

Cadre canadien des services de données nationaux

Document de travail

rédigé pour le 2^e *Sommet sur le Cadre des services de données nationaux*
(24 et 25 janvier 2019, Ottawa, ON)

Données de recherche Canada

Groupe de travail sur le CSDN
Comité de l'infrastructure

Élaboré par les intervenants de DRC, notamment :
David Baker, Eugene Barsky, Jane Burpee, Mark Leggott, Jeff Moon, Michael Sinatra, Byron Spencer, Laura Gerlitz

Version 1.0
doi : 10.5281/zenodo.2542158

Le 9 janvier 2019

Table des matières

Contexte et définitions	3
Définitions	3
Situation à l'étranger	6
Situation au Canada	6
Lacunes et possibilités de l'écosystème	7
Financement de la GD par CANARIE en 2018.....	9
Évolution du CSDN canadien	12
Prochaines étapes	14
Ressources	15
Annexe A. Déclaration de Vienne sur l'European Open Science Cloud	16
Annexe B. Diagramme du CSDN	18

Contexte et définitions

Données de recherche Canada (DRC) a commencé d'employer l'expression « cadre des services de données nationaux » (CSDN) en 2016, principalement pour alimenter le débat général sur l'infrastructure de recherche numérique (IRN) nationale et les aspects se rapportant à la gestion des données de recherche (GDR). Pour cela, l'organisme s'était appuyé sur les discussions tenues lors de sommets antérieurs sur le même sujet^{1,2,3}, mais aussi sur les travaux du National Data Services Interest Group (NDSIG)⁴ de la Research Data Alliance (RDA), co-présidé par Mark Leggott, directeur exécutif de DRC, par Christine Kirkpatrick de la NDS américaine et par Adrian Burton, de l'Australian Data Commons. Ce document se veut un document d'information destiné au 2^e Sommet sur le CSDN⁵ et servira de document de travail aux spécialistes en gestion des données scientifiques du Canada.

Définitions

L'expression « cadre des services de données nationaux » a été forgée au 1^{er} Sommet sur le CSDN. Selon l'approche canadienne au développement de services nationaux de données mieux intégrés et plus durables, l'expression suppose les trois éléments que voici :

1. un dialogue entre les intervenants, à tous les niveaux;
2. une entente sur les pratiques exemplaires, les normes et les protocoles;
3. une série de services et de ressources interopérables.

Pareille définition avait pour but de faciliter le débat national sur la meilleure manière d'épauler les membres de l'écosystème de la recherche tout en introduisant les changements d'attitude et de politiques essentiels pour que les données scientifiques deviennent réellement ouvertes, c'est-à-dire trouvables, accessibles, interopérables et réutilisables (principes FAIR – *Findable, Accessible, Interoperable, Reusable*)⁶.

Les discussions subséquentes ont débouché sur l'élaboration d'une définition plus formelle qui a été soumise au CASRAI afin qu'il l'ajoute à son dictionnaire⁷.

Définition abrégée : Service offrant une ou plusieurs fonctions liées aux données aux intervenants et aux domaines concernés dans un contexte national précis.

Définition complète : Service offrant une ou plusieurs fonctions liées aux données aux intervenants et aux domaines concernés dans un contexte national précis au moyen d'un partenariat national entre fournisseurs et agrégateurs de données, groupes

¹ <https://www.rdc-drc.ca/download/background-to-the-2011-canadian-research-data-summit/?wpdmdl=673>

² <https://www.rdc-drc.ca/download/report-of-the-canadian-research-data-summit/?wpdmdl=675>

³ <https://digitalleadership.ca/summit-2017/>

⁴ <https://www.rd-alliance.org/groups/national-data-services.html>

⁵ <https://zenodo.org/record/1035843>

⁶ <https://www.nature.com/articles/sdata201618>

⁷ https://dictionary.casrai.org/Category:Research_Data_Domain

spécifiques, éditeurs et fournisseurs de cyber-infrastructure. Du côté des fonctions liées aux données scientifiques, pareil service s'efforcerait de respecter certains critères comme l'adhésion aux principes FAIR et une accessibilité adéquate pour ceux qui reçoivent ou octroient des fonds publics destinés à la recherche dans le pays concerné. Du côté des fonctions administratives liées aux données, pareil service s'efforcerait aussi de respecter des critères spécifiques au domaine concerné. Par exemple, il s'appuierait sur les efforts déployés dans le pays en question pour garantir une meilleure gestion et un meilleur partage des données, et s'associerait de la façon appropriée à d'autres services, y compris des services internationaux. Les services de ce genre seront souvent regroupés pour former un cadre reposant sur des normes, des politiques d'exploitation ou une combinaison de ces éléments afin de faciliter et d'autoriser une prestation coordonnée des services de données nationaux⁸. [Traduction]

Cette définition a servi de point de départ à un exercice collectif de révision par le NDSIG, à l'occasion de la plénière de la RDA tenue à Berlin. Le texte a connu peu de modifications, une bonne partie de la discussion portant sur l'usage du terme « national ». En effet, ainsi qu'on pouvait s'y attendre, pareil mot facilite et freine à la fois la conversation en ajoutant à celle-ci une dimension politique ou géographique. En conclusion, les participants ont convenu que, si l'usage de « national » n'était pas idéal, il convenait pour encadrer le débat dans un contexte international (et national).

« Infrastructure de recherche numérique » (IRN) est l'expression privilégiée au Canada pour désigner les ressources et les services nécessaires à la recherche. Dans son rapport de 2015 intitulé « *Élaborer la stratégie canadienne sur l'infrastructure de recherche numérique : point de vue de la FCI* », la Fondation canadienne pour l'innovation définit cette infrastructure comme suit :

...l'ensemble des composantes qui sont gérées et exploitées collectivement (comme des installations et des services partagés) par les établissements de recherche et les utilisateurs de partout au pays, celles-ci ne pouvant être prises en charge par un seul établissement en raison de leur taille, de leur complexité et de leur coût considérables. Parmi ces composantes, mentionnons les capacités de calcul, le stockage de données, les services techniques, les logiciels de recherche, les intergiciels, les réseaux optiques à grande vitesse et les capacités de gestion des données de recherche.⁹

L'expression « infrastructure de recherche » (IR) a un sens plus large et est plus usitée en Europe.

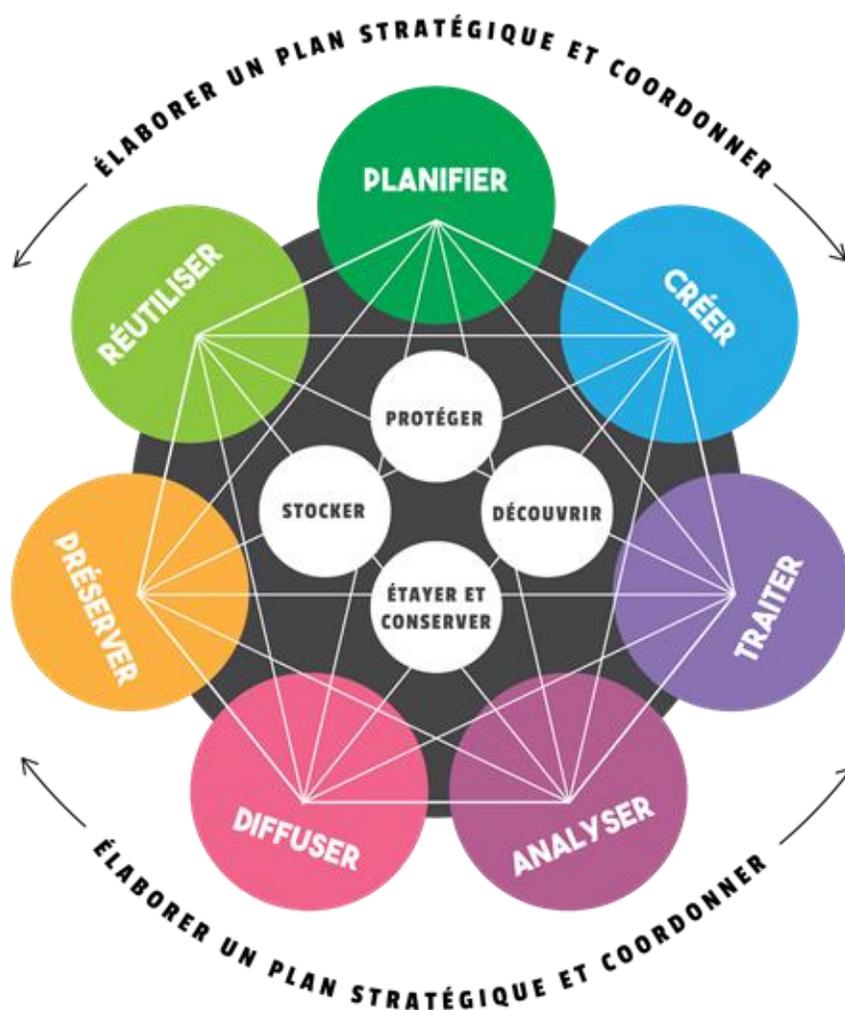
...installations, ressources et services connexes utilisés par les scientifiques pour entreprendre des recherches de pointe dans leur domaine. L'expression englobe les pièces d'équipement et les jeux d'instruments majeurs en science, les ressources

⁸ <https://forum.casrai.org/t/national-data-service/939>

⁹ <https://www.innovation.ca/sites/default/files/Funds/cyber/elaborer-strategie-irn-pour-le-canada-nov6.pdf>

articulées sur le savoir comme les collections, les archives ou les structures d'information scientifique; les infrastructures habilitantes en technologies de l'information et des communications comme les réseaux, les centres de calcul, les logiciels et les systèmes de communication, ou toute autre entité de nature unique indispensable à l'excellence en recherche. Les infrastructures de ce genre peuvent être regroupées dans un même lieu ou être réparties en un réseau organisé de ressources.¹⁰ [Traduction]

Le Research Data Lifecycle (RDL) est un outil visuel couramment employé par les spécialistes en gestion des données de recherche (GDR) pour décrire les étapes fondamentales de la majorité des projets de recherche ainsi que les ressources et les services à la disposition du chercheur à chacune de ces étapes. Le diagramme ci-dessous a été élaboré par le Conseil du leadership sur l'infrastructure de recherche numérique (CLIRN)¹¹ et illustre les étapes de la démarche scientifique (cercles extérieurs), de même que les services qui s'y associent (cercles intérieurs).



¹⁰ https://ec.europa.eu/research/infrastructures/pdf/eric_en.pdf

¹¹ <https://digitalleadership.ca>

Situation à l'étranger

Beaucoup de pays multiplient les efforts pour mettre en place des services de données nationaux (SDN). Certains développent ces services depuis des décennies, alors que d'autres n'en sont qu'à leurs débuts. Un rapport récent du NDSIG¹² décrit les efforts déployés par plusieurs d'entre eux et aboutit à diverses conclusions, parmi lesquelles les suivantes :

1. les SDN varient considérablement d'un pays à l'autre;
2. le plus souvent, les efforts ne sont pas financés assez longtemps, sont sporadiques et ne s'appuient pas sur un modèle de gestion;
3. les efforts déployés pour préciser les SDN, étape importante dans la genèse d'un cadre interopérable pour la science ouverte, se poursuivent.

Bien que le Canada puisse s'inspirer de plusieurs exemples spécifiques à l'étranger, celui de l'European Open Science Cloud (EOSC) ou « nuage européen pour la science ouverte »¹³ est celui qui s'y prête le mieux. On trouvera la description la plus succincte des principes fondamentaux et des engagements à sa base dans le document *EOSC Declaration*¹⁴. La *Vienna Declaration on the European Open Science Cloud* a été dévoilée à l'occasion du lancement de l'EOSC, le 23 novembre¹⁵. Le Canada apprendrait beaucoup de ce projet pour les raisons que voici :

1. il adopte un modèle fédéré s'appuyant sur des ressources et des services existants ou en cours de développement dans les pays membres;
2. il vise un modèle fédéré de gouvernance prenant en compte la situation dans les différents pays;
3. les principes FAIR servent de fondation à tous les efforts déployés.

Les tentatives récentes de l'Australie d'amalgamer les efforts de trois organisations d'IRN auparavant distinctes afin d'en faire l'Australian Research Data Commons (ARDC) ne manquent pas, elles non plus, d'intérêt et sont un cas d'école pour le Canada^{16,17}. Le dossier australien s'avère particulièrement intéressant, car il illustre la métamorphose d'une série d'organisations nationales vers une IRN plus cohérente et durable, ce à quoi fait exactement allusion le budget du gouvernement canadien de 2018.

Situation au Canada

L'idée d'une approche nationale à la gestion des données est sur toutes les lèvres depuis près de dix ans, au Canada. Il en est surtout question lors de rencontres et de sommets nationaux

¹² <https://drive.google.com/open?id=17iUyJ2icY0gFzMZGPWJyY5E0tUoukAtI4BFeronefv4>

¹³ <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

¹⁴ https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf#view=fit&pagemode=none

¹⁵ Voir l'annexe A

¹⁶ https://ardc.edu.au/wp-content/uploads/2018/10/ARDC_Strategic_Overview.pdf

¹⁷ <https://ardc.edu.au/strategic-planning/capital-implementation-plan>

tels ceux mentionnés précédemment, mais aussi dans les rapports de la FCI et du CLINR. Ainsi, le CLINR a rédigé trois documents (sur le calcul de pointe, sur la gestion des données et sur la coordination) à l'intention du ministère de l'Industrie, des Sciences et du Développement économique (ISDE) dans lesquels il recommande qu'on aille de l'avant en élaborant une solide approche nationale à l'IRN. Ces rapports ont été examinés durant le processus conduisant au dévoilement du budget fédéral, lequel annonçait l'injection de 572,5 millions de dollars en cinq ans, plus des fonds annuels, pour qu'on mette en place une IRN. La manière dont on procèdera pour cela est toujours à l'étude. Le document sur la gestion des données remis par le CLINR fait aussi l'objet d'une révision qui débouchera sur un autre document intitulé *Environmental Scan of the DM ecosystem in Canada* (panorama de l'écosystème GD au Canada), qui sera rendu public en 2019. L'objectif à long terme est d'actualiser ce document périodiquement pour qu'il reflète de façon dynamique les efforts en gestion des données déployés au Canada.

Dans ce contexte, les intervenants canadiens ont continué d'alimenter le débat national sur le volet « gestion des données » de l'IRN. Ils le feront notamment lors du 2^e sommet sur le CSDN qui aura lieu en janvier 2019¹⁸, dont le but, ainsi que celui des discussions et des rencontres qui s'y tiendront, consiste à développer un point de vue commun et harmonisé sur un futur et durable écosystème de la GD pour le Canada. Ces travaux sont dynamisés par les possibilités de financement actuelles (programmes de CANARIE¹⁹ et de la FCI²⁰), par les efforts de coordination et de facilitation nationaux (DRC, Portage de l'ABRC), par les activités de divers consortiums et institutions (Cybera, Ontario Library Research Cloud, Calcul Canada), par la multiplication des dépôts de données nationaux (Dataverse North, DFDR), par la création de plateformes de GDR nationales (DMP Assistant), par l'élaboration d'outils et de ressources en formation (modules en ligne des IRSC et de Portage), ainsi que par l'élaboration continue de plateformes de recherche et de dépôts spécialisés (Ocean Networks Canada, CBRAIN, répertoire de données polaires). Toutes ces réalisations visent à rendre l'IRN plus intéressante pour les chercheurs du Canada.

Lacunes et possibilités de l'écosystème

L'exercice de consultation auprès des professionnels de la gestion des données entrepris par CANARIE en juin 2018 figure parmi les tentatives les plus récentes pour identifier les lacunes et les possibilités de la GD. L'idée était d'établir ce que l'organisme devait financer en priorité lorsqu'il lancerait son premier appel à projets en gestion des données de recherche. La consultation a permis d'identifier huit aspects laissant à désirer dans l'écosystème canadien de la GD²¹. Précisons que ces aspects reflètent la situation telle qu'elle existait au Canada au milieu de 2018 et non les lacunes et possibilités de la GD, en général.

¹⁸ <https://www.rdc-drc.ca/fr/activites/esdn/>

¹⁹ <https://www.canarie.ca/DM/>, <https://www.canarie.ca/fr/logiciels/>

²⁰ <https://www.innovation.ca/fr/projets-finances>

²¹ <https://www.canarie.ca/?wpdmdl=14726>

1. Enrichissement et découverte des (méta)données²²
 - a. Les résultats des recherches se faisant de plus en plus variés, trouver des ressources et des chercheurs précis devient de plus en plus difficile, surtout à l'échelon national. On pourrait regrouper les (méta)données existantes, venant des plateformes scientifiques et des étapes de la recherche, ou trouver de nouveaux moyens pour enrichir les métadonnées (par ex., prospection des données, associations, extraction d'entités) afin qu'on découvre plus facilement les résultats des recherches canadiennes à travers un CSDN.
2. Dépôts fédérés / Interopérabilité
 - a. On recense environ 200 « dépôts de données » au Canada et un nombre plus considérable de bases de données devraient contenir des (méta)données scientifiques utiles. Or peu sont accessibles en raison d'une interface de programmation d'applications (API) mal conçue. On pourrait s'inspirer des normes internationales et des pratiques exemplaires pour mieux documenter les dépôts de (méta)données canadiens et en accroître l'accessibilité, ainsi qu'élaborer des approches mieux pensées pour exploiter et relier ces sources à l'échelon national, ce qui en rendrait la découverte et la réutilisation plus faciles entre les différentes disciplines.
3. Dépôts particuliers à un domaine
 - a. Ces dépôts proposent aux chercheurs des sources où déposer et trouver des (méta)données, de manière à mieux répondre aux exigences des organismes subventionnaires ou des diffuseurs, et soutiennent les aspects propres à un domaine, ce qui accroît l'utilité des (méta)données. On pourrait bonifier les dépôts de données du domaine Canada pour qu'ils soient interopérables à l'échelon national et que ces systèmes s'appuient sur des règles de l'art internationalement reconnues.
4. Dépôt et préservation des données
 - a. Les diffuseurs et les bailleurs de fonds adoptent des politiques afin que les chercheurs augmentent l'accessibilité et la réutilisabilité des données à l'origine des résultats qu'ils publient, mais les moyens qui le permettraient sont complexes et peuvent être difficiles à maîtriser pour un chercheur. On pourrait créer des logiciels pour qu'un chercheur puisse identifier et documenter l'approche idéale en gestion des données avec un minimum d'efforts (avec l'aide de services automatisés, par exemple), dans un contexte articulé sur des normes et des pratiques qui fera en sorte que les données adhèrent aux principes FAIR.
5. Conservation
 - a. Déterminer quelles (méta)données de recherche doivent être conservées et pendant combien de temps est parfois un défi. Cependant, il n'en reste pas moins qu'elles ne sont pas conservées, ces données ne pourront être consultées ni réutilisées à long terme. On pourrait créer des outils TI qui faciliteront la prise de telles décisions, d'une manière qui prendrait en compte les pratiques

²² La formule (méta)données n'est qu'une façon commode de combiner les termes métadonnées et données pour plus de concision.

exemplaires et ferait en sorte que les résultats des recherches demeurent accessibles et exploitables durant une période raisonnable.

6. Persistance des identifiants / Possibilité de citation
 - a. Une contrainte à laquelle on ne peut échapper pour rendre les résultats scientifiques accessibles concerne l'adoption des meilleures pratiques reconnues pour rendre les identifiants uniques, autant pour le chercheur que pour son équipe, en raison de la diversité des résultats ou de l'équipement scientifique et des services employés. On pourrait intégrer les plateformes de recherche canadiennes aux meilleures normes opérationnelles pour créer des identifiants persistants qui relieraient plus facilement les différents produits de la recherche et les attributs générés par ceux qui participent à cette activité.
7. Consultation et analyse des données
 - a. Les chercheurs recourent à des approches de plus en plus variées pour analyser et transformer leurs données, que ce soit avec des systèmes locaux ou au moyen de vastes systèmes de calcul de pointe nationaux ou internationaux. On pourrait bâtir des services qui réuniront ces sources de données disparates afin que le chercheur puisse analyser plus aisément les données à toutes les étapes du cycle de recherche.
8. Protection et sécurité des données
 - a. Dans certaines disciplines (par ex., santé humaine, biodiversité, sciences sociales), il est difficile de partager les (méta)données sans recourir à un processus aussi complexe que laborieux qui les rendra anonymes ou en assurera la protection. On pourrait créer des logiciels qui faciliteront ce travail grâce à des méthodes semi-automatiques ou à des approches qui relieront les jeux de données sur les individus sans pour autant divulguer les informations de nature personnelle, sauf si on l'autorise.

Un exemple d'une question importante qui n'a pas été abordée lors des consultations est celle des données sur les Autochtones, qui doivent être traitées différemment, en tant que propriété de la collectivité à laquelle elles se rapportent. On devra tenir compte des *Principes PCAP* (Propriété, Contrôle, Accès, Possession)²³ et des efforts déployés par l'*International Indigenous Data Sovereignty Interest Group* de la RDA quand on s'attaquera au problème des données autochtones dans le contexte canadien²⁴.

Financement de la GD par CANARIE en 2018

Le programme Gestion des données de recherche de CANARIE a débouché sur la sélection de neuf projets qui s'attaqueront à un ou à plusieurs des aspects précités, de même qu'aux principes FAIR et au CSDN. Ces projets ont été dévoilés en novembre 2018 et intègrent une IRN nationale générique (donc pouvant être utilisée par n'importe quel chercheur du pays) ou destinée à un domaine précis (utilisable par les chercheurs d'une seule discipline). Ces projets

²³ <https://fnigc.ca/fr/pcapr.html>

²⁴

<https://static1.squarespace.com/static/58e9b10f9de4bb8d1fb5ebbc/t/5b2973926d2a732f5e620744/1529443237242/2018+IDSov+Global+Progression+-+NAISA+2018+16x9+FINAL.pdf>

seront reportés sur une carte qui en illustrera les points communs, les secteurs d'interopérabilité, les chevauchements au niveau du développement des logiciels et d'autres informations importantes. Des membres de l'équipe de chaque projet se retrouveront à Ottawa le 23 janvier pour discuter de leurs travaux avant d'assister au deuxième sommet sur le CSDN.

1. **Dépôt canadien réparti de données médicales en « omique » (CanDIG CHORD) – piloté par Guillaume Bourque, Université McGill**
 - a. CanDIG est un projet national visant à faciliter l'analyse coopérative des données en génomique humaine disséminées au pays. Il autorisera une intendance complète de ces données ainsi que leur consultation d'une manière à la fois vérifiable et contrôlable. Le projet CHORD engendrera un service national fédéré pour les données de nature délicate en génomique et en santé. Grâce à lui, les membres du secteur canadien de la santé accéderont plus facilement aux technologies et aux services que CanDIG et ses partenaires internationaux de la Global Alliance for Genomics and Health ont entrepris de mettre en place.
2. **Dataverse pour le milieu de la recherche canadien – piloté par Kate Davis, Université de Toronto**
 - a. Dataverse (DV) est une plateforme ouverte de dépôt des données scientifiques élaborée par l'Institute for Quantitative Social Science de l'Université Harvard. De nombreux Canadiens, Américains et Européens ont contribué à mettre au point cette plateforme avant de l'adopter. Au départ, celle-ci était destinée aux chercheurs en sciences sociales détenant des fichiers de petite à moyenne envergure, mais l'architecture de Dataverse sera adaptée afin que le logiciel réponde aux besoins d'un grand nombre de chercheurs du Canada grâce à de meilleures capacités évolutives, au dépôt des fichiers volumineux, à la conservation des flux opérationnels et à son intégration aux services canadiens de stockage et d'authentification.
3. **DuraCloud – relier les dépôts de données aux installations de stockage pour leur préservation – piloté par Corey Davis, Council of Prairie and Pacific Research Libraries, ainsi que par Stephen Marks et Kate Davis, Université de Toronto**
 - a. Les chercheurs canadiens disposent de nombreux services de stockage qui conviennent à la préservation à long terme du contenu numérique, les données de recherche notamment. Le projet DuraCloud réunira plusieurs de ces services grâce au logiciel qu'a développé et que maintient la Fondation DuraSpace. De cette manière, les chercheurs canadiens accéderont à divers services de stockage au moyen de la même interface.
4. **Dépôt FAIR d'annotations, de corpus et de schémas (FRACS) – piloté par André Lapointe, CRIM**
 - a. Les applications misant sur l'intelligence artificielle (IA) doivent accéder à une quantité massive de données. Pour que les chercheurs des universités canadiennes haussent l'échelle de leurs projets IA et puissent concurrencer les applications du secteur privé, il est indispensable de coupler les jeux volumineux de données à des annotations détaillées. En effet, les données annotées facilitent l'entraînement des modèles que les algorithmes d'apprentissage

machine valideront par la suite. Le projet FRACS simplifiera la gestion des jeux de données massifs en facilitant les annotations, leur stockage, leur recherche, leur manipulation et leur partage.

5. **Recherche fédérée de données géospatiales au Canada** – *piloté par Eugene Barsky, Evan Thornberry et Paul Lesack, bibliothèque de l'Université de la Colombie-Britannique*
 - a. Les dépôts de recherche de données se fient depuis toujours à la recherche textuelle. Cependant, on souhaite de plus en plus avoir accès aux éléments géographiques des recherches, le trajet emprunté lors des migrations, par exemple, ou la répartition du rendement des cultures, les images infrarouges des satellites, la distribution des artefacts sur un site archéologique, voire le cheminement de l'eau qui s'écoule. Ce projet créera un logiciel évolutif de source ouverte avec lequel on pourra chercher et découvrir les données géospatiales au Canada grâce à une interface spécialement conçue pour les cartes. De cette manière, l'utilisateur découvrira les ressources géospatiales d'une façon plus intuitive dans l'espace.
6. **Utilité de l'identifiant pour suivre les données dans leur évolution (MINTED)** – *piloté par Reyna Jenkyns, Ocean Networks Canada (ONC), Université de Victoria*
 - a. L'ONC exploite des observatoires océaniques et des services dynamiques de dépôt de données reconnus dans le monde entier. Bien qu'on admette de plus en plus l'utilité et la nécessité de la citation des données, ainsi que le prouve la formulation des principes FAIR, les plateformes et les outils existants ne répondent qu'aux besoins des jeux de données statiques ou rarement actualisés. Le projet MINTED appliquera les pratiques exemplaires de la citation des jeux de données dynamiques, les identifiants d'objet numérique (DOIs) et l'identifiant ORCID des chercheurs à l'infrastructure numérique du réseau Oceans 2.0 d'ONC.
7. **Radium : logiciel de gestion des données actives en recherche** – *piloté par Kevin Schneider, Université de la Saskatchewan*
 - a. Souvent, les données scientifiques susceptibles de servir dans un autre domaine que celui qui les a produites sont disséminées entre de nombreux dispositifs, outils et plateformes de stockage. Savoir qu'un tel jeu de données existe, donc le découvrir, est un exploit en soi. Radium proposera un index des données de recherche au niveau des métadonnées du projet, peu importe la façon dont elles ont été enregistrées. Le logiciel facilitera la recherche et la citation des jeux de données scientifiques en conservant non seulement l'emplacement de ces dernières, mais aussi les métadonnées ordinaires et particulières associées au projet.
8. **Gestion des données de recherche durant leur vie utile avec Islandora** – *piloté par Donald Moses et Rosemary Le Faive, Université de l'Île-du-Prince-Édouard (UIPE)*
 - a. En collaboration avec l'Université Simon Fraser et la Fondation Islandora, l'UIPE bâtira un service de gestion des données de recherche et l'intégrera à la plus récente version d'Islandora appelée CLAW. Islandora est un cadriciel de source ouverte élaboré pour aider les organisations à gérer, découvrir et partager les actifs numériques par la collaboration, en recourant à une approche fondée sur

les pratiques exemplaires et la normalisation. Le projet permettra l'intégration des systèmes d'identification, de métadonnées, d'authentification, de stockage et de diffusion dans l'optique des principes FAIR et du cycle de vie des données de recherche.

9. **Portail sécurisé de recherche de données, de consultation et de collaboration** – piloté par Elizabeth Theriault, Ontario Brain Institute, et par Moyez Dharsee, Indoc Research

- a. L'Ontario Brain Institute (OBI) et Indoc Research ont créé Brain-CODE, une plateforme extensible en neuro-informatique conçue pour gérer la collecte, la préservation, l'analyse et le partage des données de diverse nature sur divers troubles du cerveau. Pour répondre aux besoins en GDR des chercheurs qui étudient les troubles du cerveau et d'autres maladies, l'équipe du projet développera un portail de données sous forme de logiciel avec lequel les chercheurs pourront saisir, chercher et visualiser sans difficulté et de façon sécuritaire les données sur les patients, de même qu'accéder à des sources d'aide et de formation. Le projet satisfera les besoins des équipes qui utilisent Brain-CODE, mais aussi à ceux des institutions collaboratrices et du milieu de la recherche médicale, en général.

Évolution du CSDN canadien

Le modèle illustré ci-dessous s'inspire du travail réalisé par la Commission européenne²⁵ pour donner une feuille de route à l'EOSC. On suggère de s'en servir pour poursuivre la discussion et le développement d'un CSDN canadien. Ce modèle contribuera à encadrer le débat à tous les niveaux et étayera l'élaboration d'une IRN durable pour la GD au Canada. Grâce aux travaux d'organisations comme DRC, Portage de l'ABRC et Calcul Canada, ainsi qu'à divers projets entrepris dans des domaines précis et aux communautés de praticiens, le milieu canadien de la gestion des données profite d'une longueur d'avance. Le débat sur le CSDN et l'adoption d'un modèle semblable à celui présenté ici permettront aux intervenants de la GD canadiens de faire progresser encore plus ces travaux.

²⁵ https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf

Architecture	Architecture nécessaire pour implanter un système fédéré et interopérable de plateformes, d'outils et de ressources en GD auquel tous les chercheurs canadiens auront accès.
Données	Normes et outils communs épousant les principes FAIR dans toutes les différentes et garantissant l'accès aux (méta)données enrichies dans un domaine spécifique.
Services	Services génériques ou spécifiques à un domaine qui appuient directement les chercheurs et les institutions, de même que renforcent les politiques des sources de financement, des diffuseurs et des institutions.
Accès et interface	Interfaces humaines et machine qui facilitent l'archivage et la consultation des données, ainsi que protègent les données personnelles et assurent une sécurité suffisante s'il y a lieu.
Règles	Politiques et ensemble de processus qui facilitent la participation de tous les membres de l'écosystème de la recherche, soutiennent la réglementation et engendrent la confiance.
Gouvernance	Cadre permettant à tous ceux qui concourent au développement et au maintien de l'infrastructure de recherche numérique et de l'écosystème de la recherche canadiens de se faire entendre.

Voici quelques exemples de sujets de conversation ou de projets qui pourraient servir à étoffer davantage le modèle qui précède. Nous les proposons dans le seul but de nourrir le débat au sommet sur le CSDN qui aura lieu en janvier.

1. **Architecture.** Préciser l'architecture des huit plateformes présentées au Sommet sur le CSDN de 2019 et celle des neuf projets financés par CANARIE afin d'en cerner les lacunes et d'établir les étapes qui aboutiront à une fédération nationale de services de GD.
2. **Données.** Identifier les dépôts dans différents domaines qui servent de modèle à des communautés de praticiens spécifiques et déterminer comment les présenter afin qu'ils obtiennent l'attestation *CoreTrustSeal*²⁶.
3. **Services.** Créer un répertoire de services pour que les chercheurs épousent les éléments de la politique sur la GD des trois conseils subventionnaires²⁷.
4. **Accès et interface.** Faire ressortir l'utilité des dépôts de données canadiens (en tant que service de données) qui ont adopté *SmartAPI*²⁸ et partir de quelques cas précis pour illustrer les efforts requis ainsi que l'impact de services de données canadiens plus accessibles.
5. **Règles.** Tracer l'ébauche d'une politique et d'un processus en vue d'intégrer les dépôts de données à un CSDN.

²⁶ <https://www.coretrustseal.org/>

²⁷ http://www.science.gc.ca/eic/site/063.nsf/fra/h_97610.html

²⁸ <https://smart-api.info/>

6. **Gouvernance.** Élaborer une approche à la régie d'un CSDN qui permettra à tous les intervenants de s'exprimer.

Prochaines étapes

Le Canada est loin d'en être à ses débuts en ce qui concerne la création d'un CSDN. En effet, beaucoup de travail a été abattu pour procurer une robuste série de plateformes et de services de recherche aux scientifiques du Canada. Une stratégie progressiste aurait donc pour but de tirer parti de ce qui a été fait pour combler les lacunes et veiller à ce que ces ressources continuent d'être développées et le soient de manière à demeurer longtemps.

Le Sommet sur le CSDN de janvier 2019 est la prochaine étape de cette démarche et les résultats qu'on espère en obtenir sont les suivants :

1. décrire plus en détail le contexte actuel du CSDN canadien en mettant en relief les recoupements et l'interopérabilité des projets;
2. lancer six « projets » pilotes qui entrecouperont le modèle en six parties du CSDN;
3. formuler des recommandations afin d'aller de l'avant puis faire le point au troisième sommet sur le CSDN.

Le Sommet de 2019 présentera huit plateformes et outils de GD bien implantés ou nouveaux sous l'angle des principes FAIR. Les présentations faciliteront le débat sur l'écosystème canadien de la gestion des données et la manière adéquate de progresser afin de satisfaire aux besoins des chercheurs du pays. Ces huit présentations et les projets qui s'y associent sont énumérés ci-dessous.

1. Trouvabilité

- a. DFDR : Dépôt fédéré de données de recherche (Portage/CC)
- b. Découverte de données en épigénomique grâce au portail de l'IHEC (Université McGill)

2. Accessibilité

- a. Secondary Analysis to Generate Evidence, SAGE (PolicyWise)
- b. Réseau canadien des Centres de données de recherche (RCCDR)

3. Interopérabilité

- a. Canadian Integrated Ocean Observation System (CIOOS)
- b. Centre canadien de données en astronomie (CANFAR CCDA)

4. Réutilisabilité

- a. DataStream (Fondation Gordon)
- b. Plateforme de données nationale de la Stratégie de recherche axée sur le patient (SRAP - IRSC)

Le Sommet inclura des discussions en petits groupes qui auront pour but de recommander les mesures subséquentes. Celles-ci seront approfondies par les intervenants au sortir du Sommet ce qui établira comment aller de l'avant. Ce dont on aura discuté à ces rencontres sera résumé

sous forme d'un rapport que l'on diffusera aux membres de la collectivité. Dans le cadre de cette démarche, on croisera la route tracée par ISDE dans les mois à venir. La clarté et l'orientation qui en découleront donneront une véritable longueur d'avance au milieu de la GD.

Ressources

Ressources complémentaires

- [Feuille de route de l'European Open Science Cloud](#) (en anglais)
- [Déclaration de Vienne sur l'EOSC](#) (en anglais)
- [Mapping the Open Science Ecosystem](#)
- [Rapport stratégique de la NLM](#) (en anglais)
- [Plan stratégique des NIH](#) (en anglais)

Annexe A. Déclaration de Vienne sur l'European Open Science Cloud (traduction)

Vienne, le 23 novembre 2018

Nous, ministres, délégués et autres participants, présents au lancement de l'European Open Science Cloud (EOSC) :

1. **rappelons** les difficultés que soulève la recherche articulée sur les données dans la poursuite de l'excellence en sciences, comme on l'avait souligné dans la déclaration sur l'EOSC ratifiée à Bruxelles le 10 juillet 2017;
2. **réaffirmons** le potentiel de l'European Open Science Cloud pour ce qui de transformer le paysage de la recherche en Europe et confirmons que la vision de l'European Open Science Cloud est celle d'un regroupement pérenne de données scientifiques qui englobera la totalité des disciplines dans les États membres;
3. **reconnaissons** que la mise en place de l'European Open Science Cloud est plus un processus qu'un projet, étant donné sa nature itérative reposant sur l'acquisition constante de connaissances ainsi que sur une harmonisation mutuelle, et soulignons la nécessité de poursuivre le dialogue afin d'accroître la confiance et de créer un consensus chez les scientifiques, les bailleurs de fonds, les utilisateurs et les fournisseurs de services;
4. **insistons** sur le fait que l'Europe est bien placée pour prendre la tête dans le développement et l'exploitation de services en nuage pour la science, et répétons que l'European Open Science Cloud sera à la fois européen et universel en s'ouvrant aux partenaires de recherche pertinents de l'étranger au fil du temps;
5. **rappelons** que le Conseil, dans ses conclusions du 29 mai 2018, applaudissait l'application de la feuille de route et du modèle fédéré de l'European Open Science Cloud, et avait convié la Commission ainsi que les États membres à créer un cadre de gouvernance commun qui garantira la participation des intervenants du milieu de la recherche selon des principes de transparence, d'ouverture et d'inclusion, de même d'une participation efficace des États membres;
6. **remarquons** que le Sommet sur l'EOSC de 2018 (qui a eu lieu le 11 juin 2018) réclamait que l'European Open Science Cloud devienne une réalité plus rapidement en évoquant la nécessité d'étoffer le dialogue entre institutions et intervenants, afin qu'on dévoile un nouveau cadre de gouvernance à la rencontre de Vienne, le 23 novembre 2018.

Pour ces motifs, nous :

7. **entendons** maîtriser les nombreuses activités en cours ou prévues de l'UE et des États membres de manière à coopérer pour mettre en place un partenariat inclusif qui verra la naissance de l'European Open Science Cloud, infrastructure fédérée qui bonifiera des services numériques ouverts et fiables, axés sur l'utilisateur, sur le territoire du marché commun numérique;

8. **invitons** les États membres, de même que les intervenants publics et privés d'Europe à appuyer activement ce projet conjoint et la nouvelle structure de régie de l'European Open Science Cloud afin l'initiative soit couronnée de succès;
9. **appelons** l'European Open Science Cloud à dispenser à tous les chercheurs d'Europe l'accès uniforme à une plateforme ouverte par défaut, efficace et multidisciplinaire servant au stockage, à la consultation, à la réutilisation et au traitement des données de recherche qui épousent les principes FAIR;
10. **promettons** de procurer des services de soutien à l'European Open Science Cloud en facilitant la connexion entre les nœuds nationaux et scientifiques pertinents partout en Europe;
11. **réaffirmons** la capacité de l'European Open Science Cloud à alimenter une science de premier ordre, reposant sur l'analyse des données, de même qu'à faciliter l'introduction de nouveaux modèles d'affaire dont la société et l'économie européennes bénéficieront, et **reconnaissons** que de tels services engendreront des possibilités pour les secteurs public et privé, notamment en augmentant la réutilisation de l'information dans le secteur public tout en préservant l'intégrité des données et en garantissant accessibilité et transparence à l'intérieur et en dehors des frontières.

En conséquence de quoi, nous déclarons vouloir collaborer afin que l'European Open Science Cloud réalise son plein potentiel pour le bienfait des citoyens, de la société et de l'économie.

La Déclaration de Vienne sur l'EOSC a été rédigée par Paolo Budroni, Université de Vienne, et Stefan Hanslik, ministère autrichien de l'Éducation, en étroite collaboration avec la Commission européenne. Vienne, 2018

Renseignements

Lancement de l'European Open Science Cloud (EOSC)

Université de Vienne

Bibliothèque, salle de lecture principale

Universitätsring 1

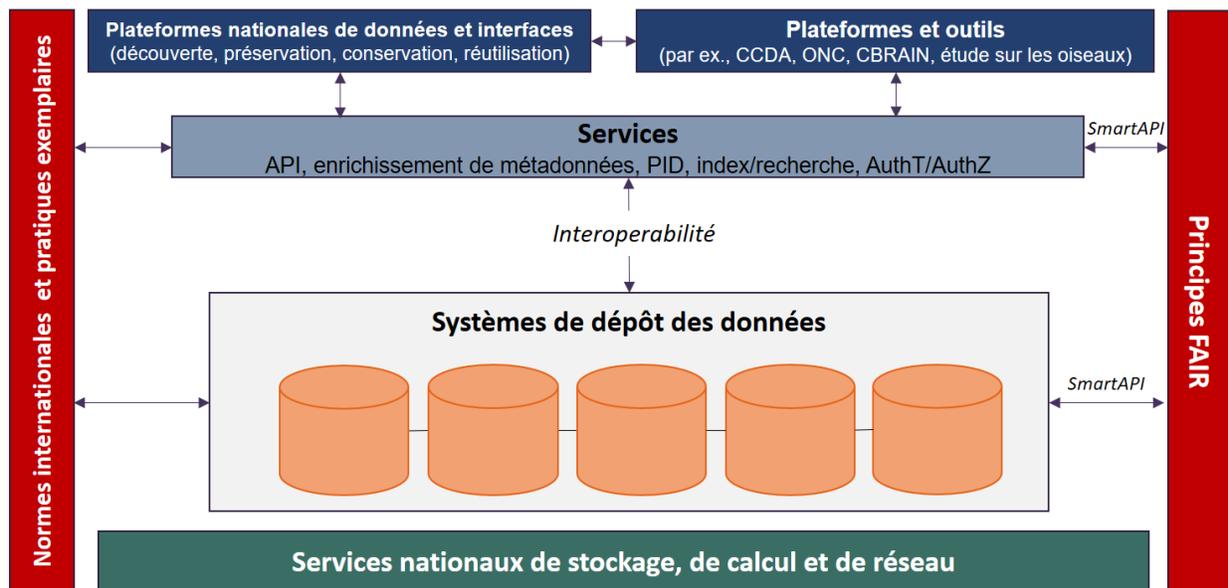
A-1010 Vienne

T: +43-1-4277-17677

congress@univie.ac.at

[Télécharger la version originale](#)

Annexe B. Diagramme du CSDN



Ce diagramme général d'un CSDN figurait dans le document d'information du premier appel à projets du programme GDR de CANARIE. Il montre les diverses couches de l'infrastructure et les services essentiels à un cadre national. Une version plus détaillée du diagramme illustrera les neuf projets financés par CANARIE, en cours de développement, qui opéreront à l'échelon national dans ce contexte.