



Novembre 2020

# État actuel de la gestion des données de recherche au Canada

## Mise à jour de l'exposé de principe du CLIRN sur la gestion des données

Rédigé par le groupe de travail sur la gestion des données de l'Alliance :

Shahira Khair, Rozita Dara, Susan Haigh, Mark Leggott, Ian Milligan, Jeff Moon, Karen Payne, Elodie Portales-Casamar, Ghilaine Roquet et Lee Wilson



**Alliance de recherche  
numérique** du Canada

**Digital Research  
Alliance** of Canada

Financé par le  
gouvernement  
du Canada

**Canada**

# Table des matières

<b>Remerciements</b> .....	2
<b>Résumé</b> .....	3
Sommaire de l'état actuel.....	4
Principaux défis et possibilités .....	6
<b>Introduction</b> .....	7
Définition de « données de recherche » .....	9
Impact et valeur de la GDR.....	11
Mouvement de science ouverte .....	17
Souveraineté des données autochtones .....	18
Gestion des données de recherche au sein de l'infrastructure de recherche numérique .....	20
Vision nationale du soutien de la GDR.....	21
<b>Soutien national actuel de la GDR</b> .....	23
Données de recherche Canada .....	23
Réseau Portage, Association des bibliothèques de recherche du Canada.....	24
<b>Analyse du paysage national d'IRN</b> .....	27
Enseignement supérieur.....	27
Organismes de recherche .....	30
Organismes de financement de la recherche .....	36
Maisons d'édition savante .....	38
Organismes para-universitaires .....	39
Fournisseurs de services tiers (commerciaux et à but non lucratif) .....	42
Organisations internationales .....	43
<b>Évaluation de l'état actuel</b> .....	47
Stockage et calcul .....	48
Interopérabilité .....	56
Services de données .....	63
Gouvernance .....	75
<b>Principaux défis et possibilités</b> .....	80
Coordination .....	80
Représentation et inclusion.....	81
Durabilité .....	82
<b>Prochaines étapes</b> .....	84
<b>Annexe A – Le cycle de vie de la recherche et les fonctions de la GDR</b> .....	85

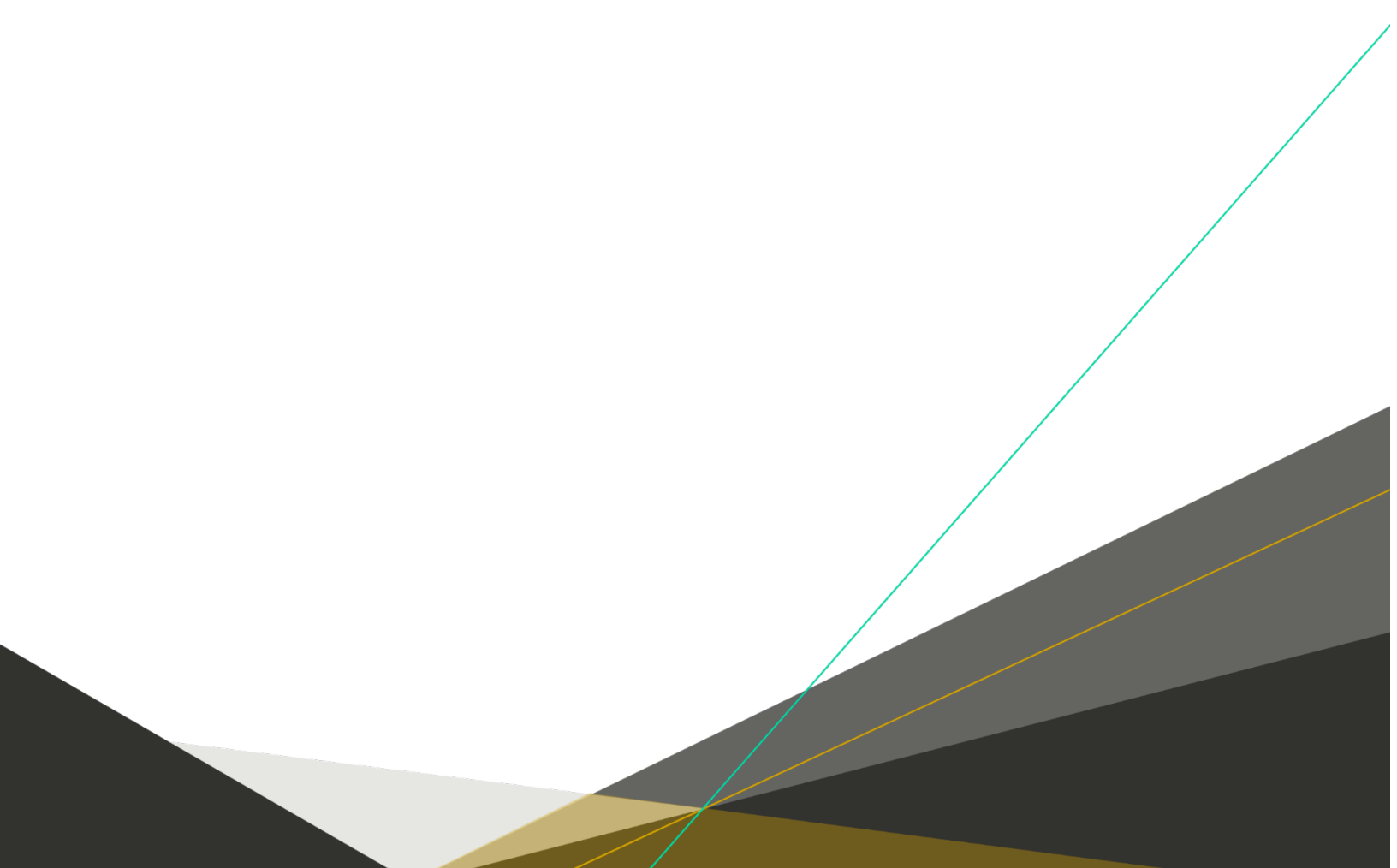
**Annexe B – Analyse environnementale des initiatives d’infrastructure de recherche numérique (IRN) de portée nationale et pan-nationale à l’appui de la gestion des données de recherche .....90**

**Annexe C – Associations internationales et nationales de gestion des données de recherche.....118**

## Remerciements

Un groupe de travail sur la gestion des données de recherche a été formé en mai 2020 pour mettre à jour l'énoncé de principe résumant l'état de la gestion des données de recherche au Canada. Il se rencontrait aux deux semaines pour superviser l'analyse de la situation et la rédaction du rapport.

Le groupe a sollicité la rétroaction de la communauté au fur et à mesure de ses travaux. Nous tenons à remercier les groupes suivants qui nous ont fait profiter de leur expertise : les analystes principaux et le conseil des chercheurs de l'Alliance, le Centre de gouvernance de l'information des Premières Nations, les coordonnateurs et coordonnatrices du réseau Portage, et la grande communauté de gestion des données de recherche convoquée par Portage et Données de recherche Canada dans leur communauté de pratique respective.



# 1 Résumé

Le présent rapport est une mise à jour de l'exposé de principe sur la gestion des données de 2017 soumis à Innovation, Sciences et Développement économique Canada (ISDE) par le Conseil du leadership sur l'infrastructure de recherche numérique (CLIRN) et est conforme au résumé public<sup>1</sup>. Le rapport résume le paysage de la gestion des données de recherche (GDR) au Canada et documente les défis et les possibilités pour l'écosystème actuel de GDR ainsi que pour l'Alliance de recherche numérique du Canada (l'Alliance). Le présent document vise à positionner l'Alliance de façon à lui permettre, en misant sur l'état actuel, de tracer la voie qui fera progresser la GDR au sein de l'infrastructure de recherche numérique (IRN) pour favoriser l'excellence en recherche.

L'expression « données de recherche » désigne toute information créée ou recueillie à titre de preuves dans la démarche scientifique, ou dont la nécessité pour valider les résultats et les constatations de la recherche est généralement acceptée dans le milieu de recherche<sup>2</sup>. Les données de recherche sont des ressources qui, si elles sont bien gérées, peuvent être réutilisées de manière innovante afin d'en tirer une plus grande valeur et de faire progresser la recherche et le savoir. La gestion de ces données fait appel à une gamme d'infrastructures et de compétences pour soutenir la documentation, le stockage, l'accès et la préservation des données au cours d'une phase de recherche et après sa conclusion. Les objectifs de la GDR sont ancrés dans le mouvement d'envergure en faveur de la science ouverte qui présente une vision d'une accélération des découvertes et des avancées scientifiques facilitées par de nouvelles technologies de l'information qui permettent que la recherche et ses données sous-jacentes soient examinées, communiquées, partagées et réutilisées de façon plus ouverte et accessible.

L'analyse de la GDR indépendamment des autres piliers du calcul informatique de pointe (CIP) et des logiciels de recherche (LR), est utile pour étudier ses fonctions, ses besoins et son incidence sur la recherche numérique; cependant, le paysage de l'IRN est, en réalité, plus complexe. Les trois piliers (GDR, CIP et LR) devraient être considérés comme étant interreliés et mutuellement habilitants, et se soutenant les uns les autres. Comme l'a fait observer le CLIRN dans son document de coordination de 2017, ces éléments individuels réalisent leur plein potentiel uniquement lorsqu'ils sont combinés afin de faciliter des flux de tâches efficaces et efficaces pour les chercheurs et les chercheuses. Une gestion efficace des données de recherche numériques repose sur un ensemble robuste d'infrastructures de soutien qui comprennent les éléments du CIP et des LR. Inversement, l'utilisation efficace du CIP et des LR exige que les données de recherche prises en charge par ces éléments soient maintenues tout au long de leur cycle de vie au moyen de pratiques de gestion efficaces, favorisant des études plus complexes, fournissant une plus grande valeur et permettant un impact accru des investissements dans l'IRN.

La croissance, au fil du temps, des initiatives, des partenariats, des réseaux et des organisations de soutien a donné lieu à un paysage canadien de la GDR de plus en plus développé et complexe. Les chercheuses et les chercheurs canadiens utilisent diverses ressources pendant le cycle de

---

<sup>1</sup> Baker, D. et al. (2019). La gestion des données de recherche au Canada : état des lieux. Zenodo. <https://zenodo.org/record/3574712#.YA7b6U-SnmE>.

<sup>2</sup> Research data [*Données de recherche*]. (s. d.) Dans *CASRAI Glossary*. Repéré en novembre 2020 à <https://casrai.org/term/research-data/> (anglais).

vie d'un projet de recherche, impliquant de multiples groupes et organisations dans la GDR. Les actrices et les acteurs impliqués concourent à l'évolution de l'écosystème de la GDR en faisant progresser les éléments au moyen d'une variété de plateformes, de services, de directives et de l'exercice de la recherche même.

Depuis le rapport du CLIRN, les initiatives du milieu de la GDR ont poursuivi leur évolution au niveau national grâce aux efforts déployés par l'Association des bibliothèques de recherche du Canada (ABRC), le réseau Portage, CANARIE, et Données de recherche Canada (DRC), qui ont réussi à faire progresser des programmes complémentaires. Le réseau Portage de l'ABRC a coordonné les efforts des partenaires institutionnels de l'ensemble du Canada et a réuni un réseau d'expertise pour développer des plateformes, des services et des directives en vue d'assurer aux établissements canadiens d'enseignement supérieur un soutien en matière de GDR. DRC a favorisé les possibilités d'un dialogue national et a convoqué une vaste panoplie de parties prenantes pour discuter d'un cadre pour les services de données nationaux et le coordonner. Miser sur les réussites de ces deux organismes pendant leur intégration dans l'Alliance exigera de demeurer pleinement conscient de la diversité des actrices et des acteurs et de leurs rôles.

## Sommaire de l'état actuel

Le présent rapport présente les entités principales intervenant dans le paysage canadien à l'échelle locale, nationale et internationale, ainsi que leurs rôles essentiels au soutien de la GDR à l'échelle nationale.

### Stockage et calcul

Du point de vue de la GDR, il existe trois configurations distinctes de l'infrastructure de calcul et de stockage qui soutiennent les phases différentes du cycle de vie des données : phase active, phase de dépôt et phase d'archivage. Ces phases diffèrent selon le but, les pratiques et le niveau de curation requis pour soutenir le stockage de données. Dans la phase de recherche active, l'infrastructure de calcul et de stockage répond aux besoins des données tout au long du processus de recherche, soit pendant la collecte, la manipulation et l'analyse. Les chercheurs et les chercheuses disposent généralement d'une infrastructure de stockage offerte par leur établissement d'attache, ainsi que d'infrastructures obtenues auprès des fournisseurs de services commerciaux ou publics. L'offre d'infrastructure est largement tributaire de la capacité institutionnelle. Les établissements et les gouvernements ont donc financé l'infrastructure de recherche à l'échelle régionale et nationale afin d'assurer un accès plus équitable à tous les groupes de recherche. Ainsi, les chercheuses et les chercheurs rattachés à des établissements d'enseignement supérieur dans l'ensemble du Canada ont aussi accès à une infrastructure qui permet le stockage des données pendant la phase de recherche active, par l'entremise de Calcul Canada et du Réseau national de recherche et d'éducation (RNRE).

Les pratiques de partage de données varient considérablement selon le domaine de recherche. En règle générale, à la conclusion d'une phase de recherche ou d'un projet financé, les chercheurs et les chercheuses prennent des décisions de curation à l'égard de la conservation et du partage des données pour répondre aux attentes de l'établissement, de l'organisme de financement ou du milieu, pour assurer la reproductibilité des résultats publiés et pour satisfaire les besoins de recherche futurs dans leurs disciplines. Ce stade du cycle de vie de la recherche repose sur le stockage des données dans un dépôt, qui favorise leur découverte et leur accès

futurs. Quelques bibliothèques universitaires au pays accordent l'accès à des dépôts conçus pour accepter divers types de données de recherche d'une vaste communauté d'utilisateurs et d'utilisatrices. Au cours des dernières années, les bibliothèques ont commencé à mettre en commun leurs ressources par l'entremise d'associations régionales et nationales en vue de développer des plateformes communes de dépôt et d'améliorer l'accès à l'infrastructure dans divers établissements universitaires présentant une capacité locale variable (p. ex. Dataverse de Scholars Portal et le Dépôt fédéré des données de recherche). À mesure que la diversité des dépôts multidisciplinaires disponibles et des modèles de service connexes prend de l'ampleur à l'intérieur et à l'extérieur des établissements, il deviendra nécessaire de se pencher sur la complexité de cette infrastructure pour soutenir les chercheurs et les chercheuses, et leurs données.

Une panoplie de dépôts de données de recherche thématiques par domaine, exploités par des organismes et des groupes de recherche situés au Canada et à l'étranger, jouent aussi un rôle important dans l'écosystème national de la GDR. Plusieurs de ces dépôts sont reconnus par leur milieu de recherche respectif, ce qui signifie que les données qu'ils contiennent sont plus susceptibles d'être découvertes par les chercheurs et les chercheuses de ces domaines, et d'inspirer la confiance de ceux-ci. De plus, vu que les données de ces dépôts sont reliées à une gamme restreinte de disciplines, des pratiques spécialisées peuvent être appliquées, multipliant ainsi les possibilités d'interopérabilité entre des ensembles de données, ainsi que la réutilisabilité des données, autant de facteurs susceptibles d'avoir un impact plus important.

Une lacune importante dans le paysage canadien de la GDR est la disponibilité et l'accessibilité d'un archivage pour favoriser la préservation à long terme des données de recherche. L'infrastructure d'archivage des données de recherche est actuellement soutenue à l'échelle locale et régionale. Une stratégie nationale en matière d'archivage qui mise sur les initiatives existantes et une approche décentralisée peut réduire les risques inhérents à la préservation responsable à long terme. Une coordination nationale entre organisations visant à superviser le stockage et la préservation des données serait nécessaire pour réaliser des économies dans un modèle décentralisé.

### Interopérabilité

L'interopérabilité entre les composantes de l'écosystème de la GDR repose sur des schémas, des normes et des protocoles communs de collecte, d'organisation et de description des données de recherche et de l'infrastructure de soutien. Afin de maximiser le potentiel des données de recherche, celles-ci doivent pouvoir être partagées de façon sécuritaire entre différents systèmes, tout en préservant leur sens et leur interprétation auprès de divers utilisateurs et utilisatrices. L'interopérabilité sémantique et technique repose sur des cadres d'exploitation qui définissent les procédures, les conditions et les relations nécessaires pour permettre un partage sans contrainte entre les composantes des infrastructures de recherche numérique. Ces infrastructures forment l'architecture de l'écosystème, laquelle permet le développement et l'intégration de nouvelles données, de nouveaux logiciels et de nouvelles infrastructures se conformant aux cadres existants.

### Services de données

L'ampleur et la complexité croissantes des besoins des chercheurs et des chercheuses en matière de GDR face aux avancées technologiques et à l'évolution des pratiques de recherche,

ainsi qu'aux nouvelles attentes des établissements, des organismes de financement et des maisons d'édition, ont stimulé le développement d'une gamme de services de soutien connexes qui s'échelonnent sur le cycle de vie des données de recherche. Un nombre croissant de services est offert au sein des établissements de recherche universitaires par l'entremise de plateformes et d'outils développés localement ou par des associations régionales et nationales. Le secteur commercial intervient aussi de plus en plus dans cette offre de services. Ces services font l'objet d'une organisation générale qui s'articule autour du cycle de vie des données de recherche, et comprennent le soutien à la planification de la gestion des données, à la curation, à la préservation, à la découverte et à l'exploration des données. Ils se développent en symbiose avec l'offre croissante d'organismes et de groupes de recherche qui implémentent leurs propres plateformes et services prenant en charge la GDR.

Le virage numérique pris par les activités de recherche a généré des besoins importants en matière de formation sur les bonnes pratiques de GDR dans les établissements d'enseignement postsecondaire. Bien que l'écart de compétences en GDR diminue suite aux efforts déployés par les établissements d'enseignement supérieur pour former les chercheurs et les chercheuses, il demeure que les ressources en GDR sont principalement concentrées dans les grandes universités. Les politiques des organismes de financement (p. ex. l'ébauche de la politique des trois organismes sur la gestion des données de recherche) représentent un des leviers utilisés pour favoriser les bonnes pratiques et le développement de ressources en GDR auprès des établissements. Les associations de recherche et les sociétés universitaires ont également un rôle important à jouer pour favoriser le développement de compétences et accroître l'offre de possibilités de formation.

## Gouvernance

Bon nombre d'organisations ont assumé un rôle de soutien dans leurs communautés de pratique en ce qui a trait à la GDR, en élaborant des directives ou des politiques, ou en développant des programmes de financement. La coordination entre ces organisations est essentielle pour favoriser la diversité des approches de soutien de la GDR qui se sont révélées efficaces. Des politiques et des exigences cohérentes pour les organismes de recherche, les infrastructures de recherche et les services connexes sont nécessaires pour garantir que les chercheurs et les chercheuses adoptent des pratiques et des cadres communs, qui, à leur tour, permettent aux systèmes soutenus à l'échelle nationale de répondre aux modèles de gouvernance nécessaires. Toutes variations sur le plan des besoins institutionnels et régionaux peuvent concourir aux défis à l'échelle nationale. Dans le paysage actuel, les répercussions des déséquilibres existants sur le plan local et régional doivent être prises en considération dans le contexte national. Il convient également d'harmoniser les stratégies nationales avec les initiatives internationales pour permettre la circulation transfrontalière des données.

## Principaux défis et possibilités

L'ampleur et le rythme croissant de la production des données de recherche, associés à la diversité des besoins et des intérêts, présentent des défis de taille en matière de soutien durable de la GDR à l'échelle nationale. Des modèles de financement durable qui tiennent compte des besoins à long terme de la GDR sont actuellement limités dans bon nombre de disciplines. Avant que l'on puisse proposer de tels modèles de financement, il convient de préciser ce que les

enveloppes du CIP, de la GD et des LR englobent au sein de l'Alliance. Parallèlement, il convient de préciser les distinctions fondamentales entre les éléments de l'écosystème de l'IRN.

Bien que le présent rapport présente un survol de haut niveau de la diversité des actrices et acteurs de l'écosystème, des infrastructures et des services soutenant la GDR au Canada, il ne brosse qu'un portrait incomplet de la situation, qui doit être peaufiné. Bon nombre des infrastructures, des plateformes et des outils existants fonctionnent en vase clos. Une meilleure intégration des infrastructures et des services nouveaux exige l'adoption de normes, de schémas et de certifications communes pour une interopérabilité fiable. Parallèlement, bon nombre des actrices et acteurs au sein de l'écosystème canadien de GDR sont relativement isolés. Des initiatives continues de consultation et de relations externes sont nécessaires pour cerner leurs besoins. Des mécanismes sont requis pour s'assurer que les fournisseurs, ainsi que les utilisateurs et les utilisatrices de tous les secteurs sont représentés et soutenus, en accordant une attention particulière à la promotion de la participation des groupes sous-représentés et des formes de recherche non occidentale. L'harmonisation avec les changements apportés dans le contexte stratégique des établissements de recherche, des organismes de financement, et des maisons d'édition est un mécanisme à utiliser pour accroître l'adoption.

Il est important d'harmoniser et d'intégrer davantage les organisations et les services, non seulement les entités canadiennes soutenues par l'Alliance, mais également leurs homologues internationaux. Déterminer la façon dont ces partenaires de l'écosystème de la GDR s'accordent est une mesure importante pour favoriser l'innovation coopérative, améliorer le soutien de la GDR, et réduire les chevauchements et le dédoublement des efforts. Le système idéal prévoirait la prestation de services à divers niveaux, ces services étant soutenus et structurés par un cadre national relié à des normes internationales et à des organisations homologues, et influencés par celles-ci.

## 2 Introduction

Nombre des jalons essentiels englobant les efforts locaux, régionaux, nationaux et internationaux font état des progrès réalisés au Canada envers la reconnaissance de la gestion des données de recherche (GDR) à titre d'élément essentiel de l'infrastructure de recherche numérique (IRN). Des initiatives comme le Comité consultatif sur l'examen du soutien fédéral à la science fondamentale<sup>3</sup>, des politiques nationales de financement comme la Déclaration de principes des trois organismes sur la gestion des données numériques<sup>4</sup>, et des groupes comme le Conseil du leadership sur l'infrastructure de recherche numérique (CLIRN) font ressortir le rôle de la GDR dans l'écosystème de l'IRN et son importance pour la promotion de l'excellence en recherche

---

<sup>3</sup> Naylor, C.D. et al. (2017). Investir dans l'avenir du Canada. Consolider les bases de la recherche au pays. <http://www.sciencereview.ca/eic/site/059.nsf/eng/home>.

<sup>4</sup> Gouvernement du Canada. (2020). Déclaration de principes des trois organismes sur la gestion des données numériques. Repéré en novembre 2020 à [https://www.ic.gc.ca/eic/site/063.nsf/fra/h\\_83F7624E.html](https://www.ic.gc.ca/eic/site/063.nsf/fra/h_83F7624E.html).



canadienne dans un environnement de recherche de plus en plus exigeant en données et en calcul.

La croissance, au fil du temps, des initiatives, des partenariats, des réseaux et des organisations de soutien a créé un paysage canadien de GDR de plus en plus développé et complexe. La situation a atteint un point décisif en novembre 2016, lorsqu'Innovation, Sciences et Développement économique Canada (ISDE) a accordé des fonds au CLIRN pour convoquer des groupes de travail se composant d'une vaste gamme de parties prenantes et charger ces groupes d'entreprendre une analyse du paysage d'IRN au Canada et de produire des exposés de principe sur l'état de la gestion des données et du calcul informatique de pointe pour la recherche, lesquels sont accompagnés de recommandations de mécanismes de coordination nationale. L'exposé de principe sur la gestion des données publié par le CLIRN en 2017 présente une vision de la coordination et de la facilitation de la GDR à l'échelle nationale, reposant sur une analyse approfondie de l'état actuel de la GDR au Canada. Cette évaluation de l'état actuel fait l'objet d'une mise à jour annuelle grâce à l'effort concerté du milieu national de GDR<sup>5</sup>.

Depuis le document produit par le CLIRN, les efforts déployés par le milieu de GDR se sont poursuivis à l'échelle nationale grâce aux initiatives du réseau Portage de l'Association des bibliothèques de recherche du Canada (ABRC) visant à développer des ressources, une expertise et une formation communes pour soutenir une communauté de pratique nationale<sup>6</sup> et à la série de Sommets sur l'encadrement des services de données nationaux (ESDN) convoqués par Données de recherche Canada<sup>7</sup>. La Déclaration de Kanata, un résultat intégrant de l'ESDN, présente une vision axée sur le milieu d'une stratégie nationale de GDR et de priorités en matière de services de données<sup>8</sup>.

À la suite du Budget fédéral de 2018 qui a engagé l'affectation de 572,5 M\$ à une stratégie nationale pour l'infrastructure de recherche numérique<sup>9</sup>, ISDE a annoncé au printemps de 2019 un programme de contribution visant à financer une nouvelle organisation d'IRN à but non lucratif, chargé de coordonner le financement et les orientations stratégiques pour les activités nationales d'IRN liées au calcul informatique de pointe (CIP) pour la recherche, à la gestion des données (GD) et aux logiciels de recherche (LR). En guise de réponse, les membres de diverses organisations de parties prenantes ont soumis une feuille de route de la GD pour 2019-2014 afin

---

<sup>5</sup> Baker, D. et al. (2019). La gestion des données de recherche au Canada : état des lieux. Zenodo. <https://zenodo.org/record/3574712#.YA7b6U-SnmE>.

<sup>6</sup> <https://portagenetwork.ca/fr/>.

<sup>7</sup> <https://www.rdc-drc.ca/fr/>.

<sup>8</sup> Participants au Sommet ESDN. (2020). Déclaration de Kanata 2020. Zenodo. <http://doi.org/10.5281/zenodo.4004831>.

<sup>9</sup> Innovation, Sciences et Développement économique Canada. (2019). Infrastructure de recherche numérique. Repéré en novembre 2020 à <https://www.ic.gc.ca/eic/site/136.nsf/fra/accueil>.

de soutenir ISDE dans ce processus. Cette feuille de route proposait les fonctions et les activités essentielles d'une organisation nationale de GDR, basées sur une vision commune<sup>10</sup>.

L'attention nationale accordée à la GDR exprime également les réorientations de la culture et des pratiques de recherche, dirigées non pas par des organisations nationales, mais par des chercheurs et des chercheuses œuvrant dans une gamme de disciplines exigeantes en données. Au Canada, il s'agit principalement de groupes de recherche en astronomie et astrophysique, en physique des hautes énergies, en sciences de la Terre, en science océanique, en sciences « omiques », en sciences biologiques et en sciences humaines numériques. Les organismes et les groupes de recherche dans l'ensemble du pays font progresser des outils et des plateformes pour gérer et analyser des données, et les partager avec des universitaires, au sein de projets de recherche complexes et dans l'ensemble de ceux-ci. Faire progresser leur propre recherche tout en recherchant les moyens d'appliquer largement leurs réussites à d'autres disciplines de recherche est un élément nécessaire de la prestation vraiment nationale de services de données. Un tel but exigera aussi des liens avec une panoplie d'organisations internationales, présentant des orientations générales et spécialisées qui favorisent des pratiques au sein des domaines de recherche et cultivent des liens solides avec le Canada.

Bien que de nombreux efforts aient commencé à se concrétiser au niveau national, la recherche est menée à une échelle internationale. Le Canada n'est pas le seul pays à reconnaître le besoin d'assurer un soutien national à l'IRN et à la GDR. La mise en œuvre de cadres communs pour les plateformes et les services afin de garantir la mobilité des données entre des infrastructures nationales et internationales sera également nécessaire pour soutenir les chercheurs et les chercheuses dans l'ensemble du Canada.

Bien que le soutien national de la GDR au Canada soit actuellement pris en charge par un groupe de base d'organisations, les initiatives de celles-ci mobilisent une diversité de collaborateurs et de collaboratrices et de fournisseurs de services qui soutiennent les chercheurs et les chercheuses et leurs données tout au long du cycle de vie de celles-ci. Le rythme auquel ce paysage évolue signifie qu'une compréhension affinée des actrices et des acteurs, de l'infrastructure et des services existants est essentielle pour les investissements futurs qui profiteront au milieu de recherche. Le présent rapport résume ce paysage complexe et décrit les mesures de soutien actuelles dans quatre principaux domaines : stockage et calcul, interopérabilité, services de données, et gouvernance. Il s'inspire de cet examen pour analyser les principaux défis et possibilités auxquels on doit faire face pour que le soutien national de la GDR porte ses fruits. Ce faisant, le présent rapport vise à positionner l'Alliance de recherche numérique du Canada (l'Alliance) pour lui permettre de miser sur les initiatives précédentes et actuelles, et de tracer une voie à suivre qui fera progresser la GDR en coordination avec d'autres éléments de l'IRN en vue de favoriser l'excellence nationale en recherche.

## Définition de « données de recherche »

Le présent rapport adopte une définition inclusive de l'expression « données de recherche », fournie par CASRAI : « *Données qui sont utilisées en tant que sources principales à l'appui de*

---

<sup>10</sup> Castle, D. et al. (2019). Position Paper – Data Management Roadmap: 2019-2024. Repéré en novembre 2020 à <https://www.rdc-drc.ca/download/position-paper-data-management-roadmap-2019-2024> (anglais).

*demandes de renseignements, de la recherche, de travaux réalisés dans le cadre d'une bourse ou d'activités artistiques touchant les sciences et la technologie, et qui sont utilisées comme éléments de preuve dans un processus de recherche ou qui sont considérées, dans la communauté de la recherche, comme nécessaires pour valider les conclusions et les résultats de recherche. Tous les autres éléments de contenu, numériques ou non, peuvent devenir des données de recherche. Les données de recherche peuvent être des données expérimentales, des données d'observation, des données opérationnelles, des données de tierces parties, des données du secteur public, des données de surveillance, des données traitées ou des données réaffectées<sup>11</sup>. »*

Bien que de nombreux types de données puissent être considérés comme des données de recherche, l'ensemble des pratiques pouvant être appliquées à leur gestion est fonction d'un ensemble de facteurs liés à la façon dont elles ont été recueillies et au but pour lequel elles l'ont été, de même qu'au sujet traité. À titre d'exemple, les données administratives dérivées de l'exploitation de systèmes administratifs saisissant des données sur les programmes, leurs opérations et des sujets connexes, sont fortement utilisées par diverses disciplines de recherche, en particulier les sciences de la santé et les sciences sociales. Par ailleurs, les données relatives aux services en ligne sont puisées dans une panoplie de sources, allant des moteurs de recherche aux transactions et aux communications en ligne, et peuvent servir à répondre à une panoplie de questions de recherche. Bien que les chercheurs et les chercheuses puissent se voir accorder l'accès à ces sources de données à des fins de recherche, leur capacité à gérer, à stocker ou à préserver ces sources est plus limitée comparativement à l'accès aux données de recherche qu'ils ou elles recueillent de première main dans le cadre d'observations ou d'expérimentations. Par conséquent, par souci de soutien des activités de recherche, il est essentiel de mobiliser et de coordonner les producteurs et les productrices de données, ainsi que les utilisateurs et les utilisatrices de données dans l'ensemble du paysage de recherche pour qu'ils et elles adoptent des pratiques rigoureuses de gestion des données.

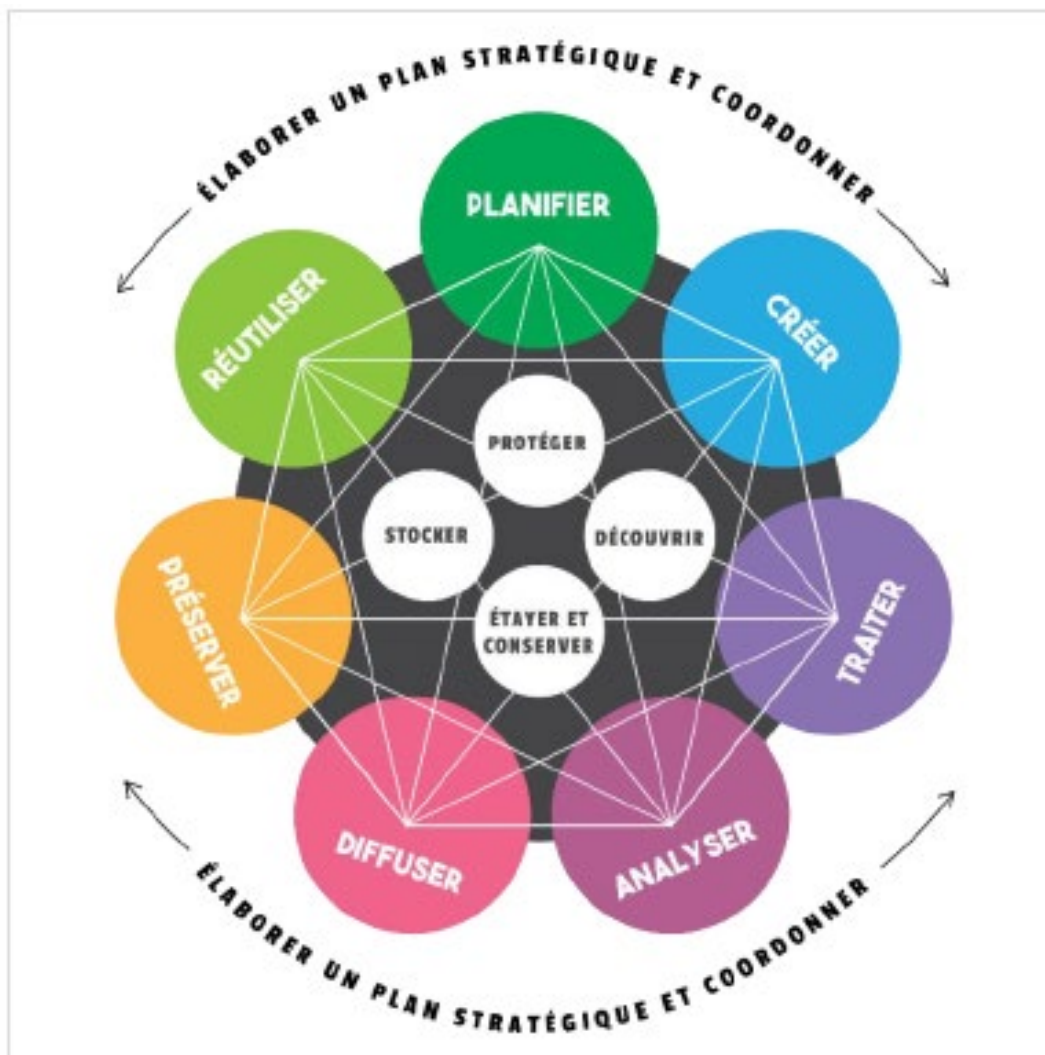
Les sources desquelles les données de recherche dérivent déterminent également comment elles devraient être gérées, en conformité avec les attentes de la communauté et les obligations connexes d'ordre éthique, juridique et commercial. En particulier, les données relatives aux communautés des Premières Nations, des Inuits et des Métis, y compris leurs peuples, leurs territoires et leurs cultures, qu'elles soient générées directement par des activités de recherche ou dérivées de sources secondaires et utilisées à des fins de recherche, doivent être gérées en conformité avec les principes de gestion des données élaborés et approuvés par ces communautés, dans le respect de la souveraineté des données des autochtones.

Par ailleurs, chaque milieu de recherche interprète à sa manière la définition de données de recherche, ce qui a une incidence sur ses propres pratiques de GDR et son attitude à l'égard de celle-ci. Favoriser une culture de recherche au Canada qui prise la gestion de toutes les formes de données de recherche réclame une approche interdisciplinaire qui soutient les domaines présentant une expérience avancée de l'utilisation des données de recherche numérique, ainsi que ceux présentant une expérience plus limitée.

---

<sup>11</sup> Research data [*Données de recherche*]. (s. d.) Dans *CASRAI Glossary*. Repéré en novembre 2020 à <https://casrai.org/term/research-data/> (anglais).

## Impact et valeur de la GDR



**Figure 1.** Fonctions de la gestion des données (cercles intérieurs) superposées aux phases du cycle de vie des données (cercles extérieurs). Dans le document *Exposé de principe sur la gestion des données : Destiné à Innovation, Sciences et Développement économique Canada*. Conseil du Conseil du leadership sur l'infrastructure de recherche numérique. Manuscrit non publié. Le 31 août 2017.

La GDR désigne la documentation, le stockage, l'accès et la préservation des données produites au cours d'une étude donnée<sup>12</sup>. Les pratiques de gestion des données portent sur le cycle de vie intégral des données de recherche, depuis la planification de l'étude jusqu'à sa tenue, depuis la sauvegarde des données au fur et à mesure qu'elles sont créées et utilisées par ceux et celles qui les ont créées jusqu'à la documentation et à la description des données en prévision de leur

<sup>12</sup> Research data management [*Gestion de données de recherche*]. (s. d.) Dans *CASRAI Glossary*. Repéré en novembre 2020 à <https://casrai-test.evision.ca/glossary-term/research-data-management/> (anglais).

partage avec d'autres collaborateurs et collaboratrices et de leur réutilisation par ceux-ci et celles-ci, et enfin, jusqu'à la préservation des documents numériques après la conclusion de l'étude de recherche. L'annexe A présente un examen plus exhaustif du cycle de vie de la GDR.

L'ère de l'information s'accompagne d'une avalanche de données, attribuable au fait que tous les principaux domaines de recherche sont devenus des producteurs et des consommateurs importants de données de recherche. Par conséquent, il est important de renforcer la capacité afin de gérer la production actuelle et de faire face à la croissance future de cette production. Cela comprend la promotion des principes FAIR, ainsi que des facteurs de préservation à long terme. Il est également important de soutenir les chercheurs et les chercheuses pendant la phase de recherche active, lorsque les processus de curation sont susceptibles de réduire le volume de données et d'en améliorer la qualité. Par exemple, Calcul Québec estime grosso modo que jusqu'à un huitième de sa capacité de stockage renferme des fichiers redondants (p. ex. fichiers en double, fichiers temporaires ou fichiers dérivés)<sup>13</sup>. Des initiatives bien financées, soutenues par un modèle coordonné à l'échelle nationale, seront essentielles pour répondre aux nombreux défis présentés par une recherche de plus en plus axée sur les données.

Un nombre croissant des initiatives entreprises par les établissements de recherche universitaires dans l'ensemble du Canada aident les chercheurs et les chercheuses à améliorer leur gestion des ressources de données de recherche<sup>14</sup>. Les organisations qui soutiennent le savoir, notamment les organismes de financement et les maisons d'édition savante reconnaissent les avantages potentiels découlant de l'élargissement de la préservation, de l'accès et de l'ouverture des données de recherche, allant de la fiabilité améliorée des résultats publiés au potentiel accru de réutilisation. À titre d'exemple, une gamme de programmes de soutien et de formation est en cours de développement par ces établissements pour aider les chercheurs et les chercheuses à planifier la gestion des données et à publier et à préserver les données<sup>15, 16</sup>. Néanmoins, des obstacles et des défis d'envergure se présentent lorsqu'il est question de traduire ce soutien descendant en pistes de solutions réelles pour la gamme diversifiée de problèmes rencontrés par les chercheurs et les chercheuses lorsqu'ils et elles créent, traitent et analysent leurs données de façon à ce que celles-ci puissent être échangées avec leurs pairs et réutilisées par eux.

Si on écarte pour l'instant les écarts et les défis réels (et très coûteux) qui accroissent la difficulté de la GDR, quels sont les avantages que l'on peut tirer des initiatives visant à surmonter ces obstacles? La « Déclaration de principes des trois organismes sur la gestion des données numériques » souligne l'importance d'une gestion efficace des données de recherche à l'appui de cet objectif :

---

<sup>13</sup> Talon, S., communication personnelle, septembre 2020.

<sup>14</sup> Cooper, A. et al. (2020). Sondage sur la capacité des services institutionnels de gestion des données de recherche. <https://doi.org/10.14288/1.0388722>.

<sup>15</sup> Tenopir, C. et al. (2014). Research data management services in academic research libraries and perceptions of librarians. *Libr Inf Sci Res*, 36(2):84-90. <https://doi.org/10.1016/j.lisr.2013.11.003> (anglais).

<sup>16</sup> Perrier, L. et al. (2017). Research data management in academic institutions: A scoping review. *PLoS One*, 12(5), e0178261. <https://doi.org/10.1371/journal.pone.0178261> (anglais).

*Les données de recherche sont recueillies au moyen de diverses méthodes, dont celles de l'expérimentation, de l'analyse, de l'échantillonnage et de la réorientation des données existantes. Elles sont de plus en plus produites ou converties dans un format numérique. Gérées adéquatement et partagées de manière responsable, ces ressources numériques permettent aux chercheurs de poser de nouvelles questions, de poursuivre de nouveaux programmes de recherche innovants, de tester d'autres hypothèses, de mettre en place des méthodologies innovantes et de collaborer avec des chercheurs dans d'autres pays et dans d'autres disciplines. La capacité à stocker, à consulter, à réutiliser et à mettre à profit les données numériques de la recherche est devenue essentielle pour l'avancement de la science et des connaissances, et l'élaboration de solutions novatrices aux défis économiques et sociaux, et offre des possibilités exceptionnelles d'améliorer la productivité, la compétitivité et la qualité de vie au Canada<sup>17</sup>.*

### Grands avantages de la gestion, du partage et de la réutilisation des données

**Accélération des progrès scientifiques** : Le partage de données permet aux chercheurs et aux chercheuses de consulter les données d'autres spécialistes de la recherche, de les comprendre et de les réutiliser à leurs propres fins scientifiques. Il accélère donc le rythme des nouvelles découvertes et prévient la collecte inutile — et dispendieuse — de données.

- De nombreux organismes de recherche au Canada ont reconnu ce potentiel et ont adopté des pratiques de recherche ouverte (p. ex. Institut de science ouverte Tanenbaum [Université McGill]<sup>18</sup>, Centre de génomique de la biodiversité [Université de Guelph]).<sup>19</sup>
- Milham et coll. (2018) utilisent l'Initiative internationale de partage des données de neuroimagerie pour fournir des données probantes directes de l'incidence du partage de données sur l'envergure des études connexes. Ils estiment que les quelque mille articles inclus dans leur analyse sont associés à des économies de coût de la production de novo de données se chiffrant entre 893 M\$ et 1 707 M\$20.
- Selon les estimations se fondant sur les statistiques publiées par le Conseil national de recherches Canada, on publie actuellement plus d'articles qui utilisent les données d'archives préservées provenant du télescope spatial Hubble, que d'articles qui utilisent de nouvelles observations<sup>21</sup>.

---

<sup>17</sup> Gouvernement du Canada. (2020). Déclaration de principes des trois organismes sur la gestion des données numériques. Repéré en novembre 2020 à [https://www.ic.gc.ca/eic/site/063.nsf/fra/h\\_83F7624E.html](https://www.ic.gc.ca/eic/site/063.nsf/fra/h_83F7624E.html).

<sup>18</sup> <https://www.mcgill.ca/neuro/open-science> (anglais).

<sup>19</sup> <https://biodiversitygenomics.net/> (anglais).

<sup>20</sup> Milham, M.P. et al. (2018). Assessment of the impact of shared brain imaging data on the scientific literature. *Nat. Commun.*, 9(1):1-7. <https://doi.org/10.1038/s41467-018-04976-1> (anglais).

<sup>21</sup> Baker, D. et al. (2019). La gestion des données de recherche au Canada : état des lieux. Zenodo. <https://zenodo.org/record/3574712#.YA7b6U-SnmE>.

**Amélioration de la collaboration** : Permet aux chercheurs et aux chercheuses de collaborer les uns avec les autres en partageant des données, des environnements de recherche et des outils.

- Les données ouvertes permettent de recombinaison les données provenant de sources hétérogènes s'étendant sur de multiples périodes et endroits afin de formuler de nouvelles questions<sup>22</sup>.
- Lorsqu'on utilise les mêmes normes pour créer, organiser, décrire et préserver les données, celles-ci sont plus interopérables et peuvent être intégrées dans des outils communs.
  - Par exemple, en tirant parti des formats communs des données neuroscientifiques, le McGill Centre for Integrative Neuroscience développe des plateformes et des outils logiciels ouvertement accessibles aux fins d'utilisation par le milieu de recherche canadien et international<sup>23</sup>.
  - Un examen effectué en 2016 des données ouvertes mises à la disposition des chercheurs et des chercheuses par l'Institut européen de bio-informatique estime un gain d'efficacité direct d'une valeur allant d'un à cinq milliards de livres sterling par an<sup>24</sup>.

**Accroissement de la visibilité et de l'impact de la recherche** : Les données dont la découverte et l'accessibilité sont assurées par un dépôt de données peuvent radicalement accroître l'impact de la recherche en question.

- La publication des données de recherche est associée à des taux de citation plus élevés. À titre d'exemple, on a constaté que les publications découlant d'essais cliniques qui partageaient les données sous-jacentes étaient citées à une fréquence de 70 % supérieure aux publications qui ne partageaient pas ces données<sup>25</sup>.
- Les initiatives comme le Dépôt fédéré des données de recherche (DFDR) stimulent la visibilité accrue des dépôts et de leurs données, et présentent un instantané national des ressources de données canadiennes<sup>26</sup>.

**Facilitation de la reproductibilité des résultats de recherche** : Lorsque les données sont archivées et partagées, les résultats peuvent faire l'objet d'un nouvel examen et les données

---

<sup>22</sup> Whitlock, M. C. (2011). Data archiving in ecology and evolution: best practices. *Trends Ecol. Evol.*, 26(2):61-65. <https://doi.org/10.1016/j.tree.2010.11.006> (anglais).

<sup>23</sup> <https://mcin.ca/> (anglais).

<sup>24</sup> Beagrie, J. et Houghton. (2016). The value and impact of the European Bioinformatics Institute: Full report. <https://beagrie.com/static/resource/EBI-impact-report.pdf> (anglais).

<sup>25</sup> Piwowar, H.A. et al. (2007). Sharing Detailed Research Data is Associated with Increased Citation Rate. *PLoS One*, 2(3), e308. <https://doi.org/10.1371/journal.pone.0000308> (anglais).

<sup>26</sup> <https://www.frdr-dfdr.ca/repo/?locale=fr>.

peuvent être utilisées pour une nouvelle analyse, améliorant ainsi la reproductibilité et la fiabilité des résultats publiés.

- La crise actuelle de reproductibilité a des répercussions tangibles et intangibles :
  - Freedman et coll. (2015) estiment que la prévalence totale de la recherche biomédicale non reproductible aux États-Unis dépasse 50 %; cela se traduit par des dépenses annuelles de 28 milliards de dollars pour une recherche préclinique qui n'est pas reproductible<sup>27</sup>.
  - Entre-temps, ce manque de reproductibilité scientifique joue un rôle important dans la réduction de la confiance du public dans la science (G7, 2019)<sup>28</sup>. Les résultats d'une enquête par sondage de Pew Research de 2019 étudiant la confiance vouée par le public américain aux experts et aux expertes scientifiques ont fait ressortir le rôle joué par les données ouvertes dans l'accroissement de la confiance dans la recherche<sup>29</sup>. Parmi les organismes de financement, le partage des données avait un impact légèrement plus important sur la confiance que la révision indépendante par les pairs.

### Gestion des données de recherche selon l'optique de la COVID-19

Le 11 mars 2020, la maladie respiratoire aiguë associée au nouveau coronavirus apparu en 2019 (COVID-19) a été officiellement déclarée pandémie mondiale par l'Organisation mondiale de la santé<sup>30</sup>. Au moment de la rédaction du présent rapport, la prévalence des sources de données numériques a permis aux chercheurs et aux chercheuses de faire face à cette crise de nombreuses perspectives, notamment médicale, sociale et environnementale. Comparativement aux urgences de santé publique antérieures, comme la flambée du SRAS de 2003, l'environnement contemporain de recherche numérique est facilité par une plus grande connectivité et une plus grande capacité de traitement, lesquelles ont donné lieu à de nombreux outils évolués de partage et d'analyse. Les chercheurs et les chercheuses adoptent aussi en plus grand nombre les pratiques de la science ouverte, grâce aux politiques de liberté d'accès comme la Politique des trois organismes sur le libre accès aux publications<sup>31</sup>. La réponse du milieu de

---

<sup>27</sup> Freedman, L. P., Cockburn, I. M., et Simcoe, T. S. (2015). The economics of reproducibility in preclinical research. *PLoS Biol*, 13(6), e1002165. <https://doi.org/10.1371/journal.pbio.1002165> (anglais).

<sup>28</sup> Summit of the G7 Science Academics. (2019). Science et confiance. Repéré en novembre 2020 à [https://www.academie-sciences.fr/pdf/rapport/Science\\_and\\_trust\\_G7\\_2019\\_FR.pdf](https://www.academie-sciences.fr/pdf/rapport/Science_and_trust_G7_2019_FR.pdf).

<sup>29</sup> Funk C., et al. (2019). Trust and Mistrust in Americans' Views of Scientific Experts. Pew Research Centre. Repéré en novembre 2020 à <https://www.pewresearch.org/science/2019/08/02/trust-and-mistrust-in-americans-views-of-scientific-experts/> (anglais).

<sup>30</sup> Organisation mondiale de la santé. (2020). Allocution liminaire du Directeur général de l'OMS lors du point presse sur la COVID-19 - 11 mars 2020. Repéré en novembre 2020 à <https://www.who.int/fr/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>.

<sup>31</sup> Gouvernement du Canada. (2016). Politique des trois organismes sur le libre accès aux publications. Repéré en novembre 2020 à [https://www.ic.gc.ca/eic/site/063.nsf/fra/h\\_F6765465.html](https://www.ic.gc.ca/eic/site/063.nsf/fra/h_F6765465.html)



recherche international pour aider à résoudre la crise constitue un exemple par excellence de l'impact de pratiques solides de gestion des données.

- Les organismes subventionnaires internationaux ont annoncé des bourses d'intervention rapide pour financer la recherche liée à la COVID-19, la majorité comportant des exigences de partage ouvert des données et des publications connexes<sup>32</sup>. Au Canada, cela comprend les programmes de financement de la recherche rapide des trois organismes<sup>33</sup>.
- Les dépôts de données de recherche et les services de curation se sont mobilisés pour soutenir la recherche liée à la COVID-19. Le plus important est le dépôt de données Zenodo pour le milieu de recherche sur la maladie COVID-19, qui offre un accès libre et gratuit aux chercheurs et aux chercheuses du monde entier et qui vise à partager les résultats de la recherche susceptibles d'être utiles pour le milieu scientifique<sup>34</sup>. La curation de ces ensembles de données est soutenue par le programme européen OpenAIRE, qui développe aussi un portail en ligne pour la COVID-19 permettant d'établir des liens entre les données de recherche sur la COVID-19 déposées dans de multiples dépôts<sup>35</sup>. Le réseau Portage de l'ABRC a également réuni les bibliothécaires et les responsables de la curation des données pour aider leurs établissements à gérer les données de recherche liées à la COVID-19 en publiant le *Guide sur le partage et le dépôt rapide des données sur la COVID-19 pour les chercheurs canadiens*<sup>36</sup>. Les milieux de recherche ont également répondu à la crise. Mentionnons, par exemple, le projet iReceptor sis à l'Université Simon Fraser, qui rend des données essentielles accessibles par le truchement de nouveaux dépôts de données et d'un pipeline de curation accélérée<sup>37</sup>.
- Les milieux de recherche collaborent à l'élaboration de pratiques exemplaires pour le partage et la réutilisation des données de santé publique sur la COVID-19. La conseillère scientifique en chef du Canada a ordonné la création d'un réseau national pour accélérer la communication et la collaboration entre les milieux des sciences, des soins de santé et des politiques pendant la crise de COVID-19. Le réseau CanCOVID qui en résulte est un réseau spécialisé de chercheurs et de chercheuses, de collaborateurs cliniciens et de collaboratrices cliniciennes, et de parties prenantes du secteur de santé de l'ensemble du Canada<sup>38</sup>. À l'échelle internationale, le COVID-19 Fast track Working Group de la Research Data Alliance a réuni des expertes et des experts internationaux issus d'un

---

<sup>32</sup> Wellcome Trust. (2020). Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak. Repéré en novembre 2020 à <https://wellcome.ac.uk/coronavirus-covid-19/open-data> (anglais).

<sup>33</sup> Instituts de recherche en santé du Canada. (2020). Coronavirus : l'intervention de recherche rapide du Canada. Repéré en novembre 2020 à <https://cihr-irsc.gc.ca/f/51890.html>.

<sup>34</sup> <https://zenodo.org/communities/covid-19/?page=1&size=20> (anglais).

<sup>35</sup> <https://www.openaire.eu/openaire-covid-19-gateway> (anglais).

<sup>36</sup> Fry, J., et al. (2020). Guide sur le partage et le dépôt rapide des données sur la COVID-19 pour les chercheurs. Zenodo. <http://doi.org/10.5281/zenodo.4047172>

<sup>37</sup> <http://ireceptor.irmacs.sfu.ca/repositories> (anglais).

<sup>38</sup> <https://cancovid.ca/fr/>.

ensemble de disciplines de recherche afin d'élaborer des directives pour la gestion, le partage et la préservation à long terme des données liées à la COVID-19<sup>39</sup>.

Il est trop tôt pour évaluer le plein impact de ces programmes et de programmes connexes sur la pandémie même, mais ces programmes signalent une réorientation évidente des attentes à l'égard de la culture de recherche et de la gestion des données de recherche présentant une importance nationale et internationale. Bien que les attentes puissent changer rapidement, il faudra du temps pour réfléchir au paysage stratégique dominant développé par les parties prenantes nationales (p. ex. EPTC 2)<sup>40</sup> jusqu'aux établissements locaux qui influent sur la façon dont les données de recherche sensibles sont gérées. Cette situation représente pour l'Alliance une occasion importante d'intervenir et d'aider à coordonner et à faire progresser les consultations.

## Mouvement de science ouverte

Les valeurs et les objectifs de la GDR sont ancrés dans le mouvement de science ouverte qui présente une vision de découvertes et de progrès scientifiques accélérés, facilités par de nouvelles technologies de l'information qui permettront le partage ouvert et accessible des publications, des résultats et des données de recherche dans le cadre d'un nouveau contrat social pour les sciences<sup>41</sup>. Le mouvement de science ouverte ne se limite pas à certains domaines de recherche, mais englobe de façon plus générale la recherche et le savoir numériques, allant des sciences humaines aux sciences physiques. La promotion de la science ouverte est motivée par diverses raisons, allant des propositions de valeur du rendement des investissements publics dans la recherche et de la maximisation des découvertes aux préoccupations liées à la reproductibilité et à la responsabilisation, et aux répercussions des nouvelles collaborations et de l'innovation<sup>42</sup>. Le soutien de ce mouvement représente une occasion de transformer tout le secteur de la recherche afin de favoriser la transparence, la responsabilisation et la confiance du public à l'égard des sciences.

Sans égard aux facteurs et aux motivations, la réalisation de la vision de la science ouverte repose sur la base d'une IRN robuste et accessible qui permet aux données sous-tendant les résultats de la recherche de respecter les principes « FAIR » selon lesquels les données doivent être faciles à trouver, accessibles, interopérables et réutilisables (d'après l'anglais, *findability*,

---

<sup>39</sup> RDA COVID-19 Working Group. (2020). RDA COVID-19 Recommendations and Guidelines on Data Sharing. <https://doi.org/10.15497/rda00046> (anglais).

<sup>40</sup> Gouvernement du Canada, Groupe en éthique de la recherche. (2018). Énoncé de politique des trois conseils : Éthique de la recherche avec des êtres humains – EPTC 2 (2018). [https://ethics.gc.ca/fra/policy-politique\\_tcps2-eptc2\\_2018.html](https://ethics.gc.ca/fra/policy-politique_tcps2-eptc2_2018.html).

<sup>41</sup> Open science [*Science ouverte*]. (s. d.) Dans *FOSTER Taxonomy*. Repéré en novembre 2020 à <https://www.fosteropenscience.eu/taxonomy/term/7> (anglais).

<sup>42</sup> Vicente-Saez, R. et Martinez-Fuentes, C. (2018). Open Science now: A systematic literature review for an integrated definition. *J. Bus. Res.*, 88:428-436 <https://doi.org/10.1016/j.jbusres.2017.12.043> (anglais).

*accessibility, interoperability et reusability*)<sup>43</sup>. Comme le font observer les auteurs des principes FAIR, une bonne gestion des données n'est pas une fin en soi, mais plutôt une condition préalable des diverses motivations du mouvement de science ouverte. À titre d'exemple, le partage et la réutilisation à grande échelle des données de recherche par le secteur universitaire sont fonction de divers facteurs : les données doivent être décrites au moyen d'une abondance de métadonnées qui exposent leur contenu, leur provenance, et leurs limitations; leur description, leur organisation et leur structure doivent être harmonisées et exploitables par machine; et elles doivent être prises en charge par des systèmes qui leur permettent d'être uniquement repérables, dûment accessibles, rattachées à des produits connexes, et préservées, le cas échéant, à long terme. La mise en œuvre de ces pratiques n'est pas une procédure statique et peut prendre diverses formes en fonction des données mêmes et de l'étendue des connaissances et des ensembles de compétences des chercheurs et des chercheuses, des responsables de la curation, des administrateurs et des administratrices de systèmes et d'autres parties prenantes. Ainsi, il convient de ne pas percevoir la GDR efficace comme case à cocher ou facteur secondaire, mais comme partie intégrante de tout travail de recherche de qualité.

La feuille de route pour la science ouverte, publiée récemment par le gouvernement fédéral, fait état de la relation entre la GDR et la science ouverte, et énonce une dizaine de recommandations visant à orienter les activités de science ouverte au Canada<sup>44</sup>. Ces recommandations reconnaissent notamment que « l'ouverture » des données ne suffit pas. Pour maximiser leurs avantages, les données de recherche doivent respecter les principes FAIR, dont une des conditions préalables est l'adoption de solides pratiques de gestion des données (voir la recommandation 5). À la lumière des liaisons avec l'IRN et la GDR, une occasion par excellence de faire preuve de leadership qui se présente à l'Alliance sera de soutenir le mouvement de science ouverte à l'échelle nationale en partenariat avec les parties prenantes pertinentes (voir la recommandation 9).

## Souveraineté des données autochtones

La souveraineté des données autochtones reconnaît les droits inhérents des communautés autochtones à gouverner la collecte, l'appropriation et l'utilisation de leurs propres données. Cette question occupe une place importante dans le paysage national de données de recherche qui doit être prise en compte l'Alliance dans ses efforts pour favoriser les pratiques exemplaires de GDR au Canada. Reconnaître la faute professionnelle passée caractérisant le traitement des communautés autochtones dans le processus de recherche, reconnaître la souveraineté des communautés autochtones sur leurs propres données et favoriser le respect des pratiques de gestion de données distinctes des communautés sont autant de facteurs qui forment un élément important de la GDR et qui favorisent également la réconciliation.

À l'échelle internationale, les efforts faits en ce sens reconnaissent que les peuples autochtones doivent gouverner leur propre information, y compris les données de recherche, en phase avec leurs intérêts propres. Selon les principes *CARE* régissant la gouvernance des données

---

<sup>43</sup> Wilkinson, M.D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18> (anglais).

<sup>44</sup> Bureau de la Conseillère scientifique en chef. (2020). Feuille de route pour la science ouverte. [http://science.gc.ca/eic/site/063.nsf/fra/h\\_97992.html](http://science.gc.ca/eic/site/063.nsf/fra/h_97992.html).

autochtones, le mouvement actuel de science ouverte ne s'harmonise pas avec les intérêts des peuples autochtones et ignore le déséquilibre des pouvoirs et les contextes historiques, et donc, énonce une série de principes complémentaires aux principes *FAIR*<sup>45</sup>. Ces principes traitent des valeurs d'avantage collectif, d'autorisation de contrôle, de responsabilité et d'éthique. Ils sont approuvés par la Global Indigenous Data Alliance (GIDA)<sup>46</sup> et complètent les principes *FAIR*, d'où l'expression consacrée en anglais « *Be FAIR and CARE* », qui signifie, littéralement, « soyez juste et attentionné ».

Les peuples autochtones au Canada font aussi valoir leurs propres principes de gouvernance des données :

### Premières Nations

Les principes PCAP® sont une ressource importante développée par les communautés des Premières Nations. Ils constituent un outil utile à la gouvernance de l'information au regard de la souveraineté des données autochtones<sup>47</sup>. Les principes de PCAP® (Propriété, Contrôle, Accès et Possession) affirment que les Premières Nations exercent le contrôle sur la collecte de données au sein de leurs communautés, qu'elles sont propriétaires de leurs données et qu'elles contrôlent la façon dont ces données peuvent être stockées, partagées et utilisées. Ces principes sont exprimés et affirmés par chaque nation selon ses protocoles, son savoir traditionnel et sa perception du monde.

### Métis

La Manitoba Métis Federation souscrit aux principes de propriété, de contrôle, d'accès et d'intendance (sigle anglais OCAS)<sup>48</sup>. Les principes reconnaissent le droit des Métis à être les propriétaires de leurs données et à prendre les décisions à l'égard de l'utilisation de celles-ci, ainsi que la responsabilité éthique liée à la planification et à la gestion appropriées des ressources de données.

### Inuits

*Qaujimaqatungit inuit* (QI) est un principe directeur qui régit les données inuites. Selon sa traduction littérale, QI désigne « *ce que les Inuits ont toujours su être vrai* » et est axé sur l'intégration d'une culture inuite traditionnelle dans des structures de gouvernance actuelles ainsi

---

<sup>45</sup> Research Data Alliance International Indigenous Data Sovereignty Interest Group. (2019). "CARE Principles for Indigenous Data Governance." *The Global Indigenous Data Alliance*. <https://www.gida-global.org/care> (anglais).

<sup>46</sup> <https://www.gida-global.org/> (anglais).

<sup>47</sup> Le Centre de Gouvernance de l'information des Premières Nations. (s. d.). Les principes de PCAP® des Premières Nations. <https://fnigc.ca/fr/les-principes-de-pcap-des-premieres-nations/>

<sup>48</sup> First Nations, Metis, and Inuit Health Research Strategic Planning Committee, University of Manitoba Faculty of Health Sciences. (2013). Framework for Research Engagement with First Nation, Metis, and Inuit Peoples. [https://umanitoba.ca/faculties/health\\_sciences/medicine/media/UofM\\_Framework\\_Report\\_web.pdf](https://umanitoba.ca/faculties/health_sciences/medicine/media/UofM_Framework_Report_web.pdf) (anglais).

que sur la réduction de la désautonomisation des Inuits<sup>49</sup>. Les six principes directeurs du QI, énoncés par les Aînés inuits et le gouvernement du Nunavut, comprennent les concepts du service, d'un processus décisionnel fondé sur le consensus, du développement de compétences et de connaissances, de collaboration, de gérance de l'environnement et d'ingéniosité pour résoudre des problèmes<sup>50</sup>.

## Gestion des données de recherche au sein de l'infrastructure de recherche numérique

L'adoption d'une approche réductrice de l'analyse de la GD à titre de pilier unique de l'IRN peut être utile pour cloisonner et comprendre ses fonctions, ses besoins, et son impact sur la recherche numérique; or, en réalité, le paysage de l'IRN est nettement plus complexe. Il convient plutôt de considérer les éléments comme étant imbriqués, s'habilitant et se soutenant les uns les autres. Comme le signale l'article de coordination de 2017 du CLIRN, ces éléments individuels réalisent leur plein potentiel uniquement lorsqu'ils sont intégrés afin de faciliter des flux de tâches efficaces et efficaces pour les chercheurs et les chercheuses. La gestion efficace des données de recherche numériques dépend d'un réseau robuste d'infrastructure de recherche numérique de soutien, qui comprend les éléments du CIP et des LR. Inversement, l'utilisation efficace du CIP et des LR exige que les données de recherche qu'ils prennent en charge soient maintenues tout au long de leur cycle de vie à l'aide de pratiques de gestion efficaces, afin de favoriser des études complexes et d'obtenir une valeur et des impacts plus importants sur la recherche et la société.

Par quels moyens la GDR soutient-elle les autres éléments de l'IRN ou compte-t-elle sur ces éléments?

### Réseau

- Le réseau soutient l'infrastructure de recherche numérique et facilite des protocoles communs de communication permettant le partage de ressources numériques entre nœuds de réseau. Ce processus dépend d'architectures, de normes et de procédures communes pour partager l'information de façon fiable et protégée.
- Pour pouvoir partager les données de recherche et les consulter sur les réseaux numériques, il faut les créer et les gérer en tenant ces buts à l'esprit. Cela a donné lieu à la création et à l'adoption d'une gamme de schémas et de normes sémantiques et techniques qui permettent de protéger, de partager, de découvrir, de préserver et de consulter les données.

---

<sup>49</sup> Tagalik, S. (2010). Inuit Qaujimagatuqangit: The role of Indigenous knowledge in supporting wellness in Inuit communities in Nunavut. <https://www.cnsa-nccah.ca/docs/health/FS-InuitQaujimagatuqangitWellnessNunavut-Tagalik-EN.pdf> (anglais).

<sup>50</sup> Nunavut Impact Review Board (s. d.) Inuit Qaujimagatuqangit. Repéré en novembre 2020 à <https://www.nirb.ca/fr/content/qaujimagatuqangit-inuit>.

### Calcul informatique de pointe pour la recherche (CIP)

- Le CIP est un hyperonyme qui désigne la panoplie des infrastructures et des services qui sont nécessaires pour une recherche exigeante en données, notamment l'accès à des ressources de calcul et à un stockage actif. Le respect des pratiques exemplaires est essentiel si l'on souhaite héberger les opérations et la gestion de grandes ressources de données, tout en protégeant les renseignements personnels et en assurant leur sécurité.
- Les infrastructures et les services de base du CIP, tel qu'ils sont définis dans l'exposé de principe de 2017 du CLIRN, n'englobent pas l'infrastructure de dépôt et d'archivage, qui sont des éléments essentiels de l'intendance et de la préservation à long terme des données de l'IRN.

### Logiciels de recherche

- La recherche numérique dépend d'une panoplie de logiciels qui aident les chercheurs et les chercheuses à recueillir, à façonner, et à analyser leurs données. Dans le même ordre d'idées, une gamme de logiciels pour la curation, la préservation, la publication et la découverte de données facilite la gestion de ces données au fil du temps.
- En contrepartie, afin de permettre une application plus vaste des logiciels de recherche et de s'assurer que ceux-ci fonctionnent avec divers intrants, les données doivent respecter des exigences dont leur gestion continue doit tenir compte, notamment le recours à des ontologies, à des normes et à des formats communs de données.

### Vision nationale du soutien de la GDR

Les efforts antérieurs déployés pour élaborer un énoncé de vision coordonné du soutien national de la GDR au Canada ont produit des énoncés de haut niveau qui expriment une panoplie d'aspirations collectives. Par exemple, l'énoncé de principe du CLIRN sur la gestion des données envisage « *Un milieu de la gestion des données de recherche novateur et bien coordonné qui proposera des ressources et des services adaptés aux chercheurs et aux chercheuses afin de les aider dans leurs travaux, indispensables à l'expansion et au maintien de la prospérité socioéconomique au Canada*<sup>51</sup>. » Par ailleurs, la Déclaration de Kanata de 2019 présente une série de déclarations des membres du milieu de la GDR, qui priorisent les exigences d'une organisation nationale de GDR ainsi que les mesures que celle-ci devrait prendre<sup>52</sup>.

Reconnaître la diversité, l'ampleur et l'étendue incroyables du milieu de recherche canadien et formuler une vision unique qui fait état de la gamme des objectifs et des capacités des organismes de recherche dans l'ensemble du pays est un défi de taille. Si on réduit les énoncés précédents pour parvenir à leurs principes de base, on constate que le milieu de recherche canadien favorise

---

<sup>51</sup> Conseil du leadership sur l'infrastructure de recherche numérique. (2017). Énoncé de principe sur la gestion des données destiné à Innovation, Sciences et Développement économique Canada, p. 26 (non publié, anglais).

<sup>52</sup> Participants au Sommet ESDN. (2020). Déclaration de Kanata 2020. Zenodo. <http://doi.org/10.5281/zenodo.4004831>.

un soutien national qui est innovateur, axé sur les chercheurs et les chercheuses, inclusif et durable. Ces quatre principes expriment également les principes directeurs de l'Alliance<sup>53</sup> qui ancrent la vision du développement futur de l'Alliance.

### Innovateur

- À mesure que la recherche devient plus exigeante en données, il convient de favoriser de nouvelles approches du soutien de la curation et de la préservation des données en vue de promouvoir la disponibilité des données à long terme. Un service national innovateur doit être flexible, être capable de reconnaître de nouvelles conditions et d'y réagir, et posséder la capacité de soutenir à partir de zéro les initiatives qui font progresser la recherche et l'écosystème de l'IRN général.

### Axé sur les chercheurs et les chercheuses

- Les besoins des chercheurs et des chercheuses devraient être au cœur du concept et de la prestation de l'infrastructure nationale et des services. Les ressources qui sont invisibles, inaccessibles et peu maniables, et qui ne s'intègrent pas dans les flux de tâches existants et avec les outils actuellement utilisés par les chercheurs et les chercheuses, afficheront une adoption médiocre. Il est essentiel de communiquer et de collaborer avec les établissements d'attache des chercheurs et des chercheuses et avec les organismes de recherche qui leur fournissent le soutien et l'infrastructure locaux auxquels ils se fient.

### Inclusif

- Le soutien national, le financement et la formation devraient être accessibles aux chercheurs et aux chercheuses, quels que soient leurs établissements et leurs domaines, tout au long du cycle de vie des données de recherche. Le soutien devrait être offert à parts égales pour les données qui permettent un partage libre, ainsi que pour diverses formes de données sensibles qui exigent des mesures de protection supplémentaires sur le plan de la protection des renseignements personnels et de la sécurité ou qui se conforment à des systèmes de gouvernance distincts.
- Une organisation nationale devrait favoriser et faciliter la communication au sein du milieu de recherche, notamment entre chercheurs et chercheuses, établissements, organismes de financement et fournisseurs de services, afin de favoriser la confiance à l'égard des priorités de la GDR et le consensus sur leur établissement, et devrait favoriser la diversité au sein de l'organisation.

### Durable

- Il convient d'inciter la mosaïque de partenariats et de modèles de financement à favoriser la résilience parmi les éléments de l'écosystème de l'IRN. Cet objectif est particulièrement valable pour les services et l'infrastructure de GDR qui soutiennent la disponibilité à long terme des données de recherche.

---

<sup>53</sup> Nouvelle organisation d'infrastructure de recherche numérique (2020). Principes directeurs. Repéré en novembre 2020 à <https://engagedri.ca/la-noirn/principes-directeurs?lang=fr>.

### 3 Soutien national actuel de la GDR

Dans un paysage décentralisé, il est essentiel que les pratiques, les politiques et les normes soient coordonnées. Au cours de la dernière décennie, Données de recherche Canada (DRC) et le réseau Portage de l'Association des bibliothèques de recherche du Canada (ABRC) ont assumé des rôles de leadership pour faciliter, à l'échelle nationale, le soutien et la coordination de la GDR, ainsi que la collaboration à son égard. Les deux organisations ont réussi à faire progresser des programmes d'action complémentaires. Le réseau Portage de l'ABRC a coordonné les efforts des partenaires institutionnels de l'ensemble du Canada et a réuni un réseau d'experts et d'expertes pour développer des plateformes, des services et des directives et pour assurer un soutien pratique aux établissements de recherche canadiens. DRC a favorisé les possibilités d'un dialogue national, convoquant une vaste gamme de parties prenantes de la GDR pour discuter de l'encadrement des services de données nationaux, et coordonner ce cadre. Ce dialogue a abouti à des déclarations ciblées, telles que la Déclaration de Kanata. Le travail et les progrès faits par ces deux organisations devraient se poursuivre sans interruption pendant que l'Alliance regroupe leurs mandats distincts en plateforme cohésive, chargée de soutenir et de faire progresser l'état de la GDR au Canada.

#### Données de recherche Canada

(<https://www.rdc-drc.ca/fr/>)

L'établissement de Données de recherche Canada (DRC) découle d'une recommandation formulée dans le rapport du Sommet de 2011 sur les données de recherche canadiennes<sup>54</sup>. DRC rassemble les principales parties prenantes afin d'échafauder une stratégie et de favoriser la communication et les partenariats en vue de faire progresser des objectifs communs, de promouvoir l'éducation et de rehausser la sensibilisation à la GDR à l'échelle nationale. DRC est une organisation pilotée et animée par les parties prenantes. Le comité directeur de DRC, qui se compose de représentants et de représentantes des organisations de parties prenantes, assure la supervision et la gouvernance des activités de DRC au nom de la communauté élargie des parties prenantes.

CANARIE héberge et soutient les activités de DRC depuis 2014, et la direction de l'organisation est assurée par un directeur exécutif depuis 2016.

DRC mène ses activités par l'entremise d'une série de comités et de groupes de travail qui ont dirigé des initiatives visant à favoriser des normes pour l'interopérabilité des données de recherche et de l'infrastructure de la GDR, et pour schématiser le paysage actuel des parties prenantes, des services et de l'infrastructure.

---

<sup>54</sup> Groupe de travail sur la stratégie des données de recherche. (2011). Portrait de la situation des données : Rapport sur le Sommet 2011 sur les données de recherche canadiennes. <https://www.rdc-drc.ca/fr/download/rapport-sur-le-sommet-2011-sur-les-donnees-de-recherche-canadiennes/?wpdmdl=1729>.



Voici les thématiques actuelles de travail :

- Communications, éducation et formation
- Infrastructure
- Politiques
- Normes et interopérabilité

DRC convoque depuis 2017 une série de sommets sur l'encadrement des services de données nationaux (ESDN) qui réunissent les membres du milieu de la GDR pour que ceux-ci et celles-ci discutent d'une action concertée visant à faire progresser l'état de la GDR au Canada, et qu'ils et qu'elles formulent des propositions à ce sujet. Les sommets ont donné lieu à divers produits essentiels, dont la Déclaration de Kanata.

DRC participe aussi activement aux initiatives internationales. Mentionnons à titre d'exemple, la collaboration avec la Research Data Alliance (RDA), dont le Conseil est actuellement coprésidé par le directeur exécutif actuel de DRC. RDA est une plateforme influente sur laquelle les expertes et les experts internationaux en données de recherche se réunissent pour échanger des opinions et promouvoir les questions liées aux pratiques exemplaires, aux normes et aux protocoles. Ce travail aboutit à une panoplie de produits, notamment les recommandations de RDA.

Pendant la transition vers l'intégration avec l'Alliance, DRC vise essentiellement à renforcer son écosystème en promouvant les mesures proposées par les sommets récents de l'ESDN, en favorisant la collaboration entre les organismes de financement de la GDR et de l'IRN, et en facilitant la collaboration internationale entre les initiatives canadiennes et mondiales de science ouverte.

## Réseau Portage, Association des bibliothèques de recherche du Canada

[\(https://portagenetwork.ca/fr/\)](https://portagenetwork.ca/fr/)

Lancé en 2015, le réseau Portage est une initiative nationale de l'ABRC et a pour objet de promouvoir une intendance partagée des données de recherche et de renforcer la capacité de GDR au Canada par le truchement d'un réseau se composant de plus de 150 membres du milieu qui forment une communauté de pratique croissante. Portage a commencé son travail en mobilisant les membres du milieu bibliothécaire universitaire et les chercheurs et les chercheuses ne relevant pas du milieu universitaire. Le but central de Portage est de coordonner et d'élargir l'expertise, les services, les outils et les plateformes pour veiller à ce que les chercheurs et les chercheuses dans l'ensemble du Canada aient accès aux mesures de soutien et aux outils dont ils et elles ont besoin pour la GDR.

La supervision et la direction de Portage sont assurées par un comité de pilotage se composant de cadres de l'ABRC, et par un comité consultatif se composant de représentants et de représentantes de groupes de parties prenantes externes essentielles.

Le réseau Portage poursuit son travail par l'entremise de groupes d'expertes et d'experts et de groupes de travail thématiques qui se penchent sur diverses étapes du cycle de vie des données de recherche. Ces groupes ont la responsabilité de mener les recherches, de produire des

directives et des pratiques exemplaires, et de superviser le développement de Portage ainsi que le soutien que le réseau accorde aux outils et aux services de GDR.

Voici les thématiques actuelles de travail :

- Curation
- Découverte de données
- Dépôts de données
- Plans de gestion des données
- Stratégies institutionnelles
- Préservation
- Renseignement sur la recherche
- Données sensibles
- Formation

Portage établit des partenariats avec des organisations alliées pour financer et soutenir le développement de ses plateformes et de ses services nationaux.

- Assistant PGD<sup>55</sup>, dont l'Université de l'Alberta est l'hôte, est un outil bilingue de préparation de plans de gestion de données. La plateforme, à laquelle les chercheurs et les chercheuses dans l'ensemble du Canada peuvent accéder gratuitement, permet aux établissements de recherche de créer des orientations et des modèles personnalisés pour leurs membres.
- Le Dépôt fédéré de données de recherche (DFDR)<sup>56</sup>, développé conjointement avec Calcul Canada, est une plateforme de dépôt fédérée et évolutive pour la gestion, la publication et la découverte de données. La plateforme sert d'interface nationale de découverte qui recueille des notices de métadonnées auprès d'autres dépôts canadiens, expose les ressources de données, et oriente le trafic vers les dépôts hôtes. Elle permet aussi aux chercheurs et aux chercheuses de déposer des ensembles de données en vue de leur curation, de leur publication et de leur préservation, et elle est conçue pour accueillir des ensembles de données qui sont trop volumineux pour être hébergés par la majorité des dépôts institutionnels ou généralistes.
- Le développement d'une instance de dépôt national Dataverse en partenariat avec le Scholars Portal du Conseil des bibliothèques universitaires de l'Ontario (CBUO)<sup>57</sup>, avec la participation de plus de 55 établissements et les directives du Groupe de travail Dataverse Nord du réseau Portage.
- Le consortium DataCite Canada<sup>58</sup>, dirigé par le réseau Portage de l'ABRC et le Réseau canadien de documentation pour la recherche (RCDR), qui offre des services de création

---

<sup>55</sup> <https://assistant.portagenetwork.ca/?locale=fr>

<sup>56</sup> <https://www.frdr-dfdr.ca/repo/?locale=fr>

<sup>57</sup> <https://dataverse.scholarsportal.info/fr/>

<sup>58</sup> <https://www.crkn-rcdr.ca/fr/consortium-datacite-canada>

de DOI (identifiants numériques d'objet) aux établissements canadiens par l'entremise de Datacite.

À ce jour, le réseau Portage a établi une base solide pour les services et les plateformes de GDR qui sont offerts par l'entremise des établissements de recherche et au moyen d'un soutien direct. Il a, en sus, aidé à rehausser le profil national de la GDR d'une façon plus générale. Jusqu'en 2019, le réseau Portage était financé par des investissements et des contributions en nature de la part des établissements membres de l'ABRC, qui subventionnaient les postes de directeur ou directrice et de gestionnaire des services chargés de superviser le développement du réseau.

En 2019, le réseau Portage a réussi à obtenir un financement de transition d'ISDE pour élargir la capacité des activités liées à la GD en prévision de la transition vers l'Alliance. CANARIE a administré ce financement au nom du réseau. À la suite de l'obtention du financement de transition, le réseau Portage a été en mesure d'étendre ses opérations grâce à une embauche stratégique pour soutenir les groupes d'expertes et d'experts essentiels et les secteurs de services connexes. Les secteurs d'intervention privilégiés comprennent le soutien du DFDR pour permettre à celui-ci d'offrir des services de découverte, de curation et de dépôt de données, le soutien du développement d'un réseau national assurant un soutien à la curation, le soutien du développement de services d'Assistant PGD, le soutien d'un programme national de formation, et la promotion du soutien d'un service national de préservation. Un deuxième cycle d'embauches prévu pour l'automne 2020 ajoutera de nouveaux postes qui assureront un soutien supplémentaire à la curation des données et à la planification de la gestion des données, ainsi qu'un soutien nouveau au renseignement sur la recherche et aux données sensibles.

## 4 Analyse du paysage national d'IRN

Une gamme d'autres parties prenantes, réparties dans l'ensemble de l'écosystème, jouent un rôle dans le soutien de la GDR nationale. Les chercheuses et les chercheurs canadiens puisent dans les mesures de soutien ou les services tout au long du cycle de vie d'un projet de recherche, ce qui a donné lieu à une pluralité de groupes et d'organisations soutenant l'intendance à long terme de nombreuses ressources de données de recherche. Il sera important de brosser un portrait exact des organisations existantes, à l'échelle locale et nationale, et de cerner la coordination internationale appropriée, pour que l'Alliance réussisse à s'intégrer dans le paysage de GDR et assure une direction et une coordination nationales appropriées.

Vu l'envergure et la complexité de l'écosystème, il est peu pratique de fournir une liste détaillée des centaines d'acteurs et d'actrices dans le contexte du présent rapport. La section qui suit expose les parties prenantes dans ce paysage qui aident à soutenir la GDR à l'échelle nationale, et précise quelques organisations à des fins contextuelles si nécessaire. Voici donc les secteurs essentiels, énumérés selon leur proximité aux chercheurs et aux chercheuses œuvrant dans les établissements universitaires.

### Enseignement supérieur

La répartition des responsabilités de l'enseignement supérieur et de la recherche au Canada entre les établissements individuels, les provinces et les territoires, et le gouvernement fédéral a produit un modèle stratifié d'organisations, d'infrastructure et de services.

Les établissements d'enseignement supérieur ont pour mission de contribuer à la société au moyen de l'éducation, de l'apprentissage et de la recherche. Une panoplie de parties travaille au sein des établissements d'enseignement supérieur pour soutenir les données de recherche créées et gérées par les chercheurs et les chercheuses dans leur milieu.

- Les facultés et les départements auxquels les chercheurs et les chercheuses sont rattachés jouent un rôle dans le soutien de la GDR. Lorsqu'ils fixent les exigences des programmes d'études et les calendriers des cours, et offrent des séminaires aux étudiants et étudiantes et au corps professoral, ils peuvent exercer une influence directe sur les pratiques de GDR adoptées par leur milieu de recherche et peuvent, de façon plus générale, favoriser une culture axée sur la science ouverte. Sur le plan de l'infrastructure, ils peuvent aussi fournir un stockage des données de recherche sur les réseaux locaux et mettre en œuvre des politiques précises régissant le stockage, la conservation et la publication de données de recherche.
- Les services des systèmes et de la TI des universités offrent un catalogue d'infrastructures et de services de calcul à l'appui de la GDR. Ces infrastructures et ces services comprennent généralement une gamme d'options de stockage et d'infrastructure de sauvegarde pour la recherche active, ainsi qu'un stockage à long terme dont les délais tiennent compte des politiques de conservation de l'établissement. Ils gèrent aussi les systèmes de gestion des identités, qui contrôlent l'accès aux fonds de données de recherche stockés dans les systèmes locaux et externes intégrés, ainsi que la sécurité de ces fonds.

- Les bibliothèques et les archives universitaires soutiennent la GDR à titre de secteur de services nouveau, en concordance avec la reconnaissance croissante que les données de recherche représentent des productions savantes importantes qui doivent être prisées et gérées de concert avec d'autres objets savants plus traditionnels. Ce soutien s'articule autour de l'hébergement de systèmes de gestion de données numériques, tels les dépôts et les plateformes de collecte numérique des établissements, ainsi que des services connexes de communications savantes et de droits d'auteur. Les services en amont du cycle de vie de la recherche, tel que le soutien à la planification de la gestion des données et à la curation des données, sont apparus pour améliorer la qualité des données archivées en aval. À mesure que de plus en plus de données de recherche sont intégrées dans leurs collections, ces établissements soutiennent aussi les données en aval au moyen de la prestation de services de préservation numérique.
- Les bureaux de recherche soutiennent la GDR en énonçant les politiques et les stratégies institutionnelles liées aux pratiques et à l'administration de la recherche. Cela comprend le soutien des comités d'éthique de la recherche qui s'assurent que la recherche impliquant des sujets humains respecte les pratiques exemplaires institutionnelles et nationales établies. Les bureaux de recherche aident aussi les chercheurs et les chercheuses à se conformer aux politiques et aux exigences en matière de GDR fixées par les organismes pertinents d'octroi de subventions et de bourses.

La coordination et la direction à l'échelle nationale de chacun de ces groupes sont assurées par des associations nationales, dont l'Association canadienne des administratrices et des administrateurs de recherche (ACAAR)<sup>59</sup>, l'Association canadienne des comités d'éthique de la recherche (ACCER)<sup>60</sup>, l'Association des bibliothèques de recherche du Canada (ABRC)<sup>61</sup>, et le regroupement des dirigeants et des dirigeantes des technologies de l'information du secteur de l'enseignement supérieur au Canada (CUCCIO)<sup>62</sup>.

La mesure dans laquelle les parties locales sont en mesure d'offrir un soutien varie énormément selon la capacité et l'intensité de recherche d'un établissement. Des fournisseurs de services régionaux et nationaux sont apparus en appui aux chercheurs et chercheuses pour coordonner les pratiques et les normes dans l'ensemble des établissements, améliorer l'accès à l'infrastructure et aux services, et tirer les avantages des économies d'échelle.

---

<sup>59</sup> <https://cara-acaar.ca/home?lang=fr>

<sup>60</sup> <https://www.careb-accr.org/fr>

<sup>61</sup> <https://www.carl-abrc.ca/fr/>

<sup>62</sup> <https://www.cuccio.net/fr/>

- En partenariat avec les organismes régionaux WestGrid<sup>63</sup>, Calcul Ontario<sup>64</sup>, Calcul Québec<sup>65</sup>, et ACENET<sup>66</sup>, Calcul Canada<sup>67</sup> déploie l'infrastructure de stockage soutenant la GD de nombreux projets de recherche actifs. Calcul Canada et le réseau Portage de l'ABRC sont également des partenaires dans la plateforme nationale du DFDR.
- Le réseau national de recherche et d'éducation (RNRE) est un regroupement essentiel d'infrastructures, d'outils et de personnes au service de la recherche et de l'enseignement supérieur. Le RNRE canadien<sup>68</sup> est formé de CANARIE et de douze partenaires. Ces ressources de réseau facilitent la connectivité et la sécurité de bon nombre de plateformes et de services de GD, ainsi que l'accès à ceux-ci. Quelques partenaires du RNRE offrent aussi à leurs établissements membres une infrastructure de stockage de longue durée.
- Les consortiums régionaux de bibliothèques universitaires, dont le Conseil des bibliothèques universitaires de l'Ontario (OCUL)<sup>69</sup> et le Council of Prairie and Pacific University Libraries (COPPUL),<sup>70</sup> offrent à leurs établissements membres un accès à une infrastructure et à des services partagés. Cette mesure de soutien comprend l'accès à des plateformes de dépôt (p. ex. Dataverse de Scholars Portal)<sup>71</sup> et à une infrastructure de stockage et de préservation (p. ex. Ontario Library Research Cloud<sup>72</sup>, Permafrost<sup>73</sup> et WestVault<sup>74</sup>).
- Le Réseau canadien de documentation pour la recherche favorise l'accès accru à des ressources savantes en négociant des licences au nom de ses membres<sup>75</sup>, comme l'accès

---

<sup>63</sup> <https://www.westgrid.ca/> (anglais)

<sup>64</sup> <https://computeontario.ca/> (anglais)

<sup>65</sup> <https://www.calculquebec.ca/>

<sup>66</sup> <https://www.ace-net.ca/> (anglais)

<sup>67</sup> <https://www.computecanada.ca/>

<sup>68</sup> <https://www.canarie.ca/fr/reseau/rnre/>

<sup>69</sup> <https://ocul.on.ca/> (anglais)

<sup>70</sup> <https://coppul.ca/> (anglais)

<sup>71</sup> <https://dataverse.scholarsportal.info/fr/>

<sup>72</sup> <https://cloud.scholarsportal.info/> (anglais)

<sup>73</sup> <https://permafrost.scholarsportal.info/> (anglais)

<sup>74</sup> <https://coppul.ca/westvault> (anglais)

<sup>75</sup> <https://www.crkn-rcdr.ca/fr>

à des outils d'identifiant permanent par le truchement du consortium ORCID-CA<sup>76</sup> et du Consortium DataCite Canada<sup>77</sup>, offert en partenariat avec l'ABRC.

## Organismes de recherche

Faciliter l'accès à des données de recherche de pointe présente une panoplie de facteurs et de défis liés à la GDR, et exige en général l'accès à des ressources d'IRN considérables. Dans l'ensemble du pays, les organismes de recherche responsables de la création et de l'intendance de nombreuses ressources de données nationales desquelles les chercheurs et les chercheuses dépendent qui les considèrent comme des sources de données essentielles, ont joué un rôle déterminant dans le développement de services et d'outils innovateurs pour faire progresser la GDR au sein de leur discipline, et ont profité à des chercheuses et à des chercheurs éloignés. En règle générale, ces organismes sont hébergés et pris en charge par les établissements universitaires et les organes de recherche du gouvernement, et accaparent une quantité importante de fonds et de ressources d'IRN.

À ce jour, un inventaire détaillé de ces organismes, de ces infrastructures et de ces ressources de données fait défaut dans une évaluation du paysage canadien, et devrait être priorisé pour favoriser la planification stratégique future de l'Alliance. Étant donné que bon nombre des initiatives canadiennes spécialisées de GDR collaborent avec leurs homologues dans d'autres pays, l'inventaire devrait préciser les liens entre les initiatives nationales et les fédérations et les agrégateurs internationaux. Il conviendrait de miser sur les efforts de divers groupes de travail du réseau Portage et de DRC pour accélérer ce processus.

La section suivante résume diverses initiatives existantes de GDR financées à l'échelle nationale, selon les disciplines de recherche essentielles. **Cette liste ne se veut pas exhaustive, mais vise plutôt à illustrer la diversité et la complexité des initiatives.** Une relation axée sur la collaboration entre l'Alliance et ces organismes et des organismes similaires sera nécessaire pour coordonner l'écosystème national de GDR et établir des partenariats mutuellement avantageux.

## La Terre, les océans et l'environnement

La vaste étendue et la diversité des climats du Canada contribuent à la production d'une multitude de données tirées de l'infrastructure de recherche d'observation et accompagnées d'une modélisation prédictive, qui soutiennent la recherche sur la conservation, la gestion environnementale et l'exploitation de ressources. La gamme d'organismes qui participe à la collecte et à la gestion de ces données se fie à des ressources importantes d'IRN et assure une direction internationale de la promotion de normes favorisant l'interopérabilité qui transcende les frontières juridiques et spatiales.

---

<sup>76</sup> <https://orcid-ca.org/fr>

<sup>77</sup> <https://www.crkn-rcdr.ca/fr/consortium-datecite-canada>

### *Système intégré d'observation des océans du Canada (SIOOC)<sup>78</sup>*

En 2019, le gouvernement du Canada a annoncé un investissement annuel de 1,5 M\$ à titre de financement continu d'un Système intégré d'observation des océans du Canada afin de promouvoir le partage de données et d'expertise à l'appui des travaux de recherche visant à mieux comprendre, surveiller et gérer les activités dans les océans du Canada<sup>79</sup>. L'initiative est financée conjointement par le Marine Environmental Observation, Prediction and Response Network (MEOPAR), qui fournit 2 M\$ sur une période de quatre ans. Les partenaires comprennent des organismes de recherche de l'ensemble du pays, notamment l'Ocean Frontier Institute, l'Université Dalhousie, le Coastal and Ocean Information Network Atlantic, le Fisheries and Marine Institute de l'Université Memorial de Terre-Neuve, l'Ocean Tracking Network, l'Observatoire global du Saint-Laurent, la Tula Foundation, et Ocean Networks Canada (Université de Victoria), qui mettront en œuvre la première phase du système. Cette initiative représente un partenariat régional avec le Système mondial d'observation de l'océan (SMOO)<sup>80</sup>, un programme réalisé par la Commission océanographique intergouvernementale (COI) de l'UNESCO.

### *Catalogue des données polaires<sup>81</sup>*

Le Canadian Cryospheric Information Network a été établi dans les années 1990 grâce à un partenariat de collaboration entre divers ministères du gouvernement du Canada, l'Université de Waterloo et le secteur privé pour faciliter l'échange d'information entre les chercheurs et les chercheuses, les localités du Nord, les programmes internationaux et le public. Le Catalogue des données polaires (PDC, d'après l'anglais *Polar Data Catalogue*) est leur dépôt des métadonnées et des données, qui décrit les ensembles de données produits par les chercheuses et les chercheurs se spécialisant dans les régions de l'Arctique et de l'Antarctique, et qui accorde l'accès à ces ensembles. Les dossiers portent sur de nombreux sujets, allant des sciences naturelles et sciences de la santé, aux sciences sociales et aux politiques. Le Catalogue des données polaires est membre du Scientific Committee on Antarctic Research (SCAR) (communément appelé en français le Comité scientifique pour les recherches antarctiques)<sup>82</sup>, qui est un organisme thématique du Conseil international des sciences (CIS). Le SCAR reconnaît le catalogue comme le National Antarctic Data Centre (NADC) pour le Canada, dans le cadre de ses obligations de rendre les données disponibles en vertu du Traité sur l'Antarctique (1959). À

---

<sup>78</sup> <https://cioos.ca/fr/accueil/>.

<sup>79</sup> Gouvernement du Canada, Pêches et Océans Canada. (2019). L'investissement du gouvernement du Canada dans la technologie d'observation des océans contribue à rendre les eaux côtières plus sécuritaires et à accroître la résilience des collectivités côtières. Repéré en novembre 2020 à <https://www.canada.ca/fr/peches-occeans/nouvelles/2019/03/linvestissement-du-gouvernement-du-canada-dans-la-technologie-dobservation-des-occeans-contribue-a-rendre-les-eaux-cotieres-plus-securitaires-et-a-a.html>.

<sup>80</sup> <https://www.gooseocean.org/> (anglais).

<sup>81</sup> <https://www.polardata.ca/> (anglais).

<sup>82</sup> <https://www.scar.org/> (anglais).



son tour, le SCAR est un partenaire de l'International Arctic Science Committee<sup>83</sup>.

### L'astronomie et l'astrophysique

L'envergure de la production et de la gestion de vastes ensembles de données astronomiques exige un investissement considérable dans l'IRN, qui est réalisable grâce à des collaborations nationales et internationales. Cette réalité a favorisé une culture de partage coopératif de données parmi les chercheurs et les chercheuses en astronomie et en astrophysique et, ce faisant, a donné lieu à la création de diverses organisations nationales et internationales se spécialisant dans la résolution des défis liés à l'interopérabilité et à la gestion à long terme de ces ensembles de données.

#### *Centre canadien de données astronomiques<sup>84</sup>*

Le Centre canadien de données astronomiques (CCDA) a été établi en 1986 par le Conseil national de recherches Canada grâce à une subvention de l'Agence spatiale canadienne. En partenariat avec Services partagés Canada, Calcul Canada, CANARIE et le milieu universitaire (financé par la Fondation canadienne pour l'innovation ou FCI), le CCDA offre des services d'informatique en nuage, de stockage géré par les utilisateurs et les utilisatrices, de gestion de groupes d'utilisateurs et d'utilisatrices et de publication de données, en sus de sa mission continue de fournir un stockage permanent d'importantes collections de données. En 2019, le CCDA a livré plus de 2 pétaoctets de données à des milliers d'astronomes au Canada et dans 80 autres pays. Le CCDA est membre d'IVOA, l'International Virtual Observatory Alliance<sup>85</sup> (Alliance internationale des observatoires virtuels) qui a créé pour les centres de données un cadre standard pour la prestation de services de données interopérables et d'analyse et un logiciel de visualisation des données dans une interface utilisateur conçue de manière à soutenir les chercheurs et les chercheuses à l'échelle mondiale.

### La physique nucléaire et la physique des particules

La recherche sur la physique des hautes énergies exige des investissements importants dans l'infrastructure, lesquels sont réalisables uniquement grâce à une collaboration internationale. Les stratégies de gestion de ces énormes quantités de données reposent sur un investissement et une collaboration d'envergure nationale.

#### *TRIUMF/ATLAS-T1<sup>86</sup>*

TRIUMF est le centre national d'accélération des particules du Canada. Il est situé à l'Université de la Colombie-Britannique et est régi par des membres universitaires de l'ensemble du pays. TRIUMF accueille aussi le Centre de données de données de niveau un ATLAS-Canada, qui est financé par la FCI. Le Centre est l'un des principaux centres de données de CERN, qui distribue

---

<sup>83</sup> <https://iasc.info/> (anglais).

<sup>84</sup> <https://www.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/fr/>.

<sup>85</sup> <http://ivoa.net> (anglais).

<sup>86</sup> <https://www.triumf.ca/atlas-group> (anglais).

les données du Grand collisionneur de hadrons (LHC) par le truchement d'une grille de calcul mondiale aux fins d'analyse par les chercheuses et les chercheurs canadiens et internationaux.

### Sciences de la vie

Les avancées technologiques dans les sciences de la vie, en particulier les sciences « omiques », ont révolutionné les approches de recherche sur les organismes vivants. La production de la majorité des données biomédicales est fortement répartie. Or, les technologies qui sont requises pour acquérir, stocker et partager les renseignements numériques ont exigé des infrastructures collectives ouvertes au milieu scientifique.

#### *Barcode of Life Data System*<sup>87</sup>

Le Barcode of Life Data System (BOLD) est un dépôt de données et un environnement de recherche sur la bio-informatique en ligne qui se consacre à l'étude des données de code-barres ADN développées au Centre de génomique de la biodiversité à l'Université de Guelph et à l'utilisation de ces données. La plateforme offre l'accès à plus de 8 millions de code-barres de plus de 300 000 espèces. Les organismes de financement comprennent Génome Canada par l'entremise de l'Ontario Genomics Institute, le Fonds ontarien pour l'innovation et le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG)<sup>88</sup>.

#### *BioGRID*<sup>89</sup>

Le Biological General Repository for Interaction Datasets (BioGRID) (référentiel général biologique des ensembles de données d'interaction) est une base de données publique qui archive et partage des données d'interaction génétique et d'interaction de protéines provenant d'organismes et d'humains modèles. Le référentiel renferme actuellement plus de 1 740 000 interactions puisées dans des ensembles de données à débit de traitement élevé et des études individuelles, et dérivées de plus de 70 000 publications dans la littérature principale. Le référentiel intègre les données provenant de dizaines d'organismes modèles et de bases de données sur les interactions. Les établissements canadiens hôtes sont le Lunenfeld-Tanenbaum Research Institute au Sinai Hospital à Toronto et l'Université de Montréal grâce à un financement fourni par les Instituts de recherche en santé du Canada (IRSC) et les National Institutes of Health aux États-Unis.

### Sciences de la santé et médecine

Une mosaïque d'organismes de recherche, d'hôpitaux de recherche et d'autorités sanitaires gouvernementales font progresser la recherche sur de nouvelles thérapeutiques, les études cliniques et la santé et la sécurité publiques par le truchement de nouvelles technologies. Ce travail exige des pratiques de gestion rigoureuses qui doivent concilier des stratégies responsables en mesure de maintenir la sécurité des données et la confiance du public.

---

<sup>87</sup> <http://v4.boldsystems.org/> (anglais).

<sup>88</sup> Ratnasingham, S. & Hebert, P.D.N. (2007). BARCODING: bold: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Mol. Ecol. Notes*, 7(3):355-364. <http://doi.org/10.1111/j.1471-8286.2007.01678.x> (anglais).

<sup>89</sup> <https://thebiogrid.org/> (anglais).

### *Brain-CODE*<sup>90</sup>

La plateforme de neuro-informatique Brain-CODE est dirigée par l'Institut ontarien du cerveau afin de favoriser l'acquisition et le stockage de données multidimensionnelles recueillies auprès de patients présentant divers troubles du cerveau, ainsi que l'accès à ces données. Le développement de Brain-CODE est soutenu par une série d'organismes du secteur public et d'organismes à but non lucratif. La base de données est hébergée par le Centre for Academic Computing.

### *iReceptor*<sup>91</sup>

iReceptor est une passerelle scientifique qui facilite la découverte, l'analyse et le partage des répertoires d'anticorps (lymphocytes B et T) (séquences du répertoire de récepteurs immuns adaptatifs ou AIRR-seq en anglais), ainsi que l'accès à ces répertoires, lesquels proviennent de nombreux laboratoires et établissements. La passerelle intègre d'importants ensembles de données répartis conformément à des normes communautaires en matière d'interopérabilité et de partage, permettant aux utilisateurs et aux utilisatrices de lancer une recherche dans deux milliards de séquences. Le projet se déroule au Centre IRMACS à l'Université Simon Fraser et est financé par le programme des plateformes sur réseau de CANARIE, la FCI, le BC Knowledge Development Fund, les IRSC, et le programme de recherche et d'innovation Horizon 2020 de l'Union européenne.

### *CanDIG*<sup>92</sup>

La Canadian Distributed Infrastructure for Genomics (CanDIG) est un projet qui construit une plateforme de génomique de santé pour la réalisation d'analyses fédérées nationales impliquant des ensembles de données privés sous contrôle local. Le projet CHORD vise à établir un service de données fédéré à l'échelle nationale pour les données sensibles sur la génomique et la santé. CanDIG est financé par le programme de cyberinfrastructure de la FCI et raccorde des sites à l'Université McGill, au Hospital for Sick Children, au UHN Princess Margaret Cancer Centre, au Michael Smith Genome Sciences Centre, à l'Hôpital général juif et à l'Université de Sherbrooke. Le projet représente aussi une collaboration avec Génome Canada, Calcul Canada et CANARIE.

## Sciences sociales et humaines (SSH)

Les infrastructures de recherche permettent la création, la manipulation et la gestion d'importants ensembles de données hétérogènes dans les disciplines des SSH. Les organismes de recherche font progresser le travail des humanistes numériques qui publient et gèrent d'importants corpus numériques et d'autres documents culturels, tout en soutenant le travail des humanistes œuvrant dans des milieux non numériques, qui font face à des attentes croissantes en matière de partage de données. Dans les sciences sociales, les organismes de recherche bien établis qui produisent et compilent les données issues de statistiques publiques, de sondages scientifiques et de sondages d'opinion sont des sources de données essentielles pour le milieu de recherche canadien.

---

<sup>90</sup> <https://www.braincode.ca/> (anglais).

<sup>91</sup> <https://gateway.ireceptor.org/> (anglais).

<sup>92</sup> <https://www.distributedgenomics.ca/> (anglais).

### *Réseau canadien des Centres de données de recherche*<sup>93</sup>

Le Réseau canadien des Centres de données de recherche (RCCDR) découle d'un partenariat établi entre un consortium d'universités canadiennes et Statistique Canada. Il a pour objet d'accorder aux chercheuses et aux chercheurs universitaires et gouvernementaux et à d'autres spécialistes de la recherche approuvés l'accès à des microdonnées confidentielles, sociales, économiques et sanitaires stockées dans des installations informatiques protégées situées sur des campus universitaires à travers le pays. Sis à l'Université McMaster depuis 2010, il comporte 32 centres et antennes et accorde aux chercheurs et aux chercheuses un accès unique et protégé aux données issues des enquêtes et des recensements de Statistique Canada, ainsi qu'à un nombre croissant de fichiers de données protégées administratives. On reconnaît de plus en plus que bien que le dépôt ou l'échange de microdonnées protégées ne soit pas possible par le truchement de dépôts ouverts, cela n'est pas le cas de la documentation et du code connexes. L'exploitation du réseau est régie par une entente officielle avec Statistique Canada. Le Conseil de recherches en sciences humaines (CRSH), les IRSC et la FCI fournissent des fonds, tandis que les universités hôtes et Statistique Canada fournissent un soutien en espèces et en nature. Dans le cadre de son Plan stratégique pour 2019-2024, le Réseau s'est engagé à renforcer les relations stratégiques avec des organisations internationales stratégiquement complémentaires<sup>94</sup>.

### *Centre de gouvernance de l'information des Premières Nations*<sup>95</sup>

Le Centre de gouvernance de l'information des Premières Nations (CGIPN) est un organisme à but non lucratif qui vise à garantir aux Premières Nations au Canada la souveraineté des données. Dans le cadre d'une collaboration avec ses partenaires, le centre supervise des initiatives importantes de collecte de données ayant pour but d'examiner la santé et le bien-être des membres et des communautés des Premières Nations. Mentionnons à titre d'exemple l'Enquête régionale sur la santé des Premières Nations et l'Enquête régionale sur la petite enfance, l'éducation et l'emploi. Le CGIPN est également responsable de l'intendance des principes de PCAP® (propriété, contrôle, accès et possession) et entreprend la formation et l'éducation connexes<sup>96</sup>.

---

<sup>93</sup> <https://crdcn.org/fr>

<sup>94</sup> Réseau canadien des Centres de données de recherche. (s. d.). Plan stratégique 2019 – 2024. [https://crdcn.org/sites/default/files/uploads/crdcn\\_stratplan\\_fr.pdf](https://crdcn.org/sites/default/files/uploads/crdcn_stratplan_fr.pdf).

<sup>95</sup> <https://fnigc.ca/fr/>.

<sup>96</sup> Le Centre de Gouvernance de l'information des Premières Nations. (s. d.). Les principes de PCAP® des Premières Nations. <https://fnigc.ca/fr/les-principes-de-pcap-des-premieres-nations/>.

## Coalition Publica<sup>97</sup> et CO.SHS<sup>98</sup>

Coalition Publica est un partenariat entre deux plateformes canadiennes de publication ouverte, Érudit<sup>99</sup> et Public Knowledge Project<sup>100</sup>, qui vise à faire progresser le milieu canadien de l'édition savante dans le domaine des SSH vers un libre accès par le truchement du développement technologique partagé et du soutien des activités de recherche portant sur l'écosystème de l'édition savante. CO.SHS, une initiative complémentaire d'Érudit, est un projet d'infrastructure de recherche ouverte qui vise à accroître la découvrabilité de la recherche diffusée sur la plateforme d'Érudit et à favoriser l'exploration de ce corpus d'études à l'aide d'outils avancés d'analyse et de visualisation.

## Organismes de financement de la recherche

Le mandat des organismes de financement, qui impose l'obligation de partager les données de recherche, influe fortement sur le comportement des chercheurs et des chercheuses et la demande d'infrastructures et de services de GDR. Au cours des dix dernières années, les organismes de financement et les gouvernements dans le monde entier ont reconnu le besoin d'élaborer des politiques nationales de GDR afin de favoriser l'accès aux données subventionnées par l'État. En 2016, les trois organismes canadiens ont publié une Déclaration de principes des trois organismes sur la gestion des données numériques qui expose leurs attentes en matière de GDR, ainsi que les responsabilités des divers acteurs et actrices à l'égard de la satisfaction de ces attentes<sup>101</sup>. En 2018, une ébauche de politique sur la GDR a été annoncée, qui obligerait tous les titulaires d'une subvention à... « *déposer dans un dépôt numérique reconnu, les données de recherche numérique, les métadonnées et les codes qui appuient directement les conclusions de la recherche... d'autres résultats de la recherche qui découlent de la recherche appuyée par les organismes subventionnaires,* » tout en respectant les exigences éthiques, juridiques et commerciales et adhérant aux principes de la souveraineté des données autochtones<sup>102</sup>.

Les établissements universitaires dans l'ensemble du pays ont déjà commencé à répondre à l'ébauche de politique sur la GDR des trois organismes pour soutenir les chercheuses et les chercheurs locaux<sup>103</sup>. Cet effort a obtenu un soutien considérable du réseau Portage, qui a

---

<sup>97</sup> <https://www.coalition-publi.ca/le-projet>

<sup>98</sup> <https://co-shs.ca/fr/>

<sup>99</sup> <https://www.erudit.org/fr/>

<sup>100</sup> <https://pkp.sfu.ca/ojs/> (anglais)

<sup>101</sup> Gouvernement du Canada. (2020). Déclaration de principes des trois organismes sur la gestion des données numériques. Repéré en novembre 2020 à [https://www.ic.gc.ca/eic/site/063.nsf/fra/h\\_83F7624E.html](https://www.ic.gc.ca/eic/site/063.nsf/fra/h_83F7624E.html).

<sup>102</sup> Gouvernement du Canada. (2020). L'Ébauche de politique des trois organismes sur la gestion des données de recherche. Repéré en novembre 2020 à [https://www.ic.gc.ca/eic/site/063.nsf/fra/h\\_97610.html](https://www.ic.gc.ca/eic/site/063.nsf/fra/h_97610.html).

<sup>103</sup> Cooper, A. et al. (2020). Sondage sur la capacité des services institutionnels de gestion des données de recherche. <https://doi.org/10.14288/1.0388722>.

développé une panoplie de plateformes, d'outils, de services et de formations pour aider les établissements et les chercheurs et les chercheuses à respecter les trois piliers de la politique. Mentionnons notamment un outil de plan de gestion des données (PGD) ainsi que les directives connexes, deux options de dépôt généraliste qui permettent le dépôt de données, et des modèles et des directives pour l'élaboration de stratégies institutionnelles. Le réseau Portage a également suivi les progrès réalisés dans le développement d'une capacité institutionnelle visant à répondre à l'ébauche de politique des trois organismes sur la GDR par le truchement de quelques sondages ciblés. Il reste à voir comment les trois organismes aideront les chercheurs et les chercheuses à respecter la politique, bien qu'une série de subventions axées sur la GDR ait été mise en œuvre avec succès par le CRSH pour aider les établissements et les chercheurs et les chercheuses à renforcer la capacité en guise de préparation<sup>104</sup>.

Les organismes provinciaux et territoriaux de financement de la recherche (p. ex. Fonds de recherche du Québec<sup>105</sup>, Research Manitoba<sup>106</sup> et Fonds pour la recherche en Ontario<sup>107</sup>) n'ont pas encore adopté des politiques ou mis en œuvre des exigences en matière de GDR. Un effort concerté de mobilisation des organismes de financement provinciaux et territoriaux afin d'éclairer la convergence des exigences et des politiques profiterait à la recherche pancanadienne. Des modèles de collaboration sont proposés par les associations de recherche existantes (p. ex. Alliance canadienne des organismes provinciaux de la recherche en santé<sup>108</sup>).

À l'échelle internationale, les organismes de financement de la recherche publique aux États-Unis et au Royaume-Uni<sup>109</sup>, ainsi que d'autres sources de financement de prestige (p. ex. Wellcome Trust<sup>110</sup> et la Bill & Melinda Gates Foundation<sup>111</sup>) obligent déjà les chercheurs et les chercheuses à pratiquer une bonne gestion des données et à partager les ensembles de données sous-jacents, à titre de condition d'obtention de fonds.

À l'avenir, l'Alliance devra se tenir au courant du contexte des organismes de financement afin de s'assurer que ses services et son infrastructure peuvent aider les chercheurs et les

---

<sup>104</sup> Conseil de recherches en sciences humaines. (2020). Initiative visant le renforcement de la capacité de gestion des données de recherche. Repéré en novembre 2020 à [https://www.sshrc-crsh.gc.ca/funding-financement/programs-programmes/data\\_management-gestion\\_des\\_donnees-fra.aspx](https://www.sshrc-crsh.gc.ca/funding-financement/programs-programmes/data_management-gestion_des_donnees-fra.aspx).

<sup>105</sup> <http://www.frqsc.gouv.qc.ca/accueil>.

<sup>106</sup> <https://researchmanitoba.ca/> (anglais).

<sup>107</sup> <https://www.ontario.ca/fr/page/fonds-pour-la-recherche-en-ontario>.

<sup>108</sup> <https://www.naphro.ca/about> (anglais).

<sup>109</sup> Data Curation Centre (s. d.) Funders' data plan requirements. Repéré en novembre 2020 à <https://www.dcc.ac.uk/resources/data-management-plans/funders-requirements> (anglais).

<sup>110</sup> Wellcome Trust. (2017). Data, software and materials management and sharing policy. Repéré en novembre 2020 à <https://wellcome.ac.uk/grant-funding/guidance/data-software-materials-management-and-sharing-policy> (anglais).

<sup>111</sup> Bill & Melinda Gates Foundation. (2011). Bill & Melinda Gates Foundation's Data Access Principles: Frequently Asked Questions. Repéré en novembre 2020 à <https://docs.gatesfoundation.org/documents/faq.pdf> (anglais).

chercheuses à respecter le nombre croissant d'attentes et peuvent s'intégrer avec les autres mesures de soutien susceptibles d'être mises au point. L'Alliance devra aussi faire en sorte que ses propres besoins de financement soutiennent les aspects particuliers uniques de ses communautés de recherche, y compris le respect de la souveraineté des données autochtones. Une sensibilisation accrue devrait également favoriser le développement de volets de financement et de programmes qui comblent les lacunes actuelles dans le paysage du financement.

## Maisons d'édition savante

Les politiques relatives au partage de données adoptées par les maisons d'édition savante sont un autre facteur qui motive l'adoption de meilleures pratiques de GDR par le milieu de recherche. Le nombre de revues et de maisons d'édition adoptant des politiques relatives aux données de recherche augmente. Ces politiques traitent d'une panoplie de questions connexes, notamment le dépôt des données dans des dépôts approuvés, les déclarations relatives à la disponibilité des données, la citation des données, les normes et les formats des données, et l'examen des données de recherche par les pairs<sup>112</sup>.

Dans de nombreux cas, les chercheurs et les chercheuses ont fait progresser le mandat de la science ouverte par l'entremise de leurs sociétés savantes, de leurs revues savantes et de politiques connexes relatives au partage de données<sup>113</sup>. L'American Geophysical Union<sup>114</sup>, la Genetics Society of America<sup>115</sup>, et la British Ecological Society<sup>116</sup> constituent quelques exemples de pionniers dans ce domaine.

Bien que ce virement de cap fasse état des grandes tendances du paysage de l'édition vers un accès élargi aux publications savantes grâce à des modèles d'accès libre, il fait aussi état de la transition de quelques maisons d'édition vers l'exercice d'une plus grande influence tout au long du cycle de vie savant, et d'une implication plus importante dans celui-ci. En reconnaissance de la demande des chercheurs et des chercheuses d'un soutien de la GDR, des maisons d'édition commerciale comme Springer-Nature<sup>117</sup> et Elsevier<sup>118</sup> ont commencé à offrir des services de soutien direct de la curation et de la publication d'ensembles de données, tandis que d'autres ont commencé à établir des partenariats avec des dépôts tiers pour aider les auteurs et les autrices

---

<sup>112</sup> Hrynaszkiewicz et al. (2020). Developing a Research Data Policy Framework for All Journals and Publishers. *Data Sci. J.*, 19 (1): 5. <http://doi.org/10.5334/dsj-2020-005> (anglais).

<sup>113</sup> Mitchener, W.K. (2015). Ecological data sharing. *Ecol Info*, 29(1):33-44. <https://doi.org/10.1016/j.ecoinf.2015.06.010> (anglais)

<sup>114</sup> AGU (2019). Position Statement on Data. Repéré en novembre 2020 à [https://www.agu.org/Share-and-Advocate/Share/PolicyMakers/Position-Statements/Position\\_Data](https://www.agu.org/Share-and-Advocate/Share/PolicyMakers/Position-Statements/Position_Data) (anglais).

<sup>115</sup> McIntyre, L.M. (2010). Data: The foundation of science. *Genetics*, 184(1):1. <https://www.genetics.org/content/184/1/1.full> (anglais).

<sup>116</sup> Norman, H. (2014). Mandating data archiving: experiences from the frontline. *Learn. Publ.*, 27(5):S35-S38. <https://doi.org/10.1087/20140507> (anglais).

<sup>117</sup> <https://www.springernature.com/gp/authors/research-data/research-data-support> (anglais).

<sup>118</sup> <https://data.mendeley.com/> (anglais).

à publier les ensembles de données sur lesquelles reposent leurs publications (p. ex. le partenariat entre Wiley et Dryad)<sup>119</sup>. L'expansion des maisons d'édition commerciale dans le domaine de la publication des données préoccupe le milieu universitaire dans une certaine mesure, compte tenu de la croissance et des répercussions connexes de la monopolisation de l'édition savante par ces sociétés (voir les ouvrages de Vincent Larivière pour obtenir le contexte<sup>120</sup>). Les études préliminaires menées par le Groupe d'experts sur la recherche et l'intelligence du réseau Portage révèlent que si de nombreuses maisons d'édition autorisent les chercheurs et les chercheuses à publier les données en vertu d'une licence générale ouverte, des modèles plus restrictifs existent (p. ex. IEEE Dataport, qui a mis en œuvre un modèle selon lequel un abonnement est nécessaire pour obtenir un accès sans restriction et qui oblige les déposants et les déposantes de données à payer des frais pour publier leurs données en vertu d'une licence générale ouverte)<sup>121</sup>. Une étude plus poussée des modèles d'octroi de licences et de prestation de services développés par les maisons d'édition savante revêt une importance croissante à mesure que ces services prennent de l'ampleur, pour que les intérêts des communautés soient respectés.

Les maisons d'édition savante ont aussi été un facteur important du développement et de la mise en œuvre de ce qu'on appelle parfois le « Graphe de la science » ou le « Graphe de la recherche » : il s'agit d'une base de données en ligne reliant les chercheurs et les chercheuses, les données primaires et la littérature savante qui en découle. Les graphes de recherche représentent un domaine de développement actif et sont en voie d'élargissement pour inclure des liens vers les organismes de financement et d'autres productions savantes, comme les flux de tâches reproductibles et les brevets. Les chercheuses et les chercheurs canadiens participent à des cadres et à des mises en œuvre de graphes de recherche, dont le Scholix Framework<sup>122</sup> mis en œuvre sur le portail OpenAIRE de l'Union européenne<sup>123</sup> et le Web of Science<sup>124</sup>.

## Organismes para-universitaires

Divers organismes de recherche évoluant hors du milieu universitaire produisent et gèrent d'importantes ressources de données qui ont de la valeur pour le milieu de recherche canadien. Bien que ces organismes puissent se fier à leur propre IRN et à leurs propres services de soutien, les données de recherche dont ils assurent l'intendance sont souvent produites en partenariat avec les chercheuses et les chercheurs universitaires ou utilisées directement par celles-ci et ceux-ci. Le milieu national de recherche a intérêt à s'assurer que ces ensembles de données sont accessibles à l'IRN nationale et qu'ils sont interopérables avec celle-ci, tout en respectant les

---

<sup>119</sup> Wiley (s. d.) Data sharing: Data sharing with Dryad. Repéré en novembre 2020 à <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-service.html> (anglais).

<sup>120</sup> Larivière, V. (s. d.) Articles de revues scientifiques ou professionnelles. Repéré en novembre 2020 à <https://crc.ebsi.umontreal.ca/publications/?person=71>.

<sup>121</sup> <https://ieee-dataport.org/> (anglais).

<sup>122</sup> <http://www.scholix.org/> (anglais).

<sup>123</sup> <https://www.openaire.eu/> (anglais).

<sup>124</sup> <https://clarivate.libguides.com/webofscienceplatform/coverage> (anglais).



besoins des communautés et la souveraineté des données autochtones. Nonobstant les aspects et les facteurs singuliers de ces organismes, l'adoption de pratiques et de cadres communs de GDR dans l'ensemble du paysage peut favoriser cette réalisation. Le renforcement des partenariats entre le secteur universitaire et les organismes para-universitaires dans le but de soutenir les données de recherche donnera lieu aussi à un paysage d'IRN plus résilient et plus durable.

### Gouvernement : ministères et centres de recherche avec une abondance de données

Plusieurs ministères et organismes du gouvernement recueillent, gèrent et publient des données sur les aspects sociaux et scientifiques de la vie canadienne, qui constituent une ressource de valeur pour le milieu de recherche canadien. Nombre de ces organismes du secteur public cultivent aussi des relations de travail très étroites avec les organismes de recherche externes et le secteur universitaire. Ainsi, la séparation entre le secteur public et l'organisme de recherche n'est pas toujours clairement définie. À titre d'exemple, le Conseil national de recherches Canada, le principal organisme de recherche et de technologie du gouvernement du Canada, exploite divers centres de recherche qui collaborent étroitement avec le secteur universitaire et l'industrie<sup>125</sup>. Statistique Canada est un autre organisme national qui a la responsabilité de donner aux chercheurs et aux chercheuses du secteur universitaire et du secteur public l'accès à des données précieuses relatives aux questions socioéconomiques et aux questions de santé par le truchement de ses partenariats avec un consortium d'établissements canadiens en vertu du RCCDR<sup>126</sup>. Voici une liste de divers exemples d'intendance de données d'importance nationale par des ministères à tous les ordres du gouvernement.

- Les centres de recherche situés au sein de ministères et d'organismes fédéraux dirigent la collecte de données scientifiques, et partagent les résultats par l'entremise de dépôts thématiques et du portail des données ouvertes du gouvernement du Canada (p. ex. le Catalogue des données de Pêches et Océans Canada<sup>127</sup>, le Système de données d'observation de la Terre de Ressources naturelles Canada [RNCa] <sup>128</sup>, la Collection de fichiers de microdonnées à grande diffusion de Statistique Canada <sup>129</sup>)<sup>130</sup>.
- Les ministères provinciaux recueillent et partagent des données liées à des secteurs essentiels, notamment la santé, l'éducation et les ressources naturelles, par le truchement de bases de données, de dépôts de données thématiques, et de portails de données

---

<sup>125</sup> <https://nrc.canada.ca/fr>

<sup>126</sup> <https://crdcn.org/fr/%C3%A0-propos-du-rccdr>

<sup>127</sup> <https://www.pac.dfo-mpo.gc.ca/science/oceans/data-donnees/index-fra.html>

<sup>128</sup> [https://www.eodms-sgdot.nrcan-rncan.gc.ca/index\\_fr.jsp](https://www.eodms-sgdot.nrcan-rncan.gc.ca/index_fr.jsp)

<sup>129</sup> <https://www150.statcan.gc.ca/n1/pub/11-625-x/11-625-x2010000-fra.htm>

<sup>130</sup> <https://open.canada.ca/fr>

ouvertes (p. ex. Alberta Geological Survey<sup>131</sup>, Santé publique Ontario<sup>132</sup> et BC Data Catalogue<sup>133</sup>).

- Les gouvernements municipaux recueillent, regroupent et publient des données civiques par le truchement de portails locaux de données ouvertes (p. ex. ville d'Ottawa<sup>134</sup>).

### Autorités sanitaires régionales et organismes de recherche en santé

Au sein d'hôpitaux dans l'ensemble du pays, les cliniciens et les cliniciennes, et les chercheurs et les chercheuses travaillent en commun pour mener une recherche vitale en vue de prévenir, de diagnostiquer et de traiter des maladies dans le cadre d'essais cliniques, et de recueillir des données probantes en vue d'améliorer les services de santé. La gestion protégée de ces données, tout en veillant à leur accès par les chercheuses et les chercheurs autorisés, repose sur des pratiques rigoureuses de GDR. Les autorités sanitaires régionales peuvent coordonner les politiques et les pratiques au sein d'une région et favoriser l'accès aux données par les chercheurs et les chercheuses externes (p. ex. BC Cancer<sup>135</sup> et BC Health Authorities<sup>136</sup>).

Les ministères de santé provinciaux peuvent aussi verser des fonds à des associations de recherche indépendantes qui fournissent aux chercheuses et aux chercheurs autorisés et aux demandes de recherche approuvées un accès contrôlé aux documents administratifs sur les services de santé (p. ex. Population Data BC<sup>137</sup>, ICES<sup>138</sup>). Les organismes à but non lucratif nationaux liés à la santé dirigent aussi la collecte et la gestion de données de santé, ainsi que l'accès contrôlé à ces données, autant de fonctions qui sont contrôlées dans l'ensemble des secteurs de compétence provinciaux et municipaux (p. ex. Réseau de recherche sur les données de santé du Canada<sup>139</sup> et Institut canadien d'information sur la santé<sup>140</sup>).

### Institutions culturelles (galeries, bibliothèques, archives et musées, ou « GLAM », d'après l'anglais *galleries, libraries, archives and museums*)

Les institutions culturelles et les institutions de conservation de la mémoire facilitent l'accès à des données relatives au patrimoine culturel et favorisent les échanges entre le secteur GLAM et les

---

<sup>131</sup> <https://ags.aer.ca/> (anglais).

<sup>132</sup> <https://www.publichealthontario.ca/fr/data-and-analysis/using-data/open-data>.

<sup>133</sup> <https://data.gov.bc.ca/> (anglais).

<sup>134</sup> <https://ouverte.ottawa.ca/>.

<sup>135</sup> <http://www.bccancer.bc.ca/health-professionals/professional-resources/bc-cancer-registry/request-registry-data> (anglais).

<sup>136</sup> <https://www2.gov.bc.ca/gov/content/health/conducting-health-research-evaluation/data-access-health-data-central/health-authorities> (anglais).

<sup>137</sup> <https://www.popdata.bc.ca/> (anglais).

<sup>138</sup> <https://www.ices.on.ca/> (anglais).

<sup>139</sup> <https://www.hdrn.ca/fr>.

<sup>140</sup> <https://www.cihi.ca/fr>.

chercheurs et les chercheuses, en particulier dans la discipline des sciences humaines. À mesure que la technologie de numérisation s'améliore et devient plus accessible, et que davantage de documents sont créés dans un format numérique, une meilleure curation et gestion numérique à long terme sera essentielle pour faciliter une exploitation efficace de ces ressources de données sous forme de données de recherche. Les institutions du secteur GLAM ont aussi beaucoup à partager avec le milieu de recherche universitaire pour faire progresser la GDR, en particulier relativement au développement (sélection et évaluation) et à la préservation de collections. Mentionnons à titre d'exemple les collections de grandes institutions comme les collections numériques de Bibliothèque et Archives Canada<sup>141</sup>, les collections numériques de musées (p. ex. Données ouvertes d'Ingenium<sup>142</sup>), et les archives institutionnelles de plus petite taille.

### Industrie et secteur privé

Les chercheurs et les chercheuses dans l'ensemble des disciplines collaborent avec les organisations de l'industrie et du secteur privé pour mener des recherches, et pour recueillir et gérer des données présentant une valeur commerciale ou stratégique. Les centres canadiens de collaboration entre l'industrie et le secteur universitaire dirigent l'innovation et les découvertes dans une gamme de disciplines (p. ex. MaRS Innovation<sup>143</sup>, le Consortium du Québec pour la découverte de médicaments<sup>144</sup> et le programme Leap de DWave<sup>145</sup>). Un nombre plus élevé de chercheurs et de chercheuses exploite aussi des données tierces recueillies par des entreprises et des fournisseurs de services technologiques commerciaux (p. ex. les médias sociaux, les fournisseurs de services Internet [FSI]). La gestion protégée de la propriété intellectuelle (PI) générée par ces partenariats, tout en répondant aux nouvelles attentes en matière d'accès aux données, présente des défis de gestion de longue durée.

### Fournisseurs de services tiers (commerciaux et à but non lucratif)

Plusieurs groupes de recherche universitaires dépendent aussi de l'infrastructure numérique commerciale pour mener leur recherche en raison de la convivialité de cette infrastructure et de leur connaissance de celle-ci, ou lorsque leurs besoins en IRN sont trop importants pour être satisfaits par les fournisseurs de services institutionnels. Ceci est particulièrement vrai de l'infrastructure de stockage et de calcul, et de nombreux projets de recherche s'appuient sur les services offerts par Amazon, Google et Microsoft. La gestion et le partage à long terme de données stockées sur une infrastructure commerciale présentent des défis d'envergure au chapitre de la GD. Ces défis comprennent, sans en exclure d'autres : la lourdeur de la migration des ressources de données existantes vers les infrastructures infonuagiques et le développement ultérieur de nouveaux flux de tâches, l'élaboration de politiques institutionnelles en matière d'approvisionnement, et l'assurance que les fournisseurs de services commerciaux se conforment aux politiques et aux règlements de l'établissement en matière de sécurité et de

---

<sup>141</sup> <https://www.bac-lac.gc.ca/fra/Pages/accueil.aspx>.

<sup>142</sup> <https://ingeniumcanada.org/fr/collection-et-recherche/donnees-ouvertes>.

<sup>143</sup> <https://www.marsdd.com/> (anglais).

<sup>144</sup> <https://cqdm.org/fr/>.

<sup>145</sup> <https://www.dwavesys.com/take-leap> (anglais).

protection des renseignements personnels. À l'avenir, l'Alliance devrait exercer un rôle dans la détermination des défis et des avantages de la collaboration avec les fournisseurs d'infrastructures numériques commerciales.

Divers organisations commerciales et à but non lucratif appuyant le mouvement de science ouverte ont également pris de l'importance dans divers milieux de recherche au cours de la dernière décennie en raison du développement d'environnements de recherche virtuels. Mentionnons, par exemple, les plateformes de collaboration et de partage comme Open Science Framework<sup>146</sup>, les plateformes de dépôt comme Dryad et Figshare<sup>147,148</sup>, les plateformes de recherche informatique comme Code Ocean<sup>149</sup>, et les nombreux fournisseurs de carnets de laboratoire électroniques. Il y a lieu de se renseigner sur les partenariats et les intégrations entre l'infrastructure financée par l'État et les services offrant des outils scientifiques ouverts largement adoptés pour soutenir la recherche nationale.

## Organisations internationales

Quelques initiatives nationales ou pannationales importantes existent pour coordonner le paysage de la science ouverte et fournir des services et une infrastructure de base afin d'aider les chercheurs et les chercheuses à gérer leurs données de recherche. À mesure que l'Alliance arrive à maturité, il lui incombe de participer à ces initiatives (quoique cette participation ait déjà commencé par l'entremise de DRC et du réseau Portage) afin d'assurer la concordance et tirer parti des similitudes.

### Initiatives gouvernementales nationales et pannationales

Diverses initiatives nationales et pannationales à divers stades de développement existent dans des administrations distinctes dans le monde entier. Elles exécutent un mandat comparable à celui de l'Alliance au Canada, à savoir offrir des services et une infrastructure de base pour aider les chercheurs et les chercheuses à gérer leurs données de recherche. La gamme d'approches retenues par ces initiatives au chapitre de l'organisation, des services et des modèles opérationnels offre à l'Alliance d'éventuelles possibilités d'apprentissage.

Une approche fédérée de l'avancement de la science ouverte se précise en Europe. En 2016, la Commission européenne a attribué des fonds à la fédération des infrastructures de données scientifiques par l'entremise d'une nouvelle entité appelée European Open Science Cloud (EOSC)<sup>150</sup>. EOSC favorisera la création d'un réseau d'organisations et d'infrastructures de divers pays et de diverses communautés qui soutiennent la création et la diffusion ouvertes de connaissances et de données scientifiques. EOSC a pour objet de soutenir la GDR dans l'ensemble de l'Europe au moyen de données, de services et d'infrastructures interopérables en

---

<sup>146</sup> <https://osf.io/> (anglais).

<sup>147</sup> <https://datadryad.org/> (anglais).

<sup>148</sup> <https://figshare.com/> (anglais).

<sup>149</sup> <https://codeocean.com/> (anglais).

<sup>150</sup> <https://www.eosc-portal.eu/> (anglais).

formant la « plus petite plateforme viable » énonçant les règles de participation afin de guider la prestation de services interconnectés et de données interopérables.

Une autre approche fédérée se précise en Afrique avec l'African Open Science Platform (AOSP), une initiative lancée en 2016 par le ministère des Sciences et de la Technologie de l'Afrique du Sud afin d'établir des liens entre les activités de science ouverte en cours dans l'ensemble du continent au moyen de mécanismes de collaboration et de coordination, et d'échange de pratiques exemplaires<sup>151</sup>.

Plusieurs initiatives nationales existantes tiennent compte également de l'importance de la GDR pour le soutien de l'IRN nationale. L'Australian Research Data Commons (ARDC) a le mandat d'assurer la cohérence nationale des données et de la capacité des plateformes de recherche en ligne, et d'accélérer la recherche australienne en développant, en mettant à l'essai et en prenant en charge des plateformes où les chercheurs et les chercheuses peuvent stocker, chercher, partager et consulter des objets numériques (données, logiciels, etc.) et interagir avec eux<sup>152</sup>. En Allemagne, l'infrastructure nationale de données de recherche adopte une approche distincte et réunit des collaboratrices et des collaborateurs au moyen d'un réseau coordonné de consortiums chargés de fournir des services de données aux milieux de recherche<sup>153</sup>. Organisés en règle générale selon la discipline ou la méthode de recherche, les consortiums visent à améliorer et à protéger l'accès aux données de recherche dans leur secteur respectif, ainsi que l'utilisation de ces données.

L'annexe B présente une analyse plus exhaustive des initiatives comparables nationales et internationales.

### Associations internationales

De nombreuses organisations internationales ayant tissé des liens avec le Canada font la promotion des pratiques exemplaires en matière de GDR en développant et en coordonnant des communautés de pratique. La plus éminente de ces initiatives est la Research Data Alliance (RDA) lancée en 2013<sup>154</sup>. En mars 2020, l'organisation comptait presque 10 000 membres répartis sur 144 pays<sup>155</sup>. La RDA offre une plateforme où les spécialistes internationaux en données de recherche se rencontrent pour échanger des opinions et discuter de questions liées aux pratiques exemplaires, aux normes et aux protocoles. Le travail est effectué par des groupes de travail munis d'une orientation fonctionnelle (p. ex. gestion de l'identité, services de vocabulaire, environnements de recherche virtuels) ou axée sur un domaine. Les produits comprennent les recommandations de la RDA (documents susceptibles d'inclure des

---

<sup>151</sup> <http://africanopenscience.org.za/> (anglais).

<sup>152</sup> <https://ardc.edu.au/> (anglais).

<sup>153</sup> [https://www.dfg.de/en/research\\_funding/programmes/nfdi/index.html](https://www.dfg.de/en/research_funding/programmes/nfdi/index.html) (anglais).

<sup>154</sup> <https://rd-alliance.org/> (anglais).

<sup>155</sup> WDS-ITO. (2019). RDA Membership Worldwide. Tableau Public. Repéré en novembre 2020 à <https://public.tableau.com/profile/littlehelper#!/vizhome/RDAMembershipWorldwide/RDAMembershipWorldwide> (anglais).

spécifications, des ontologies, des flux de tâches, des modèles de données, etc.) qui sont officiellement approuvées par la RDA.

Entre-temps, d'autres organisations internationales visent à fournir des services et un soutien aux organisations membres. À titre d'exemple, le World Data System est un organe interdisciplinaire de l'International Science Council (ISC, en français, le Conseil international pour la science) dont la mission est de favoriser l'accès aux données scientifiques fiables, aux services et aux produits de données, et à l'information sur les données, de même que favoriser l'intendance de ces divers éléments<sup>156</sup>. Les membres comprennent 125 dépôts de données scientifiques, sociétés de données, services de données et organisations connexes, répartis dans le monde entier. L'International Technology Office (ITO) de l'organisation est sis à l'Université de Victoria et est soutenu par trois organisations canadiennes : Ocean Networks Canada, Polar Data Catalogue et le Centre canadien de données astronomiques<sup>157</sup>. L'ITO soutient les organisations membres et les partenaires en leur offrant une infrastructure technique, des conseils éclairés et des services favorisant l'accès aux données scientifiques.

Le Canada est membre du Comité sur les données (CODATA) d'ISC depuis les années 1960 et deux Canadiens siègent actuellement à son comité de direction<sup>158</sup>. CODATA est représenté et coordonné au Canada par le Comité national canadien pour CODATA (CNC/CODATA)<sup>159</sup>. Le Comité international des politiques de CODATA fournit, à une panoplie d'initiatives internationales, des commentaires experts sur l'élaboration et la mise en œuvre de politiques sur les données<sup>160</sup>, notamment les Principes et lignes directrices de l'Organisation de coopération et de développement économiques (OCDE) et de CODATA pour l'accès aux données de la recherche financée sur des fonds publics<sup>161</sup>.

Étant donné que CODATA et WDS sont des organes internationaux du Conseil international pour la science, les deux soutiennent le programme Decadal, qui vise l'interopérabilité des données dans l'ensemble des disciplines<sup>162</sup>. De façon plus générale, en 2020, quatre grandes organisations internationales de données (WDS, RDA, CODATA et GO FAIR) ont officiellement énoncé leur engagement commun de collaborer afin d'optimiser l'écosystème mondial des données de recherche. Collectivement appelé « Data Together », le document oblige tous les signataires à travailler en commun pour « [Traduction] optimiser l'écosystème mondial des

---

<sup>156</sup> <https://www.icsu-wds.org/> (anglais).

<sup>157</sup> <https://wds-ito.org/> (anglais).

<sup>158</sup> <https://codata.org/> (anglais).

<sup>159</sup> <https://codata.org/canada/> (anglais).

<sup>160</sup> CODATA. (s. d.). International Data Policy Committee. Repéré en novembre 2020 à <https://codata.org/initiatives/strategic-programme/international-data-policy-committee/> (anglais).

<sup>161</sup> OCDE. (2007). Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics. <http://www.oecd.org/fr/science/inno/38500823.pdf>

<sup>162</sup> CODATA. (2019). Decadal Programme: Making data work for cross-domain grand challenges. Repéré en novembre 2020 à <https://codata.org/initiatives/strategic-programme/decadal-programme/> (anglais).

données de recherche et cerner les possibilités et les besoins qui amèneront les infrastructures fédérées à servir la nouvelle réalité des sciences axées sur les données. »<sup>163</sup>

La Global Indigenous Data Alliance (GIDA) est un réseau international de chercheuses et chercheurs, d'experts et expertes en données et de décideurs et décideuses autochtones voués à la défense du contrôle des données autochtones par les Autochtones en faisant valoir la souveraineté des données et la gouvernances des données autochtones à l'échelle internationale et à l'intérieur des États-nations. L'Alliance a généré des ressources qui soutiennent la souveraineté des données autochtones, y compris les principes *CARE* et les récentes lignes directrices relatives à la COVID-19 sur le partage des données respectant la souveraineté autochtone.<sup>164</sup>

Un nombre croissant d'organisations internationales axées sur des domaines particuliers s'occupent également de l'avancement et du soutien de la GDR au sein de leur communauté afin de coordonner les efforts de recherche mondiale. Voici deux exemples :

- Global Open Data for Agriculture and Nutrition (GODAN) comporte un réseau de plus de 1000 membres provenant de gouvernements nationaux, d'ONG, d'organisations internationales et d'organisations du secteur privé<sup>165</sup>. GODAN et ses partenaires visent à améliorer la sécurité alimentaire et les moyens de subsistance des communautés agricoles en associant une action revendicatrice touchant les données ouvertes et une consultation à des solutions innovatrices. Le secrétariat de GODAN est sis à l'Université McGill.
- Global Alliance for Genomics and Health (GA4GH) est une alliance internationale qui réunit plus de 500 organisations des secteurs de la santé, de la recherche, de la défense des droits des patients, des sciences de la vie et des technologies de l'information<sup>166</sup>. La communauté GA4GH fait progresser les cadres et les normes qui facilitent le partage responsable, volontaire et protégé de données génomiques et de données liées à la santé. Sise à Toronto, dans le MaRS Discovery District, GA4GH est financée en partie par l'initiative pancanadienne de partage international des données (Can-SHARE)<sup>167</sup>.

L'annexe C présente une liste plus détaillée des organisations internationales alliées qui soutiennent la GDR.

---

<sup>163</sup> GoFAIR. (2020). Data Together Statement. Repéré en novembre 2020 à <https://www.go-fair.org/2020/03/30/data-together-statement/> (anglais).

<sup>164</sup> RDA COVID-19 Working Group. (2020). RDA COVID-19 Recommendations and Guidelines on Data Sharing. <http://doi.org/10.15497/rda00052> (anglais).

<sup>165</sup> <https://www.godan.info/> (anglais).

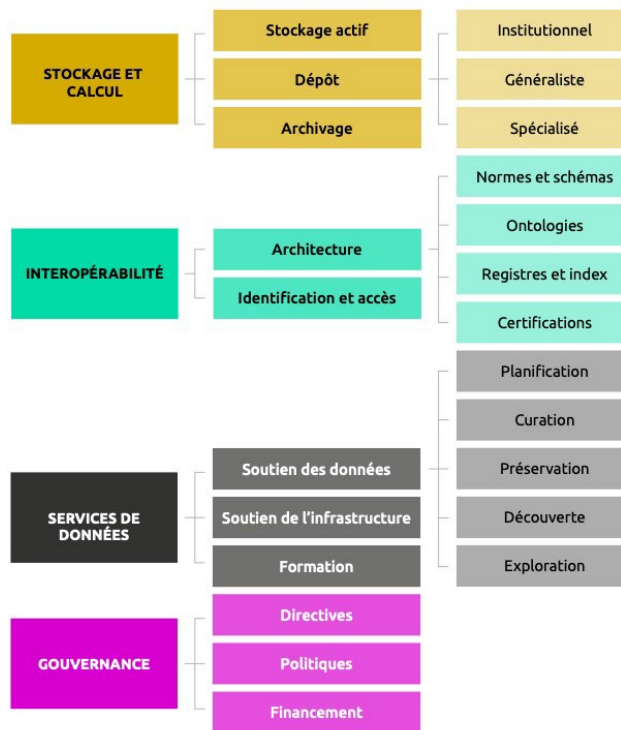
<sup>166</sup> <https://www.ga4gh.org/> (anglais).

<sup>167</sup> <https://www.genomebc.ca/projects/canadian-international-data-sharing-initiative-can-share/> (anglais).

## 5 Évaluation de l'état actuel

« L'écosystème » de la GDR se compose d'une série évolutive d'éléments interconnectés qui interagissent entre eux et se soutiennent mutuellement. Le concept d'écosystème constitue une bonne analogie, car il s'agit d'un contexte où les éléments sont reliés et interdépendants, et où ils évoluent de pair avec la technologie et les pratiques de recherche, formant un réseau croissant s'articulant autour du soutien des données de recherche sous-jacentes qui y circulent, et où une série de « niches » expriment les priorités ou les exigences spécialisées. Les relations tissées dans cet écosystème vont au-delà de celles entre un producteur ou une productrice de données et un ensemble de données particulier, et englobent divers aspects connexes qui ont une incidence sur le cycle de vie des données. Mentionnons, à titre d'exemple, les normes et les protocoles gouvernant les données, l'infrastructure sur laquelle les données sont créées, analysées et stockées, la formation suivie par les créateurs et les créatrices, et les gestionnaires, ainsi que les politiques directrices. L'éventail des organisations qui évoluent dans le paysage précité tire parti de l'évolution de l'écosystème de la GDR et y contribue en faisant progresser les éléments au moyen d'une gamme de plateformes, de services, de directives et de l'exercice de la recherche même.

L'organigramme de l'état actuel de la GDR (figure 2) organise l'écosystème en éléments distincts, qui sont nécessaires pour soutenir la GDR à l'échelle nationale : stockage et calcul, interopérabilité, services de données et gouvernance. La section suivante du présent rapport approfondit chacune de ces catégories pour examiner la façon dont elles contribuent au soutien de la GDR, et résume les mesures de soutien existantes qui sont offertes.

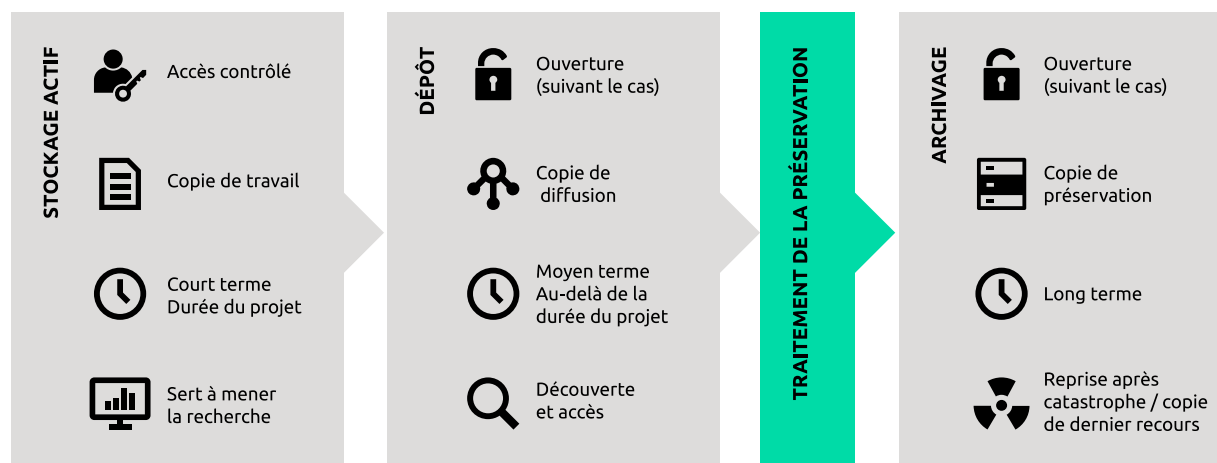


**Figure 2.** Organigramme des principales catégories d'infrastructures et de services utilisées pour décrire l'état actuel de l'écosystème de la GDR au Canada.



## Stockage et calcul

Du point de vue de la GDR, il existe trois configurations distinctes de l'infrastructure de calcul et de stockage qui soutiennent les stades distincts des données tout au long de leur cycle de vie : phase active, phase de dépôt et phase d'archivage. Il est important de faire remarquer que la principale différence entre les types de stockage que le présent modèle vise à délimiter n'est pas technique. Par exemple, une infrastructure d'archivage est fournie pour quelques projets de recherche active (p. ex. sous-systèmes de bandes magnétiques à capacité élevée pour les ensembles de données consultés à intervalles peu fréquents, qui font partie intégrante d'un projet de recherche active)<sup>168</sup>, et les dépôts et les services d'archivage qui utilisent la même infrastructure utilisée par les projets de recherche active et qui sauvegardent en même temps les données sur des systèmes d'archivage à bandes magnétiques. Le modèle du continuum met l'accent sur l'adoption des pratiques, des ressources et des politiques adaptées à ces trois stades distincts du cycle de vie des données.



**Figure 3.** Description du spectre du stockage des données de recherche (mention de source : réseau Portage de l'ABRC)

### Stockage actif (phase de recherche)

Dans la phase de la recherche active, l'infrastructure de calcul et de stockage satisfait aux besoins des chercheurs et des chercheuses et des données pendant le processus de recherche même, lorsque les données sont recueillies, modifiées ou analysées. Elle comprend entre autres le calcul de haute performance (CHP), le stockage massif, l'infonuagique, le stockage infonuagique, l'informatique répartie, le stockage décentralisé, l'informatique personnelle, le stockage personnel. Les exigences varient selon la durée du projet de recherche et la quantité de données traitées et stockées. Bien que des discussions continues de l'infrastructure de recherche active et du rôle des solutions institutionnelles, régionales et nationales de stockage à long terme s'imposent au sein de la communauté du CIP, il est important de préciser, dans le contexte du

<sup>168</sup> Consultez « /nearline », dans Terminologie des concours pour l'allocation de ressources. Repéré en novembre 2020 à

[https://docs.computecanada.ca/wiki/Technical\\_glossary\\_for\\_the\\_resource\\_allocation\\_competitions/fr](https://docs.computecanada.ca/wiki/Technical_glossary_for_the_resource_allocation_competitions/fr)

présent rapport, les liens avec la GDR pour laquelle l'infrastructure de stockage est particulièrement pertinente.

En règle générale, les chercheurs et les chercheuses auront accès à une infrastructure de stockage mise à leur disposition par leur établissement d'attache, ainsi qu'à l'infrastructure qu'ils et qu'elles sollicitent aux fournisseurs de services commerciaux ou financés par l'État. Par exemple, au sein d'un établissement universitaire, la faculté ou le département d'attache du chercheur ou de la chercheuse peut fournir un archivage de fichier en réseau, tandis que le service des systèmes de l'université peut offrir aux chercheurs et aux chercheuses de l'ensemble du campus l'accès à un CHP et à un stockage infonuagique. L'offre sera largement fonction de la capacité de l'établissement. C'est la raison pour laquelle les établissements et les gouvernements ont tenté de financer l'infrastructure de recherche active à l'échelle régionale et nationale et de tirer parti des économies d'échelle. Ainsi, les chercheurs et les chercheuses travaillant à des établissements d'enseignement supérieur dans l'ensemble du Canada ont également accès à l'infrastructure qui soutient la gestion des données pendant la phase de recherche active par l'entremise de Calcul Canada, de CANARIE et du RNRE.

Plus récemment, la prévalence de l'infonuagique et du stockage infonuagique, et la facilité d'accès ont motivé de nombreux chercheurs et chercheuses à financer leurs besoins en stockage actif par l'entremise de fournisseurs commerciaux comme Amazon, Microsoft et Google. Bien que la dépendance à ces services présente des défis de gestion des données sur le plan de la préservation des données qui peuvent être stockées sur ces serveurs ou devraient l'être, ainsi que des facteurs de risque connexes, les partenariats établis avec les fournisseurs commerciaux se sont avérés un moyen efficace pour quelques groupes de recherche de contourner les obstacles économiques et technologiques à l'accès aux ressources de CIP pour satisfaire à leurs besoins en matière de données. Les moyens de mobilisation des fournisseurs de stockage commercial et des partenaires du RNRE afin d'assurer un soutien national méritent d'être étudiés. À titre d'exemple, les National Institutes of Health (NIH) des États-Unis ont lancé leur initiative STRIDES en 2018 pour aider les chercheuses et les chercheurs subventionnés à accéder à des services infonuagiques commerciaux afin d'accélérer les découvertes biomédicales<sup>169</sup>. Les ententes de services avec les fournisseurs de services commerciaux soulèvent toutefois diverses préoccupations, qui ont fait l'objet d'une étude dans le cadre du Big Data Project de la National Oceanic and Atmospheric Administration (NOAA) entre 2015 et 2019<sup>170</sup>. La NOAA a signé cinq ententes identiques de recherche et développement coopératif (CRADA) avec Amazon Web Services (AWS), Google Cloud Platform (GCP), IBM, Microsoft Azure et Open Commons Consortium (OCC), et a conclu que les plateformes des fournisseurs de services infonuagiques offrent des avantages techniques, mais présentent aussi des obstacles éventuels pour les gestionnaires de données et les chercheurs et les chercheuses. Bien que le nombre de questions à prendre en considération soit trop élevé dans le contexte du présent rapport, on peut faire ressortir que la facilité de l'évolutivité des systèmes commerciaux s'accompagne du désavantage éventuel d'asservissement à un fournisseur et peut en fait créer des silos de données entre des nuages concurrentiels, réduisant éventuellement la capacité d'analyse des données provenant

---

<sup>169</sup> <https://datascience.nih.gov/strides> (anglais).

<sup>170</sup> Vance, T.C. et al. (2019). From the oceans to the cloud: Opportunities and challenges for data, models, computation and workflows. *Front. Mar. Sci.*, 6:211 <https://doi.org/10.3389/fmars.2019.00211> (anglais).

de divers fournisseurs. Dans tous les cas, il ressort clairement qu'un personnel qualifié est nécessaire pour gérer les politiques, la gouvernance et l'approvisionnement des ententes sur le niveau de service, et que l'infrastructure soutenue permet une intégration efficace avec les dépôts et le stockage de préservation.

### Dépôts (phase d'accès et de publication)

À la conclusion d'une étude ou d'un projet subventionné, et souvent en concordance avec une publication, les chercheurs et les chercheuses prennent des décisions de curation pour déterminer les données qu'ils ou qu'elles conserveront et partageront pour satisfaire aux exigences de l'établissement, de la maison d'édition et de l'organisme de financement, faciliter la reproductibilité des résultats publiés, et favoriser les besoins en recherche future. Ce stade du cycle de vie de la recherche se fie aux dépôts qui favorisent la découverte future et l'accès approprié<sup>171</sup>. Le but principal des dépôts est de s'assurer que les données numériques présentant une valeur pour la recherche sont stockées de façon protégée et permettent une recherche et une consultation appropriées. Par conséquent, ces dépôts doivent permettre des contrôles suffisants pour garantir la fiabilité, l'accessibilité et l'exploitabilité des données à l'avenir. Les principes TRUST (Transparence, Responsabilité, Orientation vers l'utilisateur, Durabilité, Technologie) qui viennent d'être publiés expriment les attentes du milieu en matière de gestion et d'administration responsables des dépôts de données<sup>172</sup>. Alors que les principes FAIR établissent un cadre pour les discussions sur les pratiques exemplaires en matière de gestion de données, les principes TRUST énoncent les règles des dépôts numériques fiables. Bien que la mise en œuvre d'exigences basées sur des principes soit difficile à opérationnaliser, des évaluations objectives des principes TRUST peuvent comprendre la certification des dépôts par des organisations faisant autorité (discuté à la section portant sur l'interopérabilité). Face à la complexité croissante du paysage des dépôts de données, l'évaluation de la qualité est importante pour prévenir les conséquences durables de la perte de données.

### Dépôts institutionnels et généralistes

Dans le but d'aider les chercheurs et les chercheuses à partager les résultats de leur recherche conformément aux attentes en matière de science ouverte du milieu, des organismes de financement et des maisons d'édition, les bibliothèques universitaires favorisent l'accès aux dépôts numériques gérés par les établissements depuis près de deux décennies. Plus récemment, cet effort a pris de l'ampleur pour inclure les dépôts de données institutionnelles qui visent particulièrement à satisfaire aux besoins des données de recherche (à l'opposé d'autres objets numériques qui synthétisent les données, comme les publications). Étant donné que le concept de ces dépôts leur permet de prendre en charge tous les types de données de recherche provenant d'une vaste communauté d'utilisateurs et d'utilisatrices, les dépôts sont souvent qualifiés de dépôts « multidisciplinaires » qui recueillent des métadonnées de haut niveau relatives à leurs collections au niveau de « l'ensemble de données » (plutôt qu'au niveau de la donnée) et acceptent une panoplie de formats de fichiers.

---

<sup>171</sup>Castle, D. et al. (2019). Position Paper – Data Management Roadmap: 2019-2024. Repéré en novembre 2020 à <https://www.rdc-drc.ca/download/position-paper-data-management-roadmap-2019-2024> (anglais).

<sup>172</sup> Lin, D. et al. (2020). The TRUST Principles for digital repositories. *Sci Data* 7, 144 (anglais).

En 2017, le réseau Portage de l'ABRC a convoqué un groupe d'experts et d'expertes qu'il a chargé d'étudier les possibilités de créer un service de dépôt institutionnel national en utilisant la plateforme Dataverse, qui permettrait une coordination et un développement de plateformes dans l'ensemble des établissements membres, qui tireraient parti des économies d'échelle liées à la gestion et aux opérations du dépôt. À la suite de ce travail, Scholars Portal, le volet de service du Conseil des bibliothèques universitaires de l'Ontario, a déployé un modèle de service national pour son dépôt Dataverse et a établi des partenariats avec 55 établissements universitaires dans l'ensemble du Canada<sup>173</sup>. Les coûts et les responsabilités sont partagés entre les membres. La coordination, le développement et le soutien institutionnels sont assurés par Scholars Portal, tandis que les services locaux et le soutien des chercheurs et des chercheuses sont fournis par les bibliothèques membres. Les données déposées dans le Scholars Portal sont actuellement stockées dans le Centre de données de l'Université de Toronto.

La mise à l'échelle des dépôts institutionnels pour que ceux-ci puissent accueillir de très gros ensembles de données (des centaines de Go à des To) est particulièrement difficile en raison de l'infrastructure qui est requise et des coûts élevés que les établissements devraient assumer à long terme à mesure que de plus en plus de données sont stockées. Vu ces contraintes, le réseau Portage de l'ABRC et Calcul Canada ont dirigé le développement du Dépôt fédéré de données de recherche (DFDR)<sup>174</sup>, un dépôt multidisciplinaire national exploité sur l'infrastructure de Calcul Canada, qui pourrait accueillir de très gros ensembles de données et qui serait mis à la disposition des chercheurs et des chercheuses d'établissements universitaires dans l'ensemble du pays. Un cadre gouvernant l'accès et les allocations de stockage sont en voie de développement.

En sus des dépôts généralistes financés par l'État, une série de dépôts multidisciplinaires exploités par des organismes de recherche à but non lucratif exercent maintenant un rôle important dans le paysage en raison de leur adoption par de nombreux chercheurs et chercheuses. Un avantage évident de ces dépôts est le fait qu'ils favorisent la collaboration et le dépôt de données à l'échelle internationale, quelle que soit l'affiliation institutionnelle ou nationale. Mentionnons, notamment, Zenodo, exploité par l'Organisation européenne pour la recherche nucléaire (CERN)<sup>175</sup> et Dryad, un organisme de dépôt à but non lucratif formé par des partenariats entre sociétés scientifiques et maisons d'édition savante<sup>176</sup>.

### Dépôts thématiques par domaine

Une gamme de dépôts de données spécialisés, exploités par des groupes et des organismes de recherche sis au Canada et à l'étranger, jouent aussi un rôle essentiel dans l'écosystème national de GDR. Divers dépôts thématiques par domaine sont approuvés par leur milieu de recherche respectif, ce qui signifie que les chercheurs et les chercheuses dans ces disciplines sont plus susceptibles de rechercher les données qu'ils renferment et d'y faire confiance. Par ailleurs, étant donné que les données stockées dans ces dépôts sont liées à une gamme étroite de matières, il est possible d'appliquer des métadonnées plus abondantes, ainsi que des pratiques spécialisées

---

<sup>173</sup> <https://dataverse.scholarsportal.info/fr/#partners>

<sup>174</sup> <https://www.frdr-dfdr.ca/repo/?locale=fr>.

<sup>175</sup> <https://home.cern/fr>.

<sup>176</sup> <https://datadryad.org/stash> (anglais).

en matière de normes de données et de types de fichiers, par rapport aux dépôts généralistes, ce qui accroît les possibilités d'interopérabilité entre des ensembles de données distincts et de réutilisabilité des données.

Les dépôts thématiques par domaine peuvent prendre diverses formes selon la fonction (tableau 1). La section ci-haut sur le paysage donne des exemples d'organismes de recherche canadiens exploitant des dépôts spécialisés.

Type	Attributs	Exemples
Dépôt de projets/groupes de recherche	<ul style="list-style-type: none"> <li>• Partage les données recueillies pendant la durée du projet de recherche financé</li> <li>• Le dépôt lui-même peut être le projet de recherche financé</li> <li>• Ne permet pas de dépôt ouvert de données</li> <li>• Peut fournir une application pour la manipulation et la visualisation de données, ou d'autres fonctionnalités de visite des données</li> <li>• Géré par des laboratoires individuels ou des équipes de responsables de la curation</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Mountain Legacy Project</a> (anglais)</li> <li>• <a href="#">Canadian Epigenetics, Environment and Health Research Consortium Network</a> (anglais).</li> <li>• <a href="#">Primate Cell Type DB</a> (anglais)</li> </ul>
Dépôt d'organisme de recherche	<ul style="list-style-type: none"> <li>• Fournit des données recueillies dans le cadre de multiples projets de recherche financés</li> <li>• Regroupe les données de divers projets de recherche</li> <li>• Ne permet pas le dépôt de données de chercheurs et de chercheuses externes</li> <li>• Peut fournir une application pour la manipulation et la visualisation de données, ou d'autres fonctionnalités de visite des données</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">Ocean Networks Canada</a> (anglais)</li> <li>• <a href="#">Centre de données des Premières Nations</a></li> </ul>
Dépôt du gouvernement	<ul style="list-style-type: none"> <li>• Fournit des données recueillies ou compilées par des ministères du gouvernement</li> <li>• Axé spécifiquement sur un domaine (c.-à-d. qu'il ne s'agit pas de sites de données ouvertes génériques)</li> </ul>	<ul style="list-style-type: none"> <li>• <a href="#">BC Data Conservation Centre</a></li> <li>• <a href="#">World Ozone and Ultraviolet Radiation Data Centre</a></li> <li>• <a href="#">National Climate Data Archive</a></li> <li>• <a href="#">NRCan Earth Observation Data Management System</a></li> </ul>

Dépôt thématique par domaine	<ul style="list-style-type: none"> <li>Recueille des données auprès de multiples chercheurs et chercheuses, projets et organisations liés à un domaine particulier</li> <li>Peut ou ne peut pas offrir des services de curation</li> <li>Peut ou ne peut pas regrouper les données de divers ensembles de données</li> </ul>	<ul style="list-style-type: none"> <li><a href="#">Polar Data Catalogue</a> (anglais)</li> <li><a href="#">Barcode of Life Data System</a> (anglais)</li> </ul>
Base de connaissances	<ul style="list-style-type: none"> <li>Permet l'extraction, la collecte et la curation de données portant sur une matière</li> <li>Fait appel à des ensembles de données de base pour relier un ensemble croissant de renseignements</li> </ul>	<ul style="list-style-type: none"> <li><a href="#">Avibase</a></li> <li><a href="#">DrugBank</a> (anglais)</li> <li><a href="#">BioGRID</a> (anglais)</li> </ul>
Infrastructure de dépôt fédérée	<ul style="list-style-type: none"> <li>Logiciels ou plateformes qui facilitent une recherche fédérée dans divers dépôts de données</li> </ul>	<ul style="list-style-type: none"> <li><a href="#">Données ouvertes Canada</a></li> <li><a href="#">iReceptor Commons</a> (anglais)</li> <li><a href="#">Global Biodiversity Information Facility</a> (anglais)</li> <li><a href="#">Plateforme canadienne de neuroscience ouverte</a></li> </ul>

**Tableau 1.** Catégories de dépôts thématiques par domaine au Canada

Bien que les dépôts thématiques par domaine aient un rôle vital à jouer dans le soutien de la GDR, leur administration par un ensemble diversifié de groupes et d'organismes de recherche présente un défi en matière de durabilité à long terme. Quelques organisations exploitant des dépôts thématiques par domaine ont un modèle opérationnel qui produit des revenus ou disposent d'un financement institutionnel prévisible, mais la majorité dépend d'un financement de courte durée (3-5 ans) basé sur le projet et octroyé par les organismes de subvention de la recherche. Cette période de 3 à 5 ans reflète la durée type des projets de recherche. La dépendance à ce type de source de financement est incompatible avec la mission à long terme des dépôts de données. Par conséquent, d'autres sources de financement devraient être créées pour soutenir les fonctionnalités de GDR offertes par les dépôts.

Le financement axé sur le projet destiné aux dépôts de données thématiques par domaine a également donné lieu à la prolifération de divers dépôts spécialisés adaptés aux résultats de projets de recherche particuliers. Ces dépôts en particulier peuvent être vulnérables si leurs administrateurs et administratrices ne possèdent pas une expérience solide des pratiques exemplaires de GDR afin de garantir la mise en œuvre de normes, de pratiques de documentation

et de procédures de sauvegarde éprouvées. La prolifération de nombreux petits dépôts axés sur un projet n'est pas viable dans l'optique de la préservation à long terme. Les initiatives qui établissent des pipelines permettant la migration de données de dépôts vers des plateformes partagées et plus stables font l'objet d'une étude actuelle par le Groupe d'experts sur la préservation de Portage-ABRC<sup>177</sup>. Le nombre croissant de dépôts a également des conséquences pour les modèles de financement durable. Un rapport récent de l'OCDE sur les modèles opérationnels des dépôts de données a constaté que bon nombre d'entre eux sont largement tributaires du financement public et que la majorité combine un financement structural avec d'autres flux de rentrées, dont les services à valeur ajoutée et les frais additionnels de dépôt<sup>178</sup>. Les auteurs et autrices font remarquer qu'à mesure que ce secteur prend de l'ampleur, il présentera un nombre croissant de possibilités de s'approvisionner en infrastructure, plateformes et services par l'entremise de fournisseurs de services spécialisés. Le programme DataONE est un exemple actuel de la façon dont des dépôts diversifiés peuvent obtenir des avantages d'une gamme de services et d'infrastructures adaptés à leurs besoins, tout en réalisant des économies d'échelle grâce à un modèle fédéré<sup>179</sup>.

De façon plus générale, les efforts déployés pour regrouper les dépôts multidisciplinaires et les dépôts thématiques par domaine profiteraient à l'écosystème de la GDR. Ces dépôts existent, mais fonctionnent de manière relativement isolée les uns des autres. Les efforts actuellement déployés par le groupe d'experts mixte sur les dépôts de données du réseau Portage de l'ABRC et de DRC pour assurer une coordination de haut niveau et une approche cohésive du développement de dépôts au Canada visent à combler ce fossé<sup>180</sup>. Il existe aussi la possibilité de créer une synergie avec les programmes logiciels de recherche afin de développer et de promouvoir des outils et des plateformes qui favorisent la cohésion et la collaboration entre dépôts.

### Archivage (phase de préservation)

L'archivage, aussi appelé le stockage de préservation, favorise la garde et l'accessibilité à long terme des objets numériques présentant une valeur pour la recherche. Le modèle de référence Open Archival Information System offre un cadre cohérent de principes et de terminologie pour les pratiques de gestion et de préservation d'une archive numérique<sup>181</sup>. L'archivage vise à préserver une copie à long terme qui est indépendamment vérifiable et fiable, qui ne dépend pas d'un logiciel ou d'un matériel, et qui, par conséquent, comporte nettement plus de facteurs à prendre en considération pour l'intendance de son contenu. Ainsi, l'archivage est un des facteurs dont il faut tenir compte dans la discipline générale de préservation numérique, qui vise à garantir

---

<sup>177</sup> <https://portagenetwork.ca/fr/reseau-dexperts/groupe-dexperts-sur-la-preservation-du-reseau-portage/>.

<sup>178</sup> OCDE. (2017). Business models for sustainable research data repositories. OECD Science, Technology and Industry Policy Papers. <https://doi.org/10.1787/302b12bb-en> (anglais).

<sup>179</sup> <https://www.dataone.org/> (anglais).

<sup>180</sup> <https://portagenetwork.ca/fr/reseau-dexperts/groupe-dexperts-sur-les-depots-de-donnees/>.

<sup>181</sup> <http://www.oais.info/> (anglais).

que l'information numérique de valeur durable demeure accessible et utilisable<sup>182</sup>. En sus du stockage, les activités de maintenance périodique, telles que les migrations, la régénération des supports, la recherche d'erreurs, et l'élaboration de plans de reprise après catastrophe, jouent un rôle important dans la facilitation d'une préservation durable de l'accès.

L'infrastructure d'archivage et la préservation numérique des données de recherche sont actuellement soutenues aux niveaux local et régional. Au sein des établissements d'enseignement supérieur, les services des systèmes et de la TI peuvent offrir des options de stockage protégé qui tiennent compte des politiques de conservation de l'établissement ou de l'organisme subventionnaire (en règle générale, de 7 à 10 ans). La prévalence et la capacité de cette pratique méritent une étude plus approfondie. Toutes les pratiques de préservation varient selon le type de stockage utilisé et les priorités d'administration. Les fournisseurs régionaux de services de calcul fournissent aussi une infrastructure d'archivage (p. ex. le service EduCloud Backup de BCNet<sup>183</sup> et le High Performance Storage System de SciNet<sup>184</sup>).

Au niveau régional, des consortiums de bibliothèques universitaires ont dirigé des initiatives pour fournir une infrastructure d'archivage aux établissements membres afin que ceux-ci puissent préserver leurs collections numériques, qui représentent la mémoire collective de leur établissement et de leur communauté. À titre d'exemple, le réseau décentralisé de préservation et de stockage numérique WestVault du Council of Prairie and Pacific University Libraries offre un réseau de stockage point à point présentant un taux élevé de redondance à l'ensemble des quatre provinces de l'Ouest<sup>185</sup>. Par ailleurs, le service de préservation numérique Permafrost de Scholars Portal du Conseil des bibliothèques universitaires de l'Ontario offre l'intégration du logiciel Archivemata<sup>186</sup> avec l'Ontario Library Research Cloud afin de garantir un stockage de préservation protégé de longue durée. Soutenues par cette infrastructure, les bibliothèques et les archives universitaires peuvent former des partenariats avec les chercheurs et les chercheuses pour accepter les collections de données de recherche qu'elles jugent utiles pour leur mission et leur mandat organisationnel respectif et qu'elles géreront en vertu de leurs stratégies de préservation numérique de longue durée.

Un rapport publié par DRC en 2015 a constaté que l'offre d'un archivage constitue la plus grande lacune dans le paysage canadien de la GDR<sup>187</sup>. Une des principales raisons de cette situation tient au fait qu'aucune organisation n'a la mission de financer et de soutenir la prestation de services d'archivage. Une stratégie nationale en matière d'archivage et de préservation des données de recherche doit reconnaître le rôle de la décentralisation dans la réduction des risques à titre d'élément d'une intendance responsable. La règle selon laquelle « beaucoup de copies

---

<sup>182</sup> UVic Libraries. (2017) Digital preservation framework. Repéré en novembre 2020 à <https://www.uvic.ca/library/featured/digitalpreservation/dp-framework-FINAL.pdf> (anglais).

<sup>183</sup> <https://www.bc.net/service-catalogue/educloud-backup> (anglais).

<sup>184</sup> <https://www.scinethpc.ca/high-performance-storage-system-hpss/> (anglais).

<sup>185</sup> <https://coppul.ca/westvault> (anglais).

<sup>186</sup> <https://www.archivemata.org/en/> (site Web principalement en anglais).

<sup>187</sup> Baker, D. et al. (2019). La gestion des données de recherche au Canada : état des lieux. Zenodo. <https://zenodo.org/record/3574712#.YA7b6U-SnmE>.



protègent les choses » (*lots of copies keep stuff safe*, ou LOCKSS) et la pratique consistant à utiliser des endroits de stockage dispersés sur le plan géographique aident à garantir la récupération des données en cas de catastrophe. Ce modèle favorise aussi la formation de partenariats entre organisations institutionnelles, régionales et nationales pour veiller à la résilience et à la viabilité à long terme des ressources de données. Une coordination nationale entre les organisations nouvelles et existantes visant à superviser le stockage et la préservation des données serait nécessaire pour réaliser des économies dans un modèle décentralisé.

Dans son rapport sur un modèle national de préservation des données de recherche, le groupe d'experts sur la préservation du réseau Portage de l'ABRC fait remarquer que la supervision de la prestation d'un stockage actif, de dépôts et d'un archivage dans le cadre d'une stratégie nationale de stockage fédéré permettra de réaliser d'importantes économies<sup>188</sup>. Le modèle proposé par le groupe d'experts milite contre l'homogénéité de l'architecture d'archivage, et préfère plutôt coordonner un réseau stratégique et diversifié de fournisseurs de services de préservation (FSP), qui sont fédérés par l'entremise d'une stratégie nationale qui détermine les lacunes et les chevauchements de la prestation de leurs services et définit un ensemble d'exigences en matière de maintenance et de sécurité constituant des « pratiques exemplaires », tout en laissant l'exploitation et la maintenance des FSP individuels à leur établissement d'accueil. Une telle approche miserait sur la capacité, l'expertise et l'investissement des établissements et des organisations à l'appui de la stratégie d'archivage plus vaste pour le pays. Elle permettrait aussi de reconnaître les FSP qui constituent une pratique exemplaire dans des domaines particuliers et qui peuvent fournir les mêmes services à un milieu international de chercheurs et de chercheuses. Cette reconnaissance laisse également entrevoir le besoin d'assurer à tous les FSP canadiens, tant nationaux qu'internationaux, un niveau de soutien plus durable.

Le stockage de préservation, ainsi que le rôle des solutions de stockage durable aux niveaux institutionnel, régional et national, devra être développé et déployé, ce qui exigera des discussions continues avec les milieux pertinents. À mesure que l'ensemble des partenaires et la capacité de l'écosystème actuel de la GDR se précisent, il sera possible de coordonner les besoins en matière de stockage de dépôt et de préservation. Ces efforts amélioreront non seulement l'intendance et l'accessibilité des données de recherche, mais réaliseront aussi des économies d'échelle qui aideront à créer un réseau durable d'options de stockage des données de recherche.

## Interopérabilité

L'interopérabilité des éléments de l'écosystème de la GDR repose sur des schémas, des normes et des protocoles communs qui organisent et décrivent les données de recherche ainsi que l'infrastructure auxiliaire. Selon les principes FAIR, l'interopérabilité désigne la combinaison de données d'ensembles de données distincts, et l'intégration avec des plateformes et des applications aux fins d'analyse, de stockage et de traitement<sup>189</sup>. Il convient d'élargir ce concept d'opérabilité pour qu'il comprenne d'autres artefacts créés pendant le processus de recherche,

---

<sup>188</sup> Qasim, U. et al. (2018). La préservation des données de recherche au Canada : Livre blanc. <http://dx.doi.org/10.14288/1.0371945>.

<sup>189</sup> Wilkinson, M.D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18> (anglais).

dont le code logiciel, les protocoles de laboratoire et les flux de tâches scientifiques, ainsi que les politiques générales et les pratiques d'administration qui gouvernent leur création et leur gestion. Si on veut maximiser le potentiel des données de recherche, celles-ci doivent permettre leur échange et leur intégration avec des systèmes distincts et leur interprétation correcte et appropriée par des utilisatrices et utilisateurs distincts.

L'ébauche de l'European Interoperability Framework (EIF) d'EOSC décrit quatre types d'interopérabilité qui sont nécessaires pour la gouvernance des services de données nationaux et pan nationaux : technique, sémantique, organisationnelle et légale<sup>190</sup>. (La section ci-dessous sur la gouvernance se penche sur les aspects de l'interopérabilité organisationnelle et légale.)

- L'interopérabilité technique décrit la capacité des applications et des infrastructures à échanger des données et à exécuter les tâches nécessaires sans intervention de l'exploitant. Selon l'EIF, cela peut inclure les spécifications relatives aux interfaces, les services d'intégration des données, la présentation et l'échange des données, et les protocoles de communication protégée.
- L'interopérabilité sémantique décrit la « *[Traduction]* la capacité du système informatique à transmettre des données présentant une signification commune sans ambiguïté<sup>191</sup>. » Cette capacité dépend de l'adoption d'artéfacts de recherche communs, dont les schémas et ontologies des métadonnées, par l'ensemble des milieux de recherche. Elle exige aussi la maintenance, au sein des milieux de recherche, des registres de ces artéfacts afin de faciliter leur découverte et leur utilisation.

Pour les besoins de la description de l'écosystème de la GDR, les éléments de l'interopérabilité technique et sémantique sont décrits selon les catégories d'architecture et d'accès.

## Architecture

Si on veut favoriser l'interopérabilité sémantique et technique entre les éléments de l'écosystème de l'IRN, on doit se doter de cadres d'exploitation qui définissent les procédures, les critères et les relations nécessaires pour permettre l'échange de données sans contrainte entre les infrastructures de recherche numérique. Ces cadres munissent l'écosystème d'une architecture qui permet le développement et l'intégration de nouvelles données, de nouveaux logiciels et de nouvelles infrastructures se conformant aux cadres existants.

## Normes et schémas

Le terme schéma décrit un cadre commun qui modélise les structures et les relations entre une série d'éléments d'information connexes. Une fois officialisé par une organisation agréée, un schéma devient une norme : un document qui s'applique collectivement aux codes, aux spécifications, aux pratiques recommandées, aux classifications, aux méthodes d'essai et aux guides qui ont été préparés par une organisation ou un groupe chargé de l'élaboration de normes

---

<sup>190</sup> Corcho, O. et al. (2020). EOSC Interoperability Framework (v1.0). <https://www.eoscsecretariat.eu/sites/default/files/eosc-interoperability-framework-v1.0.pdf> (anglais).

<sup>191</sup> Heikki, L. et al. (2019). D2.1 Report on FAIR requirements for persistence and interoperability 2019. Zenodo. <https://zenodo.org/record/3557381> (anglais).

et qui ont été publiés en conformité avec les procédures établies<sup>192</sup>. Favoriser l'adoption ou l'évolution des normes existantes, au lieu de contribuer à l'expansion de schémas personnalisés ou nouveaux, est un facteur important à prendre en considération lorsqu'on envisage de développer un service national. Les milieux de recherche qui sont réputés être principalement desservis par des données à longue traîne ou qui n'ont pas de riche culture de gestion des données auraient avantage à rechercher les pratiques exemplaires éprouvées au sein des milieux scientifiques axés sur les mégadonnées, à savoir la physique, l'astronomie, l'observation de la Terre et les sciences « omiques », qui ont, par nécessité, créé des normes et des systèmes éprouvés pour gérer des données complexes, ainsi que les services connexes.

- Au niveau du fichier de données de recherche, les formats de fichier et les schémas de métadonnées standard et largement adoptés au sein des milieux de recherche favorisent la collecte systématique de données et leur recombinaison dans divers ensembles de données et permettent l'élaboration de pratiques exemplaires communes en matière de curation et de préservation. Diverses normes thématiques par domaine sont utilisées par les milieux de recherche et sont souvent élaborées et proposées par des associations liées aux domaines (p. ex. GA4GH<sup>193</sup> et DDI Alliance<sup>194</sup>). La mise en correspondance de ces normes avec des cadres ou des normes de métadonnées conceptuels (p. ex. ISO 11179<sup>195</sup>) est une priorité pour assurer une interopérabilité accrue.
- L'établissement de normes pour les logiciels de recherche est également nécessaire afin de favoriser l'interopérabilité des données de recherche dans diverses plateformes logicielles et piles (p. ex. la norme REFI-QDA<sup>196</sup>).
- Au niveau de l'infrastructure, des protocoles de communication standard sont nécessaires pour échanger des données et des métadonnées (p. ex. OAI-PMH<sup>197</sup>) entre systèmes.

Dans l'ensemble de l'écosystème, les développeurs et les développeuses, ainsi que les gestionnaires de l'infrastructure adoptent des cadres et les font progresser par l'entremise d'organes de normalisation nationaux et internationaux (p. ex. l'Organisation internationale de normalisation [ISO], le Conseil canadien des normes [CCN], la National Information Standards Organization [NISO]). L'adoption de ces éléments progresse de façon naturelle au sein des milieux de recherche, pendant que les cadres ou les outils connexes les mieux adaptés se démarquent ou sont promus délibérément par les organismes de recherche. Une connaissance approfondie des normes et des schémas manque dans divers domaines et devrait être favorisée à l'échelle nationale en collaboration avec les associations des domaines pour promouvoir une adoption plus large.

---

<sup>192</sup> Standard [Norme]. (s. d.) Dans CASRAI Glossary. Repéré en novembre 2020 à [https://casrai-test.evision.ca/?page\\_id=485](https://casrai-test.evision.ca/?page_id=485) (anglais).

<sup>193</sup> <https://www.ga4gh.org> (anglais).

<sup>194</sup> <https://ddialliance.org> (anglais).

<sup>195</sup> <https://www.iso.org/fr/standard/68766.html>.

<sup>196</sup> <https://www.qdasoftware.org/> (anglais).

<sup>197</sup> <https://www.openarchives.org/pmh/> (anglais).

Établi en mai 2019, le Collectif canadien de normalisation en matière de gouvernance des données travaille actuellement sur la proposition d'une activité coordonnée de normalisation dans l'ensemble des groupes de parties prenantes au Canada (p. ex. le secteur public, le secteur universitaire et l'industrie)<sup>198</sup>. Le rôle du Collectif sera de produire une feuille de route exposant les lacunes et les besoins du paysage et de déterminer les domaines prioritaires à aborder où des normes et une conformité sont nécessaires.

### *Ontologies*

Les ontologies représentent, nomment et définissent les catégories d'entités dans un domaine donné, les propriétés de ces entités et les relations entre elles<sup>199</sup>. Une utilisation plus importante des ontologies pour représenter de façon structurée les catégories de concepts, de données et d'entités dans un ensemble de données, les propriétés de ces divers éléments et les relations entre eux est un facteur essentiel de la promotion de l'interopérabilité sémantique. Pour permettre leur interprétation et leur utilisation appropriée au sein d'un domaine et dans plusieurs domaines, les données et les métadonnées de recherche doivent avoir une signification évidente, exprimée par des moyens lisibles par machine dans le but de respecter les principes FAIR. Cet objectif peut être atteint au moyen d'artéfacts tels que les vocabulaires et les thésaurus contrôlés, et les taxonomies connexes.

Une gamme d'ontologies peut exister au sein d'un domaine donné, comme il ressort clairement de l'exploration d'une ressource comme BioPortal, maintenu par le É.U. National Centre for Biomedical Ontology<sup>200</sup>. Les artéfacts ontologiques découlent d'une série de sources, liées aux domaines de recherche, et à l'utilisation d'instruments, de systèmes ou de méthodologies, et sont promues par des associations nationales et internationales, ainsi que par des initiatives partagées du milieu. Par exemple, l'Open Biomedical and Biological Ontologies (OBO) Foundry a été lancé en 2007 par des développeurs et des développeuses d'ontologies visant la collaboration ouverte et le respect de principes communs<sup>201</sup>. Des exemples axés sur les domaines comprennent le MMI Ontology Registry and Repository pour les sciences de la mer<sup>202</sup> et l'ESIP Community Ontology Repository pour les sciences de la Terre<sup>203</sup>. En réponse à la pandémie de COVID-19,

---

<sup>198</sup> <https://www.scc.ca/fr/programmes-phares/gouvernance-des-donnees>.

<sup>199</sup> Ontologie (informatique). (s. d.). Dans Wikipedia. [https://fr.wikipedia.org/wiki/Ontologie\\_\(informatique\)](https://fr.wikipedia.org/wiki/Ontologie_(informatique))

<sup>200</sup> <https://bioportal.bioontology.org/> (anglais).

<sup>201</sup> <http://www.obofoundry.org> (anglais).

<sup>202</sup> <https://mmisw.org> (anglais).

<sup>203</sup> <http://cor.esipfed.org> (anglais).

CIDO<sup>204</sup>, une ontologie biomédicale de source ouverte axée sur le milieu dans le domaine des maladies infectieuses causées par des coronavirus, a également été lancée<sup>205</sup>.

Comme il est le cas pour les normes et les schèmes, il est essentiel de promouvoir la connaissance de l'existence de ces artefacts sémantiques et de leur importance pour l'interprétation et la réutilisation des données dans l'ensemble des domaines afin de favoriser une plus grande interopérabilité. Des langages communs et des correspondances entre domaines pour exprimer les différences sémantiques entre les concepts partagés sont également nécessaires pour favoriser une interopérabilité plus vaste. L'utilisation de l'IA et des outils d'apprentissage machine pour le développement et l'harmonisation d'ontologies est un domaine actif de recherche<sup>206</sup>.

### *Registres et index*

Les registres sont des ressources renfermant des identifiants attribués à des objets et comportant une description des termes connexes. Ils jouent un rôle important dans la connexion des utilisateurs et des utilisatrices et des appareils, et de leur orientation dans l'ensemble de l'écosystème de la GDR<sup>207</sup>. Ils profitent non seulement aux chercheurs et aux chercheuses cherchant des outils et des ressources pour leur travail, mais aussi aux développeurs et aux développeuses, ainsi qu'aux administrateurs et aux administratrices souhaitant intégrer les éléments dans des systèmes et des infrastructures. Les identifiants permanents (IP), un outil habilitant de base de ces ressources, font l'objet d'une discussion plus approfondie dans la section sur l'accès.

Les index sont similaires aux registres, mais ne comportent pas d'identifiants homologués associés aux objets énumérés. Les registres et les index existent ou sont en voie de développement pour une gamme d'éléments de l'écosystème de la GDR. Comme il a été mentionné ci-haut, BioPortal est un exemple d'une ressource qui indexe les ontologies dans le domaine biomédical. Fairsharing.org est un autre index organisé de normes de métadonnées, qui sont interreliées avec les bases de données et les politiques relatives aux données<sup>208</sup>. PRONOM est un registre de formats de fichiers et de renseignements techniques connexes<sup>209</sup>.

L'Interoperability Service Reference Framework développé par ELIXIR (l'infrastructure européenne pour l'information biologique), un organisme de GDR en biosciences soutenu par

---

<sup>204</sup> CIDO: Coronavirus Infectious Disease Ontology, plateforme GitHub. Repéré en novembre 2020 à <https://github.com/CIDO-ontology/cido> (anglais).

<sup>205</sup> He, Y. et al. (2020). CIDO, a community-based ontology for coronavirus disease knowledge and data integration, sharing, and analysis. *Sci Data* 7, 181. <https://doi.org/10.1038/s41597-020-0523-6> (anglais).

<sup>206</sup> Gromann, D., Espinosa Anke, L., et Declerck, T. (2019). Special Issue on Semantic Deep Learning. 815 – 822. <https://content.iospress.com/articles/semantic-web/sw190364> (anglais).

<sup>207</sup> Registry [Registre]. (s. d.) Dans CASRAI Glossary. Repéré en novembre 2020 à <https://casrai-test.evision.ca/glossary-term/registry/> (anglais).

<sup>208</sup> <https://fairsharing.org/> (anglais).

<sup>209</sup> <https://www.nationalarchives.gov.uk/PRONOM/Format/proFormatSearch.aspx?status=new> (anglais).

EOSC, fournit un modèle du développement d'un service de registre qui aide les personnes et les machines à découvrir, à consulter, à intégrer et à analyser les données de recherche<sup>210</sup>. Il convient de soutenir des ressources similaires à l'échelle nationale et à l'échelle des domaines, afin de soutenir une gamme d'éléments de l'écosystème de l'IRN.

### *Certifications*

Les capacités de l'IRN, les niveaux de confiance en celle-ci et les données de recherche qu'elle soutient varient selon les exigences des systèmes et des utilisateurs et des utilisatrices. Les certifications fournies par les autorités auxquelles un milieu fait confiance garantissent la conformité avec les dispositions des normes, des codes ou des autres exigences utilisés<sup>211</sup>. Une certification d'assurance de la qualité est la certification ISO, qui dépend de la disponibilité des normes ISO sous-jacentes par rapport auxquelles l'évaluation est réalisée<sup>212</sup>.

Dans le milieu des dépôts de données, CoreTrustSeal s'est dégagé comme étant la certification dominante des dépôts de confiance<sup>213</sup>. CoreTrustSeal octroie à tout dépôt de données intéressé une certification de base se basant sur le catalogue des exigences des dépôts de données fiables de base (Core Trustworthy Data Repositories Requirements)<sup>214</sup>, développé par Data Seal of Approval et World Data System sous l'égide de la Research Data Alliance afin d'harmoniser leurs propres certifications de dépôts de données. Aider les dépôts de données de recherche canadiens à atteindre les normes élevées de CoreTrustSeal est une initiative prévue du Réseau Portage, qui permettra d'améliorer la qualité et les normes des dépôts canadiens à l'échelle nationale.

Au niveau des données de recherche et des méthodes de recherche, de nouveaux modèles de certification sont à l'étude pour affirmer la qualité et la reproductibilité de la recherche publiée. Mentionnons à titre d'exemple, la Certification Agency for Scientific Code and Data (CASCAD), soutenue par la Fondation nationale pour les sciences en France, et un consortium d'établissements de recherche français<sup>215</sup>. L'évaluation de la reproductibilité et de la qualité des données confidentielles et des analyses connexes est particulièrement difficile pour les données sensibles. C'est pour cette raison que CASCAD et le Centre d'accès sécurisé aux données (homologue du RCCDR) ont collaboré à la conception d'un processus de certification de la

---

<sup>210</sup> <https://elixir-europe.org/platforms/interoperability> (anglais).

<sup>211</sup> Certified product [Produit certifié]. (s. d.) Dans CASRAI Glossary. Repéré en novembre 2020 à <https://casrai-test.evision.ca/glossary-term/certified-product/> (anglais).

<sup>212</sup> <https://www.isoqsltd.com/fr/faq/> (FAQ partielle).

<sup>213</sup> <https://www.coretrustseal.org/> (anglais).

<sup>214</sup> ICSU World Data System. (2016). Core Trustworthy Data Repositories Requirements v01.00. [https://www.coretrustseal.org/wp-content/uploads/2017/01/Core\\_Trustworthy\\_Data\\_Repositories\\_Requirements\\_01\\_00.pdf](https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf) (anglais).

<sup>215</sup> <https://www.cascad.tech/> (anglais).

reproductibilité pour les données confidentielles<sup>216</sup>. Des entités et des relations similaires pourraient être envisagées dans le contexte canadien.

### Identification et accès

En raison de l'envergure et de la complexité de la recherche nationale et internationale, les contrôles qui définissent les relations entre les entités et les permissions d'accès jouent, eux aussi, un rôle important dans la mise en communication de l'écosystème de la GDR.

Les identifiants permanents (IP) se dégagent comme des éléments importants du soutien de la GDR. Les IP sont des références durables à des objets uniques qui fournissent une information sur l'objet en question indépendamment des établissements individuels ou de l'infrastructure<sup>217</sup>. Les systèmes d'IP deviennent ainsi le moyen préférentiel de renvoi et d'accès non ambigus et durables. La gamme d'objets pour laquelle des systèmes d'IP pourraient être développés est vaste et englobe les chercheurs et les chercheuses, les ensembles de données, les instruments scientifiques, et d'autres éléments.

Bien que les systèmes d'IP puissent être déployés et gérés à l'échelle locale au sein des établissements de recherche, la mise en œuvre de systèmes largement adoptés et administrés par des tiers présente une approche plus cohérente d'établissement de liaisons entre les données et l'infrastructure dans l'ensemble de l'écosystème. Au Canada, deux systèmes d'IP ont été approuvés à l'échelle nationale.

- ORCID (Open Research and Contributor ID) est un système d'IP qui vise l'identification unique des chercheurs et des chercheuses et qui a été adopté par d'importants organismes de financement, maisons d'édition et établissements de recherche à l'échelle mondiale<sup>218</sup>. L'IP d'ORCID est en mesure de relier les chercheurs et les chercheuses à l'écosystème de la recherche au moyen de leur dossier de publications, de leur historique d'emploi, de leurs bourses, de leurs collaborations et d'autres résultats. Le système d'identification des chercheurs et des chercheuses ORCID est approuvé par le Consortium ORCID au Canada, qui offre aux organismes de recherche une adhésion à ORCID à un prix réduit, et offre aussi un soutien et une mobilisation du milieu<sup>219</sup>. Il compte actuellement 37 membres.
- DataCite est un fournisseur mondial d'un système d'IP d'identifiants numériques d'objet pour les ensembles de données, qui a été mis en œuvre par de nombreux dépôts de données<sup>220</sup>. Le Consortium DataCite Canada, une initiative récente dirigée par le Réseau canadien de documentation pour la recherche (RCDR) et le réseau Portage de l'ABRC,

---

<sup>216</sup> Pérignon et al. (2019). Certify reproducibility with confidential data. *Science*, 365(6449):127-128. <https://science.sciencemag.org/content/365/6449/127> (anglais).

<sup>217</sup> Koster, L. (2020). Persistent identifiers for heritage objects. *Code4Lib*, (47). <https://journal.code4lib.org/articles/14978> (anglais).

<sup>218</sup> <https://orcid.org/> (site partiellement en français)

<sup>219</sup> <https://orcid-ca.org/about> (anglais).

<sup>220</sup> <https://datacite.org/> (anglais).

partage les frais liés à l'adhésion à DataCite entre les 45 organisations membres, tout en assurant une administration centrale et un soutien au milieu<sup>221</sup>.

Parmi les autres types de systèmes d'IP internationaux émergents, mentionnons les identifiants des activités de recherche (RAID)<sup>222</sup>, les identifiants des instruments scientifiques (PIDINST)<sup>223</sup> et les identifiants d'organismes de recherche (ROR)<sup>224</sup>.

Pour être efficaces, les IP doivent pouvoir être intégrés dans tous les systèmes et toutes les interfaces qu'on utilise pour décrire les objets en question, aussi bien pour les utilisatrices et les utilisateurs humains que les machines. Cependant, même l'IP le plus largement adopté est aussi permanent que peut l'être l'administration de son système. La promotion délibérée de l'adoption des systèmes d'IP recommandés et la recherche de modèles de coûts durables, tels que les approches de consortium décrites ci-haut, sont essentielles pour assurer la disponibilité à long terme des systèmes d'IP.

Lorsque les IP sont interopérables avec les systèmes d'identité, ils peuvent également soutenir les systèmes d'authentification et d'autorisation. Au sein des établissements d'enseignement supérieur canadiens, les ressources de réseau offertes par CANARIE et le RNRE du Canada soutiennent les solutions de gestion de l'identité et de l'accès pour les établissements d'enseignement et de recherche canadiens et les fournisseurs de contenu. À titre d'exemple, la Fédération canadienne d'accès de CANARIE est un environnement de gestion d'accès fiable qui offre aux chercheurs et aux chercheuses dans l'ensemble du Canada une connectivité aux plateformes tierces en utilisant les justificateurs d'identité de leur établissement de recherche d'attache<sup>225</sup>. Le service d'identité fédéré du Canada prend une dimension internationale en raison de sa participation à eduGAIN, un service d'interconnexion qui relie les fédérations d'identité à l'échelle mondiale<sup>226</sup>. La cybersécurité est un élément connexe de l'écosystème de la GDR, y compris les mécanismes et les stratégies utilisés pour prévenir l'accès non autorisé aux ressources numériques, depuis le niveau des données jusqu'au niveau des réseaux. La cybersécurité est un dossier promu conjointement par l'Alliance et CANARIE et fera l'objet de discussions dans d'autres produits.

## Services de données

À mesure que les besoins des chercheurs et des chercheuses ont pris de l'ampleur en raison des avancées technologiques et des progrès réalisés dans le domaine des pratiques de recherche, ainsi que des nouvelles attentes des établissements hébergeurs, des organismes de financement et des maisons d'édition de revues, une panoplie de services connexes a été développée pour

---

<sup>221</sup> <https://www.crkn-rcdr.ca/fr/consortium-datacite-canada>

<sup>222</sup> <https://www.raid.org.au/> (anglais).

<sup>223</sup> <https://www.rd-alliance.org/group/persistent-identification-instruments/case-statement/persistent-identification-instruments> (anglais).

<sup>224</sup> <https://ror.org/> (anglais).

<sup>225</sup> <https://www.canarie.ca/fr/identite/fca/>.

<sup>226</sup> <https://technical.edugain.org/status> (anglais).



aider les chercheurs et les chercheuses à accéder à l'IRN afin d'améliorer la GDR grâce à l'adoption de nouvelles pratiques et de nouveaux outils que ces spécialistes de la recherche intègrent dans leurs flux de tâches de recherche. Des services sont offerts au sein des établissements de recherche universitaires par les fournisseurs d'infrastructures en conjugaison avec les plateformes et les outils particuliers qu'ils offrent par l'entremise des associations et des fournisseurs de service régionaux et nationaux, et de plus en plus, par des entités commerciales. La gamme de services offerts s'étend sur le cycle de vie des données de recherche et cible la gestion des données de recherche dans toutes leurs formes.

Comme il a été décrit dans la section ci-haut traitant du stockage et du calcul, bon nombre de groupes de recherche exploitent aussi une infrastructure de GDR, en particulier les dépôts spécialisés axés sur un projet ou un domaine, ce qui rend leurs données de recherche repérables et accessibles au milieu de recherche général. Ainsi, des modèles sont en voie de développement pour aider les organismes de recherche à développer et à exploiter l'IRN à des fins de GDR.

## Soutien des données de recherche

### *Planification*

Un plan de gestion des données (PGD) est un énoncé officiel qui décrit la façon dont les données seront gérées et documentées tout au long d'un projet de recherche, ainsi que les conditions du dépôt des données dans un dépôt à des fins de gestion, de partage et de préservation à long terme<sup>227</sup>. L'élaboration d'un PGD au début d'un projet de recherche peut aider à cerner, et par conséquent à atténuer, les difficultés liées à la gestion des données de recherche générées tout au long d'un projet, et est donc réputée être une pratique exemplaire de GDR. Les PGD constituent une attente des organismes octroyant des subventions et des bourses de recherche et des établissements universitaires aux États-Unis, au Royaume-Uni et en Australie<sup>228</sup>. Il est prévu que la politique prochaine des trois organismes sur la GDR les désigne comme exigence du financement.

En 2015, le réseau Portage de l'ABRC a dirigé la création d'une plateforme en ligne pour aider à créer un PGD. Appelée Assistant PGD, la plateforme est bilingue et est mise librement à la disposition des chercheurs et des chercheuses dans l'ensemble du pays<sup>229</sup>. University of Alberta Libraries est l'organisation qui héberge et maintient la plateforme, mais ce sont les établissements et les groupes de recherche individuels qui gèrent les modèles locaux pour leur milieu de recherche respectif. Le 7 juillet, la plateforme comptait 52 comptes institutionnels et 12 489 chercheurs et chercheuses étaient titulaires d'un compte individuel. Les responsabilités en matière de relations externes, de soutien et de mobilisation sont partagées entre le secrétariat du réseau Portage et les établissements qui ont mis l'outil en service au niveau local. Au sein des établissements universitaires, le soutien accordé à l'échelle locale aux chercheurs et aux

---

<sup>227</sup> Data Management Plan [Plan de gestion de données]. (s. d.) Dans CASRAI Glossary. Repéré en novembre 2020 à <https://casrai.org/term/data-management-plan> (anglais).

<sup>228</sup> Unsworth K., et Smale, N. (2017) Overview of Australian institution and UK/USA funding body data management plan mandates. University of Melbourne. Jeu de données. <https://doi.org/10.4225/49/5986bde74f8f5> (anglais).

<sup>229</sup> <https://assistant.portagenetwork.ca/fr>.

chercheuses utilisant Assistant PGD et élaborant des PGD est souvent partagé entre les bibliothèques et les bureaux de recherche.

Bien que les PGD constituent une pratique établie depuis plus d'une décennie, ils sont principalement des documents à lisibilité humaine, ce qui limite leur utilité au-delà du groupe de recherche les ayant rédigés. Il est prévu que la prochaine génération de PGD soit exploitable par des machines, ce qui permettra aux plans d'assurer un soutien plus efficace aux activités de recherche grâce à leur interopérabilité avec les autres systèmes de recherche<sup>230</sup>. Mentionnons à titre d'exemple des avantages prévus, l'intégration avec les applications des comités d'éthique de la recherche, la notification des fournisseurs de stockage relativement aux besoins en matière de capacité et la prévision des processus de préservation.

Attendue à l'automne 2020, la mise à niveau d'Assistant PGD utilisera les codes de base fusionnés d'autres plateformes existantes de PGD (appelée la feuille de route de la PGD) qui comportera une fonctionnalité d'interface de protocole d'application, ouvrant ainsi la voie à une meilleure lisibilité par machine et une meilleure interopérabilité<sup>231</sup>.

### *Curation*

Une gamme de services de curation des données est apparue dans l'ensemble de l'écosystème de la GDR en réponse aux avancées touchant l'infrastructure et la disponibilité de celle-ci, ainsi qu'aux changements qui s'opèrent dans l'attitude des chercheurs et des chercheuses et les exigences en matière de gestion et de partage des données à long terme. La curation des données se définit le mieux comme processus itératif d'optimisation des ensembles de données en vue de leur exploitation actuelle ou de leur découverte et de leur réutilisation futures, qui est guidé par le cadre conceptuel des principes FAIR<sup>232</sup>. De bonnes pratiques de curation favorisent la gestion des données tout au long de leur cycle de vie, pendant que les données et les collections sont nettoyées, documentées, normalisées, interreliées, stockées et partagées. Le processus fait appel à de nombreux ensembles de compétences, notamment une connaissance interdisciplinaire, une bonne connaissance du processus de recherche, une connaissance des normes gouvernant les métadonnées et des pratiques exemplaires en matière de gestion des données, ainsi qu'une capacité à exploiter diverses technologies.

Les chercheurs et les chercheuses sont encouragés à documenter les processus de collecte, de manipulation et d'analyse des données pendant la partie active de leur processus de recherche pour veiller à la collecte de métadonnées exactes et complètes en vue de favoriser les activités ultérieures de gestion, telles que la publication, la préservation et la réutilisation. Cependant, plusieurs ensembles de données actuellement publiés font état d'une documentation irrégulière, et la curation est soulevée uniquement à la fin du projet de recherche. Vu la portée de la curation des données, les chercheurs et les chercheuses peuvent obtenir un soutien de la curation des

---

<sup>230</sup> Simms, S. et al. (2017). Machine-actionable data management plans (maDMPs). *Research Ideas and Outcomes* 3: e13086. <https://doi.org/10.3897/rio.3.e13086> (anglais).

<sup>231</sup> DMPRoadmap, plateforme GitHub. Repéré en novembre 2020 à <https://github.com/DMPRoadmap> (anglais).

<sup>232</sup> Réseau Portage. Introduction à la curation de données. [https://portagenetwork.ca/wp-content/uploads/2019/09/Curation\\_Primer\\_Aug2019\\_FR.pdf](https://portagenetwork.ca/wp-content/uploads/2019/09/Curation_Primer_Aug2019_FR.pdf).

données de diverses sources. En règle générale, le processus est éclairé par le besoin d'assurer la fonctionnalité d'un ensemble de données particulier avec un autre élément de l'écosystème de la GDR ou de l'écosystème plus vaste de l'IRN (p. ex. l'intégration avec un autre ensemble de données, le respect d'une norme des métadonnées, le fonctionnement avec un logiciel de recherche, le dépôt dans un dépôt de données, le partage avec des collègues). Bien que certaines activités de curation puissent être exécutées au moyen d'un logiciel de recherche (méthode que l'on doit continuer à étudier comme moyen évolutif d'aider les chercheurs et les chercheuses), vu la complexité du processus, elle a généralement besoin d'une certaine forme d'intervention humaine.

Au sein des établissements d'enseignement supérieur, le soutien de la curation des données de recherche est assuré à l'intérieur des groupes de recherche. À titre d'exemple, bon nombre de groupes avec des ressources de données importantes retiennent les services d'un gestionnaire ou d'une gestionnaire de données pour favoriser la curation. Un soutien est également sollicité au service des systèmes ou de la TI pour organiser les données de sorte à permettre leur intégration avec l'infrastructure disponible de calcul pour la recherche. Les bibliothèques et les archives offrent aussi un soutien de curation aux chercheurs et aux chercheuses qui déposent des données dans les dépôts institutionnels ou d'autres systèmes hébergés de gestion des ressources numériques, ainsi que dans des dépôts de données externes. Un sondage récent sur la capacité actuelle des établissements canadiens d'enseignement supérieur à offrir des services de GDR a constaté qu'un peu plus de la moitié des 77 répondants offraient une certaine forme de soutien de la curation par l'entremise de leurs bibliothèques<sup>233</sup>. Le même sondage a révélé qu'entre un quart et un tiers des répondants offrent des services informatiques de données et/ou un soutien technique (p. ex. chiffrement/anonymisation des données) par l'entremise de leurs services des systèmes ou de la TI. Une lacune importante qui est relevée dans bon nombre d'établissements de recherche est le soutien limité que ceux-ci accordent à la curation et au stockage de données sensibles, lesquelles sont particulièrement pertinentes à la lumière du financement récent de programmes liés à la COVID-19, qui exigent souvent un partage ouvert des résultats et des données<sup>234</sup>.

Les discussions tenues par les participants du Forum canadien sur la curation des données de 2019 ont fait ressortir le besoin de cerner le niveau de curation que les établissements sont disposés à offrir<sup>235</sup>. Le Data Curation Network (DCN) est un modèle de mise à l'échelle de la capacité et de la spécialisation qui sont nécessaires pour soutenir la curation de données dans de nombreux domaines représentés au sein des établissements d'enseignement supérieur<sup>236</sup>. Les établissements qui adhèrent au DCN partagent les ressources humaines avec d'autres établissements membres, mutualisant ainsi leur temps et la diversité de leur expertise. Cela permet aux dépôts de données de tous les membres de veiller à la curation collective et plus

---

<sup>233</sup> Cooper, A. et al. (2020). Sondage sur la capacité des services institutionnels de gestion des données de recherche. <https://doi.org/10.14288/1.0388722>.

<sup>234</sup> Tayler, F. et Ripp, C. (2020). FAQ : Partage et dépôt de données en appui à une intervention rapide face à la COVID-19. <https://doi.org/10.5683/SP2/522KV2>.

<sup>235</sup> Brodeur, J., Sawchuk, S., et Newson, K. (2020). Materials from the 2019 Canadian Data Curation Forum. Zenodo. <http://doi.org/10.5281/zenodo.3899401> (anglais).

<sup>236</sup> <https://datacurationnetwork.org> (anglais).

efficace d'une plus grande variété de types de données (p. ex. disciplines, types de fichiers, logiciels, etc.) au-delà de ce que la majorité des établissements individuels seraient en mesure d'offrir sans aide.

Le type de soutien de la curation le plus commun que les chercheurs et les chercheuses obtiennent est lorsqu'ils prennent la décision de publier leurs données dans un dépôt. La majorité des dépôts de données, qu'ils soient généralistes ou spécialisés, offrent un certain niveau de soutien de la curation aux chercheurs et aux chercheuses qui exploitent l'infrastructure du dépôt. Les dépôts exploitent un nombre diversifié de plateformes logicielles, qui permettent ou automatisent quelques activités de curation pendant le processus de dépôt. L'auteur d'une étude récente décompose ces fonctions selon le modèle Repository Interest Group de la RDA<sup>237</sup>.

Les dépôts de données offrent aussi un soutien plus direct aux chercheurs et aux chercheuses par l'entremise de leurs responsables de la curation des données, et de leurs équipes de soutien. Le soutien offert peut aller d'une instruction sur le mode d'emploi de leurs plateformes et l'interprétation des schémas de métadonnées pris en charge, à une curation plus exhaustive des fichiers de l'ensemble de données, ainsi qu'au codage et à la documentation connexes. Mentionnons par exemple le partenariat récent aux États-Unis entre les National Institutes of Health (NIH) et figshare pour piloter un nouveau dépôt généraliste destiné à la recherche financée par les subventions des NIH sans dépôt spécialisé désigné pour ses données<sup>238</sup>. Dans ce pilote, les déposants et déposantes sont associés à des bibliothécaires de données qualifiés qui examinent les métadonnées et les licences<sup>239</sup>. Ce niveau de curation s'oppose au soutien à intervention plus directe, généralement offert par les dépôts plus spécialisés. À titre d'exemple, les chercheurs et les chercheuses souhaitant déposer des données dans l'International Neuroimaging Data-Sharing Initiative (INDI) sont priés de communiquer en premier lieu avec le dépôt pour obtenir des instructions détaillées sur la méthode à suivre pour contribuer des données<sup>240</sup>. Dans ce modèle, les responsables de la curation sont accessibles pour aider à diagnostiquer et à résoudre les problèmes pendant le processus de dépôt, développer des scripts personnalisés pour les chercheurs et les chercheuses qui automatiseront la préparation des données, ou préparer les données pour eux et les elles si leur comité d'éthique de la recherche l'autorise<sup>241</sup>.

Les modèles précisant le moment de l'intervention de curation varient également. À titre d'exemple, les ensembles de données déposés dans le DFDR ou Dryad, qui sont tous les deux des dépôts généralistes, sont mis en file d'attente aux fins d'examen par un ou une responsable

---

<sup>237</sup> Kim, S. (2018). Functional requirements for research data repositories. IJKCDT, 8(1), 25-36. <http://ijkcdt.net/xml/13281/13281.pdf> (anglais).

<sup>238</sup> <https://nih.figshare.com> (anglais).

<sup>239</sup> Hyndman, A. (2019). Figshare announces data repository partnership with the National Institutes of Health to store and reuse research data. Repéré en novembre 2020 à [https://figshare.com/blog/Figshare\\_announces\\_data\\_repository\\_partnership\\_with\\_the\\_National\\_Institutes\\_of\\_Health\\_to\\_store\\_and\\_reuse\\_research\\_data/518](https://figshare.com/blog/Figshare_announces_data_repository_partnership_with_the_National_Institutes_of_Health_to_store_and_reuse_research_data/518) (anglais).

<sup>240</sup> [http://fcon\\_1000.projects.nitrc.org](http://fcon_1000.projects.nitrc.org) (anglais).

<sup>241</sup> INDI. (s. d.). Data Contribution Guide v3.1. [http://fcon\\_1000.projects.nitrc.org/indi/indi\\_data\\_contribution\\_guide.pdf](http://fcon_1000.projects.nitrc.org/indi/indi_data_contribution_guide.pdf) (anglais).

de la curation avant la publication finale. Les responsables de la curation examinent l'ensemble de données afin de vérifier l'intégralité des métadonnées, ainsi que l'organisation et la réutilisabilité des fichiers de données, et de s'assurer que les fichiers ne sont pas altérés et ne renferment pas de renseignements sensibles, et qu'ils respectent les conditions d'utilisation à tout autre point de vue<sup>242</sup>. Entre-temps, on propose aux chercheurs et aux chercheuses qui déposent des données dans le Qualitative Data Repository (QDR) de faire un dépôt initial comportant le strict minimum de renseignements au début de leur projet afin de déclencher un processus de consultation continue<sup>243</sup>.

Les dépôts de données qui possèdent l'infrastructure et l'expérience nécessaires pour se charger d'ensembles de données susceptibles de renfermer des renseignements sensibles jouent un rôle important dans la curation des données en conformité avec les normes et les lois établies ainsi qu'avec les exigences des comités d'éthique. À titre d'exemple, les chercheurs et les chercheuses qui soumettent des données à la Cancer Imaging Archive sont priés d'anonymiser et de chiffrer les données avant de les soumettre. Tous les ensembles de données déposés sont traités comme s'ils renfermaient des renseignements sensibles et sont saisis, en premier lieu, dans un système protégé, où ils sont examinés par des responsables de la curation ayant suivi une formation sur la réglementation des renseignements de santé et de protection des renseignements personnels<sup>244</sup>. D'autres dépôts capables d'accueillir des données sensibles entreprennent des examens similaires du risque lié à la communication. À titre d'exemple, les responsables de la curation à ICPSR proposeront des méthodes de modification des données afin de limiter les risques, ou suggéreront le partage des données au sein de leur dépôt à un niveau de restriction plus élevé<sup>245</sup>.

Un plus grand nombre d'initiatives de coordination de la curation des données dans l'ensemble des organisations et de l'écosystème de l'IRN profiteraient considérablement aux données de recherche, aux plateformes connexes et aux compétences des responsables de la curation même. Au Canada, le réseau Portage dirige une initiative de coordination nationale de la curation des données au moyen d'un réseau modèle de responsables de la curation canadiens. Les possibilités d'une coordination nationale ont été discutées au Forum canadien de la curation des données de 2019. Les discussions ont abouti à la publication d'un rapport sommaire avec des recommandations destinées à l'Alliance concernant la promotion du soutien national de la curation des données. Ces recommandations comprennent l'investissement dans le développement de la capacité humaine au sein des organismes de recherche et de l'infrastructure

---

<sup>242</sup> FRDR (s. d.). Documentation : Après le dépôt. Repéré en novembre 2020 à [https://www.frdr-dfdr.ca/docs/fr/apr%C3%A8s\\_le\\_d%C3%A9p%C3%B4t](https://www.frdr-dfdr.ca/docs/fr/apr%C3%A8s_le_d%C3%A9p%C3%B4t) ; et Dryad (s. d.) Dryad Submission and Publication Process. Repéré en novembre 2020 à [https://datadryad.org/stash/submission\\_process#curation](https://datadryad.org/stash/submission_process#curation) (anglais).

<sup>243</sup> <https://qdr.syr.edu> (anglais).

<sup>244</sup> Cancer Imaging Archive. (2020). Submission and De-identification Overview. Repéré en novembre 2020 à <https://wiki.cancerimagingarchive.net/display/Public/Submission+and+De-identification+Overview> (anglais).

<sup>245</sup> ICPSR. (s. d.). Data Confidentiality. Repéré en novembre 2020 à <https://www.icpsr.umich.edu/web/pages/datamanagement/confidentiality/index.html>

de calcul partagée, qui permettrait de faire progresser la coordination nationale, la communication et la représentation au moyen d'une approche de réseau national<sup>246</sup>.

### Préservation

À l'opposé de la curation, la préservation numérique fait appel à une gamme de stratégies pour favoriser le maintien à long terme de l'accès à la documentation au-delà de la limite de défaillance des supports ou de changement technologique<sup>247</sup>. Il n'y a pas d'état final auquel on peut affirmer qu'un objet numérique est enfin préservé. Une gamme d'activités de préservation est nécessaire pour gérer la diversité des objets numériques associés à un ensemble de données particulier ou à un projet de recherche de plus grande portée. L'analyse de ces activités isolément des autres éléments du paysage d'IRN est difficile, car la préservation numérique dépend de nombreux éléments travaillant en commun pour garantir la préservation réussie des renseignements numériques.

La préservation numérique représente un spectre de tâches, qu'il s'agisse de la protection de l'intégrité des fichiers et de la fonctionnalité des composants logiciels ou de la protection contre l'obsolescence de l'infrastructure de stockage. La National Digital Stewardship Alliance (NDSA) présente sa matrice de niveaux de préservation comme une série de domaines fonctionnels prévoyant des phases de mesures de préservation, allant de la connaissance du contenu numérique (niveau 1), à la protection de ce contenu, à sa surveillance et enfin à son soutien (niveau 4)<sup>248</sup>.

Le développement de logiciels et leur intégration dans l'IRN peuvent faciliter les services de préservation. À titre d'exemple, Archivemata est une série d'outils de source ouverte qui exécutent des tâches distinctes sur les objets numériques ingérés dans le but de produire des résultats prêts à l'archivage<sup>249</sup>. Ce logiciel est actuellement intégré dans les deux options de dépôt nationales du Canada (DFDR et Dataverse de Scholars Portal), et permet aux responsables de la curation d'appliquer un traitement de préservation aux ensembles de données et d'assurer la migration des résultats vers l'archivage.

Cependant, bien que les logiciels puissent faciliter les processus de préservation, il s'agit, en dernier ressort, d'une pratique exigeante en ressources humaines, qui réclame une attention, une vérification et une maintenance. Les services qui soutiennent actuellement la préservation des données de recherche sont limités au sein des établissements de recherche universitaires. Dans l'ensemble des campus, peu de bibliothèques et d'archives au Canada pourvoient des postes

---

<sup>246</sup> Clary, E. et al. (2020). Conceptualisation d'une approche nationale quant aux services de curation de données au Canada. Zenodo. <https://zenodo.org/record/3895031#.YAi36jmSnmE>.

<sup>247</sup> Digital Preservation [*Préservation numérique*]. (s. d.) Dans *CASRAI Glossary*. Repéré en novembre 2020 à <https://casrai.org/term/digital-preservation/> (anglais).

<sup>248</sup> NDSA. (2019). Levels of Digital Preservation. Repéré en novembre 2020 à <https://nds.org/publications/levels-of-digital-preservation/> (anglais).

<sup>249</sup> <https://www.archivemata.org/fr/> (site Web partiellement en français).

pour soutenir la préservation des données de recherche<sup>250</sup>. Cependant, des consortiums de bibliothèques universitaires ont proposé des modèles de soutien partagé à l'échelle régionale. Le Digital Stewardship Network de COPPUL et le service Permafrost d'OCUL fournissent tous les deux des outils, une formation et un soutien à leurs établissements membres pour les aider à préserver leurs collections numériques.

Alors que les organismes de recherche exploitant des dépôts de données spécialisés s'intéressent principalement à la préservation à long terme de leurs propres objets numériques, des modèles de service intéressants sont proposés pour répondre aux besoins de leurs communautés. À titre d'exemple, Population Data BC offre un service d'archivage des versions plus anciennes des données à des fins de référence historique<sup>251</sup>. Le service est également en mesure de stocker des copies d'archive d'extraits de données dans le but de soutenir les chercheurs et les chercheuses tout en respectant les exigences des comités d'éthique de la recherche en matière de période de conservation approuvée<sup>252</sup>. Le Centre canadien de données astronomiques prend en charge des flux de tâches complexes afin d'archiver les données brutes obtenues de divers instruments ainsi que les produits de données découlant du traitement par des pipelines scientifiques pour transformer des quantités énormes de données en produits utilisables par les chercheurs et les chercheuses<sup>253</sup>.

L'élargissement des services existants afin de les intégrer dans un modèle soutenu à l'échelle nationale est le fondement du modèle de fournisseurs de services de préservation distribués du réseau Portage<sup>254</sup>, qui propose une infrastructure partagée proportionnelle tout en facilitant une curation locale des collections individuelles et une prise de décision à leur égard. Cela permettrait la réalisation d'économies d'échelle liées au développement et à la maintenance de l'infrastructure matérielle, sans négliger les connaissances et les objectifs particuliers des collections locales. Bien que le modèle en soit encore à l'étape de la validation de principe, une exigence essentielle à satisfaire pour le déploiement durable d'efforts de préservation à l'échelle nationale sera de faciliter l'interopérabilité des systèmes et des pratiques existants, ce qui dépendra de la création et de l'adoption de cadres et de normes sur lesquels les communautés concernées se mettent d'accord.

## *Découverte*

Un des principaux motifs qui sous-tend la GDR est de permettre le partage et la réutilisation d'ensembles de données existants, tandis qu'un service important, mais souvent méconnu est d'aider les chercheurs et les chercheuses et les organisations à découvrir les données de

---

<sup>250</sup> Cooper, A. et al. (2020). Sondage sur la capacité des services institutionnels de gestion des données de recherche. <https://doi.org/10.14288/1.0388722>.

<sup>251</sup> <https://www.popdata.bc.ca> (anglais).

<sup>252</sup> Population Data BC. (s. d.). Archival back up and storage of data. Repéré en novembre 2020 à <https://www.popdata.bc.ca/dataproviders/services/archivalstorage> (anglais).

<sup>253</sup> Centre canadien de données astronomiques (2020). Une archive comme service. Repéré en novembre 2020 à <https://www.cadc-ccda.hia-ihp.nrc-cnrc.gc.ca/fr/doc/AaaS/>.

<sup>254</sup> Qasim, U. et al. (2018). La préservation des données de recherche au Canada : Livre blanc. <http://dx.doi.org/10.14288/1.0371945>.

recherche qui les intéressent. Le concept de découverte englobe le soutien de la recherche, de l'identification, de l'interprétation et de la consultation d'ensembles de données publiés présentant un intérêt. Au sein des établissements d'enseignement supérieur, les bibliothèques de recherche jouent, depuis plusieurs années, un rôle important dans la prestation aux chercheurs et aux chercheuses d'un soutien de l'identification et de l'utilisation des données d'intérêt accessibles au public. Un jalon important au Canada a été la formation de l'Initiative de démocratisation des données<sup>255</sup>, un partenariat entre Statistique Canada et les établissements d'enseignement supérieur ayant pour objet de favoriser l'accès aux ensembles de données publiés par Statistique Canada.

En raison du nombre croissant de dépôts de données, un effort concerté est nécessaire pour inventorier les dépôts et leurs ressources de données dans l'ensemble du Canada et à l'échelle internationale. Au Canada, la plateforme du Dépôt fédéré de données de recherche (DFDR) se veut une couche nationale de découverte des données et recueille des métadonnées de certains dépôts de données canadiens pour un moteur de recherche national. Ce faisant, le DFDR offre une exposition aux dépôts de données, quelle que soit leur taille, et oriente le trafic d'utilisateurs et d'utilisatrices vers les hébergeurs. Le Groupe d'experts sur la découverte et les métadonnées du réseau Portage dirige une initiative visant à identifier les dépôts de données situés dans l'ensemble du Canada et à sensibiliser le milieu à leur existence en les intégrant dans le service de découverte du DFDR<sup>256</sup>.

À l'échelle internationale, les initiatives visant à indexer les ensembles de données publiés relèvent de Data Citation Index, un outil commercial de Clarivate<sup>257</sup> et de l'outil Dataset Search de Google<sup>258</sup>, qui recueille de l'ensemble du Web les données de recherche (entre autres) décrites au moyen de la norme schéma.org. L'outil ResearchGraph d'OpenAIRE Explore (graphe de recherche) va un pas plus loin en rattachant les métadonnées d'une source massive de résultats scientifiques, dont les ensembles de données, les logiciels et les articles, à une information sur les organismes de recherche, les organismes de financement et les partenaires<sup>259</sup>. Le Dépôt fédéré des données de recherche de Portage est en cours d'intégration avec OpenAIRE et le Data Citation Index pour garantir que les données de recherche canadiennes sont plus facilement repérables à l'échelle internationale.

Un des principaux enjeux liés aux initiatives visant à indexer les données et à veiller à ce qu'elles soient largement repérables tient au fait qu'elles tendent inéluctablement à se fier à des normes simplifiées de métadonnées pour indexer le contenu. À l'opposé, les métadonnées thématiques par domaine sont nettement plus détaillées et abondantes, ce qui aide les chercheurs et les chercheuses à évaluer la pertinence des données et à déterminer si elles conviennent à leurs fins. À l'avenir, on mettra probablement davantage l'accent sur l'indexation de métadonnées plus abondantes ou sur l'élargissement des normes de métadonnées simples. Par exemple, plusieurs

---

<sup>255</sup> <https://www.statcan.gc.ca/fra/idd/idd>.

<sup>256</sup> <https://portagenetwork.ca/fr/reseau-dexperts/groupe-dexperts-sur-la-decouverte-et-des-metadonnees/>.

<sup>257</sup> <https://clarivate.com/webofsciencgroup/solutions/webofscience-data-citation-index/> (anglais).

<sup>258</sup> <https://datasetsearch.research.google.com/> (anglais).

<sup>259</sup> <https://explore.openaire.eu> (anglais).



groupes du milieu créent des extensions de schéma.org pour accroître l'utilisabilité des données dans les biosciences<sup>260</sup> et les sciences de la Terre<sup>261</sup>.

### Exploration

Au lieu de simplement fournir aux chercheurs et aux chercheuses des fichiers de données qu'ils ou qu'elles peuvent télécharger et réutiliser localement, les intégrations logicielles dans les dépôts de données permettent aux utilisateurs et aux utilisatrices d'afficher, de manipuler et d'interpréter les données au sein de la plateforme existante. Ce concept de « visite des données » offre de nombreux avantages au chapitre de la curation et de la préservation à long terme. Par exemple, en réponse à la crise de COVID-19, un ensemble de points de données FAIR (FDP, d'après l'anglais *Fair Data Points*) a été créé sous l'égide de l'initiative de Virus Outbreak Data Network (VODAN)<sup>262</sup>. Les FDP sont des dépôts de données FAIR présentant des capacités « d'arrimage » qui acceptent les machines virtuelles qui viennent « visiter » les données localement, dans le but de poser une question précise ou d'exécuter une tâche particulière<sup>263</sup>. Le résultat de la tâche, et pas les données, est renvoyé au client ou à la cliente. Étant donné que les données ne quittent pas le dépôt, la prolifération d'ensembles de données dérivés qui doivent être organisés et préservés au fil du temps est évitée. Ainsi, la visite de données permet de reproduire les configurations ou les analyses de données au sein de la plateforme de dépôt. D'autres exemples de dépôts qui permettent la visite de données comprennent le système d'accès à distance en temps réel (ADTR) de Statistique Canada<sup>264</sup>, le Barcode of Life Data System<sup>265</sup>, OpenNeuro<sup>266</sup>, l'enclave Virtual data d'ICPSR<sup>267</sup>, et les centres des Services d'accès aux données et aux informations (DIAS) pour le programme Copernicus d'observation de la Terre de l'Union européenne<sup>268</sup>. Dans les domaines qui produisent des quantités énormes de données, dont la manipulation dans le but de créer des résultats pertinents avant l'analyse exige une formation spécialisée, la visite des données présente aussi des possibilités d'accroître l'accessibilité des données et d'augmenter le nombre de résultats scientifiques<sup>269</sup>.

---

<sup>260</sup> <https://bioschemas.org> (anglais).

<sup>261</sup> <https://github.com/ESIPFed/science-on-schema.org> (anglais).

<sup>262</sup> Mons, B. (2020). The VODAN IN: support of a FAIR-based instrument for COVID-19. *Eur. J. Hum. Genet.* 28, 724–727. <https://doi.org/10.1038/s41431-020-0635-7> (anglais).

<sup>263</sup> FAIR Data Point, plateforme GitHub. Repéré en novembre 2020 à <https://github.com/FAIRDataTeam/FAIRDataPoint> (anglais).

<sup>264</sup> <https://www.statcan.gc.ca/fra/adtr/adtr>.

<sup>265</sup> <https://www.boldsystems.org/index.php> (anglais).

<sup>266</sup> <https://openneuro.org> (anglais).

<sup>267</sup> <https://www.icpsr.umich.edu/web/pages/NACJD/virtual-data-enclave.html> (anglais).

<sup>268</sup> <https://www.copernicus.eu/fr/acces-aux-donnees>.

<sup>269</sup> Kern, J., Glendenning, B., et Robnett, J. (2019). The Science Ready Data Products Revolution at the NRAO. [https://science.nrao.edu/science/astro2020/apc-white-papers/136-78e9d3c08d7e3e149ce77c6fdd6b9366\\_KernJeffreyS.pdf](https://science.nrao.edu/science/astro2020/apc-white-papers/136-78e9d3c08d7e3e149ce77c6fdd6b9366_KernJeffreyS.pdf) (anglais).

L'adoption croissante de programmes comme Jupyter et R markdown, qui produisent des documents de calcul dynamiques renfermant un code en temps réel et un texte descriptif, et qui sont associés à des outils de conteneurs comme Docker<sup>270</sup> et Binder<sup>271</sup>, permet aussi aux réutilisateurs et aux réutilisatrices de données d'interagir directement avec les ensembles de données publiés. La prise en charge de ces outils logiciels peut être assurée de plus en plus par les dépôts afin de favoriser la reproductibilité informatique (p. ex. l'intégration Git — Zenodo — Binder)<sup>272</sup>. Le soutien d'un environnement national de calcul interactif Jupyter est actuellement assuré par Calcul Canada et Cybera<sup>273</sup> et est intégré dans plus de 20 établissements d'enseignement supérieur membres<sup>274</sup>.

### Soutien de l'infrastructure

Une gamme d'infrastructures de calcul et d'outils logiciels facilite la GDR. Les initiatives de GDR décrites ci-haut exigent aussi du soutien pour leur propre développement. Cet élément de l'écosystème de la GDR présente un chevauchement considérable avec les portefeuilles du CIP et des LR de l'Alliance. Au sein des établissements universitaires, les services des systèmes et de la TI offrent des initiatives de soutien, hébergées localement, tandis que les fournisseurs de services nationaux Calcul Canada et CANARIE offrent un soutien aux initiatives hébergées à l'échelle nationale. Les possibilités de développement propres à la GDR sont relativement peu fréquentes. Un exemple récent est le Programme de gestion des données de recherche de CANARIE<sup>275</sup> qui finance le développement de nouveaux outils logiciels pour aider les chercheurs et les chercheuses à adopter les pratiques exemplaires de GDR ou à les intégrer dans leurs flux de tâches.

Les modèles de soutien national peuvent également être recueillis auprès d'organisations internationales. Par exemple, l'International Technology Office de World Data System offre aux membres des services et des infrastructures leur permettant de soutenir la GDR et d'examiner les fonds de données<sup>276</sup>. Si on veut fournir des services de données d'envergure nationale, on devra déterminer les possibilités d'aider les organismes de recherche à promouvoir leurs propres outils et plateformes de GDR tout en recherchant les moyens d'appliquer leurs réussites de façon plus large dans l'ensemble de l'écosystème.

### Formation

Le virage numérique du secteur de la recherche a créé des besoins importants en matière de formation et d'amélioration des compétences des chercheurs et des chercheuses et des professions de soutien à la recherche. La gamme de compétences numériques recherchées

---

<sup>270</sup> <https://www.docker.com> (anglais).

<sup>271</sup> <https://mybinder.org> (anglais).

<sup>272</sup> <https://blog.jupyter.org/binder-with-zenodo-af68ed6648a6> (anglais).

<sup>273</sup> <https://www.cybera.ca> (anglais).

<sup>274</sup> <https://syzygy.ca/accueil.html>.

<sup>275</sup> <https://www.canarie.ca/fr/gdr/>.

<sup>276</sup> <https://wds-ito.org/> (anglais).

englobe la science des données et l'intendance des données. Le rapport de la Commission européenne sur les principes FAIR, intitulé *Turning FAIR into Reality*, définit la science des données comme étant « [Traduction] la capacité à manipuler, à traiter et à analyser les données en vue d'en dégager des aperçus <sup>277</sup> », en faisant appel aux compétences en informatique, en développement logiciel et en statistique, tandis que l'intendance des données est définie comme étant « [Traduction] les compétences requises pour garantir la gestion, le partage et la préservation appropriés des données tout au long du cycle de vie de la recherche. » Bien que tous les chercheurs et chercheuses aient besoin de compétences de ces ensembles de compétences, l'inclusion de postes de spécialistes dans les projets de recherche est de plus en plus reconnue comme mécanisme utile de soutien de la recherche numérique. À titre d'exemple, selon des estimations récentes, une personne sur 20 d'un effectif de recherche devrait être spécialisée en soutien numérique<sup>278</sup>.

La croissance de la recherche exigeante en données dans divers domaines a mis en lumière la lacune importante dans la formation sur l'adoption, dans les établissements d'enseignement postsecondaire, de bonnes pratiques de GDR, entre autres ensembles de compétences numériques<sup>279</sup>. Le rapport récent de l'OCDE sur le renforcement de la capacité en matière de compétences en sciences numériques soutient que les universités ont un rôle essentiel à jouer en leur qualité de principaux fournisseurs de formation en recherche<sup>280</sup>. Les bibliothèques universitaires en particulier sont les mieux placées pour soutenir le développement des compétences en intendance des données, tandis que les services informatiques ont beaucoup à apporter au regard des compétences logicielles et informatiques qui sous-tendent la science des données. Bien que l'écart de compétences en GDR diminue grâce aux efforts déployés par les établissements d'enseignement supérieur pour investir dans une formation supplémentaire destinée aux chercheurs et aux chercheuses, la capacité à effectuer ce travail relève principalement des grandes universités<sup>281</sup>. Les mandats et incitatifs du gouvernement forment un ensemble de mécanismes utilisés pour encourager les établissements à entreprendre ce développement.

Les associations de recherche et les sociétés savantes ont un rôle important à jouer dans la création de communautés et dans l'amplification de la prestation de possibilités de formation. À titre d'exemple, les possibilités de combler les écarts d'accès dans l'ensemble des établissements ont été entreprises par les partenaires régionaux de Calcul Canada (p. ex. l'école d'été/d'hiver de

---

<sup>277</sup> Commission européenne. (2018). *Turning FAIR into reality: Final report and action plan from the European Commission Expert Group on FAIR Data*. [https://ec.europa.eu/info/sites/info/files/turning\\_fair\\_into\\_reality\\_1.pdf](https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf) (anglais).

<sup>278</sup> Mons, B. (2020). Invest 5% of research funds in ensuring data are reusable. *Nature*, 578 :491. <https://www.nature.com/articles/d41586-020-00505-7> (anglais).

<sup>279</sup> Conseil du leadership. (2014). *Canadian DI Environmental Scan: A Supplement to the Background Précis Document Provided to DI Summit 2012* (comme cité par Baker et al. (2019), accessible à <https://zenodo.org/record/3574685> (document en anglais)).

<sup>280</sup> OCDE. (2020). *Building digital workforce capacity and skills for data-intensive science*. OECD Science, Technology and Industry Policy Papers. <https://doi.org/10.1787/e08aa3bb-en> (anglais).

<sup>281</sup> Cooper, A. et al. (2020). Sondage sur la capacité des services institutionnels de gestion des données de recherche. <https://doi.org/10.14288/1.0388722>.

WestGrid<sup>282</sup> et le catalogue de formation d'ACENET<sup>283</sup>) et par le programme de formation du réseau Portage de l'ABRC<sup>284</sup>. Les approches nationales ou régionales de la formation peuvent adopter une gamme de modèles. À titre d'exemple, Portage offre aux chercheurs et aux chercheuses et aux professionnels et professionnelles des données des possibilités de formation en personne ou en ligne pour étendre la portée de leurs programmes. Le réseau produit aussi des formations que les professionnels et professionnelles des données peuvent dispenser aux communautés de recherche locales dans le cadre du modèle de « formation des formateurs et formatrices ». Quelques organismes de recherche exploitant des services de données offrent aussi une formation spécialisée sur la GDR qui cible les chercheurs et les chercheuses, les gestionnaires de données et les administrateurs et les administratrices. Mentionnons à titre d'exemple, le cours appelé Fondamentaux des principes de PCAP® du Centre de gouvernance de l'information des Premières Nations<sup>285</sup>, et la série d'éducation et de formation de Population Data BC<sup>286</sup>. À l'échelle internationale, divers modèles de formation sur la GDR pourraient être adaptés en fonction de caractéristiques démographiques ciblées. Par exemple, le Registre de formation pour les sciences de la vie d'ELIXIR<sup>287</sup>, la School of Research Data Science de CODATA-RDA qui vise les chercheurs et les chercheuses en début de carrière<sup>288</sup>, ou la Research Data Management Librarian Academy<sup>289</sup>. Les initiatives du milieu qui surgissent en raison de lacunes dans la programmation ordinaire devraient aussi être reconnues comme éléments importants du paysage de la formation. DRC et le réseau Portage de l'ABRC collaborent au développement d'une approche nationale de la formation sur la GDR. Une action concertée dans l'écosystème s'impose pour constituer et maintenir l'effectif du personnel hautement qualifié qui est nécessaire pour favoriser la recherche numérique et les objectifs de la science ouverte.

## Gouvernance

Bon nombre d'organisations ont assumé un rôle de soutien des communautés de pratique de la GDR en élaborant des orientations et des politiques ou en développant des possibilités de financement. La coordination entre ces organisations est essentielle pour favoriser la diversité des approches réussies de la GDR. Dans le paysage actuel, les répercussions des déséquilibres existants au niveau local et régional doivent être prises en considération dans le contexte national. Il convient aussi d'envisager l'harmonisation avec les initiatives internationales pour permettre la circulation transfrontalière des données.

---

<sup>282</sup> <https://www.westgrid.ca/support/training> (anglais).

<sup>283</sup> <https://www.ace-net.ca/training/> (anglais).

<sup>284</sup> <https://portagenetwork.ca/fr/outils-et-ressources/ressources-de-formation-de-portage/>.

<sup>285</sup> <https://fnigc.ca/fr/les-principes-de-pcap-des-premieres-nations/suivre-le-cours/>.

<sup>286</sup> <https://www.popdata.bc.ca/etu> (anglais).

<sup>287</sup> <https://tess.elixir-europe.org/> (anglais).

<sup>288</sup> <https://codata.org/initiatives/strategic-programme/research-data-science-summer-schools> (anglais).

<sup>289</sup> <https://rdmla.github.io/> (anglais).

Plusieurs organisations ont fourni des recommandations, des conseils ou d'autres renseignements en réponse aux enjeux importants qui émergent dans la gestion des données de recherche. À l'échelle internationale, quelques organisations munies de mandats de recherche générale et de recherche axée sur un domaine particulier ont produit d'importantes ressources adoptées par les milieux de recherche (voir l'annexe C). Au Canada, ce mouvement est largement dirigé par des associations nationales représentant les entités de soutien de la recherche, en particulier le réseau d'experts Portage de l'ABRC qui a produit des rapports et des résultats plus appliqués au profit des établissements de recherche et du milieu global de la GDR<sup>290</sup>. Au sein des milieux de recherche, un nombre relativement faible d'associations ou de sociétés canadiennes propres à un domaine de recherche ont produit des ressources pour leurs membres afin de guider la gestion améliorée des données de recherche — en particulier dans les domaines de recherche moins exigeants sur le plan du calcul. Les pratiques et les facteurs de GDR varient énormément d'un domaine à l'autre et il y a donc lieu d'encourager ces groupes à s'impliquer davantage. Les efforts déployés par Données de recherche Canada pour réunir les parties prenantes de l'ensemble du paysage dans le cadre des Sommets sur l'encadrement des services de données nationaux (ESDN) sont un exemple des initiatives visant à stimuler l'implication accrue des associations de recherche.

Ce bloc grossissant du renseignement sur la valeur de la GDR confronte les chercheurs et les chercheuses et leurs organisations à un ensemble croissant de politiques connexes; à titre d'exemple, l'ébauche influente de la politique des trois organismes sur la gestion des données de recherche, ou les politiques adoptées par les maisons d'édition savante en matière de partage et d'accessibilité des données. En réponse, les établissements de recherche dans l'ensemble du Canada élaborent petit à petit des politiques traitant des données de recherche produites par la recherche financée<sup>291</sup>. Au sein des établissements de recherche, les politiques existantes relatives à la recherche, à l'éthique et à la propriété intellectuelle ont déjà des conséquences pour les modalités de gestion des données de recherche. Des initiatives visant à dresser l'inventaire des politiques institutionnelles existantes relatives à GDR sont en cours au Canada, où elles sont dirigées par le Groupe d'experts sur la recherche et l'intelligence de Portage, et à l'échelle internationale par l'entremise de la RDA<sup>292</sup>.

Les organismes de recherche, les infrastructures de recherche et les services connexes ont besoin de politiques et d'exigences cohérentes pour garantir que les chercheurs et les chercheuses adoptent des pratiques et des cadres communs. Les variations des exigences institutionnelles et régionales, y compris la nécessité de respecter les droits autochtones, ajoutent aux défis qui se dressent à l'échelle nationale. Par exemple, les différences entre provinces dans la gestion des renseignements personnels peuvent miner le partage et la collaboration transfrontaliers. À l'opposé des autres provinces, la Colombie-Britannique et la Nouvelle-Écosse limitent la capacité des organisations publiques et de leurs fournisseurs de services à transférer

---

<sup>290</sup> <https://zenodo.org/communities/portage-network>.

<sup>291</sup> Cooper, A. et al. (2020). Sondage sur la capacité des services institutionnels de gestion des données de recherche. <https://doi.org/10.14288/1.0388722>.

<sup>292</sup> <https://www.rd-alliance.org/group/research-funders-and-stakeholders-open-research-and-data-management-policies-and-practices-ig> (anglais).

des renseignements personnels hors du Canada<sup>293</sup>, ce qui a des répercussions sur les secteurs de la santé et des sciences sociales en particulier.

La gestion des données de recherche pendant et après un projet de recherche donné utilise des ressources financières considérables pour soutenir l'infrastructure humaine et technique. Cependant, le financement direct accessible aux chercheurs et aux chercheuses menant des projets financés est limité, car la GDR est souvent réputée ne pas faire partie du processus de recherche standard, ou du budget de recherche normal<sup>294</sup>. Plus récemment, des possibilités de financement spécifique de la GDR sont apparues pour encourager les chercheurs et les chercheuses à améliorer la gestion de leurs données de recherche et pour résoudre les problèmes connexes dans le cadre d'activités de renforcement de la capacité. Mentionnons à titre d'exemple, les subventions Connexion du CRSH pour le renforcement de la capacité de GDR<sup>295</sup>, ou l'Initiative sur les données de recherche ouvertes du Centre de recherches pour le développement international (CRDI)<sup>296</sup>.

Pour quelques projets de recherche financés par subvention, les frais de curation des données engagés pendant la période de recherche peuvent être compris dans le budget du projet. Cependant, la gestion continue des données au-delà de la vie d'un projet de recherche devrait être incluse dans les programmes des coûts indirects (p. ex. le Fonds de soutien à la recherche des trois organismes<sup>297</sup>). Les établissements de recherche admissibles présentent une demande aux programmes des coûts indirects afin de compenser les frais engagés par la gestion de ces bourses, et les montants qu'ils reçoivent sont déterminés par un exercice de calcul annuel. Cependant, bon nombre de ces bourses sont déjà réservées, car d'autres secteurs d'un établissement de recherche en dépendent, et les calculs des coûts peuvent être lents à changer par rapport au soutien croissant dont la GDR a besoin<sup>298</sup>.

Alors que le contexte des politiques auquel les chercheurs et les chercheuses sont confrontés continue à prendre de l'ampleur, ces politiques ne précisent ni qui est responsable de la gestion continue des données (p. ex. le chercheur ou la chercheuse, l'établissement, l'organisme de financement) ni qui doit en assumer les coûts. Avant de pouvoir régler ces questions, il convient

---

<sup>293</sup> Commissariat à la protection de la vie privée du Canada. (2017). Aperçu des lois sur la protection des renseignements personnels au Canada. Repéré en novembre 2020 à [https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/lois-sur-la-protection-des-renseignements-personnels-au-canada/02\\_05\\_d\\_15/](https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/lois-sur-la-protection-des-renseignements-personnels-au-canada/02_05_d_15/).

<sup>294</sup> Knowledge Exchange Research Data Expert Group and Science Europe Working Group on Research Data. (2016). Funding research data management and related infrastructures. [https://www.scienceeurope.org/media/uuqf0i03/se-ke\\_briefing\\_paper\\_funding\\_rdm.pdf](https://www.scienceeurope.org/media/uuqf0i03/se-ke_briefing_paper_funding_rdm.pdf) (anglais).

<sup>295</sup> [https://www.sshrc-crsh.gc.ca/funding-financement/programmes-programmes/data\\_management-gestion\\_des\\_donnees-fra.aspx](https://www.sshrc-crsh.gc.ca/funding-financement/programmes-programmes/data_management-gestion_des_donnees-fra.aspx).

<sup>296</sup> <https://www.idrc.ca/fr/financement/initiative-sur-les-donnees-de-recherche-ouvertes>.

<sup>297</sup> <https://www.rsf-fsr.gc.ca/administer-administrer/depenses-fra.aspx>.

<sup>298</sup> Erway, R. et Rinehart, A. (2016). If You Build It, Will They Fund? Making Research Data Management Sustainable. OCLC Research. <https://www.oclc.org/content/dam/research/publications/2016/oclcresearch-making-research-data-management-sustainable-2016.pdf> (anglais).

de clarifier les considérations et les conséquences plus générales de la GDR. À titre d'exemple, quels résultats d'un projet de recherche doivent être stockés et préservés au fil du temps? Quels sont les critères qui différencient le spectre du stockage (phase active, phase de dépôt, phase d'archivage)? Quelles sont les conditions de la préservation des résultats de la recherche (p. ex. durée de conservation, responsabilités, et critères de sélection, de maintenance et de désélection)?

Les coûts de la GDR varient selon les particularités du projet (p. ex. stockage requis, sensibilité des données, durée de la préservation)<sup>299</sup> et les besoins en matière de curation tout au long du cycle de vie de la recherche<sup>300</sup>. Le Groupe d'experts de haut niveau sur le Nuage européen pour la science ouverte estime que selon une règle empirique générale, 5 % des dépenses totales d'un projet de recherche devraient être affectés à la GDR<sup>301</sup>. Cependant, le financement de l'infrastructure de calcul, de stockage et de logiciels de laquelle la GDR dépend — lorsque ce financement existe — est principalement destiné au développement (p. ex. le Fonds d'innovation de la FCI<sup>302</sup>, les Concours d'allocation des ressources de Calcul Canada<sup>303</sup>, et les appels de développement de logiciels de CANARIE<sup>304</sup>) plutôt qu'aux opérations et à la maintenance, ce qui donne lieu à de courtes périodes de financement qui sont incompatibles avec la GDR. Même le Fonds des initiatives scientifiques majeures de la FCI, qui contribue aux besoins continus en matière de fonctionnement et de maintenance des établissements de recherche nationaux, est attribué pour trois ou cinq ans<sup>305</sup>. De nouveaux modèles d'une infrastructure de financement durable de la GDR sont nécessaires pour garantir la longévité.

L'incidence du financement durable des dépôts de données sur les pratiques de gestion des données des chercheurs et des chercheuses ne peut être minimisée. Cette incidence est évidente dans les milieux de recherche munis de plateformes vigoureuses de GDR, financées à l'échelle nationale, qui sont devenues des ressources intégrales de ces milieux (p. ex. le milieu canadien de l'astronomie et le Centre canadien de données astronomiques, financés par l'entremise du Conseil national de recherches Canada et l'Agence spatiale canadienne<sup>306</sup>).

Dans son examen des sources de financement destinées aux dépôts de données institutionnels exploités par les bibliothèques universitaires, OCLC (Online Computer Library Center) a relevé

---

<sup>299</sup> OpenAIRE (s. d.). What will it cost to manage and share my data? Repéré en novembre 2020 à <https://www.openaire.eu/rdm-researcher-costs-infographic/view-document> (anglais)

<sup>300</sup> Westerhof, A. et al. (s. d.). Research Data Management: more than just storage. [https://www.lcrdm.nl/files/lcrdm/2020-04/RDM%20and%20Costs\\_v20160218\\_EN.pdf](https://www.lcrdm.nl/files/lcrdm/2020-04/RDM%20and%20Costs_v20160218_EN.pdf) (anglais, traduction du document néerlandais).

<sup>301</sup> LCRDM. (s. d.). xFinancials. Repéré en novembre 2020 à <https://www.lcrdm.nl/en/xfinancials> (anglais).

<sup>302</sup> <https://www.innovation.ca/fr/le-financement/fonds-dinnovation>.

<sup>303</sup> <https://www.computecanada.ca/page-daccueil-du-portail-de-recherche/acces-aux-ressources/concours-dallocation-des-ressources/?lang=fr>.

<sup>304</sup> <https://www.canarie.ca/fr/logiciels/>.

<sup>305</sup> <https://www.innovation.ca/fr/le-financement/fonds-initiatives-scientifiques-majeures>.

<sup>306</sup> <http://www.cadc-ccda.hia-ihc.nrc-cnrc.gc.ca/fr/>.

un mélange de modèles de coûts; cependant, presque la moitié des entités contactées dépendait exclusivement du financement de base<sup>307</sup>. Un mélange de modèles de financement (p. ex. financement structural, services à valeur ajoutée, paiement à l'utilisation) est nécessaire pour développer une résilience aux changements survenant dans le contexte de gouvernance. Au Canada, le réseau Portage de l'ABRC travaille sur le développement d'une infrastructure de dépôt partagée par le truchement de Dataverse North et du DFDR, lesquels sont des exemples de mécanismes de soutien national en mesure de surmonter les obstacles liés aux coûts et de favoriser un accès plus équitable aux établissements dans l'ensemble du Canada. Le financement par le biais de partenariats avec les organisations du secteur privé peut offrir d'importantes possibilités, mais doit tenir compte des moyens d'assurer l'accès aux données, en particulier si un tel accès ne crée aucun désavantage commercial réel<sup>308</sup>.

Le financement de la préservation et de l'archivage à long terme est particulièrement difficile et mérite une étude plus approfondie afin d'obtenir des mécanismes de financement durable. Lorsque les ensembles de données sont déposés dans un dépôt, il est impossible de prédire avec certitude les ensembles qui présentent une valeur de longue durée ou les organisations qui en profiteront le plus et qui devraient donc assumer une plus grande part des coûts. Entre-temps, la dépendance aux critères axés sur la demande du nombre de citations ou du nombre de téléchargements utilisés pour déterminer la valeur et l'impact des ensembles de données contrevient aux principes de la science ouverte. L'expérience vécue par le milieu d'archivage de ces questions difficiles devrait être prise en considération, vu le soutien à long terme dont la GDR a besoin.

---

<sup>307</sup> Erway, R. et Rinehart, A. (2016). If You Build It, Will They Fund? Making Research Data Management Sustainable. OCLC Research.

<https://www.oclc.org/content/dam/research/publications/2016/oclcresearch-making-research-data-management-sustainable-2016.pdf> (anglais).

<sup>308</sup> Knowledge Exchange Research Data Expert Group and Science Europe Working Group on Research Data. (2016). Funding research data management and related infrastructures.

[https://www.scienceeurope.org/media/uuqf0i03/se-ke\\_briefing\\_paper\\_funding\\_rdm.pdf](https://www.scienceeurope.org/media/uuqf0i03/se-ke_briefing_paper_funding_rdm.pdf) (anglais).



## 6 Principaux défis et possibilités

La présente section, qui s'inspire de l'évaluation ci-haut de l'état actuel de la GDR au Canada, présente une série de défis auxquels l'Alliance est confrontée pendant qu'elle assume la direction nationale du soutien et de l'avancement de la GDR au Canada. Les enjeux soulevés ne sont pas exhaustifs et sont délibérément présentés à un haut niveau pour favoriser la sensibilisation et stimuler la discussion, et font ressortir les possibilités de les régler et les efforts actuellement déployés à cette fin. Bon nombre des recommandations et des suggestions qui suivent sont issues des conversations facilitées avec la grande communauté et concrétisées par la Déclaration de Kanata et les rapports sur les Sommets sur l'encadrement des services de données nationaux (ESDN).

Cette connaissance des défis et des possibilités actuels sera amplifiée par le processus prochain d'évaluation par l'Alliance des besoins des chercheurs et des chercheuses. De plus, l'intégration ultérieure des rapports sur l'état actuel du CIP et des LR favorisera une évaluation plus stratégique de la façon dont l'Alliance relèvera et priorisera les défis, à titre de chef de file de l'IRN au Canada et en collaboration avec les partenaires dans l'ensemble du Canada et à l'échelle internationale.

### Coordination

L'importance d'une meilleure coordination et communication se dégage comme principale priorité des conversations avec le milieu de GDR, et présente dès le début à la fois un défi et une possibilité pour l'Alliance. Bien que le présent rapport présente un survol de haut niveau des divers acteurs et actrices, des infrastructures et des services soutenant la GDR au Canada, il ne brosse qu'un portrait partiel qui doit être peaufiné. Bon nombre des infrastructures, des outils et des plateformes existants fonctionnent de manière relativement isolée les uns des autres. Voilà bien un défi. Une meilleure intégration des infrastructures et des services nouveaux et existants exige l'adoption de normes, de schémas et de certifications communs pour assurer une interopérabilité fiable. Une telle intégration est particulièrement difficile lorsqu'elle porte sur divers domaines, qui se caractérisent par des différences énormes sur le plan des pratiques et des outils disponibles pour gérer les données tout au long de leur cycle de vie. Les initiatives existantes mentionnées tout au long du présent rapport peuvent servir de modèles pour favoriser l'amélioration de l'interopérabilité entre les éléments de l'écosystème.

Parallèlement, bon nombre d'actrices et d'acteurs au sein de l'écosystème canadien de la GDR existent de façon relativement isolée. Des initiatives continues de consultation et de relations externes sont nécessaires pour cerner leurs besoins. Une mobilisation continue du milieu de la GDR et des discussions avec celui-ci sont nécessaires pour renforcer la crédibilité et la confiance, favoriser une coordination et une intégration plus importante, et communiquer l'évolution de l'écosystème de la GDR aux chercheurs et aux chercheuses, aux organisations alliées et au grand public. La création de canaux de communication bilatérale entre les organisations et la promotion de possibilités de collaboration mutuellement avantageuses sont susceptibles d'améliorer l'état actuel. Dans le secteur de l'enseignement supérieur, le travail du réseau Portage sert d'exemple des avantages et de l'efficacité d'un modèle de communauté de pratique. Il y a lieu de réunir les diverses communautés décrites dans le présent rapport pour renforcer la confiance, la compréhension et le consensus. À titre d'exemple, on pourrait réunir les

représentants et les représentantes des organisations clés pour élaborer et tenir à jour une feuille de route de haut niveau de la GDR afin d'aider à prioriser les investissements et le développement. De plus, la constitution d'un comité consultatif de l'Alliance sur la GDR, composé de représentants et de représentantes clés, pourrait aider l'Alliance à s'acquitter de ses responsabilités en matière d'orientation et de supervision.

Le Canada doit poursuivre sa mobilisation à l'échelle internationale, et travailler de façon coopérative pour faire progresser et adopter des pratiques et des normes communes afin de favoriser le développement de l'écosystème international de la GDR. Cette mobilisation doit être dotée de ressources suffisantes pour que le Canada puisse se faire entendre sur la scène internationale. À mesure que les écosystèmes canadien et mondial de la GDR continuent à évoluer, des possibilités peuvent se présenter pour adapter les outils mis au point ailleurs (comme il a été le cas d'Assistant PGD) ou développer des paramètres systématiques d'évaluation et d'analyse comparative, qu'on se doit de soutenir afin d'assurer une intégration plus vaste.

## Représentation et inclusion

Le milieu de recherche générale connaît peu les infrastructures et les services existants de GDR qui sont offerts au niveau local, régional ou national, quoiqu'on doive reconnaître que des variations existent entre les disciplines et les organisations. Afin de davantage sensibiliser le milieu, il est essentiel de prioriser la création d'un index détaillé des services et des infrastructures disponibles au niveau local, national et international. Des mécanismes évidents pour représenter et désigner la gamme des éléments de l'écosystème sont également nécessaires. Cela pourrait inclure le soutien d'un usage élargi d'identifiants permanents ou du développement de registres répertoriant les éléments de l'écosystème, au moyen d'un effort mondial concerté. Naturellement, les registres et les index ne suffisent pas à sensibiliser le milieu, et il y a donc lieu d'envisager des mécanismes de promotion active. Les organisations nationales ne sont pas toujours bien placées pour entreprendre ce travail et il convient donc de faire intervenir divers collaborateurs et collaboratrices (p. ex. sociétés savantes, organismes de recherche et établissements d'enseignement supérieur) qui ont un accès plus direct aux chercheurs et aux chercheuses.

Outre la simple sensibilisation, la promotion de la mise en application des pratiques exemplaires, des infrastructures et des services existants représente un enjeu essentiel. L'harmonisation avec les changements qui s'opèrent dans le contexte des politiques des établissements de recherche, des organismes de financement et des maisons d'édition est une stratégie qui permettrait de favoriser l'adoption. L'ébauche de la politique des trois organismes sur la gestion des données de recherche en particulier représente une possibilité importante au Canada de favoriser la confluence des modèles de gouvernance des données institutionnelles avec les mesures de soutien offertes, par le truchement de son exigence en matière d'adoption de stratégies de GDR institutionnelles. Le travail accompli par Portage pour développer des plateformes qui aident de façon générale les établissements et les chercheurs et les chercheuses à mettre en œuvre des éléments stratégiques, tels les plans de gestion des données et le dépôt de données est d'une importance fondamentale, mais devrait être élargi pour également faire ressortir la confluence avec les infrastructures thématiques par domaine et les infrastructures internationales.

Toute discussion de la représentation ou de la promotion des éléments de l'écosystème doit s'inspirer d'une réflexion approfondie sur les pratiques d'inclusion. Des mécanismes sont nécessaires pour que les fournisseurs de services et les utilisateurs et les utilisatrices de tous les

secteurs et de toutes les disciplines soient bien représentés et soutenus, en accordant une attention spéciale à la promotion de la participation des groupes sous-représentés, lesquels comprennent les femmes, les communautés racialisées, et les chercheurs et les chercheuses œuvrant dans des domaines de recherche de moindre importance et présentant une orientation technique moindre, et en envisageant l'adoption de mesures d'adaptation pour ce faire. Les formes de recherche non occidentale doivent également être incluses. En vertu de la responsabilité du Canada de favoriser la réconciliation, on doit s'assurer que les communautés des Premières Nations, des Métis et des Inuits participent pleinement au processus.

Dans le cadre d'un effort de plus grande portée, l'Alliance a un rôle à jouer dans le soutien de mécanismes pour reconnaître et récompenser les chercheurs et les chercheuses qui adoptent de bonnes pratiques de GDR, car cette adoption peut favoriser des changements culturels plus vastes touchant l'attitude envers la GDR. La formation et le soutien des chercheurs et des chercheuses en début de carrière sont particulièrement importants pour faciliter ce changement.

## Durabilité

La vision d'un écosystème national de la GDR, qui est coordonné et évolutif, où la circulation des données entre les systèmes des éléments est soumise à un nombre minimal de restrictions depuis leur création jusqu'à leur préservation à long terme, et qui fournit des services rapides et des ressources adaptées aux besoins des chercheuses et des chercheurs, exige une mosaïque de partenariats et de modèles de financement pour renforcer la résilience. La GDR durable a besoin d'infrastructures de stockage appropriées à toutes les phases de recherche active, de dépôt et d'archivage, qui sont non seulement reliées, mais aussi coordonnées et qui prévoient la demande future. La vision impose aussi la création et la curation des données tout au long de leur cycle de vie dans le respect des principes *FAIR* et *CARE*, ainsi que la prestation d'un soutien suffisant aux chercheurs et aux chercheuses en matière d'outils, de formation et de personnel, tant à l'intérieur qu'à l'extérieur de leur établissement, afin de favoriser la préservation et l'utilisabilité à long terme de leurs données pendant que celles-ci circulent dans l'ensemble de l'écosystème tout au long de leur cycle de vie.

L'ampleur et le rythme de la génération des données de recherche, associés à la diversité des besoins et des intérêts, présentent de nombreux défis au soutien durable de la GDR à l'échelle nationale. Vu la stabilité relative de l'enveloppe de financement de l'Alliance, on doit prévoir une allocation proportionnelle des fonds aux infrastructures et services de GDR afin de favoriser la réalisation des aspirations de plus grande portée de l'IRN. Bon nombre de milieux de recherche sont actuellement aux prises avec un manque de financement durable nécessaire pour répondre à leurs besoins à long terme en matière de GDR. Des cadres de financement durable des infrastructures de base de la GDR, qui sont largement accessibles par les milieux de recherche qui leur font confiance (intergiciels et systèmes de dépôts et d'archivage) sont particulièrement nécessaires pour la prise de décisions.

Avant de proposer de tels modèles de financement, il convient de clarifier au sein de l'Alliance ce que les enveloppes de financement du CIP, de la GD et des LR couvrent, et parallèlement, de clarifier les distinctions fondamentales entre les éléments de l'écosystème de l'IRN. À titre d'exemple, le CIP et la GDR présentent quelques besoins en matière d'infrastructure qui se chevauchent, mais qui visent des objectifs distincts. Bien que les dépôts et les archives prennent en charge les données au-delà de la durée de vie de projets, les délais et les mécanismes

connexes ne sont pas bien établis. Dans ce spectre du stockage, le développement d'une compréhension commune entre chercheurs et chercheuses, fournisseurs de services, décideurs politiques et autres parties prenantes est nécessaire pour assurer la cohérence, et aura des répercussions importantes sur les décisions à l'égard des politiques et du financement. Quelques activités proposées par la Déclaration de Kanata comprennent l'élaboration de feuilles de route visant à intégrer les éléments du CIP et de la GD (p. ex. langage commun pour les flux de tâches, protocoles de collecte et de consultation) et de feuilles de route visant le financement des mécanismes et des fournisseurs des systèmes de dépôts et d'archivage. Toute nouvelle organisation court le risque de « repartir à zéro » au lieu de miser sur les modèles et les programmes existants, aux dépens de la continuité et de la durabilité. Comme l'expose le présent rapport, des progrès importants ont été réalisés dans l'écosystème de la GDR, même durant le court laps de temps écoulé depuis la publication des derniers documents du CLIRN. L'Alliance se doit d'exploiter la capacité existante dans l'écosystème pour optimiser les avantages pour les chercheuses et les chercheurs canadiens. Une plus grande harmonisation et une meilleure intégration des organisations et des services s'imposent, non seulement des entités canadiennes soutenues par l'Alliance, mais aussi de leurs homologues internationaux. Déterminer les rôles de ces partenaires dans l'écosystème de l'IRN à tous les niveaux — local, régional, national, international — est une étape importante qui stimulera l'innovation coopérative, améliorera le soutien de la GDR et réduira le chevauchement et le dédoublement des efforts. Dans l'idéal, le système prévoirait la prestation de services à toute une gamme de niveaux, soutenue et structurée par un cadre national qui est rattaché à des normes internationales et à des organisations homologues, et qui est influencé par celles-ci.

## 7 Prochaines étapes

La capacité et la volonté de relever les défis et les possibilités actuels réclament une compréhension rigoureuse du paysage, ainsi que des cibles précises pour les objectifs à long terme. Le présent rapport résume l'état actuel du paysage de la GDR au Canada afin de favoriser chez les membres de l'Alliance une vision commune de l'étendue et de la complexité de la mobilisation dans ce domaine, et sert de base permettant à l'Alliance de tracer une voie à suivre vers le soutien national de la GDR au Canada. Le processus d'élaboration du plan stratégique de l'Alliance comporte plusieurs jalons qui aideront à définir et à clarifier le rôle de l'Alliance dans l'écosystème.

### *Évaluation des besoins des chercheurs et des chercheuses (octobre — mai 2021)*

L'Alliance consultera le milieu de recherche pour évaluer les services actuels et cerner les besoins en matière de calcul, de données et de logiciels selon les priorités. Le Conseil des chercheurs, nouvellement constitué, supervisera le processus d'évaluation des besoins des chercheurs et des chercheuses, qui mobilisera une gamme diversifiée de communautés de ces spécialistes et d'associations disciplinaires, et fournira à l'Alliance des conseils indépendants sur les questions liées à la prestation de services et de programmes pour le milieu de recherche.

À la suite du processus d'évaluation des besoins, les résultats seront intégrés avec les conclusions des rapports sur l'état actuel du CIP, des LR et de la GD dans un énoncé de principe cohésif et dirigé sur l'IRN, qui formulera des recommandations visant à soutenir l'effort de planification stratégique de l'Alliance.

### *Modèle de prestation de services (mai 2021)*

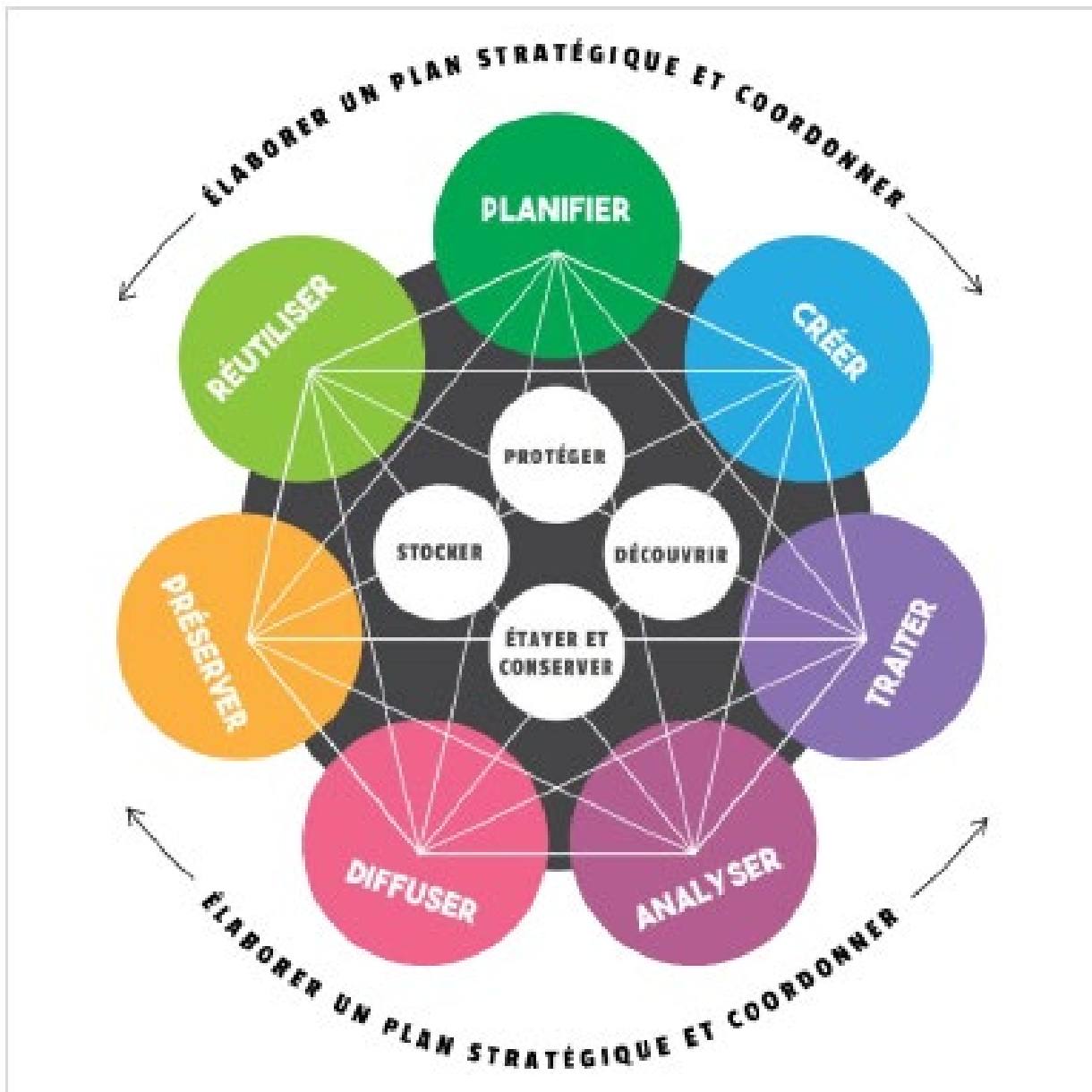
En collaboration avec les partenaires de l'IRN, l'Alliance mettra au point un nouveau modèle de prestation de services qui définit les services nationaux, régionaux et locaux pour la GD, le CIP et les LR, notamment les niveaux de service prévus, les nouveaux modèles de financement, et les rôles et les responsabilités.

### *Plan stratégique pour 2021-2024 (septembre 2021)*

L'Alliance présentera une stratégie et une vision nationales pour le CIP, les LR et la GD, qui intégreront les conclusions issues de ses activités d'évaluation et de relations externes. Le plan stratégique exposera une feuille de route pour transformer l'écosystème en remplaçant l'état actuel axé sur des piliers, où le CIP, les LR et la GD sont traités comme des entités distinctes, par un état futur plus intégré, où la recherche est soutenue par une IRN vigoureuse tout au long de son cycle de vie.

# Annexe A – Le cycle de vie de la recherche et les fonctions de la GDR

L'extrait qui suit est reproduit avec la permission de Baker, D., Bourne-Tyson, D., Gerlitz, L., Haigh, S., Khair, S., Leggott, M., Moon, J., Tourangeau, R. et Whitehead, M. (13 décembre 2019). *La gestion des données de recherche au Canada : état des lieux*. Zenodo. <https://zenodo.org/record/3574712#.YA7b6U-SnmE>.



**Figure A1.** Activités relatives aux données dans le processus de recherche (créée par le Leadership Council for Digital Research Infrastructure. In *Advanced Research Computing (ARC)*)

Le cycle de vie de la recherche sert de feuille de route utile aux chercheurs et chercheuses pour suivre les éléments relatifs à leurs données qu'ils ou qu'elles doivent examiner à chaque étape. Les mesures à prendre, à chaque étape, en matière de gestion des données (GD) couvrent cinq grands axes : 1) stratégies, normes et protocoles, 2) politiques et procédures, 3) leadership, conseils, soutien et formation, et 5) outils et plateformes.

### Planifier

À l'étape de la planification du cycle de vie, le chercheur ou la chercheuse s'organise lui-même ou elle-même et organise ses données en vue d'en permettre plus tard la découverte, la réutilisation et l'archivage. En principe, il ou elle devrait se familiariser avec les lignes directrices en matière de GD et les mandats de leur institution subventionnaire, de leur établissement postsecondaire ou autre et recenser les normes et les protocoles applicables qui correspondent aux pratiques exemplaires en GD de son organisation ou de son domaine. À cette étape, il ou elle devrait établir un plan de gestion des données (PGD) et choisir le dépôt qui convient pour le stockage et l'archivage des données.

C'est à cette étape que le chercheur ou la chercheuse devrait demander l'aide d'experts ou d'expertes en GDR qui peuvent l'orienter dans sa prise de décision. Ces experts ou expertes peuvent apporter leur soutien sous forme de formation en GD, clarifier les processus de l'université (ou autre organisation) qui peuvent se recouper avec les mandats national ou de son domaine et offrir des conseils sur la planification concernant le dépôt, le partage et la réutilisation des données.

### Créer

L'étape de la création consiste à identifier, acquérir et créer les données et les métadonnées de recherche. C'est, on s'en doute bien, à cette étape que les chercheurs et les chercheuses doivent se renseigner au sujet de toute politique propre à l'établissement ou au domaine définissant la procédure de collecte des données. La gestion des données fait son entrée à cette étape en ce qui concerne les pratiques exemplaires assurant la qualité, l'intégrité, les versions et la provenance des données; les dites pratiques au niveau local recoupent souvent celles en vigueur au niveau international. Les pratiques exemplaires en matière de métadonnées doivent aussi être appliquées à cette étape pour garantir l'interopérabilité et la découverte, par l'utilisation de schémas et de protocoles. Les établissements et le personnel de la GD peuvent offrir de la formation et tenir des activités au regard de la qualité et de l'intégrité en se concentrant sur des approches propres aux domaines. Les données peuvent être partagées et transformées pendant leur création au moyen de plateformes de logiciels de recherche tels les environnements de recherche virtuels, les passerelles scientifiques et les plateformes d'e-science.

### Traiter

À l'étape du traitement, les données sont préparées pour analyse (vérification, validation du nettoyage, description, etc.); le respect des ontologies des domaines est vérifié à cette étape. C'est aussi à cette étape que le code est géré pour qu'il soit repérable et réutilisé, souvent à l'aide de plateformes telles que *GitHub* et *Jupyter*. La normalisation des flux de travail est un autre

élément à considérer, de même que la documentation. Les experts et les expertes en GD peuvent aider à trouver les outils utiles pour ces tâches (p. ex. outils pouvant servir pour la réutilisation des flux de travail, tels que Taverna, Galaxy et Kepler), et contribuer à établir les politiques et les procédures relatives aux querelles de données (« *data wrangling* ») propres au domaine.

### Analyser

L'analyse des données suit naturellement la préparation et le traitement. À cette étape, il est encore important de gérer le code et le flux de travail et documenter les processus, de même que de créer et de promouvoir des politiques propres au domaine visant à faciliter l'analyse, les produits, le rapprochement de données, la reproductibilité et la confidentialité. Encore ici, les experts et expertes en GD peuvent conseiller les chercheurs et les chercheuses, donner de la formation sur l'utilisation de la modélisation et des logiciels de données. Des ressources informatiques spécialisées, telles que le calcul de haute performance et l'infonuagique, offertes par Calcul Canada, Amazon et Microsoft Azure, pourraient aussi être requises pour les chercheurs et les chercheuses.

### Diffuser

L'étape de la diffusion est celle du partage. Avant que les données soient transférées dans un dépôt, il faut s'occuper des accords de dépôt, des licences, des conditions de réutilisation (ou accès) et des méthodes de découverte et de préservation. Le cadre stratégique national est pris en compte dans les politiques universitaires (p. ex. éthique et protection des renseignements personnels) et généralement recoupe les politiques des éditeurs et les grandes stratégies des universités. Le code logiciel et les codes, et autres paramètres du système, doivent être mis à disposition pour permettre la reproduction.

Il y a lieu de suivre les pratiques exemplaires en matière de durabilité, d'interopérabilité, de découvrabilité et de réutilisation; parmi ces pratiques : l'utilisation d'identifiants permanents de données, le bon formatage des fichiers et la conformité aux pratiques internationales. Les dépôts et les autres plateformes de partage des données, telles que le Dépôt fédéré de données de recherche (DFDR), Scholars Portal Dataverse et le Centre canadien de données astronomiques, sont des ressources utiles pour la création des métadonnées et l'assurance de la qualité. Les experts et les expertes en GD peuvent offrir du soutien aux chercheurs et aux chercheuses, d'une part, en créant et en encourageant les pratiques exemplaires pour le partage et la reproductibilité et, d'autre part, en les conseillant sur la protection des données (ou métadonnées).

### Préserver

La préservation est l'avant-dernière étape du cycle de vie de la recherche et elle consiste à faire passer les données de l'état actif à l'état d'archive. Pour protéger les données, il faut mettre en œuvre des politiques nationales et universitaires de préservation à long terme (qui souvent reflètent ou recourent les directives internationales). Il faut préparer les données, les métadonnées, la documentation, les codes et les copies de sauvegarde en vue d'assurer l'accès à long terme et la réutilisation; dans certains cas, il faut migrer les données dans des formats qui se prêtent mieux à la préservation. La caractérisation et la normalisation des fichiers sont d'autres processus de préservation numérique. Des dépôts de données sécurisés sont d'excellents services qui ont reçu une accréditation prouvant à la communauté de la recherche que l'infrastructure numérique du dépôt est fiable et durable, et qui offrent une plateforme sur laquelle il est possible de stocker les données et d'y accéder à long terme. À ce niveau, les experts et les



expertes en GD peuvent offrir de la formation sur les pratiques exemplaires en archivage et en préservation numérique, et examiner et mettre en œuvre des accords ou des mandats de dépôt de données.

### Réutiliser

La réutilisation de recherche est l'étape finale du cycle de vie de la recherche et, pour ce qui est des données, elle signifie en assurer la découvrabilité et l'accès pour qu'elles puissent être combinées pour former de nouveaux ensembles de données, et référencées ou analysées par d'autres chercheurs et chercheuses. Au plus haut niveau, se trouvent des règles et des cadres législatifs de domaines, nationales et internationales portant sur le partage et le dépôt. L'application des principes *FAIR* à cette étape facilite la réutilisation, tout comme les outils qui permettent la réutilisation de la documentation (Lab Books) et des logiciels (GitHub). Les experts et les expertes en GD peuvent continuer d'assurer le soutien en matière de querelles de données et de règles (ou politiques) concernant l'attribution, la provenance et les licences, de même que l'interrogation et l'analyse secondaire.

### Stocker

Le stockage des données varie selon l'étape du cycle de vie de la recherche et de l'état (actif ou non). Peu importe l'état, les données, ou à tout le moins les métadonnées, devraient être rendues quelque peu accessibles à d'autres chercheurs et chercheuses. Ainsi, le stockage comporte donc des éléments à considérer quant au dépôt et à l'extraction de données (souvent facilités par les normes ouvertes telles que SWIFT et ORE) dans les plateformes de stockage en ligne (et le cas échéant, dans les supports physiques). L'utilisation de l'archivage demande un arrêt de réflexion sur l'accès à long terme et la protection de l'intégrité numérique du contenu, de même que la diffusion. Les règles des domaines, des universités et les politiques nationales concernant la confidentialité et la sécurité et le partage des données doivent être prises en compte dans la décision sur l'accès.

Pour que les données (métadonnées) soient accessibles, des plateformes de stockage à long terme devraient être ouvertes et durables; il existe un certain nombre d'options : OpenStack, FRDR Globus, Open Science Framework (OSF) du Centre for Open Science et autres services de domaines). Les experts et les expertes en GD peuvent faire profiter la communauté de la recherche de leurs connaissances au sujet de services spécifiques à des domaines et des calendriers de stockage à établir selon les pratiques locales, nationales, internationales et de gouvernance des données des domaines. Les experts et les expertes devraient aussi tenir compte de l'intégration des environnements de bureau et des processus, tels que la synchronisation des fichiers, pour faciliter le stockage des données pour les chercheurs et les chercheuses.

### Découvrir

Les chercheurs et les chercheuses devraient s'efforcer de rendre leurs données découvrables à toutes les étapes du cycle de vie de la recherche, entendu que « découvrir » dans ce contexte ne se limite pas à chercher des données mais englobe aussi leur mobilisation, leur localisation, leur interprétation et leur évaluation, ce qui, en retour, permet aux autres chercheurs et chercheuses de compiler et de créer de nouvelles données (métadonnées). Il y aurait lieu de tenir compte d'un certain nombre de pratiques exemplaires pouvant mener à une découvrabilité de qualité plus tard dans le cycle de vie : décider des schémas et des ontologies appropriés et comprendre les

possibles concordances, tenir compte des protocoles de collecte appropriés pour tous les types de métadonnées (OAI-ORE) et adopter des PID (DOI, ORCID). Connaître le dépôt dans lequel les données seront déposées est aussi une étape importante, étant donné que cela permettra au chercheur ou à la chercheuse de se préparer à suivre les normes du dépôt (tel que le modèle des données *SHARE*), ou de tenir compte des couches de découverte spécialisées, telles que des registres, qui facilitent une approche de découverte fédérée.

Suivre les principes *FAIR* renforce l'accessibilité et la découvrabilité des métadonnées de toutes les étapes du cycle de vie de recherche, y compris de celles qui sont dérivées de l'information et des données de recherche. Les experts et les expertes en GD peuvent conseiller la communauté de recherche sur les principes *FAIR* ainsi que la former sur différents services et approches de découverte. Les experts et les expertes devraient aussi envisager le développement de services en se basant sur des cas d'utilisation générale.

### Étayer et conserver

Les tâches consistant à étayer (documenter) et conserver les données (métadonnées) devraient être planifiées tôt et être exécutées tout au long du cycle de vie de la recherche pour maximiser l'interopérabilité et la découvrabilité, ce qui demande de décrire le contexte et le flux de travail entourant les données – code et autres documents, par exemple – et appliquer les normes relatives aux métadonnées appropriées (trouvées grâce à des ressources telles que [FAIRsharing.org](http://FAIRsharing.org)) afin de fournir des descriptions riches aux niveaux appropriés. Bien sûr, les données elles-mêmes devraient aussi être décrites, identifiées et expliquées à des fins de préservation.

Les chercheurs et les chercheuses devraient connaître les principes *FAIR* et les politiques de gestion de données nationales appropriées, de même que les politiques des domaines et des revues, mais si ce n'est pas le cas, ils et elles peuvent faire appel aux experts et expertes de leur établissement ou à des ressources externes telles que le glossaire de GDR de DRC-CASRAI pour faciliter la formation. Les registres des stratégies de format (p. ex. registres PRONOM et RDA) fournissent aussi des services de normalisation utiles.

### Protéger

Le consentement relatif au partage des données est un autre aspect de la GD à considérer : se demander si le consentement ou l'anonymat des données est requis ou non, établir les limites quant au partage des données et faire en sorte que les conditions juridiques et éthiques sur l'utilisation des données soient suivies et que l'intégrité et la provenance soient maintenues. Les chercheurs et les chercheuses devraient être prêts à se protéger contre les divulgations non intentionnelles tout en permettant un accès approprié aux données. Il est nécessaire de connaître les politiques en matière d'éthique à tous les niveaux et, dans de nombreux cas, un chercheur ou une chercheuse trouve des pratiques exemplaires internationales ou propres à son domaine (p. ex. *Health Insurance Portability and Accountability Act* et *Federal Information Security Act*) susceptibles de le ou la guider. D'autres normes telles que les normes *W3C security standards* peuvent guider le chercheur ou la chercheuse au chapitre de la protection de ses données. Des plateformes de sécurité éprouvées, telles que RedCap et DataSHIELD, ou l'utilisation d'installations sécurisées, telles que le Réseau canadien des centres de données de recherche (RCCDR), sont des ressources qui devraient être mises de l'avant à cette étape. Il est particulièrement important pour le chercheur ou la chercheuse de comprendre que les exigences

de sécurité changent tout au long du cycle de vie de la recherche et d'adopter des pratiques exemplaires et des procédures adaptées à ces changements. Les bureaux chargés de la confidentialité, les services de sécurité des TI des universités et les communautés dans lesquels la question de protection des renseignements personnels et d'accès revêt une importance particulière peuvent apporter conseils et assistance au chapitre de la sécurité des données de recherche.

## **Annexe B – Analyse environnementale des initiatives d'infrastructure de recherche numérique (IRN) de portée nationale et pan-nationale à l'appui de la gestion des données de recherche**

### **DOCUMENT DE TRAVAIL**

#### **Introduction**

Des initiatives de portée nationale et pan-nationale visant à coordonner les interventions dans le paysage de la science ouverte et à apporter des services et l'infrastructure de base aux chercheurs et chercheuses pour la gestion de leurs données de recherche sont en cours. Au Canada, l'Alliance de recherche numérique du Canada (l'Alliance) jouera un rôle crucial dans la défense et la coordination de l'IRN pour la recherche au Canada. En collaboration avec des partenaires et des parties prenantes de partout au pays, cette nouvelle organisation fournira aux chercheuses et chercheurs canadiens les outils, les services et l'infrastructure nécessaires à l'excellence en recherche, à l'innovation et aux avancées dans toutes les disciplines.

Le présent rapport dresse le portrait d'un certain nombre d'initiatives en cours dans d'autres pays qui visent à développer des infrastructures et des services communs pour soutenir et faire avancer l'état de la gestion des données de recherche, aux fins d'information et de comparaison avec l'Alliance.

## Critères d'examen

Les initiatives sont examinées en fonction d'une combinaison des éléments suivants :<sup>309</sup>

Élément	Description
Mandat	Mission et orientation de l'organisation
Administration	Structure organisationnelle et de gouvernance
Mise en œuvre	Forme, fonction, et/ou structure de l'organisation <ul style="list-style-type: none"><li>• Découverte et accès (découvrable, accessible)</li><li>• Services de données (interopérables, réutilisables, qualité +, Préservation)</li><li>• Compétences et formation</li></ul>

## Initiatives pan-nationales

### European Open Science Cloud (EOSC)<sup>310</sup> (nuage pour la science ouverte de l'Europe)

#### Mandat

Se dessine en Europe une approche fédérée d'avancement de la science ouverte. En 2016, la Commission européenne a accordé 260 M€ (note de traduction : 380 M\$ CA) pour que soient fédérées les infrastructures de données scientifiques sous une nouvelle entité appelée « European Open Science Cloud » (EOSC). Cette entité accueillera un réseau d'organisations et d'infrastructures de divers pays et communautés, lequel réseau soutiendra la libre création et la diffusion du savoir et des données scientifiques. Son objectif est de montrer la voie à l'Union européenne (UE) en gestion des données de recherche grâce à [trad.] « *un environnement virtuel avec des services gratuits, ouverts et transparents pour le stockage, la gestion, l'analyse et la réutilisation des données de recherche, au-delà des frontières et des disciplines scientifiques* ». <sup>311</sup>

La vision de l'EOSC pour des données, des services et des infrastructures interopérables prendra du temps à se réaliser, mais les mesures de départ comprennent la formation d'une « plateforme rudimentaire viable », assortie de règles de participation, qui orientera la prestation de services

<sup>309</sup> Adapté de CODATA (2019). Coordinating Global Open Science Commons Initiatives. Repéré en novembre 2020 à [https://conference.codata.org/CODATA\\_2019/sessions/155](https://conference.codata.org/CODATA_2019/sessions/155)

<sup>310</sup> <https://www.eosc-portal.eu>

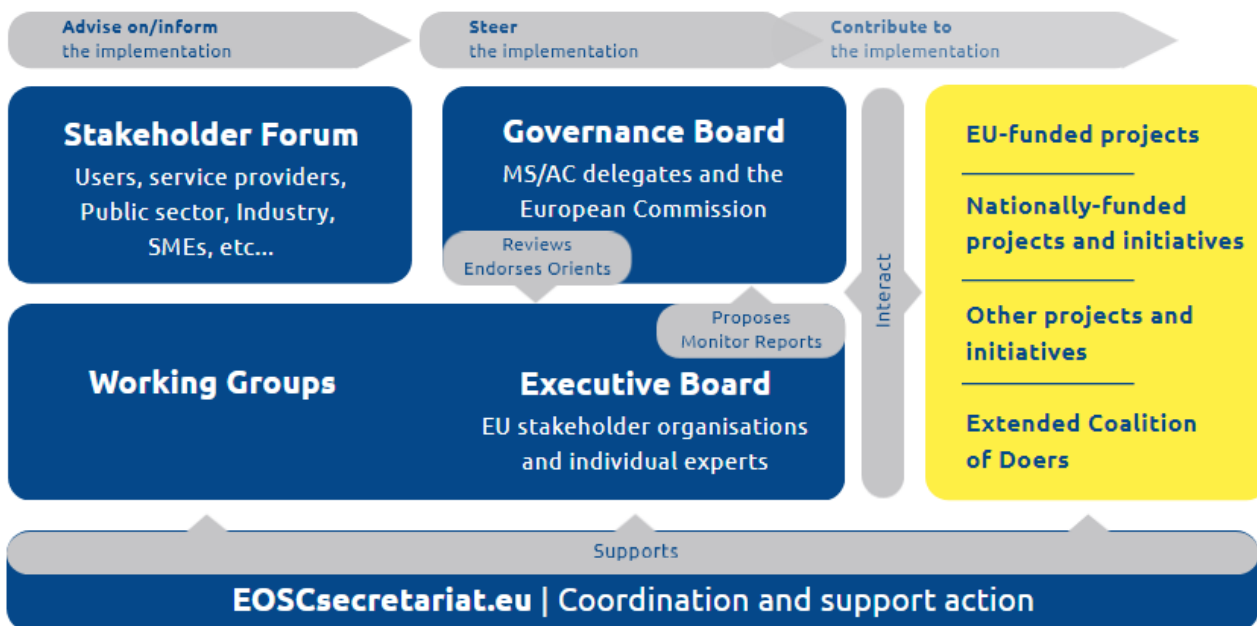
<sup>311</sup> Commission européenne. (2018). Implementation Roadmap for the European Open Science Cloud, SWD(2018)83. [https://ec.europa.eu/transparency/documents-register/detail?ref=SWD\(2018\)83&lang=en](https://ec.europa.eu/transparency/documents-register/detail?ref=SWD(2018)83&lang=en)

et l'établissement d'un plan d'action pour l'interopérabilité des données afin de rendre les principes *FAIR* opérationnels.

### *Administration*

Pour la période 2019-2020, le modèle de gouvernance de l'EOSC comprend un conseil de gouvernance, un conseil de direction et un forum des parties prenantes.

- Le conseil de gouvernance est composé de représentant(e)s des États membres et des États associés et il est présidé par des représentant(e)s de la Commission européenne.
- Le conseil de direction est composé de 11 membres choisis au terme d'un appel de candidatures.
- Le forum des parties prenantes regroupe des utilisateurs et utilisatrices, des projets d'envergure européenne et nationale, des fournisseurs de services, le secteur public, des PME, le secteur privé, etc.
- Cinq groupes de travail coordonnent l'avancement des priorités choisies par le conseil de gouvernance selon les propositions faites par le conseil de direction. Ils sont formés de représentant(e)s des parties prenantes de l'EOSC.
  - Paysage : cartographier les infrastructures de recherche existantes pressenties pour faire partie de la fédération de l'EOSC;
  - FAIR : mettre en œuvre les principes régissant les données *FAIR* en définissant les exigences correspondantes pour le développement des services de l'EOSC afin de favoriser l'interopérabilité entre les disciplines;
  - Architecture : définir le cadre technique requis pour rendre possible et soutenir une fédération des systèmes en continuel changement;
  - Règles de participation : concevoir les règles de participation qui définiront les droits, les obligations régissant les transactions de l'EOSC entre les utilisateurs et utilisatrices, les fournisseurs et les opérateurs;
  - Durabilité : soumettre une série de recommandations concernant la mise en œuvre d'une fédération de l'EOSC opérationnelle, modulable et durable après 2020.



**Figure B1.** Modèle de gouvernance de l' EOSC.<sup>312</sup>

### Mise en œuvre

Une feuille de route décrit six mesures à prendre pour mettre en œuvre l'EOSC:<sup>313</sup>

- Architecture
  - L'EOSC serait formée d'un noyau et de diverses infrastructures de données de recherche fédérées engagées à fournir des services dans le cadre de l'EOSC. Le noyau doit être constitué de ressources partagées de l'EOSC et d'un cadre de conformité, y compris les règles de participation.
  - Le processus de fédération comprend deux activités interreliées :
    - Développer des ressources partagées pour le noyau
    - Connecter au noyau un grand nombre d'infrastructures de données de recherche
- Données

<sup>312</sup> Image tirée de <https://www.eoscsecretariat.eu>

<sup>313</sup> Commission européenne. (2018). Implementation Roadmap for the European Open Science Cloud, SWD(2018)83. [https://ec.europa.eu/transparency/documents-register/detail?ref=SWD\(2018\)83&lang=en](https://ec.europa.eu/transparency/documents-register/detail?ref=SWD(2018)83&lang=en)

- Favoriser le développement de pratiques professionnelles de gestion des données de recherche et d'intendance en Europe en :
  - instaurant un culture axée sur la gestion des données de recherche et en perfectionnant les compétences pratiques de la communauté de recherche de l'UE;
  - développant des outils, des spécifications, des catalogues et des normes régissant les données et l'offre de services en appui aux chercheurs et chercheuses;
  - encourageant la cohérence dans les mandats et les mesures incitatives en faveur de données *FAIR* de la part des bailleurs de fonds et des établissements de recherche en Europe.
  
- Services
  - L'EOSC planifie d'offrir cinq grands services à la communauté de recherche européenne, sans égard à l'appartenance disciplinaire ou aux frontières des pays :
    1. un service unique d'identification et d'authentification et un point d'accès et un système d'acheminement aux ressources de l'EOSC;
    2. un espace/milieu de travail protégé et personnalisé;
    3. un accès à l'information de service utile (p. ex. liste des infrastructures de données fédérées, information sur les politiques) et à des directives spécifiques (directives pour les données FAIR, homologation de dépôts);
    4. services pour trouver, consulter, réutiliser et analyser des données de recherche générées par d'autres, accessibles à l'aide de catalogues et services de données (p. ex. analyse, fusion, exploration, traitement);
    5. services pour créer ses propres données FAIR, pour les stocker et en assurer la préservation à long terme.
  
- Accès et interface
  - De multiple points d'entrée pour accéder aux services de l'EOSC sont recommandés pour soutenir une transition harmonieuse des anciens systèmes, plutôt que d'imposer un point d'accès unique.
  - Les points d'entrée seraient une application frontale qui s'adapterait aux besoins des communautés d'utilisateurs et d'utilisatrices, qui se trouverait sur une plateforme commune donnant accès aux ressources partagées de l'EOSC.
  
- Règles
  - Des règles partagées entre les parties prenantes participantes établiront les droits, les obligations et la responsabilité.

- Les règles devront concerner l'utilisation des ressources partagées de l'EOSC (outils, spécifications, normes, catalogues), la réglementation des transactions dans l'EOSC et les cadres juridiques applicables.
- La conformité aux règles variera selon le rôle, l'endroit, la maturité de l'organisation et le contexte disciplinaire.
- Gouvernance
  - Un cadre opérationnel de gouvernance et de coordination assorti d'initiatives nationales pertinentes.
  - Un cadre de financement pour garantir la viabilité.

Les premiers projets menant à des résultats qui agiront comme couche de base initiale de l'EOSC ont été financés par le programme Horizon 2020.<sup>314</sup> Exemples :

- EOSCPilot établit le cadre de gouvernance et les règles initiales de participation.<sup>315</sup>
- EOSC-Hub soutient la fédération des principales infrastructures électroniques en créant un système d'intégration et de gestion qui agira comme point de contact unique pour les chercheurs et les chercheuses leur permettant de découvrir, de consulter et d'utiliser les ressources d'IRN.<sup>316</sup>
- Les projets eInfraCentral et EOSC-Hub ont collaboré sur le portail de découverte des ressources de l'EOSC.<sup>317</sup>
- HNSciCloud a développé une plateforme infonuagique hybride connectant les fournisseurs de services infonuagiques commerciaux et les organisations de recherche financées par le secteur public.<sup>318</sup>
- GÉANT est le fournisseur de réseau pour la fourniture par l'EOSC d'un accès approprié aux services, données, infrastructures de recherche infonuagiques et de beaucoup d'autres ressources et services de l'EOSC.<sup>319</sup>

---

<sup>314</sup> Pour la liste complète des projets financés contribuant à l'EOSC : <https://www.eosc-portal.eu/about/eosc-projects>

<sup>315</sup> <https://cordis.europa.eu/project/id/739563>

<sup>316</sup> <https://cordis.europa.eu/project/id/777536>

<sup>317</sup> <https://cordis.europa.eu/project/id/731049>

<sup>318</sup> <https://cordis.europa.eu/project/id/687614>

<sup>319</sup> <https://ec.europa.eu/digital-single-market/en/geant-project-european-success-story> (nota : non accessible le 17 juin)



- OpenAIRE soutient une gamme d'initiatives favorisant l'adoption par la communauté de recherche de la science ouverte.<sup>320</sup>

Le catalogue des services actuels de l'EOSC donne accès aux ressources intégrées, qui sont découvrables par domaine scientifique, catégorie de service ou fournisseur. Voir <https://www.eosc-portal.eu/services-resources>

## African Open Science Platform (AOSP)<sup>321</sup> (plateforme africaine pour la science ouverte)

### *Mandat*

La plateforme africaine pour la science ouverte (« AOSP ») est une initiative lancée en 2016 par le ministère de la science et de la technologie sud-africain dans l'objectif de rapprocher les activités de science ouverte en Afrique par des mécanismes de collaboration et de coordination, et l'échange de pratiques exemplaires. Elle vise à soutenir la recherche avancée en science ouverte partout en Afrique et à s'aligner sur les programmes existants dans les systèmes de recherche régionaux et nationaux.

La vision de l'AOSP :

1. Un système fédéré fournissant à la communauté de recherche et autres acteurs de la société les moyens de trouver, de déposer, de gérer, de partager et de réutiliser des données, des logiciels et des métadonnées dans leur travail
2. Un réseau rapprochant les acteurs dispersés, permettant l'adoption d'outils numériques et développant les capacités des personnes et des établissements

### *Administration*

La période d'essai de 2017-2019 pour l'AOSP a été soutenue par le ministère des sciences et de l'innovation de l'Afrique du Sud et gérée par la National Research Foundation of South Africa (NRF) et l'Academy of Science of South Africa, avec des partenaires du International Science Council et CODATA. En avril 2020, on annonçait que le bureau de projet de l'AOSP serait hébergé par la NRF.<sup>322</sup>

### *Mise en œuvre*

La phase suivante de l'AOSP consistera à développer un cadre de gouvernance et un modèle de financement à long terme durable et à officialiser le modèle d'activité de l'AOSP.

---

<sup>320</sup> <https://cordis.europa.eu/project/id/731011>

<sup>321</sup> <http://africanopenscience.org.za> (nota : lien non valide le 18 juin)

<sup>322</sup> <https://www.nrf.ac.za/media-room/news/nrf-south-africa-host-aosp-project-office>

L'équipe de gestion provisoire sera formée d'un directeur, de quatre agents de plateforme (sciences des données, intendance des données, formation et compétence, et érection, communication et rayonnement du réseau) et un agent administratif.

Les membres de l'AOSP seront les universités et les organismes représentatifs, les académies de science, les conseils subventionnaires et les fournisseurs de services.

Pendant la période d'essai de l'AOSP, la plateforme a été lancée au moyen d'ateliers, de réunions et de présentations visant à sensibiliser les parties prenantes sur la nécessité de protéger de manière sécurisée les données scientifiques en Afrique. Parmi les produits : une étude du paysage mettant en lien toutes les initiatives de recherche à forte intensité de données, ainsi qu'un travail d'exploration des cadres et des feuilles de route en vue de préparer une politique et une infrastructure pour la science ouverte et renforcer les capacités à cet égard.

Les activités décrites dans la stratégie de l'AOSP sont :<sup>323</sup>

- Fournir des installations d'infonuagique pour le calcul, l'accès aux données et les outils d'analyse en réseau
- Fournir des logiciels et des conseils pratiques sur la GDR, les politiques et les pratiques en matière de science ouverte
- Créer et soutenir des capacités de recherche concurrentielles en analytique et en intelligence artificielle
- Créer des programmes de recherche à forte intensité de données en appliquant des technologies de données aux grands domaines de recherche

The Arab States Research and Education Network (ASREN)<sup>324</sup> (réseau d'éducation et de recherche des États arabes)

### *Mandat*

L'ASREN vise à mettre en œuvre, à gérer et à étendre les infrastructures électroniques panarabes dédiées aux communautés de recherche et d'éducation et à stimuler la coopération et la recherche scientifiques dans les pays membres en fournissant des infrastructures et des services électroniques de classe mondiale.

### *Administration*

L'ASREN est une organisation internationale à but non lucratif, enregistrée à Dusseldorf, en Allemagne, le 3 juin 2011, sous l'égide de la Ligue des États arabes. Il est composé de

---

<sup>323</sup> The African Open Science Platform (s. d.). The future of science and science for the future. <https://www.nrf.ac.za/sites/default/files/documents/AOSP%20Strategy%20Final%20HR.pdf>

<sup>324</sup> <http://asrenorg.net>

l'association de réseaux nationaux d'éducation et de recherche (« NREN ») du monde arabe ainsi que de leurs partenaires stratégiques.

### *Mise en œuvre*

L'ASREN s'est d'abord employé à créer des communautés de passerelles scientifiques fournissant l'accès haute vitesse aux applications scientifiques et aux ressources informatiques, en :

- Développant des réseaux de communications de données haute vitesse, notamment
  - un point de présence (« PoP ») à Londres fournissant les terminaux de l'UE et l'appairage aux connexions de son PoP à la NREN arabe et permettant l'interconnexion avec le réseau GÉANT en Europe, Internet2 aux États-Unis et d'autres réseaux régionaux dans le monde.<sup>325</sup>
  - une contribution à un bâti régional sur l'infrastructure de réseau EUMEDCONNECT et AFRICACONNECT, co-fondée par la Commission européenne.
- Développant l'infrastructure électronique de grille EUMEDGRID et en favorisant le transfert de nouvelles applications sur la plateforme de grille

L'ASREN fournit des services d'internet haute vitesse et des services d'authentification par le truchement, respectivement, d'eduroam et d'eduGAIN.

### *Collaborations*

L'ASREN collabore en ce moment avec EGI.eu pour coordonner et harmoniser ses infrastructures électroniques en définissant un modèle opérationnel et organisationnel qui est interopérable avec les infrastructures électroniques des pays de l'UE et comme passerelles vers d'autres régions.<sup>326</sup>

L'ASREN collabore aussi avec LIBSENSE et WACREN à la création d'une revue en libre accès et un dépôt de données au service de l'Afrique du Nord et de Moyen-Orient.<sup>327</sup>

Le projet MAGIC (Middleware for collaborative Applications and Global Virtual Communities) de l'ASREN établit des ententes pour l'Europe, l'Amérique latine et d'autres régions participantes pour créer un lieu d'échange de services et d'applications en temps réel pour des groupes de recherche internationaux.

---

<sup>325</sup> <http://asrenorg.net/?q=content/london-pop>

<sup>326</sup> <http://asrenorg.net/?q=content/chain-reds>

<sup>327</sup> <http://asrenorg.net/?q=content/libsense-iii-workshop-agenda>

## Nordic e-Infrastructure Collaboration (NeIC)<sup>328</sup> (collaboration nordique en matière d'infrastructure électronique)

### *Mandat*

NeIC a été créé en 2012 dans l'objectif de devenir un modèle de collaborations transfrontalières en matière d'infrastructure électronique durable et décentralisée. Elle explore, évalue et met en place des services d'infrastructure innovateurs pour répondre aux besoins des fournisseurs d'infrastructure électronique nationaux, de leurs utilisateurs et utilisatrices, et des projets d'intérêt commun pour le Nord.

Des projets de NeIC sont en cours dans les domaines des sciences de la physique et de l'ingénierie, des sciences de l'environnement, des sciences sociales, de la culture et société, des sciences de la vie et de la cyber-science.

### *Administration*

NeIC est hébergé par NordForsk, la division de financement de la recherche et de l'infrastructure de recherche du conseil nordique, qui est l'organe officiel de la coopération interparlementaire entre les pays nordiques.

Le conseil de NeIC est formé d'un représentant de chaque fournisseur d'infrastructure électronique nationale : CSC (Finlande), SNIC (Suède), UNINETT Sigma2 (Norvège), DeIC (Danemark), RH Net (Islande) et ETAIS (Estonie).

Le conseil de NeIC a le pouvoir de prendre des décisions stratégiques au sujet de l'infrastructure de calcul et de stockage de données. Il recommande au NordForsk l'embauche du directeur de NeIC.

NeIC est géré par une équipe de direction présidée par le directeur de NeIC. L'équipe de direction coordonne les activités et participe aux groupes de pilotage de projets en tant de responsables des projets.

En collaboration avec les fournisseurs d'infrastructure électronique et les représentants de la communauté des utilisateurs, NeIC mandate des experts pour participer aux projets et aux activités opérationnelles.

Le financement des activités de NeIC provient d'organismes de financement nationaux, de NordForsk et de partenaires de projet participants.

### *Mise en œuvre*

NeIC donne aux chercheurs et chercheuses des pays membres l'accès à un centre de soutien commun, à des outils communs pour le partage et l'analyse de données, à des mécanismes de partage entre les ressources de calcul de haute performance et des ateliers pour les chercheurs et les chercheuses sur l'analyse des données, le développement de logiciels et la gestion des données de recherche.

---

<sup>328</sup> <https://neic.no>

NeIC finance des projets d'infrastructure électronique de collaboration avec les partenaires nordiques.

NeIC fournit une gamme de services connexes, qui sont inscrits avec un DataCite PID (10.23673/kpyv-1k13).<sup>329</sup>

Les services en matière de gestion des données comprennent une plateforme en ligne pour la création de plans de gestion de données, des plateformes de dépôt pour le stockage, le partage et la gestion des données, et un environnement de bloc-notes pour travailler avec les données et les programmes, une plateforme pour recueillir, stocker, analyser et partager des données sensibles et des ateliers sur les données *FAIR* et les logiciels de recherche.

NeIC bâtit aussi des réseaux professionnels pour rapprocher les experts et les expertes qui travaillent sur les mêmes sujets dans différentes organisations nordiques. Pour ce faire, NeIC organise des ateliers et bâtit des espaces virtuels pour le travail d'équipe décentralisé.

## Initiatives nationales

---

### Pays-Bas

[Data Archiving and Networked Services \(DANS\)](#)<sup>330</sup> (archivage de données et services en réseau)

#### *Mandat*

Pour promouvoir l'accès durable aux données de recherche numériques, DANS prodigue des conseils avisés et des services aux chercheurs et aux chercheuses qui veulent rendre leurs données de recherche *FAIR*, notamment des services d'archivage à long terme et de réutilisation des données de recherche venant de recherches terminées et de soutien en gestion des données pendant l'exécution de projets.

#### *Administration*

DANS est une initiative menée conjointement par le Royal Netherlands Academy of Arts and Sciences et la Netherlands Organization for Scientific Research. Un comité directeur supervise les affaires courantes, la gestion et les politiques. Trois conseils consultatifs guident les plateformes et les services : le comité consultatif en sciences, le comité consultatif du portail NARCIS (système national d'information sur la recherche et les collaborations universitaires) et le conseil consultatif de DataverseNL. Un groupe d'utilisateurs formé de chercheurs et chercheuses et d'employé(e)s des universités néerlandaises donne de la rétroaction et des idées sur les services.

---

<sup>329</sup> Nordic e-Infrastructure Collaboration (2020). Nordic e-Infrastructure Services. EOSC-Nordic.  
<https://search.datacite.org/works/10.23673/kpyv-1k13>

<sup>330</sup> <https://dans.knaw.nl/>

Un secrétariat central est chargé de développer l'infrastructure, les services et les politiques, de même que les activités.

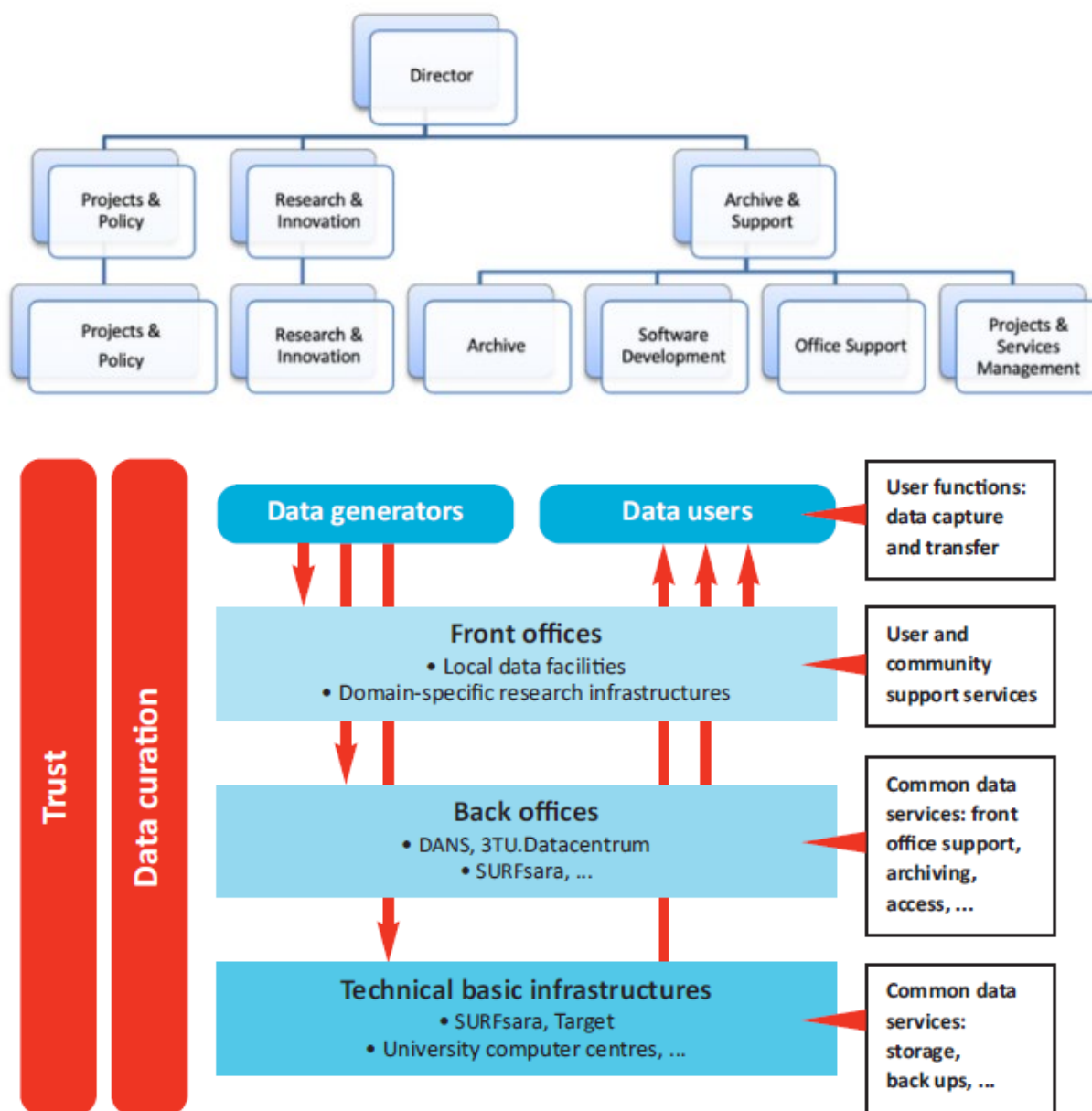


Figure B2 et B3. Structure organisationnelle et flux de travail de DANS.<sup>331</sup>

<sup>331</sup> Images tirées de <https://dans.knaw.nl/en/about/organisation-and-policy/organigram>, et <https://dans.knaw.nl/en/about/organisation-and-policy/information-material/DANSstrategienota20152020UK.pdf>

## Mise en œuvre

DANS bâtit l'infrastructure et les services fédérés en appui à la communauté de recherche. Services offerts :<sup>332</sup>

- DataverseNL : les données de recherche peuvent déjà être stockées, partagées et publiées via DataverseNL pendant la recherche. Il gère le réseau alors que les établissements participants gèrent les données déposées dans leurs dépôts locaux. Chaque établissement participe au conseil consultatif de Dataverse, qui détermine les politiques du service. Les établissements paient un montant fixe pour leur participation à DataverseNL, et les coûts de stockage des données leur sont facturés en sus.
- EASY : après les recherches, les données peuvent être stockées et partagées en permanence via EASY, le système d'archivage en ligne, qui offre aussi un accès aux microdonnées sécurisées de Statistics Netherlands. EASY est certifié CoreTrustSeal et Nestor Seal.
- NARCIS : l'information au sujet des projets de recherche, des publications ouvertes et des logiciels de recherche peuvent être partagées via le portail scientifique NARCIS.
- Formation et conseil : DANS prodigue des conseils et donne des formations dans le domaine de la durabilité numérique, de la durabilité logicielle, de la gestion des données, des données *FAIR* et de la gestion des données de recherche.

Aux Pays-Bas, DANS collabore avec les parties prenantes de l'IRN et divers dépôts de domaines. Ils interviennent aussi dans la communauté internationale de la GDR par divers réseaux et projets d'infrastructure.<sup>333</sup>

En collaboration avec le centre des sciences électroniques des Pays-Bas, DANS a lancé le site Web *fair-software.nl* en appui à la formation sur la durabilité des logiciels de recherche.<sup>334</sup>

## Allemagne

---

<sup>332</sup> <https://dans.knaw.nl/en/about/services>

<sup>333</sup> <https://dans.knaw.nl/en/projects>; <https://dans.knaw.nl/en/about/organisation-and-policy/collaboration>

<sup>334</sup> <https://fair-software.nl>

National Research Data Infrastructure (NFDI)<sup>335</sup> (infrastructure nationale de données de recherche)

### *Mandat*

L'objectif de l'infrastructure nationale de données de recherche (« NFDI ») est de gérer systématiquement les données scientifiques et de recherche, d'assurer le stockage, la sauvegarde et l'accessibilité à long terme des données, et de mettre en réseau les données à l'échelle nationale et internationale. Le NFDI rassemblera de multiples intervenants en un réseau coordonné de consortiums chargés de fournir des services de données scientifiques aux milieux de la recherche.

### *Administration*

- Les consortiums sont généralement organisés par domaine ou méthode de recherche. Ils visent à améliorer et protéger l'accès et l'utilisation des données de recherche dans leur domaine.
- Les porte-paroles de chaque consortium forment l'assemblée des consortiums. Celle-ci prend les décisions concernant la recherche, les opérations et les services au nom des consortiums de la NFDI et de cette dernière, et décide de l'introduction de procédures, de services et/ou de normes interdisciplinaires en lien avec la NFDI dans le cadre des principes approuvés par le conseil scientifique de la NFDI.
- Le conseil scientifique est l'organe stratégique de la NFDI. Il prodigue des conseils sur les questions liées au développement continu de l'infrastructure nationale des données de recherche pour que la NFDI puisse se connecter aux infrastructures nationales et internationales.
- Une direction locale réunit un directeur à temps plein et un bureau de gestion. Elle veille à ce que les sujets interdisciplinaires, tels que l'éducation, la protection des données et l'éthique des données, soient discutés, communiqués et coordonnés entre les consortiums de la NFDI.

### *Mise en œuvre*

LA NFDI sera établie en tant que réseau coopératif de consortiums en trois étapes, sur une période de trois ans (2019 à 2021). À chacune des étapes, de nouveaux consortiums peuvent être ajoutés à la NFDI dans le processus axé sur la recherche. Le but est de créer un cadre exhaustif de consortiums interconnectés offrant un service national de données de recherche à la communauté de la recherche. En tout, 22 propositions des consortiums de la NFDI ont été reçues en réponse à l'appel de propositions le 15 octobre 2019. Un total de 142 établissements étaient concernées par les propositions. D'autres appels sont prévus pour 2020 et 2021.

Le programme de la NFDI pour les consortiums :

---

<sup>335</sup> [https://www.dfg.de/en/research\\_funding/programmes/nfdi/index.html](https://www.dfg.de/en/research_funding/programmes/nfdi/index.html)



- Établissement de normes, de procédures et de lignes directrices en matière de traitement des données en étroite collaboration avec la communauté d'intérêt
- Élaboration de normes interdisciplinaires en matière de métadonnées
- Élaboration de mesures et de services de gestion des données fiables et interopérables adaptés aux besoins de la communauté d'intérêt
- Réutilisation accrue des données existantes, également au-delà des limites du sujet
- Amélioration du réseautage et de la collaboration avec des partenaires extérieurs au système de recherche universitaire allemand ayant une expertise dans la gestion des données de recherche
- Participation à l'élaboration et à l'établissement de services et de normes génériques et inter-consortiums en matière de gestion des données de recherche avec d'autres consortiums

## Chine

China Science and Technology Cloud (CSTCloud)<sup>336</sup> (nuage de la science et de la technologie de la Chine)

### *Mandat*

La fédération China CSTCloud fournit aux communautés des domaines de l'éducation, de la recherche, des sciences et des techniques, aux ministères gouvernementaux et aux entreprises de haute technologie de la Chine une panoplie de services de cyber-infrastructure et d'internet tels que la gestion de l'accès réseau et des identités de même que la puissance informatique, le stockage infonuagique et les logiciels de recherche.<sup>337</sup>

### *Administration*

Le CSTCloud est géré par le centre d'information du réseau informatique qui fait partie de l'académie des sciences de la Chine.<sup>338</sup> Le centre d'information a pour mission de bâtir la technologie de l'information et des communications en appui aux innovations scientifiques et aux activités de gestion, de promouvoir la R et D en technologie de l'information, d'intégrer des ressources électroniques de sciences et de gestion et de renforcer la propagation des idées scientifiques.<sup>339</sup>

---

<sup>336</sup> <https://www.cstcloud.net/>

<sup>337</sup> [http://english.cnic.cas.cn/patform/202001/t20200106\\_228938.html](http://english.cnic.cas.cn/patform/202001/t20200106_228938.html)

<sup>338</sup> <http://cstcloud.net/index.html> (*nota* : lien non valide le 19 juin)

<sup>339</sup> <https://www.egi.eu/about/newsletters/egi-and-the-chinese-academy-of-sciences-collaborate-to-boost-science-beyond-national-boundaries/>

Un comité des sciences et de la technologie conseille le centre d'information sur les priorités du CSTCloud,<sup>340</sup> et un comité consultatif en stratégie présente des avis sur les stratégies de développement.<sup>341</sup>

### *Mise en œuvre*

L'initiative du CSTCloud a été annoncée à l'occasion de la quatrième conférence mondiale sur l'Internet en 2017, à la suite de quoi la première plateforme était lancée en 2018. La version la plus récente date de 2019.<sup>342</sup> En théorie, les composants du CSTCloud couvrant les services de réseau, d'informatique et de données sont livrés en quatre niveaux de service :

- Infrastructure : accès internet (CSTNET), infonuagique, VPN, CHP (réseau CNGrid) et nuage de l'IA
- Opérations relatives à l'archivage des données, la préservation à long terme et la reprise après sinistre
- Collaboration pour les téléconférences par vidéo, web et informatique et services courriel et mobiles
- Série de plateformes disciplinaires pour la recherche en sciences de la matière, en physique des hautes énergies, en chimie computationnelle et en microbiologie accompagnées de plateformes réservées à la visualisation des données, au développement des logiciels de recherche et à la modélisation de l'IA

CSTCloud fournit aussi des services à d'autres initiatives de l'académie des sciences : CASEarth, CAS Space Science Missions, the Five-hundred-meter Aperture Spherical Telescope,<sup>343</sup> et le Large High-Altitude Air Shower Observatory.<sup>344</sup>

On s'emploie en ce moment à développer 20 centres de données nationaux, couvrant tous les types de données de recherche. Ces centres devraient être intégrés au CSTCloud.<sup>345</sup>

## **Japon**

---

<sup>340</sup> <http://english.cnpc.cas.cn/about/stc/>

<sup>341</sup> <http://english.cnpc.cas.cn/about/sac/>

<sup>342</sup> [http://english.cnpc.cas.cn/patform/202001/t20200106\\_228938.html](http://english.cnpc.cas.cn/patform/202001/t20200106_228938.html)

<sup>343</sup> <https://www.nature.com/articles/d41586-019-02790-3>

<sup>344</sup> <http://english.ihep.cas.cn/lhaaso/>

<sup>345</sup> <https://researchdata.springernature.com/posts/54209-is-china-ready-for-open-data>

National Institute for Informatics (NII) Research Data Cloud <sup>346</sup> (nuage des données de recherche de l'institut national d'informatique (« NII »))

### Mandat

La politique de promotion de la science ouverte est énoncée dans le 5<sup>e</sup> plan de base pour la science et la technologie (2017-2021), qui définit la politique de base sur les initiatives scientifiques et technologiques au Japon.<sup>347</sup> En appui à l'engagement en faveur de la science ouverte énoncé dans le plan, le NII a créé le nuage des données de recherche en tant qu'infrastructure électronique où les données de recherche et les autres fichiers connexes sont gérés, stockés et découverts. « *Notre infrastructure fonctionnera à tout moment comme si les girafes agissaient sans repos, et gèrera divers types de données qui diffèrent comme le motif de pelage des girafes. Les utilisateurs pourront accéder à l'un des besoins à partir d'une grande quantité de données comme si les girafes trouvant la nourriture nécessaire d'un point de vue élevé dans la savane* ». <sup>348</sup>

### Administration

Le nuage des données de recherche est bâti et tenu par le centre de recherche pour la science ouverte et la plateforme de données (« RCOS »). Un des composants, le nuage JAIRO, est une initiative menée conjointement par le NII et le consortium japonais pour les dépôts en libre accès (« JPCOAR »).<sup>349</sup>

### Mise en œuvre

Le nuage des données de recherche du NII est entré en opération en 2020. Il utilise la version 5 du réseau de l'information en sciences (« SINET5 ») pour fournir le service d'authentification, l'infrastructure infonuagique et le contenu universitaire pour promouvoir la science ouverte. Il comporte trois plateformes :

- Une plateforme de gestion de données de recherche (GakuNin RDM<sup>350</sup>) pour la gestion des données actives pendant l'exécution des projets; montée sur l'Open Science Framework.<sup>351</sup>
- Une plateforme de dépôts (WEKO3), pour stocker les données, les publications dérivées et leurs relations pour bâtir le graphique du réseau des communications savantes. Montée

---

<sup>346</sup> <https://rcos.nii.ac.jp/en/service/>

<sup>347</sup> <https://rcos.nii.ac.jp/en/about/>

<sup>348</sup> <https://rcos.nii.ac.jp/en/about/mascot/>

<sup>349</sup> <https://rcos.nii.ac.jp/en/service/weko3/>

<sup>350</sup> <https://rdm.nii.ac.jp/>

<sup>351</sup> <https://www.cos.io/our-products/osf>

sur le nuage JAIRO, service infonuagique de dépôts institutionnels offert à l'échelle nationale. Elle comprend un service DOI.

- Une plateforme de découverte (« CiNii Research ») qui recueille les métadonnées de dépôts institutionnels et d'autres bases de données ouvertes.<sup>352</sup>

Ces plateformes sous-tendent le centre de recherche pour les mégadonnées médicales du NII, une plateforme d'initiatives de R et D, y compris la technologie de l'IA sur le nuage pour analyser les images médicales.

L'accès aux services par les établissements au Japon est soutenu par Shibboleth et géré par la Fédération universitaire de gestion de l'accès, aussi appelée « GakuNin », chargée d'arriver à l'accès entre les fédérations dans le monde.<sup>353</sup>

## Corée

[KISTI/Korean Research Data Platform](#) <sup>354</sup> (KISTI/plateforme de données de recherche coréenne)

### *Mandat*

L'institut coréen de l'information scientifique et technologique (« KISTI ») est un établissement de recherche financé par le gouvernement qui soutient la compétitivité de la R et D de la Corée par la recherche et la collecte de renseignements, l'élaboration de services et de normes de portée nationale et la mise à disposition d'une infrastructure d'IRN avancée, y compris un réseau de recherche haute vitesse, le calcul de haute performance et un réseau national de dépôts.

La Corée bâtit sa plateforme de données de recherches (« KRDP ») depuis 2018 avec l'intention d'amener les chercheurs à partager, gérer, chercher, analyser et utiliser les données de recherche.

### *Administration*

Pour l'heure, le développement de la KRDP est dirigé par le centre de partage des données de recherche du KISTI, division des données nationales en sciences et en technologie.<sup>355</sup> Il y a peu de détails en ligne.

### *Mise en œuvre*

Une fois prête, la KRDP remplira les fonctions suivantes :<sup>356</sup>

---

<sup>352</sup> <https://rcos.nii.ac.jp/en/service/research/>

<sup>353</sup> <https://www.gakunin.jp/en>

<sup>354</sup> <https://www.kisti.re.kr/eng/rnd/pageView/250>

<sup>355</sup> <https://www.kisti.re.kr/eng/rnd/pageView/250>

<sup>356</sup> <https://github.com/pragmagrid/pragma-meetings/blob/master/pragma36/24/talk10-krdp.pdf>

- Environnement intégré de GDR pour préserver, partager et utiliser les données de recherche
- Outil fédéré de recherche pour les données de recherche décentralisées
- Environnement d'analyse collaborative en temps réel
- Intégration de l'outil de planification de la gestion des données

La KRDP intégrera également d'autres initiatives de science ouverte, par exemple les politiques d'accès ouvert et les dépôts et les archives d'accès ouvert coréens (« KoarXiv »).<sup>357</sup>

## Australie

**Australian Research Data Commons (ARDC)** <sup>358</sup> (réserve commune de données de recherche de l'Australie)

### *Mandat*

L'ARDC a été formée en 2018, faisant fond sur d'anciennes initiatives [l'Australian National Data Service (ANDS), les National eResearch Collaboration Tools (Nectar) et les Research Data Services (RDS)], dans l'objectif de rapprocher les gens, les données, les compétences et les ressources afin de permettre une recherche à forte intensité de données de classe mondiale. L'objectif primordial de l'ARDC est d'accélérer la recherche en Australie en développant, en testant et en soutenant des plateformes où les enquêteurs peuvent stocker, découvrir, partager, consulter des objets numériques (données, logiciels, etc.) et interagir avec eux. Son mandat est d'apporter de la cohérence à l'échelle nationale aux données et à la capacité des plateformes de recherche électronique, y compris :

- Le calcul de haute performance – l'infrastructure computationnelle nationale (« NCI ») et le centre de calcul de haute performance Pawsey
- Les réseaux de recherche – Le réseau de la recherche universitaire de l'Australie (« AARNet »)
- L'accès et l'authentification - Australian Access Federation (AAF)

L'ARDC a retenu cinq thèmes d'importance stratégique pour encadrer la mise en œuvre de sa vision :

- Thème 1 - Coordination et cohérence : permettre de constituer une réserve commune de données de recherche pour l'Australie
- Thème 2 – Personnes et politiques : Transformer la culture et la communauté

---

<sup>357</sup> <https://zenodo.org/record/3232912>

<sup>358</sup> <https://ardc.edu.au/>

- Thème 3 – Données et services : Maximiser la valeur de l’actif que représentent les données pour l’Australie
- Thème 4 – Logiciels et plateformes : Produire des résultats de recherche
- Thème 5 – Stockage et calcul : fournir une infrastructure de base

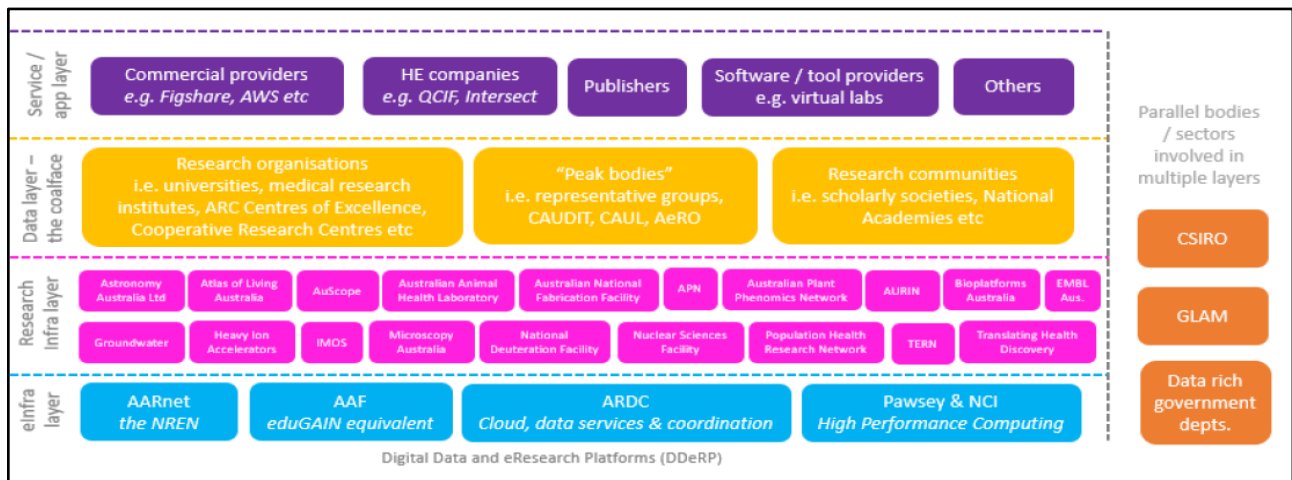
### Administration

L’ARDC dispose d’un budget de fonctionnement pour la période de son plan stratégique de 2019-2023 de 110 M\$ (note de traduction : 102 M \$CA) et d’un budget de dépenses en capital pour la même période de 72 M\$ (note de traduction : 67 M \$CA).

Une équipe de direction dirige les activités de base, en consultation avec la communauté sur les directives de travail et en soutien des activités de cocréation.

### Mise en œuvre

L’ARDC est d’avis que les partenariats avec les parties prenantes à divers niveaux dans chaque secteur de services sont essentiels à la défense de la notion de réserve commune de données. Un rapport de Sarah positionne les parties prenantes par rapport à la réserve commune (figure 4).



**Figure B4.** Position des parties prenantes dans l’ARDC (Sarah Jones, DCC)<sup>359</sup>

L’ARDC a recensé 3 secteurs clés dans son plan stratégique de quatre ans :<sup>360</sup>

<sup>359</sup> Image tirée de <https://op.europa.eu/en/publication-detail/-/publication/be6c8944-216c-11ea-95ab-01aa75ed71a1/language-en>

<sup>360</sup> Australian Research Data Commons. (2019). ARDC Strategic Plan 2019-2023. <https://ardc.edu.au/wp-content/uploads/2019/05/ARDC-Strategic-Plan-2019-2023.pdf>

## Logiciels et plateformes

- Permettre la recherche par l'utilisation de plateformes et de logiciels de recherche avancés
- Soutenir la création et la tenue d'un écosystème de logiciels et de services *FAIR*
- Transformer les habitudes en encourageant de nouvelles politiques chez les bailleurs de fonds, les éditeurs et les établissements et la création d'une culture de logiciels qui donne une place aux questions de gouvernance, de citation, d'intendance et d'attribution
- L'ARDC exécute un programme d'investissement dans les plateformes. Un des buts explicites du programme sera d'augmenter le nombre des chercheurs et des chercheuses ayant accès aux plateformes, tant en nombre absolu qu'en diversité de disciplines. Le programme cherchera à soutenir une communauté d'opérateurs de plateformes.

## Stockage et calcul

- Tirer le maximum de bénéfices des ressources existantes
- Veiller à ce que l'évolution des ressources informatiques répondent aux besoins des parties prenantes
- Élaborer des mesures fiables de l'incidence
- Élaborer des modèles de financement durable

Rafraîchir le Nectar Research Cloud et l'infrastructure de stockage de l'ARDC qui a atteint sa fin de vie utile est une initiative clé pour se doter des capacités requises pour répondre à la demande future pour :

- héberger les applications de recherche dans un environnement infonuagique souple et modulable
- accéder à une ressource computationnelle qui complète les installations de calcul de haute performance existantes et nouvelles
- déployer rapidement et partager des applications de recherche innovatrices

## Données et services

Deux objectifs stratégiques seront visés sous ce thème :

- Permettre de la recherche nouvelle à partir des données existantes
- Améliorer l'intégrité et la reproductibilité des recherches de l'ensemble du système de recherche en appliquant les principes *FAIR* aux données découlant de la recherche.

Pour le premier objectif, l'ARDC s'associera aux communautés de recherche, aux installations, aux organismes gouvernementaux et aux organismes de recherche pour trouver les collectes de données dont la valeur de réutilisation est élevée, dont l'appropriation par la communauté est forte et qui est de portée nationale. Le rôle de l'ARDC sera de fournir des ressources susceptibles

de faire germer ces collections nationales et communautaires et de développer leur durabilité à long terme.

Pour le second objectif, l'ARDC encouragera la capacité institutionnelle de protection et de gestion des données ainsi qu'une plus grande cohérence des collectes et des services de données entre les secteurs de l'écosystème national de données de recherche. Le rôle de l'ARDC sera de soutenir, faciliter et communiquer les pratiques exemplaires en matière de données, l'interopérabilité, la capacité et la capacité entre les secteurs. Cet objectif sera assorti de la volonté d'aligner l'Australie sur les normes et les initiatives internationales et d'augmenter les données *FAIR*.

Les secteurs prioritaires :

- Collections de données transformationnelles
- Données sensibles et approches, plateformes et services pour gérer, collaborer et partager ces données
- Rôles et approches des établissements dans la réserve commune de données

Services

- Nectar Research Cloud
  - Federated Research Cloud
- Research Data Australia
  - Trouver les données de recherche chez les organisations de recherche, les organismes gouvernementaux et les établissements culturels de l'Australie
- Services d'identification
  - Créer et gérer les ID pour les données
- Research Vocabularies Australia
  - Trouver et utiliser les vocabulaires contrôlés

### *Collaborations*

National Data Assets Program (le programme national des avoirs en données)

- L'ARDC a lancé une initiative appelée « National Data Assets » avec l'objectif d'établir un portefeuille des avoirs en données à l'échelle nationale. Six programmes ont d'abord été proposés, chacun procédant à son propre appel de participation avec ses propres exigences relatives aux manifestations d'intérêt et aux demandes de propositions.
  - Programme national des avoirs en données - NCRIS



- S'associer aux groupes d'installations de la NCRIS (stratégie de l'infrastructure de recherche collaborative) pour établir des collections interconnectées
- National Data Partnerships (partenariats nationaux de données)
  - S'associer à des intendants de données, des établissements, des communautés de recherche et des secteurs public et privé pour établir des avoires nationaux en données
- Pont entre le secteur public et le secteur de la recherche
  - S'associer aux intendants des données du secteur public et des utilisateurs et utilisatrices pour améliorer l'utilisation et la consultation des données du secteur public à des fins de recherche au-delà des activités de base
- Collections émergentes
  - S'associer avec les communautés et les organismes de recherche pour préparer le développement de collections de données nationales émergentes
- Bases institutionnelles
  - S'associer avec les établissements pour stimuler l'adoption des principes *FAIR*
- Avoires australiens en données nationales des études en santé
  - S'associer avec les consortiums des établissements de recherche en santé pour bâtir un avoires nationale en données décentralisées à même les résultats de la recherche nationale financée sur la santé

### *Effectif compétent*

L'ARDC établira des partenariats pour faciliter la collaboration et la coordination sur le développement des compétences, axé sur :

- la progression du changement culturel au moyen de cadres stratégique et de financement
- la planification d'un effectif spécialisé pour le secteur
- le développement de communautés de pratique de base, y compris de connexions avec les communautés et les initiatives internationales.

## Fragments d'initiatives nationales

---

Les services nationaux de données de recherche au Royaume-Uni (R.-U.) sont soutenus par une diversité de bailleurs de fonds et d'organismes du secteur de l'enseignement supérieur engagés dans la recherche nationale où le centre d'intérêt premier est le développement des capacités et le stockage des données en dépôts.

## **UK Research and Innovation (UKRI)** <sup>361</sup> (recherche et innovation R.-U.)

### *Administration*

L'UKRI est un organisme indépendant soutenu par le ministère des entreprises, de l'énergie et de la stratégie industrielle du R.-U. Elle rassemble les sept conseils de recherche du pays et constitue un cadre général d'élaboration de politiques sur les données pour chaque conseil de recherche.

### *Mise en œuvre*

Plusieurs conseils de recherche exploitent des dépôts de données et participent au développement des plateformes et des services en soutien aux données résultant de recherches financées ou de tiers. Le conseil de recherche sur l'environnement naturel exploite un certain nombre de centres de données et élabore une approche de réserve commune, qui fournira à la fois du stockage et de la puissance informatique pour permettre aux chercheurs et aux chercheuses d'amener les données et les calculs aux archives des bases de données.<sup>362</sup> Le conseil de recherche en économie et en sciences sociales soutient le UK Data Service,<sup>363</sup> qui collabore avec un certain nombre d'organismes, dont les organismes et ministères gouvernementaux, pour donner aux utilisateurs et utilisatrices l'accès à une vaste gamme de ressources en données. Il produit des directives détaillées à l'intention des chercheurs et des chercheuses, et participe à des projets dont le but est de faire progresser les questions relatives aux métadonnées, à l'interopérabilité et aux technologies du Web.

## **Jisc** <sup>364</sup> (comité mixte des systèmes d'information)

### *Mandat*

Le Jisc fournit un réseau et une infrastructure numérique partagée aux établissements d'enseignement supérieur du R.-U.

### *Mise en œuvre*

Le Jisc soutient la science ouverte au moyen de plusieurs initiatives, dont la plateforme de recherche ouverte (*Open research hub*). Construite en partenariat avec le secteur de la recherche du R.-U., la plateforme permettra aux chercheurs et chercheuses de gérer les résultats

---

<sup>361</sup> <https://www.ukri.org/>

<sup>362</sup> <https://nerc.ukri.org/>

<sup>363</sup> <https://ukdataservice.ac.uk/>

<sup>364</sup> <https://www.jisc.ac.uk/>

numériques en un seul endroit, offrira de la gestion et du stockage sécurisé, et archivera et préservera les données de recherche.

### Digital Curation Centre (DCC)<sup>365</sup> (centre de protection numérique)

#### *Mandat*

Lancé en 2004, le DCC est né des recommandations énoncées dans la stratégie de préservation numérique et d'accès continu du JISC, en tant qu'organisme national chargé de relever les défis que pose la protection numérique des données de recherche.

#### *Administration*

Les principaux partenaires sont l'université d'Édimbourg, HATII et UKOLN.

#### *Mise en œuvre*

Le DCC prodigue des conseils d'experts et une aide pratique aux organismes de recherche qui veulent stocker, gérer, protéger et partager leurs données de recherches numériques. Il les soutient aussi sur les questions de développement de politiques et de planification de gestion des données et exécute certains programmes de formation pour développer les compétences des chercheurs et des chercheuses et des gestionnaires de données en appui au principes *FAIR*.

## États-Unis

Aux États-Unis, il n'existe pas d'organisation nationale comparable à l'Alliance chargée de soutenir l'infrastructure de recherche numérique ou de fournir des services de gestion de données. Cependant, des organisations soutiennent la coordination de la gestion des données de recherche à l'échelle nationale et pour diverses communautés.

### Coordination des infrastructures/établissements

NIST RDaF<sup>366</sup> (cadre des données de recherche de l'institut national des normes et de la technologie)

#### *Mandat*

La mission de l'institut national des normes et de la technologie (« NIST ») est de promouvoir l'innovation et la compétitivité des É.-U. en faisant avancer la science, les normes et la technologie de la mesure. L'élaboration et l'application de normes est une compétence de base. Un certain nombre de normes en matière de réseaux et de systèmes de données scientifiques, de cybersécurité et de protection des renseignements personnels et de la science de la mesure ont été élaborées dans des domaines particuliers ou généraux.

---

<sup>365</sup> <https://www.dcc.ac.uk/>

<sup>366</sup> <https://www.nist.gov/news-events/events/2019/12/research-data-framework-rdaf-workshop>

### Mise en œuvre

Récemment, la NIST a réuni des parties prenantes de diverses provenances en un groupe de travail chargé d'explorer la création d'un cadre de données de recherche (« RDaF »), consciente qu'il faut assurer une plus grande coordination à l'échelle nationale pour pouvoir soutenir la planification, le lancement et l'exploitation d'une infrastructure de recherche. Le RDaF devrait fournir aux organisations une approche structurée d'élaboration d'une stratégie de gestion des données cohérente en constituant un langage commun et une base de coordination. Le RDaF touchera à tous les aspects des pratiques de gestion des données dans toutes les phases du cycle de vie des données. Il s'adressera en priorité aux établissements et aux organisations (p.ex. DPI, DPD), à la personne ayant la responsabilité des données de recherche dans l'ensemble de son organisation. Le plan de développement à long terme du RDaF prévoit une étude exploratoire, des projets pilotes dans des disciplines ciblées telles que l'astronomie, la science du matériau, l'agriculture ou l'économie, ou dans des communautés de parties prenantes telles que celles des bibliothèques universitaires ou laboratoires de recherche.

### Soutien et coordination de domaines ciblés

NIH New Models of Data Stewardship<sup>367</sup> / Office of Data Science Strategy<sup>368</sup> (nouveaux modèles d'intendance des données des instituts de santé nationaux / bureau de la stratégie en science des données)

### Mandat

Les instituts de santé nationaux (« NIH ») représentent la principale organisation des É.U. à financer la recherche biomédicale et la recherche en santé publique. Le programme des nouveaux modèles d'intendance des données (« NMDS ») a reçu le soutien du fonds commun (« Common Fund ») qui finance les investissements stratégiques dans les domaines émergents dont aucun institut des NIH ne peut s'occuper de manière autonome. Le NMDS a pris fin en 2018 et les initiatives ont été transférées au bureau nouvellement formé de la stratégie en science des données.

### Mise en œuvre

En 2017-2018, le NMDS soutenait deux projets visant à développer et tester de nouvelles stratégies de gestion des données :

1. La phase pilote de la réserve commune de données des NIH, l'objectif était d'explorer de nouvelles manières de stocker, de consulter et de partager des données biomédicales, et des outils associés dans le nuage pour qu'elles soient *FAIR*. Les sujets explorés à cette phase étaient les lignes directrices et les mesures, les identifiants uniques, les espaces de travail partagés pour trouver des données et interagir avec elles, la fonction d'indexage et

---

<sup>367</sup> <https://commonfund.nih.gov/data>

<sup>368</sup> <https://datascience.nih.gov/>

de recherche, l'aide en matière d'éthique en recherche, de protection des renseignements personnels et de sécurité, et l'enseignement et la sensibilisation.

2. L'initiative de l'infrastructure de recherche en science et en technologie pour la découverte, l'expérimentation et la durabilité (« STRIDES ») a mené à des associations avec des fournisseurs de services commerciaux d'infonuagique pour réduire les obstacles économiques et technologiques nuisant à l'accès et au calcul dans de volumineuses bases de données biomédicales afin d'accélérer les progrès dans ce domaine.

Le bureau de la stratégie en science des données soutient des éléments du NMDS, dont l'initiative STRIDES, qui a collaboré avec Google Cloud et Amazon Web Services dans des projets visant à préparer, migrer, charger et calculer des données dans le nuage. Les NIH soutiennent aussi un service d'authentification des chercheurs et des chercheuses, qui devrait lancer à la fin de 2020 un ensemble d'IP qui permettra d'ouvrir une session et d'accéder sans difficulté dans des dépôts de données intégrés de domaines. Les NIH soutiennent un certain nombre de dépôts de données propres à des domaines, mais ils font également l'essai d'un dépôt généraliste sur figshare. Ils ont récemment ouvert deux sources possibles de financement de dépôts de données biomédicales et de bases de connaissances. Enfin, un certain nombre d'initiatives de formation et de sensibilisation visant à encourager l'adoption des outils (p. ex. Fast Healthcare Interoperability Resources) et des pratiques *FAIR* sont menées grâce au bureau.

### Soutien et coordination des services

Plusieurs organisations vouées au soutien du développement et de la coordination des services de données fournis par des établissements de recherche universitaires, en général par leurs bibliothèques, ont vu le jour. Mentionnons le Data Curation Network et le CURE Consortium.

#### Data Curation Network (DCN) <sup>369</sup> (le réseau de curation des données)

##### *Mission*

Le réseau de conservation des données (« DCN ») est la « couche humaine » dans la superposition des dépôts de données et connecte les bases de données locales aux curateurs de données au moyen d'un modèle de recrutement partagé entre les établissements. Il vise à :

- fournir des services experts en conservation des données aux partenaires et aux utilisateurs finaux du réseau,
- créer et partager ouvertement des procédures et des pratiques exemplaires en curation de données,
- soutenir les possibilités de formation et de perfectionnement pour une communauté professionnelle émergente de curateurs de données.

##### *Administration*

Le réseau de curation des données est financé par des subventions de la fondation Alfred P. Sloan et de l'institut des services des musées et des bibliothèques. L'équipe de projet du réseau

---

<sup>369</sup> <https://datacurationnetwork.org/>

est formée de représentants et représentantes de 10 établissements partenaires. Les représentants et représentantes du réseau sont des gestionnaires et des directeurs ou directrices de leurs propres services de curation et assument souvent des responsabilités de supervision pour les conservateurs du réseau qui consacrent des heures de travail au projet.

### *Mise en œuvre*

Les experts et expertes en curation qui se trouvent dans les établissements partenaires partagent leur temps et leur expertise à la curation de bases de données avec d'autres établissements membres. Ainsi, tous les dépôts de données des établissements membres peuvent ensemble et plus efficacement conserver une plus grande variété de types de données (p. ex. discipline, format de fichier, etc.), ce qui dépasse ce qu'un seul établissement pourrait offrir.

## CURE Consortium <sup>370</sup>

### *Mission*

Soutenir la curation des données de recherche et l'examen du code et des objets savants numériques associés dans le but de faciliter la préservation numérique, la réutilisation et la reproductibilité des allégations scientifiques.

### *Mise en œuvre*

Le CURE Consortium développe un réseau d'experts et expertes en curation pour établir les normes, partager les pratiques et promouvoir l'adoption d'une procédure d'examen de la qualité des données. Les établissements universitaires, les organisations et les personnes qui soutiennent et encouragent les principes de la curation à des fins de reproductibilité peuvent adhérer à CURE et ils ou elles s'engagent à mettre en œuvre des flux de travail de recherche transparente et reproductible, si ce n'est déjà fait ou en voie de l'être.

---

<sup>370</sup> <http://cure.web.unc.edu/>

# Annexe C – Associations internationales et nationales de gestion des données de recherche

Le présent document fait un survol des organisations à l'intérieur et à l'extérieur du Canada qui se mobilisent pour soutenir et faire avancer la gestion des données de recherche en développant des communautés de pratique et qui pourraient se mobiliser stratégiquement dans le volet de la gestion des données de l'Alliance pour soutenir la communauté de gestion des données au Canada. Le document montre comment ces organisations font avancer l'état de la gestion des données de recherche et pourraient guider la sensibilisation et la coordination des groupes de travail établis de Portage et de DRC ainsi que des futurs groupes de travail de l'Alliance.



## Research Data Alliance (RDA)

<https://www.rd-alliance.org/>

La RDA a été lancée en 2013 par la Commission européenne, la National Science Foundation et la National Institute for Standards and Technology des États-Unis et l'Australian Department of Innovation. En mars 2020, elle comptait près de 10 000 membres venant de 144 pays. Il s'agit d'une plateforme où des experts internationaux en données de recherche se rencontrent pour échanger des points de vue et faire avancer des sujets liés aux pratiques exemplaires, aux normes et aux protocoles. Elle produit des recommandations, des documents – spécifications, taxonomies ou ontologies, flux de travail, schémas, modèles de données, etc. – approuvés par la RDA.

- Les groupes de travail ont généralement un mandat de courte durée (18 mois). Ils étaient au nombre de 36 en mai 2019.
- Les groupes d'intérêt ont un mandat plus long. On en comptait 66 en mai 2019.
- Il y a deux formes de groupe :
  - Des groupes par domaine (p. ex. chimie, santé, peuples autochtones, agriculture, biodiversité)
  - Des groupes par fonction (p. ex. gestion fédérée des ID, dépôts de domaine, PGD actifs, politique de données, mécanismes de récompense et de crédit dans le processus de partage, vocabulaires, environnements de recherche virtuels)

## RDA North America

<https://www.rd-alliance.org/groups/rda-north-america>

L'objectif de RDA North America est de bâtir des relations entre les membres de RDA et les autres collaborateurs potentiels sur le continent nord-américain.



### World Data Systems / International Programme Office (WDS-IPO)

<https://www.icsu-wds.org/organization/international-programme-office>

WDS est un organe interdisciplinaire du Conseil international des sciences dont la mission est de soutenir l'accès et l'intendance des services de données et des données scientifiques, des produits et de l'information sécurisés. Il soutient le programme CoreTrustSeal.

WDS-IPO coordonne les opérations de WDS sous l'égide du directeur général de WDS et suivant les conseils du comité scientifique. Le Bureau est hébergé et soutenu par l'institut national des technologies de l'information et des communications du Japon.

### World Data Systems / International Technology Office (WDS-ITO)

<https://wds-ito.org/what-we-do>

Le WDS-ITO est basé à l'Université de Victoria, au Canada, et soutenu par trois organisations hôtes canadiennes : Ocean Networks Canada (ONC), le Centre canadien de données astronomiques du CNRC et le Réseau canadien d'information cryosphérique/Catalogue de données polaires de l'Université de Waterloo.

Le Bureau est sous la surveillance du comité scientifique de WDS. Il appuie les organisations membres et les partenaires au moyen d'une infrastructure technique et des services soutenant l'accès aux données scientifiques. Il soutient l'infrastructure nécessaire pour les dépôts avec l'analytique et la visualisation.

- Le comité consultatif technique du Bureau conseille sur les stratégies en matière d'infrastructure, les feuilles de route, les tendances et les activités en matière de technologies dans la communauté mondiale de la GD, il assure la défense des programmes du Bureau.





## International Science Council - Committee on Data (« CODATA »)

<https://codata.org/>

(conseil international de la science – Comité sur les données)

Le but du comité sur les données du Conseil international de la science est de promouvoir la collaboration mondiale afin d'améliorer la disponibilité et la facilité d'utilisation des données pour la recherche, de même que les changements politiques, technologiques et culturels. Il est constitué de comités permanents, de groupes de travail et de groupes d'action. Comparé à RDA, c'est une organisation à l'approche descendante, en lien avec d'autres organisations internationales influentes.

- Les groupes de travail se penchent sur les besoins immédiats à satisfaire à court terme : vocabulaires (glossaire IRIDIUM), formation, modèle d'affaires de dépôt, données *FAIR* et certains sujets propres à un domaine (p. ex. nanomatériau, matériau).
- Les groupes d'action sont formés aux deux ans et couvrent une série de sujets relatifs aux besoins relatifs aux données ou aux problèmes de politique, p. ex. la représentation numérique des unités de mesure, l'amélioration de l'accès et de la réutilisation des données, la science citoyenne, les données ouvertes liées pour la recherche mondiale sur les catastrophes, la préservation des données dans les pays en développement.



CNC/CODATA

<https://codata.org/canada/> (ancien <https://www.codata.info/canada/about.shtml>)

Organisation canadienne membre de l'organisation mère CODATA. Auparavant préoccupée davantage par les normes de mesure expérimentales dans les domaines de la science et de la technologie. Accent mis sur les problèmes communs de gestion des données pour les données utilisées en dehors du domaine dans lequel elles étaient générées. CNC/CODATA était parrainé auparavant par CISTI. Soutenue en ce moment par un engagement de financement de deux ans du CNR pour sa revitalisation. Le comité directeur du CNR revoit les modalités de service. Les listes actuelles du site Web portent sur le développement de la culture de GDR, le soutien des travaux du CNR et de Portage en faveur d'une GDR d'un point de vue national, d'une approche par établissement pour les services et l'infrastructure de GDR, des scientifiques en début de carrière et de la communication scientifique. Le [mandat](#) est mis à jour en ce moment (non encore publié / approuvé par l'hôte CNR).



## GOFAIR

<https://www.go-fair.org/>

Nouvelle initiative lancée par un comité mixte de CODATA, RDA et WDS. Modèle ouvert et inclusif ou personnes et organisations collaborant en faveur des principes *FAIR* par des réseaux de mise en œuvre, autofinancés et autogouvernés et se concentrant sur l'édification de la technologie, de la culture et de la formation dans le cadre d'un internet imaginé de données et de services *FAIR*.

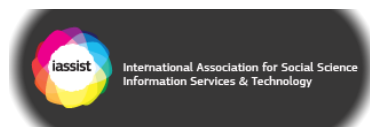


## FAIRsFAIR

<https://www.fairsfair.eu/>

Intervenant dans le développement de normes mondiales pour l'homologation *FAIR* des dépôts et de leurs données. Vise à fournir une plateforme pour l'utilisation et la mise en place des principes *FAIR* dans le travail des fournisseurs et les dépôts de données de recherche européens. Soutient ce qui suit :

- Homologation des dépôts (p. ex. DFDR)
- Formation dans les universités et professionnalisation de l'intendance des données
- Politiques (p. ex. enquête auprès des communautés sur les politiques *FAIR* dans les universités de l'UE, sémantique et interopérabilité)



## International Association for Social Science Information Services & Technology (IASSIST)

(association internationale pour les technologies et les services d'information en sciences sociales)

<https://iassistdata.org/>

Organisation internationale de plus de 300 membres professionnels de l'information soutenant les services de données dans les sciences sociales. Réunit des groupes d'action prêts à exécuter

des tâches spécifiques et solutionner des problèmes, de même que des groupes d'intérêt qui échangent de l'information sur des sujets rejoignant un plus grande nombres d'adhérents.

- Groupes d'intérêt actifs : géospatial, données qualitatives, données de santé
- Les anciens groupes d'intérêt s'intéressaient à la gestion des données au sens large : visualisation des données, outils en code source libre, gestion et curation des données, citations de données



## OPENAIRE

<https://www.openaire.eu/>

Openaire a pour mission de fournir un accès libre et ouvert aux produits de recherche financés par les fonds publics en Europe. Programmes de services : financer les bureaux nationaux de libre accès (« NOAD ») – un réseau de 34 experts dans des établissements en Europe soutenant la science ouverte et l'offre de possibilités de formation.

- Les NOAD et les partenaires forment des groupes de travail pour faire avancer la science ouverte, avec un groupe de travail en GDR
- Réunit une communauté de pratique de coordonnateurs de formation, un réseau informel d'échanges d'expériences et de recensement des activités connexes pour renforcer les capacités en formation.
- Développe des logiciels ouverts de GDR (p. ex. Amnesia, ARGOS) et des tableaux de bord pour statistiques des dépôts et des bailleurs de fonds
- Réunit des communautés de domaine pour recueillir des données ouvertes, des logiciels et des publications en tant que « passerelles communautaires » (p. ex. DARIAH-EU, EPOS).

## FORCE11

<https://www.force11.org/>

- A pour mission d'améliorer les pratiques de recherche en soutenant les innovations dans les manières de créer et de partager le savoir entre les disciplines et les communautés de recherche. S'acquitte de cette mission en connectant les communautés mondiales intéressées par les communications dans la recherche; en construisant un espace engageant pour la discussion et le travail collaboratif; en facilitant le développement de nouvelles approches et de nouveaux outils pour des communications numériques efficaces dans la recherche.
- A plusieurs groupes de travail mixtes avec RDA, notamment FAIRsharing, et d'autres groupes sur l'attribution et la citation de logiciels.

## Confederation of Open Access Repositories (COAR)

(confédération des dépôts en libre accès)

<https://www.coar-repositories.org/>

Association internationale de 157 membres qui regroupent leurs dépôts et leurs réseaux de dépôts pour renforcer les capacités, aligner les politiques et les pratiques et agir comme porte-parole mondial de la communauté des dépôts. Ses programmes couvrent la gestion des données de recherche, les métadonnées et les vocabulaires, la formation et le renforcement des capacités.

- Le groupe d'intérêt de la GDR vise à aider la communauté à étendre ses activités, sert de forum de discussions pour la communauté au sujet de la gestion des données de recherche, des pratiques exemplaires et des stratégies, du renforcement des capacités pour la GDR dans la communauté des dépôts.

## Initiatives alliées

Des initiatives internationales alliées soutiennent les communautés de pratique dans les domaines liés à la gestion des données de recherche, qui pourraient aussi être stratégiquement mobilisées dans le volet de la gestion des données de l'Alliance en soutien à la communauté canadienne de la gestion des données :

### Logiciels/outils



#### Research Software Alliance

<https://www.researchsoft.org/>

ReSA vise à rassembler les communautés et les organisations de logiciels de recherche pour développer une communauté de pratique apte à relever les défis dans la productivité, la qualité, la reproductibilité et la durabilité des logiciels, réaliser leur objectif à long terme partagé de logiciels de recherche appréciés en tant qu'éléments de base vitaux de la recherche.

ReSA dirige un certain nombre de groupes de travail sur les principes *FAIR* pour les logiciels de recherche, la preuve de l'importance des logiciels de recherche, les possibilités de financement et l'analyse du paysage

### Gestion de l'information de recherche



#### Metadata2020\*

<http://www.metadata2020.org/>

Metadata 2020 est une initiative de collaboration qui défend l'importance de métadonnées enrichies, connectées et réutilisables pour les produits de recherche. Réunit des groupes de communautés qui sont chargés de définir les défis, les barrières et les possibilités en lien avec les métadonnées dans leur domaine de communication savante. Les communautés sont les chercheurs, les éditeurs, les bibliothécaires, les éditeurs et les dépôts de données, les services, les plateformes et les outils, et les bailleurs de fonds.



## ORCID-CA

<https://www.crkn-rcdr.ca/fr/orcid-ca-accueil>

ORCID-CA est le consortium ORCID au Canada. Offre aux organisations et aux établissements canadiens une adhésion à ORCID à un coût moindre, ainsi qu'un accès à des services de soutien communautaire.



## DataCite

<https://datacite.org/steering.html>

DataCite est une organisation mondiale de premier plan qui fournit des identifiants permanents (DOI) pour les résultats de recherche. Elle peut compter sur trois groupes directeurs pour l'aider dans son développement : durabilité et entreprise, services et technologie, et communauté et mobilisation. Les groupes directeurs offrent un lieu de participation ouverte des membres de la communauté intéressés et nous aident à définir des stratégies liées à la planification, aux services et à la sensibilisation en matière de durabilité.

## Collaborations de recherche mondiales

Dans divers domaines de recherche, un certain nombre de collaborations mondiales font avancer le soutien et les pratiques exemplaires en GDR. Voici des exemples de liens étroits dans le paysage de recherche canadien :



## Global Research Collaboration for Infectious Disease Preparedness (GloPID-R)

(collaboration mondiale de recherche pour la préparation aux maladies infectieuses)

<https://www.glopid-r.org/>

GloPID-R une alliance qui réunit des organismes de financement de la recherche afin de faciliter une recherche efficace et rapide sur une éclosion importante d'une maladie infectieuse nouvelle ou qui refait surface.

Elle forme des groupes de travail qui s'intéressent à deux volets principaux : l'état de préparation et la réponse. En lien avec la GDR, un groupe de travail sur le partage de données coordonne des initiatives visant à soutenir le partage de données opportun et transparent dans les urgences de santé publique, et il a publié une feuille de route à ce sujet.



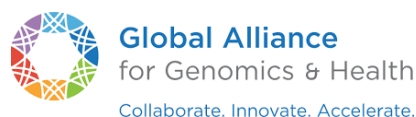
## Global Open Data on Agriculture and Nutrition (GODAN)

(données ouvertes dans le monde sur l'agriculture et la nutrition)

<https://www.godan.info/>

Réseau de plus 1 000 membres de gouvernements nationaux, d'ONG, d'organismes internationaux et du secteur privé. Combinant la défense des données ouvertes et la consultation à leur sujet avec des solutions et des produits innovateurs, GODAN et ses partenaires cherchent à améliorer la sécurité alimentaire pour les prochaines générations afin d'enrayer la faim dans le monde et améliorer la vie et les moyens de subsistance des communautés agricoles dans le monde.

Les partenaires visent à renforcer les politiques générales et le soutien du secteur privé en matière de données ouvertes. Ils encouragent la collaboration et la coopération entre les activités existantes dans les secteurs de l'agriculture, de la nutrition et des données ouvertes et les parties prenantes afin de résoudre des problèmes mondiaux qui remontent à loin.



## Global Alliance for Genomics and Health (GA4GH)

(alliance mondiale pour la génomique et la santé)

<https://www.ga4gh.org/>

Alliance internationale réunissant plus de 500 organisations œuvrant dans les soins de santé, la recherche, la défense des droits des patients, les sciences de la vie et les technologies de l'information. La communauté de GA4GH travaille ensemble pour créer des cadres et des normes permettant le partage responsable, volontaire et sécurisé des données génomiques et liées à la santé.

Les volets de travail de fondation et de technique portent sur les domaines de la réglementation, de l'éthique et de la sécurité des données en génomique. Le volet de travail de fondation englobe

la sécurité des données et la réglementation et l'éthique; le volet de travail technique, la saisie des données cliniques et phénotypiques, le nuage, l'utilisation des données et l'identité des chercheurs, la découverte, les normes du savoir en génomique et la génomique à grande échelle.



International Virtual Observatory Alliance (IVOA)  
(alliance internationale des observatoires virtuels)

<https://ivoa.net/>

Alliance internationale de plus de 20 centres internationaux de données astronomiques formée dans l'objectif de faciliter la coordination des outils, des systèmes et des structures organisationnelles par le développement d'une série partagée de normes. Elle forme des groupes de travail dont le mandat est de faire des recommandations sur les normes et les technologies d'interopérabilité. IVOA peut aussi compter sur des groupes d'intérêt qui discutent d'expériences avec les technologies d'observatoire virtuel en vue de les améliorer.