles accompagnées de directives à l'attention des universités et des chercheurs eux-mêmes. Si l'on se fie à la *Déclaration de principes* et à ce qui s'est passé chez nos voisins du Sud, on peut raisonnablement s'attendre à ce que ces exigences comprennent la rédaction d'un plan de gestion des données ainsi que l'archivage et le partage (sauf exception) des données de recherche.

À la veille de la mise en place de ces exigences, il importe de mieux comprendre les besoins et les attitudes des chercheurs canadiens quant à la gestion des données qu'ils produisent dans le cadre de leur recherche. Jusqu'à tout récemment, il existait peu d'études canadiennes à ce sujet, particulièrement pour les institutions de taille moyenne comme l'Université Concordia.

Cet article présente donc les principaux résultats d'une enquête auprès des chercheurs de l'Université Concordia à Montréal en matière de gestion des données de recherche. L'étude, qui s'est déroulée de l'automne 2015 à l'hiver 2016, consistait en un sondage en ligne suivi d'une série d'entrevues. Le sondage a été envoyé à l'ensemble des professeurs de l'Université à l'exception de trois départements (Géographie, Science politique, Sociologie) puisque ceux-ci avaient déjà été consultés lors d'une première étude effectuée en 2013. La première partie de l'article porte sur les résultats du sondage, tandis que la seconde présente un résumé des informations obtenues grâce à des entrevues avec les chercheurs.

Revue de la littérature

Bien que nous ayons répertorié un nombre relativement modeste d'études empiriques sur l'attitude et les pratiques des chercheurs en matière de GDR au Canada, plusieurs enquêtes de ce genre ont été menées aux États-Unis et en Grande-Bretagne. En général, ces études portent sur une seule université, mais il existe aussi quelques exemples d'enquêtes multi-institutionnelles. La méthodologie employée consiste le plus souvent en un sondage Web envoyé aux chercheurs ou, à quelques occasions, à la combinaison d'un sondage et d'une série d'entrevues dans une perspective similaire à notre propre recherche. Cette revue de littérature ne se veut pas exhaustive, mais présente simplement quelques études récentes, en Amérique du Nord, qui pourront servir de points de comparaison avec notre enquête à l'Université Concordia. Aux lecteurs qui souhaiteraient consulter davantage d'études empiriques, nous recommandons d'examiner le tableau comparatif qui apparaît dans le récent article d'Elizabeth Berman (2017, 3).

À notre connaissance, une des plus grandes enquêtes quantitatives est celle effectuée par Carol Tenopir et ses collègues (Tenopir *et al.* 2011) d'octobre 2009 à juillet 2010. Ce sondage a rejoint 1 329 répondants (on estime que 15 000 personnes ont reçu le sondage, donc le taux de participation s'établirait autour de 9 %) provenant majoritairement de l'Amérique du Nord (75 %) ainsi que, dans une moindre mesure, des pays européens. Parmi les résultats les plus intéressants de cette étude, on remarque qu'environ 75 % de chercheurs disent partager leurs données. Par contre, seulement 36 % de répondants indiquent que leurs données sont « facilement accessibles ». En pratique, cela signifie que dans la plupart des cas les chercheurs partagent leurs données uniquement sur demande et qu'il existe des barrières importantes au partage plus systématique par voie électronique. Les principaux obstacles cités à ce propos sont le manque de temps pour préparer les données (54 %) et l'absence de fonds pour faire ce travail (40 %). Un autre aspect intéressant des résultats se situe au niveau du soutien institutionnel fourni par les universités. Ici, les auteurs séparent les résultats en deux catégories : les répondants provenant d'institutions axées sur la recherche et ceux qui sont affiliés à des institutions se concentrant sur l'enseignement. Comme on pourrait s'y attendre, le niveau de soutien semble plus fort dans les universités axées sur la recherche. Par exemple, 19 % des répondants des institutions de recherche sont fortement d'accord (*agree strongly*) avec l'énoncé « *My organization or project has a formal established process for managing data during the life of the project* », contre seulement 8,9 % pour les répondants venant des institutions d'enseignement. Pour l'énoncé « *My organization or project provides the necessary funds to support data management beyond the life of the project* », ces chiffres sont respectivement de 18,6 % (recherche) et de 9,4 % (enseignement). Finalement, les répondants qui se considèrent comme très satisfaits (*strongly agree*) pour ce qui est de la formation offerte par les organisations en matière de GDR sont de 24,9 % (recherche) contre 18,3 % dans les institutions d'enseignement.

Plus récemment, quelques études portant sur une seule université ont retenu notre attention. En 2014, Buys & Shaw (2015) ont mené une enquête à l'Université Northwestern.

1. <www.science.gc.ca/eic/site/063.nsf/fra/h_83F7624E.html>.

Des 12 940 personnes contactées — la population cible était constituée de tous les chercheurs, étudiants aux cycles supérieurs et candidats au postdoctorat —, 833 ont répondu à certaines questions et 788 ont complété le sondage, pour un taux de participation d'environ 6,4 %. Parmi les résultats intéressants, on remarque une grande diversité de pratique en matière de stockage des données :

> *Sixty-six percent use computer hard drives, 47 % use external hard drives, 50 % use departmental or school servers, 38 % store data on the instrument that generated the data, and 27 % use flash drives. Additionally, 31 % use cloud-based storage services. When asked to name their cloud-based storage (180 written responses) Dropbox (Dropbox, Inc., 2007) was the most popular choice (63 %). Only 6 % of the respondents use external data repositories.* (Buys & Shaw 2015, 12)

Notons que les résultats ne font pas de distinction entre stockage durant le projet et archivage à la conclusion de celui-ci. Par ailleurs, pour ce qui est de la période de conservation, une pluralité de répondants veut préserver indéfiniment ses données brutes (181 répondants sur 644) et ses données publiées (230 sur 631). Les autres choix de réponses offerts pour la durée de préservation étaient : moins d'un an ; de 1 à 5 ans ; de 5 à 10 ans ; et plus de 10 ans.

La volonté de partager ses données varie selon les personnes avec qui on partage les données et selon que le partage se fait avant ou après la publication d'un article : avec les membres du groupe de recherche (avant publication 47 %, après 14 %) ; avec les collègues de Northwestern (avant publication 15 %, après 4 %) ; avec les chercheurs du même domaine (avant publication 35 %, après 47 %) ; avec le public général (avant publication 3 %, après 35 %). Ce dernier chiffre montre que le partage libre des données avec le grand public (via un dépôt de données par exemple) reste problématique tant que les chercheurs n'ont pas publié leurs résultats.

Il nous paraît intéressant de signaler une récente étude menée à l'Université du Vermont (Berman 2017) qui, comme notre propre enquête, repose sur la combinaison d'une approche qualitative — dans ce cas, une analyse du contenu de 35 plans de gestion des données (PGD) et une série d'entrevues — et d'une méthode quantitative, soit un sondage auprès des chercheurs ([N=319], taux de participation 26,8 %). L'auteur utilise une méthode exploratoire séquentielle (*exploratory sequential mixed method*) qui débute par une analyse des données qualitatives (entrevues et analyse de contenu des PGD) afin d'en dégager les thèmes principaux. Ceux-ci servent ensuite de base au développement du questionnaire utilisé pour le sondage. Dans notre étude à Concordia, nous avons procédé à l'inverse, c'est-à-dire en débutant par le sondage et en explorant certaines questions plus en profondeur lors des entrevues en personne.

Voici un aperçu des résultats les plus intéressants issus de l'étude à l'Université du Vermont. On remarque, aussi bien au niveau des résultats qualitatifs que quantitatifs, le manque de métadonnées créées pour décrire les données de recherche. L'auteur souligne que cette observation se retrouve dans de nombreuses études comme celle de Tenopir (2011). Pour ce qui est du partage des données, comme la plupart des études sur la GDR le rapportent, il y a bel et bien un désir de dissémination, mais celui-ci se heurte à des obstacles :

> *For researchers who have submitted DMPs, the issues focused around fear of misinterpretation, intellectual property concerns, and a variety of legal issues, including confidential, proprietary, or classified information. The major limitations from the quantitative phase of the study focused on the ability to maintain confidentiality, the lack of time, personnel, and tools/infrastructure to make data available, and intellectual property concerns.* (Berman 2017, 12)

Comme dans les autres études mentionnées plus haut, l'auteur remarque que l'utilisation des dépôts de données est encore le fait d'une minorité : dans la phase qualitative de l'étude, 17 des 35 PGD mentionnent l'utilisation de dépôts ; pour ce qui est des répondants au sondage, seuls sept chercheurs sur 218 ont recours à ce genre de plateforme pour partager leurs données.

Nous terminerons cette revue de littérature en examinant ce qui s'est fait sur la scène canadienne au cours des dernières années. Les enquêtes portant sur la GDR sont relativement rares au pays, ce qui pourrait s'expliquer par le nombre modeste d'universités axées sur la recherche (comparativement aux États-Unis) et surtout par le retard qu'a pris le Canada sur les États-Unis et l'Europe au niveau des politiques et du soutien institutionnel pour la gestion des données de recherche. À notre connaissance, une des premières études est celle de Carol Perry (2008) qui portait spécifiquement sur l'archivage et le partage des données par les récipients de bourses de recherche du CRSH. En 2013, Mowers, Humphrey & Perry (2013) ont publié les résultats sommaires d'une étude auprès de plus de 300 chercheurs canadiens. À ces études, s'ajoute la première mouture de notre enquête à l'Université Concordia qui a été menée en 2013 (Guindon 2014).

En 2015, un groupe de cinq universités canadiennes (l'Université de Toronto, l'Université de la Colombie-Britannique, l'Université de Waterloo, l'Université de l'Alberta et l'Université Queen's) a adopté un questionnaire commun (créé par l'Université de Toronto) pour sonder les chercheurs des facultés de science et de génie (Sewerin *et al.* 2016). Depuis, quatre autres universités ont annoncé qu'elles se joindraient à ce groupe. L'objectif de ce projet était de créer un ensemble de données sur la GDR qui permettrait de comparer les universités canadiennes. Le sondage s'est déroulé d'avril à décembre 2015 et a permis de rejoindre 780 participants. En matière de documentation des données, 38,5 % des répondants indiquent qu'ils croient avoir fourni assez

d'information pour que d'autres chercheurs, à l'extérieur de leur laboratoire ou équipe de recherche, puissent réutiliser les données contre 61,5 % qui répondent par la négative ou qui ne savent pas. Ces résultats vont donc dans la même direction que ceux de l'Université du Vermont et de plusieurs autres études. À la question qui portait sur la méthode de stockage des données (notons que les réponses multiples étaient acceptées), 60,9 % des chercheurs indiquent utiliser le disque dur d'un ordinateur, mais une proportion étonnamment élevée (59,6 %) mentionne utiliser un dépôt de données tel que Protein Data Bank, Cambridge Structural Database, GitHub, Dryad ou Figshare. Il est fort probable que ce résultat élevé soit lié au fait que les répondants provenaient uniquement des disciplines scientifiques. L'offre de dépôts spécialisés en science est assez importante et il semble que la pratique d'archivage formel des données de recherche y ait évolué plus rapidement que dans d'autres disciplines. En matière de partage des données, 57,6 % des répondants indiquent que la méthode de dissémination la plus fréquente est « *by personal request only* ». Une proportion significative (mais tout de même minoritaire) de chercheurs indique ne pas partager ses données, mais on note des variations assez importantes selon les domaines : « *Disciplines with the highest number of respondents that indicated they were not currently sharing data were computer science (44.9 %, n=22) and the engineering areas (civil/mineral/environmental, 34.2 %, n=25 ; biological/chemical/materials/mechanical, 24.1 %, n=20 ; electrical/computer, 25.0 %, n=3).* » Sur une note encourageante, quand on leur demande d'évaluer leur volonté de partager leurs données dans le futur, 30,3 % des répondants mentionnent qu'ils vont considérer l'utilisation de dépôts de données et seuls 11,5 % répondent qu'ils ne prévoient pas partager leurs données. Au chapitre des raisons invoquées pour justifier la décision de ne pas partager leurs données, les chercheurs indiquent à 48,1 % le désir de publier avant de disséminer les données ; on indique aussi des problèmes de propriété intellectuelle (23 %), des obligations contractuelles envers une tierce partie (19,7 %) et des questions de confidentialité dans 17 % des cas.

Méthodologie

L'équipe de recherche était composée d'Alex Guindon, bibliothécaire responsable des services de données statistiques et géospatiales, de Danielle Dennie, bibliothécaire de référence, ainsi que de Jennifer McGrath, professeure associée au Département de psychologie. Nous voudrions également remercier Dubravka Kapa, directrice associée à la Bibliothèque de Concordia et Muhammed Idris étudiant au postdoctorat pour leur aide à la création du questionnaire et pour la planification générale du projet de recherche.

Bien que l'élaboration du questionnaire ait été le fruit du travail de tous les collaborateurs mentionnés ci-dessus, la partie qualitative de la recherche — les entrevues effectuées auprès des chercheurs ainsi que leur analyse — relevait exclusivement de Danielle Dennie et d'Alex Guindon.

Le sondage en ligne

L'objectif de notre projet de recherche était double : d'abord d'obtenir un portrait général des pratiques courantes et de l'attitude des chercheurs de l'Université Concordia dans la gestion de leurs données ; ensuite, d'identifier les services de GDR qui répondraient aux besoins de ces mêmes chercheurs. Le questionnaire[2] est une adaptation libre du Data Asset Framework (DAF) créé par le Humanities Advanced Technology & Information Institute (HATII) de l'Université de Glasgow avec l'aide du Digital Curation Centre[3]. Nous y avons ajouté plusieurs questions afin de mieux évaluer les pratiques et besoins des chercheurs en ce qui a trait à la manipulation des données, à leur traitement statistique et à leur visualisation. Puisque nous souhaitons obtenir des résultats quantitatifs, le sondage est composé de questions fermées avec, dans plusieurs cas, la possibilité de choisir la catégorie « autre » et d'ajouter des précisions. La partie qualitative de l'enquête repose sur les entrevues effectuées par la suite.

L'Université Concordia compte environ 28 000 étudiants (équivalent temps plein), dont 4 000 aux études supérieures. On y compte approximativement 1 000 professeurs à temps plein. La population du sondage, comme mentionné plus haut, était constituée de tous les professeurs à temps plein sauf ceux des départements de Géographie, Science politique et Sociologie qui avaient déjà été consultés lors d'un précédent sondage sur la GDR. Le sondage a été diffusé de décembre 2015 à janvier 2016 via la plateforme Survey Monkey. Le taux de participation s'est élevé à 19 %, soit 132 participants sur une liste de distribution de 696. Cependant, les participants n'étaient pas tenus de répondre à chacune des questions donc le taux de participation réel varie d'une question à l'autre. Ce taux de réponse est satisfaisant pour une enquête en ligne de ce type, mais il faut tout de même souligner qu'il est possible qu'un « biais d'autosélection » influence les résultats, c'est-à-dire que les chercheurs ayant choisi de répondre au sondage sont davantage intéressés par les questions de GDR que l'ensemble de la population.

La répartition de l'échantillon selon les grandes disciplines reflète assez bien la population de chercheurs à Concordia, sauf pour ce qui est des Arts où l'échantillon sous-représente clairement la population de chercheurs dans ce domaine. Il est probable que, malgré la définition très large du terme « données de recherche » que nous avons adoptée (qui pourrait être interprétée comme inclusives des productions artistiques), les professeurs de ce domaine ne se soient pas sentis concernés par cette étude.

2. <dx.doi.org/10.11573/spectrum.library.concordia.ca.00982722>.

3. <www.dcc.ac.uk/resources/tools/data-asset-framework>.

TABLEAU 1

Répartition des répondants par discipline

Discipline		
Science & Engineering	44	34 %
Social sciences	32	25 %
Humanities	19	15 %
Business	21	16 %
Arts	12	9 %

Les entrevues

Dix entrevues ont été effectuées auprès de chercheurs. Ceux-ci avaient indiqué dans le sondage qu'ils accepteraient d'être interviewés. Les chercheurs ont été sélectionnés d'après leur niveau d'intérêt relatif au partage et à l'archivage de leurs données, puisque nous considérons que les chercheurs ayant réfléchi à la question de la GDR peuvent apporter des réponses plus riches lors des entrevues. Un chercheur de chacun des départements suivants a été interviewé : Génie du bâtiment, civil et environnemental, Design et arts numériques, Informatique, Physique, Sciences de l'exercice, Gestion, Gestion de la chaîne d'approvisionnement et des technologies d'entreprises, Communications, Arts plastiques, Études françaises.

Les entrevues étaient d'une durée d'une heure, de format semi-structuré, et comme le sondage, basées sur une adaptation du Data Asset Framework du Humanities Advanced Technology & Information Institute (HATII) à l'Université de Glasgow. L'analyse des entrevues s'est faite en deux étapes : un codage descriptif était appliqué lors d'un premier cycle de codage (Saldana 2009, 70), suivi d'un codage sélectif (*focused coding*), lors d'un deuxième cycle (Saldana 2009, 155).

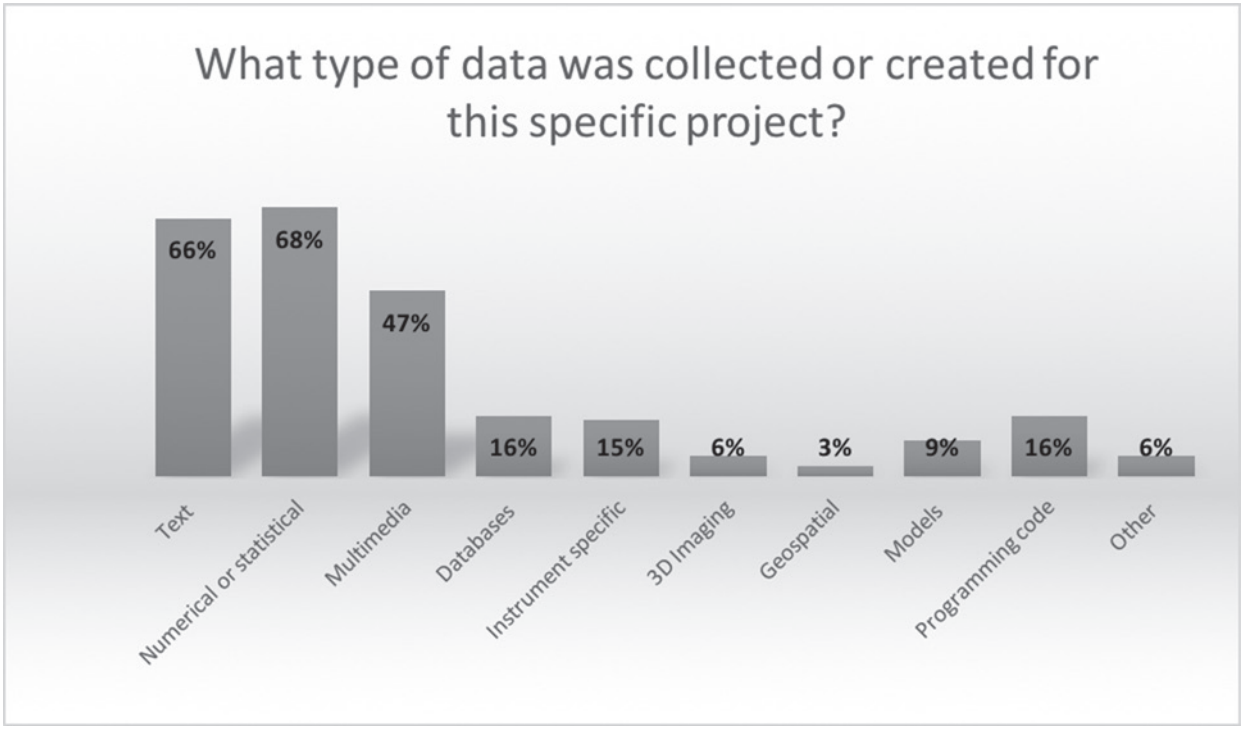
Résultats du sondage

Nous ne présentons ici que quelques-uns des résultats qui nous apparaissent les plus importants. L'ensemble complet de nos données agrégées est disponible en ligne[4].

Pour bien comprendre nos résultats, une précision méthodologique s'impose : sauf avis contraire, les pourcentages rapportés reflètent le nombre de participants ayant donné une réponse X. Comme pour la plupart des questions, les répondants pouvaient choisir plusieurs réponses, le total des pourcentages pour une question donnée est supérieur à 100 %. En d'autres termes les différentes catégories (réponses) ne sont pas mutuellement exclusives.

FIGURE 1

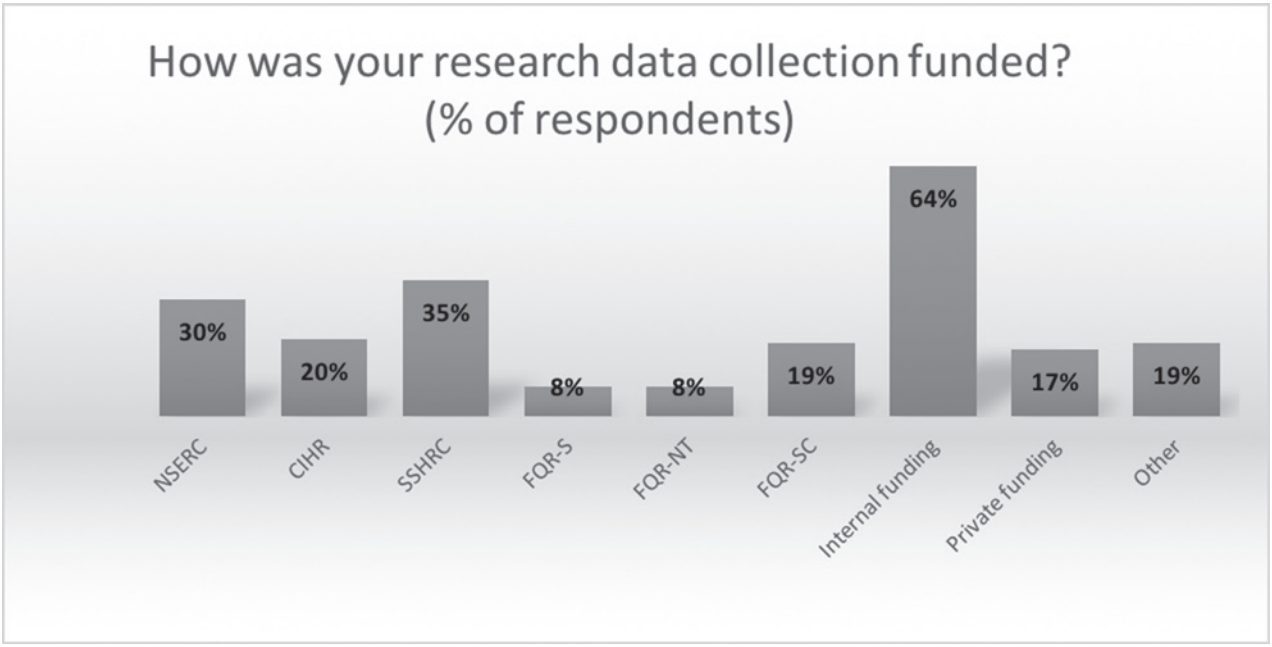
Quel type de données avez-vous collectées ou créées pour ce projet ?



4. <dx.doi.org/10.11573/spectrum.library.concordia.ca.00982722>.

FIGURE 2

Comment votre collecte de données a-t-elle été financée ?



Les graphiques et tableaux sont présentés en anglais puisque c'est dans cette langue que le sondage a été proposé. Par contre, nous avons ajouté un titre en français qui apparaît au-dessus de chaque figure.

Type de données

La Figure 2 démontre la grande diversité des données utilisées par les chercheurs. Les plus communes étant les données numériques ou statistiques (68 %), les données textuelles (66 %) et les données multimédias (images, audio, vidéo) (47 %). Sur cette base, on peut conclure que les services et outils offerts pour soutenir la GDR devraient être très flexibles et pouvoir s'adapter à divers types de données. Néanmoins, les services les plus développés ainsi que la plupart des fonctionnalités des outils devraient probablement se concentrer sur les statistiques et textuelles. Les objets multimédias sont d'une nature très différente et nécessiteront sans doute des outils particuliers.

Financement des projets de recherche

Le financement de la recherche à Concordia repose sur une multitude de sources. On remarque un fort taux de chercheurs (64 %) ayant obtenu un financement interne et un pourcentage significatif (17 %) ayant eu recours à des fonds privés.

Comparons maintenant le financement des trois grandes agences fédérales (*tri-agency*) avec celui obtenu d'autres sources. Pour ce faire, il nous semble utile de présenter les résultats en fonction du nombre de réponses obtenues pour chaque catégorie plutôt que selon le nombre de participants ayant choisi chacune des catégories. Le nombre de réponses est une meilleure approximation de la distribution des subventions puisque le même chercheur peut avoir indiqué plusieurs sources de financement (les réponses multiples étant acceptées) pour des projets de recherche différents voire même pour un seul projet.

On remarque qu'uniquement 39 % des réponses indiquent un financement provenant des trois grands organismes (CRSH, CRSNG, IRSC) contre 32 % pour les autres agences et 29 % de financement interne (subventions provenant de l'Université Concordia) pour un total de 61 % pour les sources autres que le *tri-agency*. Comme les premières politiques de GDR et les exigences qui s'y rapportent proviendront du *tri-agency*, seule une minorité de projets à Concordia y serait soumise. Si cela peut donner davantage de marge de manœuvre et de temps à l'Université pour se préparer, il faudra par contre se poser la question du financement de la GDR pour les projets qui ne relèvent pas du *tri-agency*. En effet, ceux-ci ne pourront pas compter sur d'éventuelles subventions spécifiques à la gestion des données. Par ailleurs, le nombre important de projets financés par des fonds privés pose un problème particulier, car des ententes de confidentialité interviennent fréquemment entre chercheurs et entreprises qui financent leur recherche. Dans ces conditions il est souvent impossible de partager les données publiquement, voire même avec d'autres chercheurs de la même institution.

FIGURE 3

Financement en fonction du nombre total de réponses

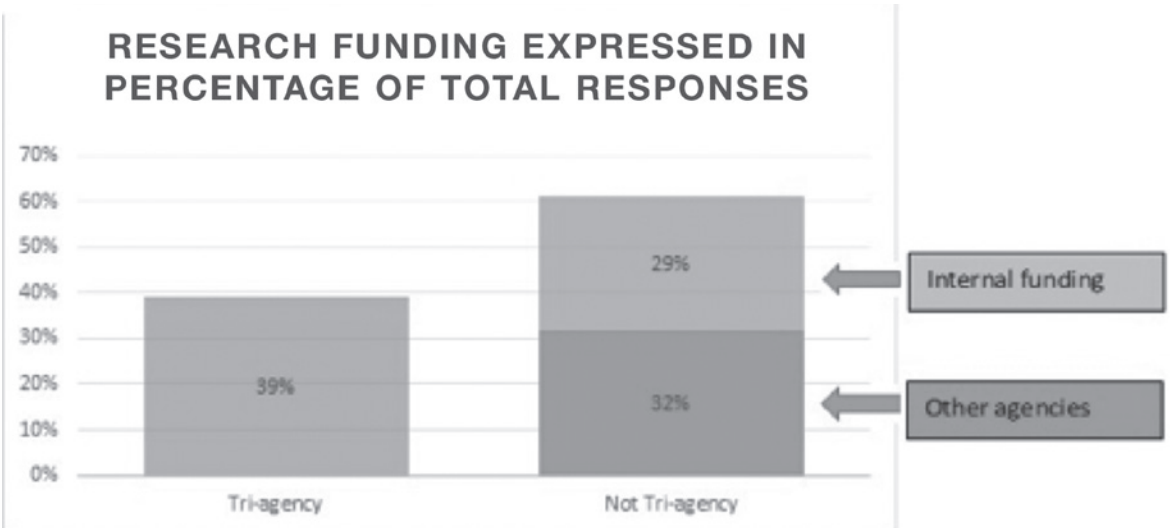
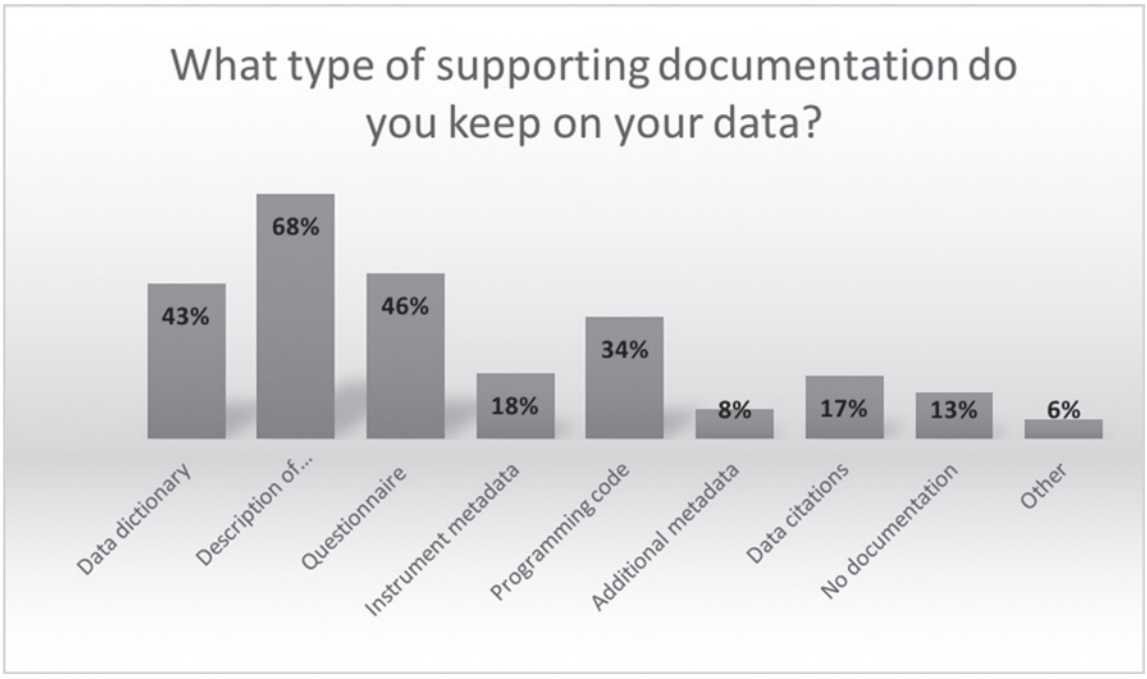


FIGURE 4

Quel type de documentation avez-vous créée pour décrire vos données ?



Documentation des données

Les résultats montrent que les chercheurs développent une large gamme de documents pour décrire leurs données. Seuls 13 % d'entre eux indiquent ne pas garder de documentation. Cette grande variété de documentation rappelle encore une fois l'importance pour les universités de disposer de dépôts de données flexibles qui peuvent accueillir différents types de fichiers. Cela étant, lorsque ceux-ci sont disponibles dans un domaine scientifique, les dépôts disciplinaires spécialisés (GenBANK, ICPSR, etc.) seront souvent

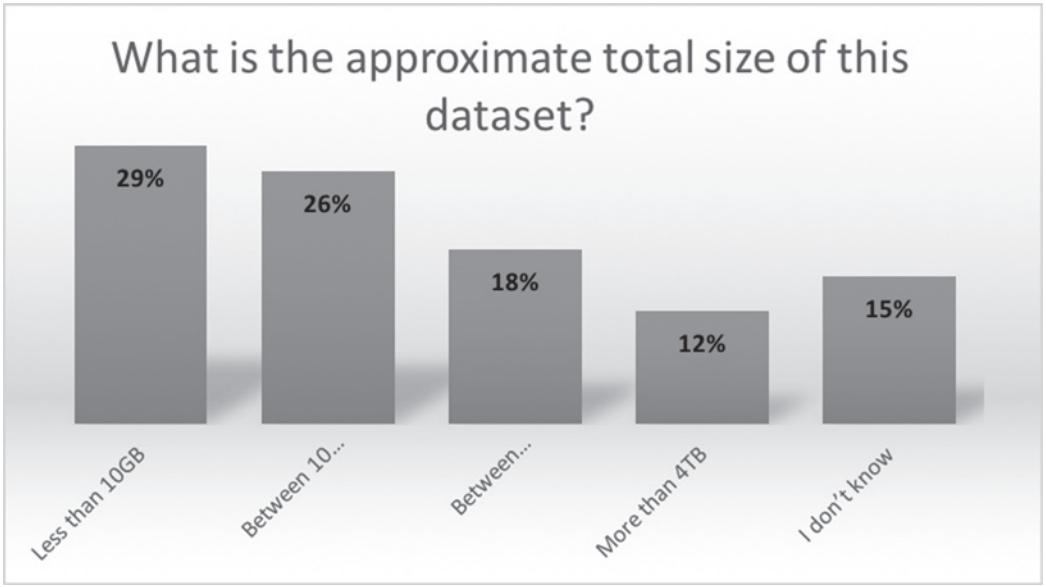
plus appropriés que des dépôts multidisciplinaires puisqu'ils sont développés afin de répondre aux types de données et de fichiers de documentation ou de métadonnées existant dans cette discipline.

Taille des ensembles de données

Nous avons demandé aux chercheurs d'évaluer la taille de l'ensemble de données qu'ils considèrent comme étant leur plus important. Cinquante-cinq pour cent (55 %) des répondants rapportent que leur ensemble de données fait moins de 100 Go. Par contre, 30 % des chercheurs possèdent un gros ensemble de données (plus de 100 Go) et 15 % indiquent ne pas savoir quelle est la taille de leur ensemble de données. Ces résultats sont plutôt encourageants, car ils indiquent que la majorité des ensembles de données pourraient sans doute être archivés par des dépôts de données reposant sur une infrastructure technologique relativement modeste. Quant aux données massives (plus de 100 Go), nous croyons qu'elles proviennent davantage des disciplines scientifiques (des analyses supplémentaires seraient nécessaires pour confirmer cette impression) dans lesquelles il existe souvent des dépôts spécialisés qui peuvent les accueillir.

FIGURE 5

Quelle est la taille approximative de votre ensemble de données ?



Stockage des données

Les réponses à cette question reflètent la très grande diversité d'options disponibles pour les chercheurs. On peut toutefois conclure, puisque 91 % des répondants utilisent le disque dur de leur ordinateur et que seulement 20 % mentionnent utiliser un serveur départemental et 1 % celui des TI (IITS est le service des technologies de l'information à Concordia), qu'ils n'utilisent pas des moyens de stockage idéaux. En effet, les disques durs employés seuls ne sont pas une façon sécuritaire de stocker les données ou de les partager avec les collaborateurs. Par ailleurs, l'utilisation de solutions infonuagiques commerciales (41 %) peut faciliter le partage avec les membres de l'équipe de recherche, mais peut poser problème en termes de sécurité et de confidentialité. Ces résultats démontrent le besoin d'une solution de stockage ou d'un *environnement de recherche virtuel* basé à Concordia et qui répondrait aux besoins de sécurité et de partage des données[5].

Archivage des données

Presque un tiers (30 %) des chercheurs déclarent ne pas vouloir archiver leurs données à la conclusion du projet de recherche. Cette attitude peut s'expliquer de plusieurs façons, notamment par une interprétation restrictive des obligations en matière d'éthique ou une méconnaissance des avantages liés à l'archivage et au partage des données. Il est aussi possible que, dans certains cas, les données ne présentent effectivement que très peu de valeur à long terme. Cependant, il faut noter que, si l'on additionne les résultats pour Spectrum (dépôt institutionnel de Concordia) (21 %) et ceux pour la catégorie « dépôt généraliste ou disciplinaire » (10 %), un nombre légèrement plus élevé

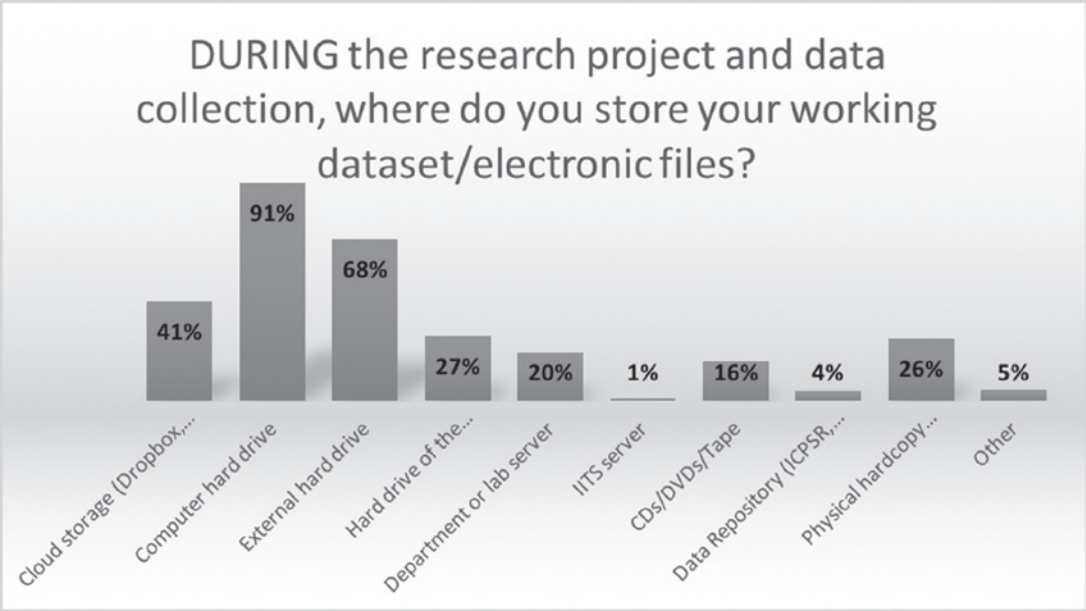
5. Ce besoin a aussi été identifié lors d'un sondage interne effectué par IITS auprès des grands laboratoires de recherche de Concordia.

(31 %) est intéressé par l'archivage des données. Il est intéressant de noter cette préférence pour une option de stockage locale plutôt que l'utilisation d'un dépôt externe. On peut penser que cela reflète une confiance en l'habileté de Concordia à fournir un service sécuritaire ou alors un attachement des chercheurs à leur institution. Par contre, une fraction encore plus importante des répondants (42 %) ne connaît pas les dépôts de données, mais souhaiterait en apprendre davantage. Cela indique sans doute une méconnaissance des dépôts spécialisés comme ICPSR, GenBANK, PsychData, etc. De toute évidence, il y a beaucoup de travail à effectuer pour faire connaître les différents dépôts disponibles pour l'archivage et le partage des données.

Finalement, notons que certains résultats sont très semblables à ceux obtenus lors du sondage de 2013 (Guindon 2014) dans lequel 32 % des répondants indiquaient qu'ils n'étaient pas intéressés à archiver leurs données. Par contre, le choix de réponses pour les options d'archivage était différent lors du premier sondage ce qui rend la comparaison un peu boiteuse, mais mentionnons tout de même que 39 % des répondants avaient exprimé un intérêt pour Spectrum et 29 % pour des dépôts externes. Encore ici, on remarque donc une préférence pour la solution locale.

FIGURE 6

Durant votre projet de recherche et pendant la collecte de données, où gardiez-vous vos jeux de données/fichiers de travail ?



Période de préservation des données

Nous notons une forte préférence (41 %) pour un archivage de durée indéfinie. Bien que cela puisse être souhaitable pour plusieurs ensembles de données, ce n'est pas toujours nécessaire ou même techniquement réalisable. En effet, certains jeux de données peuvent perdre leur pertinence après un certain nombre d'années comme l'indiquent aussi les réponses à cette question. Nous croyons qu'il faudra établir des politiques de rétention claires et alignées sur les éventuelles exigences des agences de financement.

Méthodes de partage des données

Il est très intéressant de voir que seuls 15 % des répondants indiquent qu'ils préfèrent ne pas partager leurs données. Vingt-trois pour cent (23 %) des chercheurs disent quant à eux ne pas partager leurs données jusqu'à maintenant, mais voudraient en apprendre davantage sur les options de partage qui s'offrent à eux. Par contre, lorsque l'on s'attarde aux méthodes de partage, on voit que seule une minorité des chercheurs utilisent des dépôts de données pour la dissémination (7 % en accès libre et 16 % avec un contrôle sur l'accès). À ceux-ci, on peut ajouter les répondants qui partagent leurs jeux de données via des sites Web spécialisés. Les autres préfèrent des méthodes plus « manuelles » de partage, telles que le transfert physique (46 %) (via un disque dur ou une clé USB par exemple) ou par courriel (51 %). S'il est évident que ces méthodes de partage ne sont pas idéales au niveau de la sécurité en plus de ne pas être très pratiques, elles reflètent le désir des chercheurs de garder un contrôle direct sur la dissémination de leurs données. Enfin, si l'on compare ces résultats avec ceux obtenus en 2013 (Guindon 2014), on remarque que le taux de répondants qui ne sont pas intéressés à partager leurs données est assez constant : 15 % en 2013 contre 13 % en 2016.

FIGURE 7

À la conclusion de votre projet de recherche, dans quel dépôt de données considéreriez-vous archiver votre ensemble de données ?

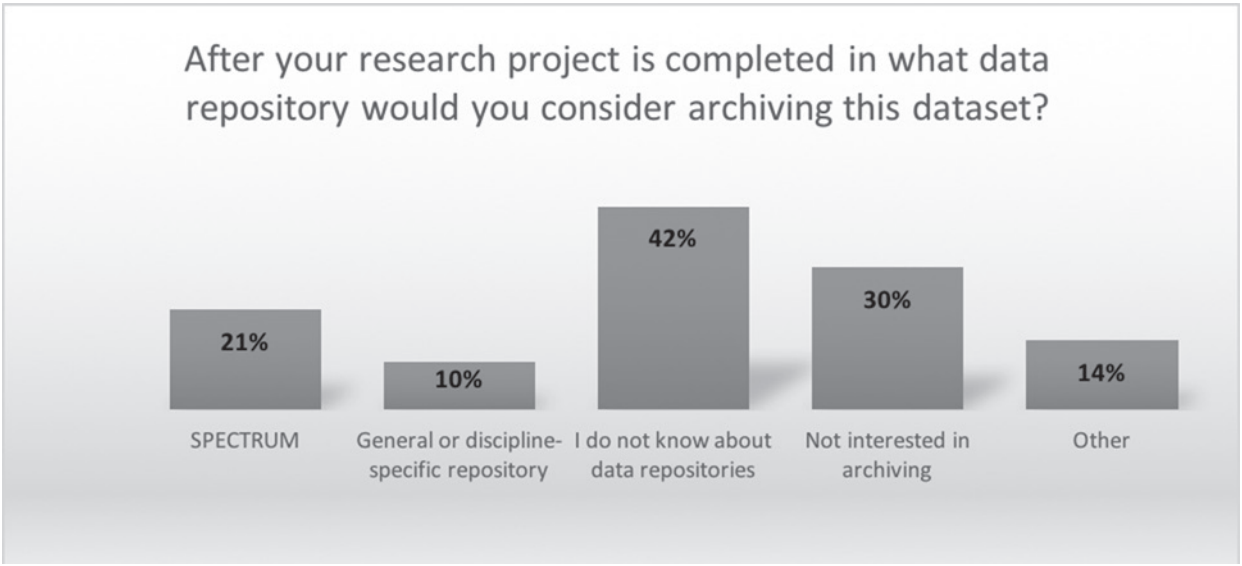
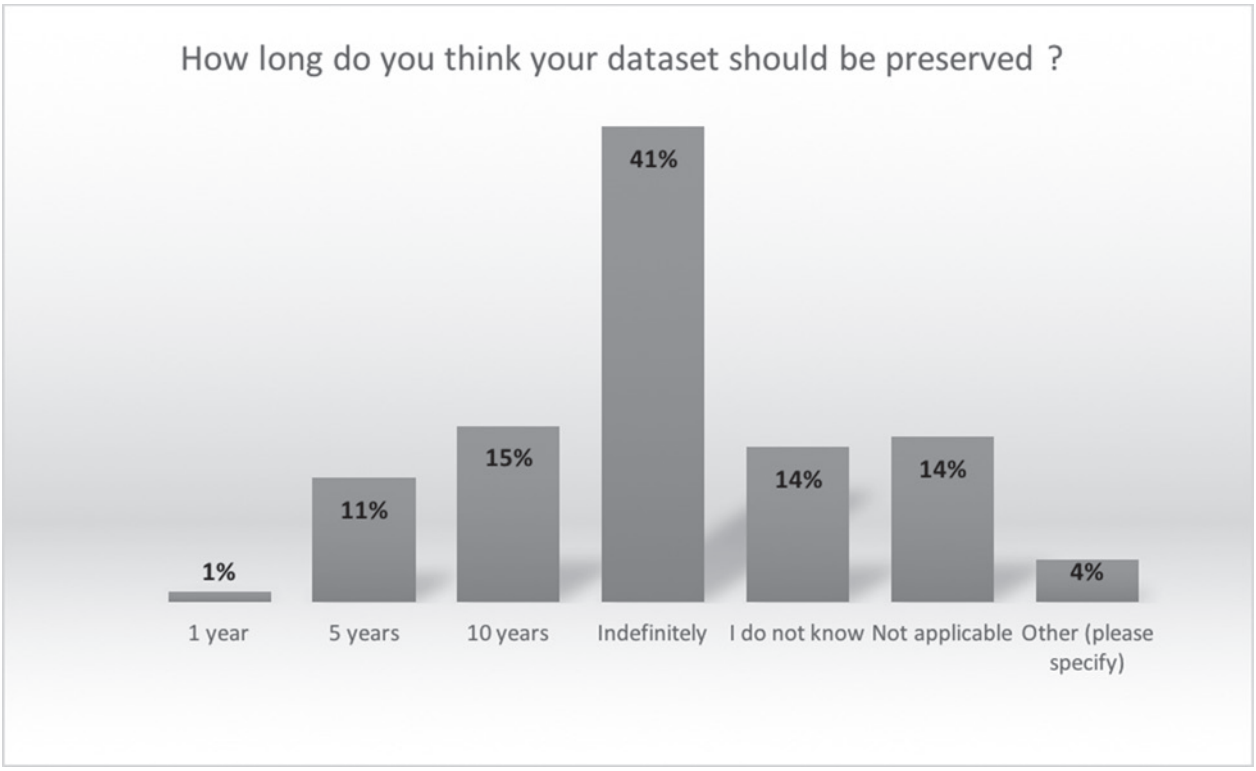


FIGURE 8

Combien de temps vos données devraient-elles être préservées ?



Expérience avec les périodiques qui exigent le dépôt des données

Le nombre de périodiques qui exigent le dépôt des données soutenant les articles scientifiques est en pleine croissance. Rappelons que ces revues ne demandent pas nécessairement que les données soient offertes en libre-accès. En effet, dans certains cas, on accepte le dépôt sur une plateforme qui permet de contrôler l'accès aux données. Nous observons que 28 % des répondants ont en effet soumis des manuscrits à de telles revues. Cependant, un plus grand nombre de chercheurs (34 %) ont jusqu'à maintenant évité de publier dans ces périodiques. Il sera intéressant de voir comment l'attitude des chercheurs évoluera au fur et à mesure que les politiques de dépôt de données deviendront de plus en plus fréquentes.

Offre de services de GDR

Un des objectifs importants de notre étude était de cerner les besoins des chercheurs en matière de soutien aux activités de GDR. Nous avons demandé aux chercheurs d'indiquer leur niveau d'intérêt pour une série d'ateliers ou de services d'aide en GDR. Les résultats du tableau 2 ci-dessous sont présentés en ordre décroissant d'intérêt pour les répondants (selon la colonne « *very interested* »). L'élément qui suscite, de loin, le plus d'intérêt est « *data management activities that will be required by grant funding agencies* ». Cela n'est pas une surprise, mais confirme que la GDR prendra beaucoup d'importance aux yeux des chercheurs une fois que des exigences concrètes seront mises de l'avant par les agences de financement. De façon générale, il semble que les répondants soient intéressés par une vaste gamme d'activités. Ainsi, même l'activité la moins populaire selon la colonne « *total interested* », « *help in ensuring the confidentiality/anonymity of data on human participants* », obtient tout de même un score de 49 %. Il sera difficile de développer des services répondant à tous ces besoins et il faudra sans aucun doute compter sur plusieurs partenaires au sein des universités (bibliothèque, services des TI, bureau de la recherche, services juridiques).

FIGURE 9

Comment partagez-vous vos données avec d'autres utilisateurs ?

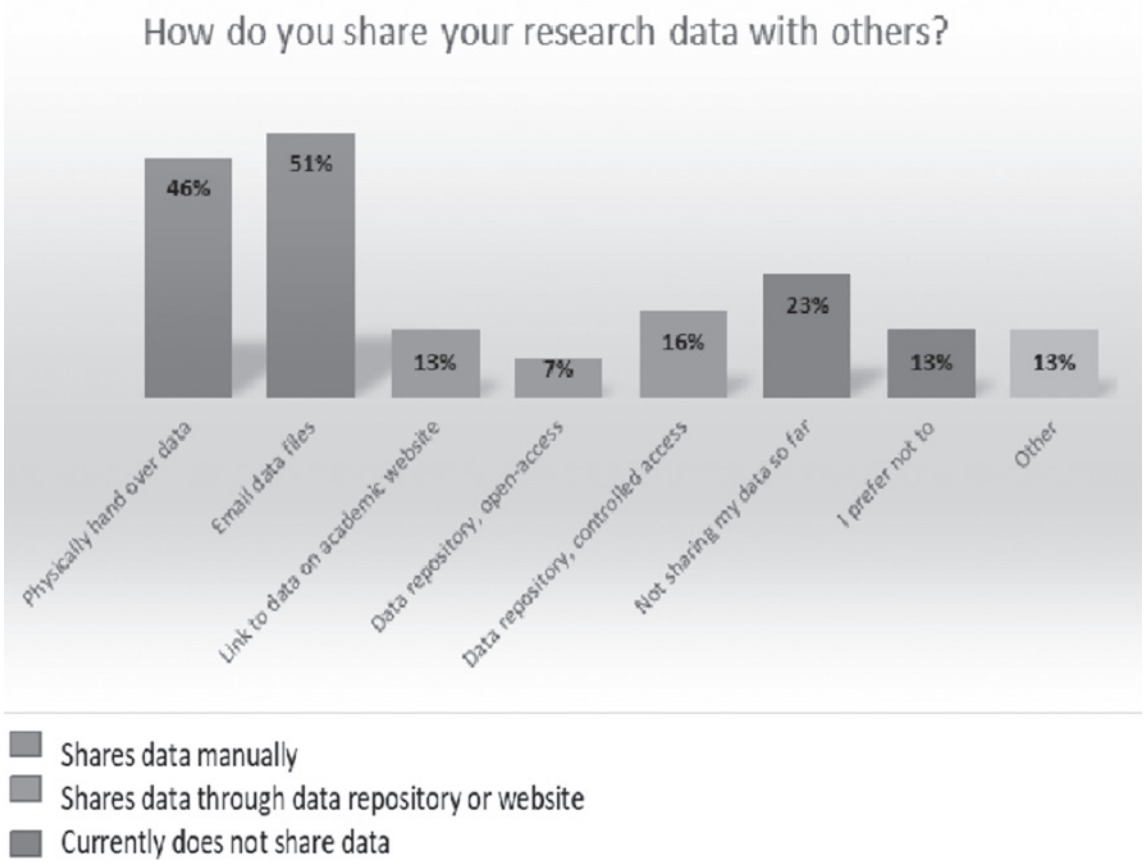


FIGURE 10

Quelle est votre expérience avec les périodiques qui exigent le dépôt des données ?

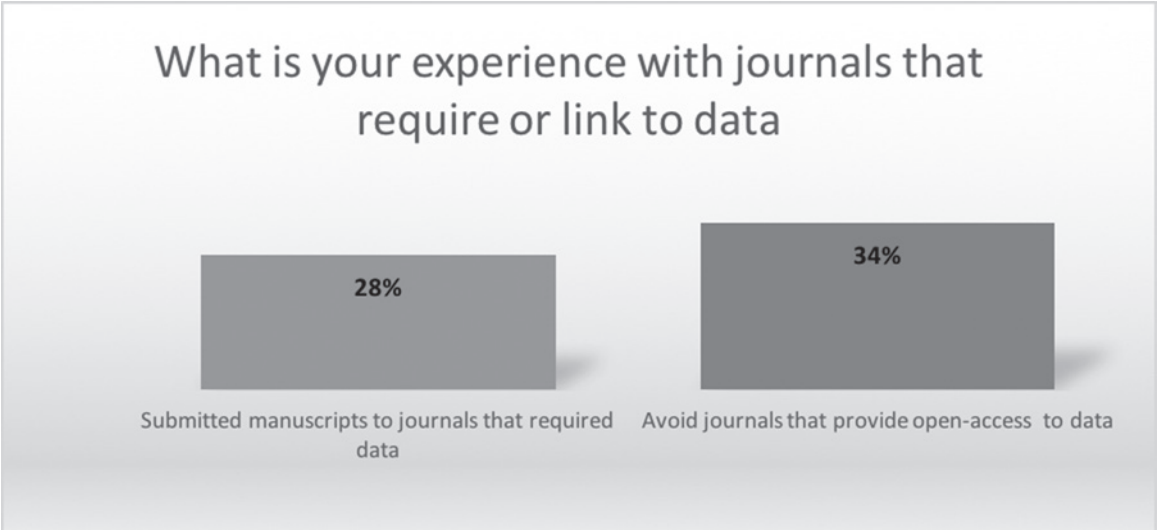


TABLEAU 2

Niveau d'intérêt pour des services de soutien en GDR

Workshop/Resource	Very interested	Somewhat interested	Total interested	Not interested
Data management activities that will be required by grant funding agencies	43 %	34 %	77 %	15 %
Data management plan	30 %	42 %	72 %	21 %
Research collaboration software tools	29 %	39 %	69 %	25 %
Finding and accessing existing data sources	29 %	34 %	63 %	23 %
Data archiving	27 %	43 %	69 %	22 %
Data visualization	27 %	44 %	71 %	22 %
Digitization of paper records	27 %	39 %	65 %	26 %
Data sharing and access	23 %	40 %	63 %	30 %
Help in ensuring the confidentiality/anonymity of data on human participants	21 %	28 %	49 %	29 %
Data management workshop	19 %	49 %	68 %	25 %
Data documentation and metadata	18 %	38 %	56 %	33 %

Sommaire des entrevues

Documentation décrivant les données

Quoique la majorité des chercheurs aient indiqué dans le sondage qu'ils produisaient de la documentation décrivant leurs données, plusieurs participants aux entrevues affirment pour leur part qu'ils n'en ont pas. Les seuls à détenir des documents sont deux chercheurs en sciences pures. Pour les autres, les raisons évoquées pour expliquer ce manque de documentation sont multiples : certains croient que leurs données sont suffisamment explicites par elles-mêmes ou que la section méthodologie de leurs publications décrit suffisamment bien leurs données ; d'autres chercheurs, en sciences humaines, croient qu'il n'y avait pas d'incitatifs ou de récompenses liés à la documentation ou la gestion des données. D'après une chercheure : « [I]*f there was acknowledgement of sustainability and there were rewards for that, that would be fantastic. There's more rewards for bringing new stuff in then there is rewards for careful preservation. And that's my observation from 25 years here.* » Finalement, certains chercheurs ont indiqué qu'ils n'avaient pas assez de temps pour bien documenter leurs données. Un ingénieur a dit : « […] *documentation is gonna be* […] *extra work for me and my team. And so,* […] *I'm not really sure if you want to go that way*

because I prefer students to do scientific work and publish papers then, you know, be data managers. »

Stockage des données

La majorité des chercheurs interviewés éprouvent de la difficulté à trouver de l'espace pour le stockage de leurs données de recherche en cours ou complétée. Plusieurs d'entre eux aimeraient avoir un serveur institutionnel, géré par l'université, avec des politiques et des directives d'utilisation claires, qui pourrait servir à stocker les données de recherche et qui serait doté d'un logiciel pour gérer ces données. Les chercheurs en sciences et en génie sont de ceux pour qui ce besoin était le plus criant. Un des scientifiques a bien résumé ce besoin :

> *We need an institution server* […]. [W]*e should have secured access with secured account for each research team, for each lab.* […] [T]*here should be a flexibility [to accommodate] people with space needs.* [We] *need* […] *a technical team supporting this server* […] *someone to* […] *check that there is enough available space, see if there's problems, to fix the problem, to manage the access, the accounts, the codes, and make sure [it is] regularly backed up and whether there should be some* […] *duplicate somewhere, to make sure everything's safe*.

Les chercheurs ayant des données confidentielles provenant de sujets humains ou de sources de données secondaires (données obtenues de tierces parties) insistent sur le fait que ce serveur se devrait d'être sécurisé et que l'université devrait créer des politiques et des directives indiquant où et comment les données de recherches devraient être stockées. Quoique 41 % des répondants du sondage utilisent un serveur commercial infonuagique, quelques-uns des chercheurs interviewés ont exprimé leurs inquiétudes vis-à-vis la sécurité des données sur ce type de serveurs basés aux États-Unis. D'après un chercheur en sciences humaines : « *I think we need our own indigenous resources, so that* […] *we're under the injunctions of the agencies that are federally funded and mandated, yet the problem of data leakage or data surveillance or data theft, once it is outside of our borders, it's in some ways beyond our control* […]. »

Une petite minorité des interviewés ne veut pas stocker ses données sur un serveur institutionnel afin de pouvoir en garder un contrôle absolu et de pouvoir répondre rapidement aux problèmes techniques potentiels.

Partage des données

Les résultats du sondage démontrent que la majorité des chercheurs partagent leurs données, mais ne nous permettent pas de savoir avec qui ces données sont partagées. Lors des entrevues, la majorité des chercheurs, peu importe leur discipline, ont indiqué qu'ils partageaient leurs données de recherche seulement avec des collaborateurs ou des étudiants directement engagés dans leur projet de recherche. Certains ne partagent pas leurs données à l'extérieur de ce groupe malgré le fait qu'ils aient reçu des fonds d'organismes subventionnaires ayant des politiques de partage de données. De tous les chercheurs interviewés, seulement trois partagent leurs données avec des personnes autres que leurs collaborateurs immédiats ou leurs étudiants. Ils le font soit parce qu'ils croient aux bienfaits du partage ou parce qu'ils doivent se conformer aux politiques des organismes subventionnaires.

Les règles de confidentialité dans les formulaires de consentement relatif à l'éthique de la recherche sont la raison principale évoquée par les chercheurs pour expliquer la raison du non-partage. Une chercheure en gestion craint que des politiques de partage des données imposées par les organismes subventionnaires peuvent faire entrave à la recherche puisque certains participants refuseraient alors de se faire interviewer. Finalement, quelques chercheurs indiquent aussi qu'ils ne croient pas que le partage des données brutes soit nécessaire dans leur communauté de recherche.

Cela étant, la majorité des chercheurs sont disposés à partager leurs données puisqu'ils y voient des avantages, tels que la possibilité de vérifier ou de reproduire la recherche et la possibilité pour l'ensemble de données d'être cité. Un chercheur en gestion a dit : « […] [G]*et*[ting] *citations* […] *is* […] *how my university recognizes the value of what I've done.* […] *I realize that if your article is valuable, people might cite it, but a lot of the time, the most valuable thing about an article is the data that was collected.* »

Par contre, certaines conditions de partage sont vues comme nécessaires, telles que l'anonymisation, un accès contrôlé, des embargos, ou même des systèmes de gestion des droits numériques (GDN). Quelques chercheurs soulèvent aussi la complexité liée au partage des données secondaires (données obtenues de tierces parties).

Archivage des données

De tous les chercheurs interviewés, un seul, en sciences, archive ses données dans un dépôt de données. Les chercheurs ont tout de même une opinion sur les caractéristiques jugées essentielles pour ce genre de dépôt.

La première est la capacité de stocker des données pour de très longues durées. Cette caractéristique est considérée comme particulièrement importante pour des données qui ont été difficiles à récolter, qui pourraient être utilisées pour des études longitudinales, qui pourraient être liées avec des données d'une autre recherche, ou pour l'exploration de données (*data mining*). La deuxième caractéristique essentielle est la présence d'un financement stable — préférablement en provenance des organismes subventionnaires — afin d'assurer la pérennité du dépôt de données. Malheureusement, d'après un des chercheurs, « […] *the funding doesn't work that way.* [They] *pay you to do the project, and once the project's over,* [y]*ou* [get] *zero support. And if the*

data is still there when someone starts up the machine the next time, good […] ». Des politiques claires de préservation et d'accès sont d'autres caractéristiques importantes des archives. Certains chercheurs ayant des données confidentielles prônent des mécanismes qui permettraient de sécuriser les données. Une petite minorité de participants ne croient pas à la nécessité de l'archivage des données puisque leur communauté de recherche n'aurait pas, selon eux, le besoin de réutiliser ou d'analyser à nouveau les données.

Conclusion

Les résultats du sondage et des entrevues démontrent que la majorité des chercheurs voudraient adopter une meilleure approche pour gérer, préserver et partager leurs données de recherche. La majorité reconnaît les avantages de la GDR en général et du partage des données en particulier, à la fois pour les chercheurs qui produisent les données et pour la communauté scientifique en général.

Par contre, à l'heure actuelle, peu de chercheurs ont adopté de bonnes pratiques de gestion des données. Lorsqu'il y a de la documentation, elle respecte rarement les normes acceptées ; lorsqu'il y a des procédures de sauvegarde, elles sont rarement systématiques ou n'adhèrent pas à des mesures de sécurité appropriées ; lorsque les données sont préservées à la fin d'un projet, c'est souvent sur un disque dur ou, au mieux, sur le serveur du laboratoire ou du département. Et lorsque les données sont partagées, cela est fait de façon informelle, soit en personne (transfert sur disque dur ou clé USB) ou par courriel, ce qui corrobore ce que d'autres études ont observé (Tenopir *et al.* 2011 ; Sewerin *et al.* 2016).

Les obstacles à l'adoption des meilleures pratiques de GDR sont multiples. Les participants à notre étude ont quant à eux mentionné, entre autres :

- le manque d'incitatifs dans le monde universitaire pour la gestion et le partage des données ;
- le manque de ressources — à la fois humaines et technologiques — nécessaires à la conservation des données ;
- les problèmes associés à la confidentialité et de l'éthique ;
- la volonté de garder un certain niveau de contrôle sur l'utilisation de leurs données.

Ces éléments concordent bien avec ceux relevés dans la littérature (Berman 2017 ; Sewerin *et al.* 2016).

Nous croyons que l'annonce prochaine des politiques des trois organismes subventionnaires ainsi que des exigences qui y seront associées constituera un moment charnière dans l'évolution des attitudes et des pratiques des chercheurs en matière de GDR. Ainsi, plus de 75 % des répondants à notre sondage ont indiqué qu'ils souhaitent suivre des formations sur la gestion des données reliées aux exigences des organismes subventionnaires. En conséquence, nous pouvons nous attendre à une plus grande demande d'aide pour la GDR dans les universités canadiennes lorsque ces nouvelles exigences seront adoptées. Les récentes études sur les pratiques et attitudes des chercheurs en matière de GDR entreprises par plusieurs universités canadiennes peuvent constituer un point de départ pour évaluer les besoins les plus immédiats de nos chercheurs. Cependant, il reste beaucoup de travail à faire et le succès de l'entreprise dépendra de la collaboration de plusieurs acteurs au sein des universités : les chercheurs eux-mêmes, bien sûr, mais aussi les bibliothèques, les services de TI, les bureaux de la recherche, les comités d'éthique et les services juridiques. De plus, il est illusoire de croire que le développement de services de GDR pouvant répondre aux exigences des agences de financement de la recherche puisse se faire à coût nul. Il faudra donc compter sur du financement et des infrastructures qui proviendront des universités et des différents paliers de gouvernement, mais aussi, espérons-le, des agences subventionnaires elles-mêmes.

SOURCES CONSULTÉES

Berman, E. A. 2017. An exploratory sequential mixed methods approach to understanding researchers' data management practices at UVM : Integrated findings to develop research data services. *Journal of eScience Librarianship* 6 (1) : e1104. doi : 10.7191/jeslib.2017.1104

Buys, C., & P. Shaw. 2015. Data management practices across an institution : Survey and report. *Journal of Librarianship and Scholarly Communication* 3 (2) : p.eP1225. doi : 10.7710/2162-3309.1225

Data's shameful neglect. 2009. *Nature* 461 (7261). doi : 10.1038/461145a

Guindon, A. 2014. Research data management at Concordia University : A survey of current practices. *Feliciter* 60 (2) : 15-17. <search.proquest.com/docview/1526973990>.

Hickson, S., K. A. Poulton, M. Connor, J. Richardson & M. Wolski. 2016. Modifying researchers' data management practices : A behavioural framework for library practitioners. *IFLA Journal* 42 (4) : 253-265.

Mowers, S., C. Humphrey & C. Perry. 2013. *Summary report : Survey of researchers' needs and practices regarding research data management in Canada*. Consulté le 1er août 2017. <gsg.uottawa.ca/data/open/aa-interim-survey-report/20130801-en.pdf>.

Saldana, Johnny. 2009. *Coding Manual for qualitative researchers*. Los Angeles : Sage.

Sewerin, C. *et al.* 2016. From coast to coast : Canadian collaboration in a changing RDM seascape. Consulté le 1er août 2017. <hdl.handle.net/1807/72802>.

Tenopir, C. *et al.* 2011. Data sharing by scientists : Practices and perceptions. *PloS One* 6 (6) : e21101. doi :10.1371/journal.pone.0021101

Van Tuyl, S. & G. Michalek. 2015. Assessing research data management practices of faculty at Carnegie Mellon University. *Journal of Librarianship and Scholarly Communication* 3 (3). doi : 10.1184/RDL/01092015