



Évaluation des besoins de la communauté de recherche: résumé des commentaires reçus

Septembre 2021



Digital Research
Alliance of Canada

Alliance de recherche
numérique du Canada

Auteurs et auteurs :

Felipe Pérez-Jvostov, Karey Iron, Shahira Khair, Seppo Sahrakorpi, Qian Zhang

Collaboratrices et collaborateurs :

⁵Alexandra Cooper (Queen's U.)
⁴Amol Verma (U. de Toronto; Unity Health Toronto)
⁴Anne Martel (Sunnybrook Research Institute)
⁴Benjamin Fung (U. McGill)
⁴Benoît Dupont (U. de Montréal)
²Brian Corrie (Simon Fraser U.)
⁴Bruno Blais (Polytechnique de Montréal)
^{2,3}Catherine Lovekin (U. Mount Allison)
³Chris Loken (Calcul Ontario)
¹Elodie Portales-Casamar (U. de la Colombie-Britannique; Hôpital pour enfants de la Colombie-Britannique)
¹Emmanuel Chateau-Dutier (U. de Montréal)
⁴Erin Dickie (Centre de lutte à la toxicomanie et de santé mentale)
⁴Girma Bitsuamlak (U. Western)
¹Ian Milligan (U. Waterloo)
²James Colliander (U. de la Colombie-Britannique)
¹Jeff Moon (Alliance)
¹Karen Payne (WDS - Bureau international de technologie)
Kestrel McNeill (Calcul Ontario)

⁴Laura Estill (U. St. Francis Xavier)
⁵Laurence Horton (U. de Toronto)
¹Lee Wilson (Alliance)
⁵Lucia Costanzo (Alliance; U. Guelph)
²Marc-Étienne Rousseau (U. McGill)
⁴Marie-Jean Meurs (U. du Québec à Montréal)
^{1,2}Mark Leggott (Alliance)
³Maxime Boissonneault (Calcul Québec)
¹Centre de gouvernance de l'information des Premières Nations
⁴Pekka Sinervo (U. de Toronto)
⁴Philippe Després (U. Laval)
⁴Randall Sobie (U. Victoria)
⁴Rebecca Davis (U. de Manitoba)
⁴Rebecca Pillai Riddell (U. York)
¹Rozita Dara (U. de Guelph)
²Scott Henwood (CANARIE)
^{3,4}Susan Brown (U. de Guelph)
¹Susan Haigh (Association des bibliothèques de recherche du Canada)

1- Groupe de travail sur la gestion des données de recherche
2- Groupe de travail sur les logiciels de recherche
3- Groupe de travail sur le calcul informatique de pointe
4- Conseil des chercheurs de l'Alliance
5- Groupe de travail sur les renseignements de recherche Alliance-Portage



Table des matières

Préambule	2
L'évaluation canadienne des besoins en matière de recherche numérique	2
Sondage.....	2
Exposés de position et documentation actuelle.....	3
Assemblées générales virtuelles.....	3
Constatations	4
1. Sensibilisation et accessibilité	4
Bilinguisme/internationalisation	4
Financement équitable pour le soutien institutionnel	4
Sensibilisation au sujet de l'infrastructure de recherche numérique (IRN) existante et des services offerts aux chercheuses et chercheurs du Canada.....	4
Supprimer les barrières administratives	5
2. Gouvernance et politiques.....	5
Barrières des compétences, législatives et réglementaires au partage des renseignements permettant d'identifier une personne et des données sur la santé	5
Collaboration avec l'industrie et le secteur privé	5
Collaboration et relations de confiance avec les peuples autochtones, et souveraineté des données autochtones	5
Confidentialité des données	6
Mandats de la recherche ouverte	6
3. Opérations.....	6
Calcul informatique de pointe	6
Collaboration avec les trois organismes subventionnaires et les fournisseurs de services (nationaux et régionaux).....	7
Cybersécurité.....	7
Équité, diversité et inclusion (EDI) pour l'ensemble des utilisateurs et utilisatrices de l'IRN, des disciplines de recherche, des établissements et des régions	7
Infonuagique	7
Participation de la communauté	8
Participation internationale	8
Stockage à long terme, curation et préservation des données	8
4. Soutien.....	9
Éducation, formation et maintien en poste du personnel de soutien professionnel.....	9
Soutien par couche : des services nationaux au soutien des établissements	9
Annexe 1	
Annexe 2	
Annexe 3	



Préambule

Partenaire incontournable dans l'application de la stratégie nationale de l'[Infrastructure de recherche numérique](#), l'Alliance de recherche numérique du Canada (l'Alliance) — auparavant nommée la Nouvelle organisation d'infrastructure de recherche numérique (NOIRN) — s'est vu accorder par le gouvernement du Canada une somme de 375 millions de dollars pour coordonner et financer les activités en rapport avec le calcul informatique de pointe (CIP), la gestion des données de recherche (GDR) et les logiciels de recherche (LR). À cet effet, l'Alliance collabore étroitement avec les parties prenantes locales, régionales et nationales pour que les chercheuses et chercheurs canadiens, toutes disciplines confondues, aient accès aux outils, aux services et à l'infrastructure numériques qui leur sont nécessaires pour effectuer des recherches de pointe.

Pour pouvoir fournir un service véritablement centré sur les attentes des chercheuses et des chercheurs, l'Alliance a lancé une vaste consultation dans le but de mieux comprendre les besoins et les défis de la communauté de recherche canadienne dans le paysage de l'IRN au Canada. Le présent document expose les principales constatations de la première [Évaluation canadienne des besoins en matière de recherche numérique](#). Les annexes au présent document les décrivent plus en détail.

L'évaluation canadienne des besoins en matière de recherche numérique

Sondage

Un sondage à l'échelle nationale a été effectué en février et mars 2021. Cette étape de la consultation visait à comprendre comment les chercheuses et chercheurs appartenant à une diversité de disciplines universitaires, de régions géographiques et d'établissements utilisaient l'IRN. Le sondage portait sur des besoins généraux en IRN et sur des aspects techniques de la GDR, des LR et du CIP.

Plus de 1 380 chercheuses et chercheurs ont répondu et donné une rétroaction au sujet de l'utilisation des outils et des services de l'IRN ainsi que sur leurs besoins et priorités en la matière. Les résultats ont montré que les provinces les plus représentées sont l'Ontario (41,8 %) et le Québec (26,5 %), suivies de la Colombie-Britannique (12,5 %) et de l'Alberta (6,8 %). La représentation par discipline varie aussi. Près de la moitié des personnes ayant répondu viennent des sciences et du génie (55,5 %); suivent les chercheuses et chercheurs des sciences sociales et humaines (23,8 %) et, enfin, ceux de la recherche en santé (20,7 %). Le détail des constatations tirées du sondage est présenté à [l'Annexe 1](#).



Exposés de position et documentation actuelle

En octobre 2020, l'Alliance a lancé l'Évaluation canadienne des besoins en matière de recherche numérique en demandant que lui soient envoyés des exposés de position et la documentation actuelle (p. ex. plans stratégiques, feuilles de route et autres documents d'opinion). L'appel a été fait à la grande communauté de recherche : chercheuses et chercheurs individuels, groupes ou centres de recherche, établissements universitaires et autres intervenant(e)s concerné(e)s par l'IRN. Les exposés de position présentent une pluralité de points de vue au sujet de l'état actuel et futur de l'IRN au Canada, accompagnés de recommandations applicables dans le cadre du mandat de l'Alliance. Le détail des constatations tirées des exposés de position est présenté à [l'Annexe 2](#).

La réponse de la communauté de recherche a été remarquable. Plus de 350 autrices et auteurs de 112 organisations ont présenté un total de [107 exposés de position uniques](#) sur une panoplie de sujets : stockage des données et accès aux données, formation et rétention du personnel hautement qualifié (HQP)¹ et du personnel de soutien professionnel, gouvernance des données et politiques, entre autres choses. Certains documents exposent un point de vue d'intérêt pour une région ou une discipline en particulier, alors que de nombreux autres exposent un point de vue de portée canadienne sans référence à une discipline en particulier.

Assemblées générales virtuelles

La consultation s'est terminée par une série de quatre assemblées générales virtuelles sous forme de séances tenues en anglais et en français. Les quatre thèmes abordés correspondaient aux grands thèmes du présent rapport : Sensibilisation et accessibilité, gouvernance et politiques, Opérations et Soutien. Un total de 477 chercheuses et chercheurs, bibliothécaires, administratrices et administrateurs de recherche et des représentant(e)s d'organismes de financement et du gouvernement y ont participé. Chaque assemblée générale a réuni en moyenne 250 participant(e)s. Ce fut pour eux l'occasion de discuter avec le personnel de l'Alliance de leurs besoins et de présenter directement leurs recommandations susceptibles de faire avancer la stratégie nationale de l'Alliance. Le détail des commentaires recueillis lors de ces assemblées est présenté à l'Annexe 3.

¹ Aux fins du présent rapport, la définition donnée à PHQ (personnel hautement qualifié) est celle de [Statistique Canada](#) : « personnes titulaires d'un baccalauréat ou d'un grade universitaire de niveau supérieur ».



Constatations

Les principales constatations de l'évaluation canadienne des besoins en matière de recherche numérique sont présentées sous la forme de quatre grands thèmes se rapportant à l'infrastructure de recherche numérique générale plutôt que sous la forme de renseignements ciblés ou poussés concernant la GDR, les LR ou le CIP. Ces constatations ne représentent pas une liste exhaustive des priorités pour les différentes communautés et disciplines, elles sont plutôt les principaux points identifiés tout au long de la consultation. Une description plus détaillée et structurée est présentée dans les annexes. Les quatre thèmes abordés ici sont volontairement généraux et comprennent une fusion de besoins et de recommandations qui sont présentés par ordre alphabétique.

1. Sensibilisation et accessibilité

Bilinguisme/internationalisation

Il est essentiel d'assurer le soutien bilingue dans tous les aspects de l'IRN. La mise en œuvre de services et de soutien en français devrait s'étendre à la conception de l'interface et à la formation.

Le développement des ressources et des services de l'IRN devrait être bilingue dès l'étape de la conception (c'est-à-dire dès le début) pour s'assurer d'une qualité et d'un message uniformes.

Financement équitable pour le soutien institutionnel

Des différences institutionnelles importantes, au niveau des organismes de recherche et des établissements d'enseignement supérieur, font obstacle à un accès national équitable aux fonds et à l'aide à la recherche. Des volets de financement additionnels devraient permettre aux établissements d'embaucher du personnel de soutien professionnel et de renforcer leur prestation de services à l'échelle locale.

Sensibilisation au sujet de l'infrastructure de recherche numérique (IRN) existante et des services offerts aux chercheuses et chercheurs du Canada

L'IRN canadienne est un système complexe autour duquel gravitent de nombreux intervenant(e)s dont les services et les mandats s'entrecoupent. Les ressources et les services ne sont pas suffisamment bien décrits et mis en valeur, ce qui crée de la confusion et de l'incertitude parmi les chercheuses et chercheurs. Un catalogue national des ressources d'IRN disponibles pourrait leur permettre de trouver rapidement la solution susceptible de répondre à leurs besoins.

L'utilisation des ressources d'IRN est très inégale d'une discipline à l'autre. Un effort spécial de sensibilisation et de mobilisation auprès des disciplines qui n'ont adopté ces outils que récemment devrait être fait (p. ex. sciences sociales et humaines) de sorte que leurs besoins soient satisfaits.



Supprimer les barrières administratives

Les processus permettant actuellement d'accéder aux fonds sont fastidieux et demandent du temps. Le fardeau administratif imposé aux chercheuses et chercheurs pourrait être allégé et les processus pourraient être facilités avec la mise en place d'un processus national et simplifié grâce à la collaboration des intervenant(e)s (p. ex. organismes de financement et fournisseurs de services) et à l'utilisation de plateformes communes, de systèmes à authentification unique et d'identifiants pérennes.

2. Gouvernance et politiques

Barrières des compétences, législatives et réglementaires au partage des renseignements permettant d'identifier une personne et des données sur la santé

L'absence de dossiers administratifs de mise en corrélation entre les sources de données fédérales et provinciales est une barrière importante pour la recherche en santé. Il est crucial de réfléchir à des solutions qui faciliteraient l'analyse, le stockage et le partage de données sensibles en conformité avec les règlements provinciaux et fédéraux pour faire avancer une stratégie d'IRN nationale et prendre avantage de la masse de données administratives et de santé disponibles au Canada.

L'accès à des infrastructures de calcul et de stockage gouvernées de manière appropriée en conformité avec la *Loi sur la protection des renseignements personnels sur la santé* devrait faciliter les collaborations entre les établissements au Canada et à l'international.

Des investissements dans des modèles de gouvernance de la sécurité, dans la surveillance et la validation régulière de l'infrastructure permettraient aux chercheuses et chercheurs canadiens de participer à des collaborations internationales; par exemple, les données qui sont gouvernées par le *Règlement général sur la protection des données* (RGPD).

Collaboration avec l'industrie et le secteur privé

Les collaborations entre l'industrie et les universitaires sont d'importants leviers d'innovation dans de nombreuses disciplines. Il est important de favoriser de telles interactions sans compromettre la propriété intellectuelle (PI), l'accès, l'utilisation et la réutilisation des résultats de la recherche et pour promouvoir la confiance et la reddition de comptes. De la même manière, garantir que les intérêts privés respectent les conditions normales de concurrence dans la prise des décisions relatives à l'IRN financée par des fonds publics devrait être une priorité dans les collaborations avec l'industrie.

Collaboration et relations de confiance avec les peuples autochtones, et souveraineté des données autochtones

Les Premières Nations, les Métis, les Inuits et les autres peuples autochtones sont les propriétaires de leurs données à titre de titulaires de droits ancestraux et de partenaires



importants dans l'écosystème de l'IRN. Il y aurait lieu de prévoir une réserve à même les fonds alloués à l'IRN pour le soutien de la recherche dirigée par les autochtones ou de la recherche menée en collaboration avec eux, à intégrer dans une stratégie d'IRN nationale sans porter atteinte à la souveraineté des données.

La reconnaissance de la valeur du savoir, de la vision, des langues et de la culture générale des autochtones devrait être priorisée. Il y aurait lieu d'encourager la collaboration entre l'Alliance, les peuples autochtones, les divers paliers de gouvernement et les établissements universitaires dans un souci de préservation à long terme pour les générations futures.

Confidentialité des données

La protection des données personnelles et sensibles est essentielle dans de nombreuses disciplines. Un cadre national de politiques, de gouvernance et de l'IRN garantissant la confidentialité par des pratiques de sécurité et d'authentification plus serrées dans les établissements simplifierait les pipelines et augmenterait l'efficacité de tels flux de travail.

Mandats de la recherche ouverte

Comme il est mentionné dans la publication [Feuille de route pour la science ouverte](#), le partage des produits générés par de la recherche financée par des fonds publics demande de la structuration et de la coordination à l'échelle nationale. Les mandats et les politiques de la recherche ouverte devraient être développés en collaboration avec des parties prenantes nationales pour s'assurer que les exigences en matière de conformité sont en phase avec la prestation de services de tous les niveaux. Alors que toutes les données ne peuvent, ou ne devraient, pas être partagées librement (p. ex. potentiel commercial ou renseignements permettant d'identifier une personne), des mécanismes appropriés pour que les données restent consultables de manière sécuritaire devraient être mis en place.

3. Opérations

Calcul informatique de pointe

La complexité et la taille grandissantes des bases de données et des flux de travail informatiques ont accru la demande pour des systèmes informatiques comprenant une unité centrale de traitement (« CPU ») et une unité de traitement graphique (« GPU »), pour des calculs parallèles à grande échelle et des environnements sécurisés pour l'analyse de données sensibles. Pour certaines disciplines, cette infrastructure de calcul doit permettre le traitement en temps réel et dans certains cas 24/24, 7/7. De plus, on prévoit que les besoins de calcul et de stockage continueront d'augmenter bien au-delà des capacités actuelles. L'Alliance devrait continuer de collaborer avec des fournisseurs de services et de solutions de domaines spécifiques qui ciblent et fournissent des systèmes et des solutions à grande échelle adaptés aux besoins des disciplines.



Pour certaines disciplines, cependant, l'amélioration du CIP ne signifie pas nécessairement une plus grande utilisation d'informatique CPU ou GPU, mais plutôt des flux de travail nouveaux et améliorés et des structures de stockage, de logiciels et de calcul complexes adaptées et intégrées. C'est particulièrement le cas dans les sciences humaines numériques où les flux de travail incluent souvent l'hébergement web d'applications spécialisées, le développement de logiciels et les services de préservation de données. De tels flux de travail, plateformes et développement d'applications dépendent grandement du soutien de PHQ et du personnel de soutien professionnel sur place.

Collaboration avec les trois organismes subventionnaires et les fournisseurs de services (nationaux et régionaux)

Le manque de concordance entre le financement des trois organismes subventionnaires et la disponibilité des ressources de calcul pour mener à bien les projets financés entraîne des pertes d'efficacité et retarde les avancées. Il faut absolument s'aligner sur les trois organismes subventionnaires et s'assurer que l'IRN nécessaire est mise à disposition de manière synchronisée avec les allocations de fonds.

Cybersécurité

Il faut affecter plus de personnel à la sécurité informatique et éduquer les chercheuses et chercheurs sur la cybersécurité pour protéger la propriété intellectuelle des chercheuses et chercheurs. Ces derniers ont besoin de plus de soutien pour assurer la sécurité de l'accès aux données sensibles et de leurs transferts. Ils doivent pouvoir accéder aux ressources en matière de sécurité et connaître les pratiques exemplaires de sécurisation des données de recherche.

Équité, diversité et inclusion (EDI) pour l'ensemble des utilisateurs et utilisatrices de l'IRN, des disciplines de recherche, des établissements et des régions

L'EDI accuse un retard dans les nombreux niveaux de l'écosystème de l'IRN national. De la faible représentation des minorités visibles, des femmes et de la 2ELGBTQIA+, à la sous-utilisation de l'infrastructure du CIP par de nombreuses disciplines, il y aurait lieu de donner un accès équitable à l'IRN et de promouvoir l'inclusion et la diversité au sein de la communauté des utilisatrices et utilisateurs.

Infonuagique

En raison de la baisse des coûts et de leur facilité d'utilisation, les services d'infonuagique commerciaux sont devenus populaires auprès des chercheuses et chercheurs universitaires. Peut-être est-ce à cause du manque de services de la part de l'infrastructure nationale et donc de la nécessité d'améliorer l'accès à faible coût aux ressources informatiques à distance, au stockage, à la curation et à l'analyse des données et à l'informatique en rafale? Soutenir l'accès à des solutions d'infonuagique, par la voie de plateformes commerciales ou nationales, devrait être accompagné d'une meilleure intégration entre les ressources informatiques des établissements et les équipes de soutien national. Un écosystème national d'IRN mieux intégré



facilitera pour les chercheuses et chercheurs universitaires l'accès en temps opportun aux ressources dont ils ont besoin.

Participation de la communauté

L'incertitude de la communauté à propos de l'Alliance et son futur rôle dans l'écosystème de l'IRN national, offre l'occasion de renouveler le dialogue entre les fournisseurs de services, les bailleurs de fonds, les établissements universitaires et la grande communauté de recherche. Une attention particulière devrait être mise sur le développement de partenariats avec les Premières Nations, les Inuits et les Métis, et tout autre groupe historiquement sous-représenté. Procéder à des consultations sur une base récurrente pourrait satisfaire et conforter les utilisatrices et les utilisateurs tout en garantissant que les services offerts répondent aux besoins de la communauté.

Participation internationale

La collaboration à l'international est au centre de nombreuses disciplines. Pour être concurrentielle à l'échelle mondiale, l'IRN nationale devrait pouvoir s'intégrer à des systèmes internationaux et permettre la coopération internationale par l'adoption de normes et de politiques communes dans un souci d'interopérabilité.

Stockage à long terme, curation et préservation des données

Les sources courantes de financement fournissent rarement de l'aide au-delà de la période des activités liées au projet de recherche, ce qui exclut l'aide requise pour la préservation et le stockage à long terme. Ce manque de fonds pour le soutien des aspects de base de la curation des données est un défi constant dans toutes les disciplines de recherche.

Les difficultés que pose l'interopérabilité des bases de données (p. ex. couplage des données, ontologies communes) et des plateformes sont répandues, particulièrement dans les sciences de la santé et de la vie ainsi que dans les disciplines utilisant des données quantitatives, pour lesquelles des normes consensuelles font souvent défaut. De la même manière, l'absence de solutions sûres et normalisées pour la gestion des données sensibles est un défi connexe qui demande une solution fédérative à cause des multiples règlements relatifs à la protection des renseignements personnels.

Aider les chercheuses et les chercheurs pour la gestion des données tôt dans leurs projets (p. ex. élaboration des plans de gestion des données, mise en œuvre des normes régissant les métadonnées et amélioration des pratiques de stockage) améliorera sur le long terme les perspectives pour la préservation et la réutilisation des données. Des volets de financement pour l'infrastructure et le personnel de soutien professionnel sont nécessaires pour que les chercheuses et chercheurs puissent effectuer leurs travaux. Dans l'écosystème, le soutien national et la coordination des plateformes de préservation et de dépôt de données, ainsi que les logiciels de base et les systèmes de stockage sont nécessaires pour la gestion à long terme et l'optimisation de l'utilité et de l'incidence des fonds consacrés à la recherche.



Il faut une infrastructure de stockage permettant de traiter les données sensibles et pour que les chercheuses et chercheurs respectent les exigences de leurs partenaires de recherche, de leurs communautés et des établissements.

4. Soutien

Éducation, formation et maintien en poste du personnel de soutien professionnel

Le personnel de soutien professionnel est un élément important de l'écosystème de l'IRN national, qu'il soit local ou national. Il se retrouve dans les services de bibliothèque et d'informatique des établissements, ainsi que dans les équipes techniques d'organismes nationaux. Les chercheuses et chercheurs comptent sur des curateurs de données, des analystes et des archivistes, des experts en métadonnées, des développeurs de logiciels et des administrateurs système pour le soutien et la gestion du matériel de base. Assurer la viabilité à long terme du personnel de soutien professionnel devrait être une priorité inscrite dans une stratégie nationale en matière d'IRN. Des volets de financement spéciaux pourraient aider les établissements et les équipes de recherche à maintenir leur personnel en poste et à offrir la sécurité d'emploi.

Soutien par couche : des services nationaux au soutien des établissements

Le recours croissant à l'IRN par les chercheuses et chercheurs pour les flux de travail universitaires doit être accompagné d'un soutien institutionnel et national. Il faut donc plus de personnel de soutien professionnel pour aider les chercheuses et chercheurs à utiliser l'IRN. Ce personnel devrait être bien intégré dans les différentes couches de soutien, notamment les établissements et l'équipe de soutien régional ou national. Il est primordial de pouvoir compter sur des moyens bilingues et un catalogue de compétences diverses. Il y aurait aussi lieu d'encourager auprès des étudiant(e)s du premier cycle et des cycles supérieurs une formation de PHQ avec, à la clé, des plans de carrière clairs et des rôles à jouer dans l'IRN national.

Une approche de formation par couche pourrait demander que toutes les cohortes soient initiées à l'IRN, y compris les étudiants des cycles supérieurs, et de toutes les disciplines; une formation personnalisée pourrait améliorer les flux de travail informatiques des chercheuses et chercheurs et l'utilisation générale de l'IRN.



Annexe 1

Résultats du sondage mené par l'Alliance auprès des chercheuses et chercheurs dans le cadre de l'évaluation des besoins

Table des matières

1.	Contexte	1
2.	Population des répondant(e)s.....	1
2.1.	Identité sociale.....	3
3.	Utilisation de l'infrastructure de recherche numérique	4
4.	Soutien et formation.....	11
5.	Besoins et priorités en matière d'IRN	14
5.1	Besoins	14
5.2	Priorités.....	15
6.	Recommandations	18



1. Contexte

Partenaire incontournable dans l'application de la stratégie nationale de l'[Infrastructure de recherche numérique](#), l'Alliance de recherche numérique du Canada (l'Alliance) s'est vu accorder par le gouvernement du Canada une somme de 375 millions de dollars pour coordonner et financer les activités en rapport avec le calcul informatique de pointe (CIP), la gestion des données de recherche (GDR) et les logiciels de recherche (LR). Pour pouvoir fournir les ressources et les services dont la communauté de recherche canadienne a besoin pour mener à bien ses travaux, l'Alliance a lancé une vaste consultation en octobre 2020. L'évaluation des besoins des chercheuses et chercheurs a consisté à solliciter des exposés de position, à effectuer un sondage et à tenir une série d'assemblées générales. La présente annexe décrit les résultats du sondage.

L'objectif du sondage était de recueillir de l'information pouvant permettre de comprendre comment les chercheuses et chercheurs de différentes disciplines universitaires, de différentes régions et de différents établissements utilisaient l'IRN. Le questionnaire comportait deux sections : les aspects généraux de l'utilisation de l'IRN et les aspects techniques de la GDR, des LR et du CIP, le tout en lien avec le cycle de vie de la recherche. Les questions portaient sur la collecte, le stockage, le calcul et l'analyse des données et la mobilisation du savoir. Des questions relatives à des besoins spécifiques, telles que le financement et la formation, ont été ajoutées. Le présent rapport fournit des informations sur la première section du sondage, c'est-à-dire les aspects généraux de l'utilisation de l'IRN par les chercheuses et chercheurs canadiens.

2. Population des répondant(e)s

Plus de 1 380 chercheuses et chercheurs ont répondu au sondage et donné de la rétroaction au sujet de l'utilisation des outils et des services de l'IRN ainsi que sur leurs besoins et priorités en la matière. Les provinces les plus représentées ont été l'Ontario (41,8 %) et le Québec (26,5 %), suivies de la Colombie-Britannique (12,5 %) et de l'Alberta (6,8 %). Personne des Territoires n'a répondu. La plupart des personnes qui ont répondu étaient affiliées à une université (96,1 %), quelques personnes venaient de collèges ou cégeps (1,1 %) et de centres et instituts de recherche (1,3 %). Les autres répondant(e)s (1,3 %) appartenaient à des organismes gouvernementaux et à des organisations de l'IRN. La répartition des répondant(e)s par établissement universitaire est présentée dans le tableau 1.

La représentation par discipline était variable. Plus de la moitié des répondant(e)s (55,5 %) venaient des Sciences et du Génie. Les autres venaient dans une moindre proportion des Sciences sociales et humaines (23,8 %) et de la Recherche en santé (20,7 %) (figure 1). La majorité des répondant(e)s étaient des chercheuses et chercheurs du corps professoral (62,1 %),



dont 24,4 % étaient professeurs à temps plein, la proportion diminuant avec le rang universitaire. Les étudiant(e)s des cycles supérieurs étaient les plus représenté(e)s dans le groupe des chercheuses et chercheurs en dehors du groupe professoral (15,4 %) comme le montre la figure 2.

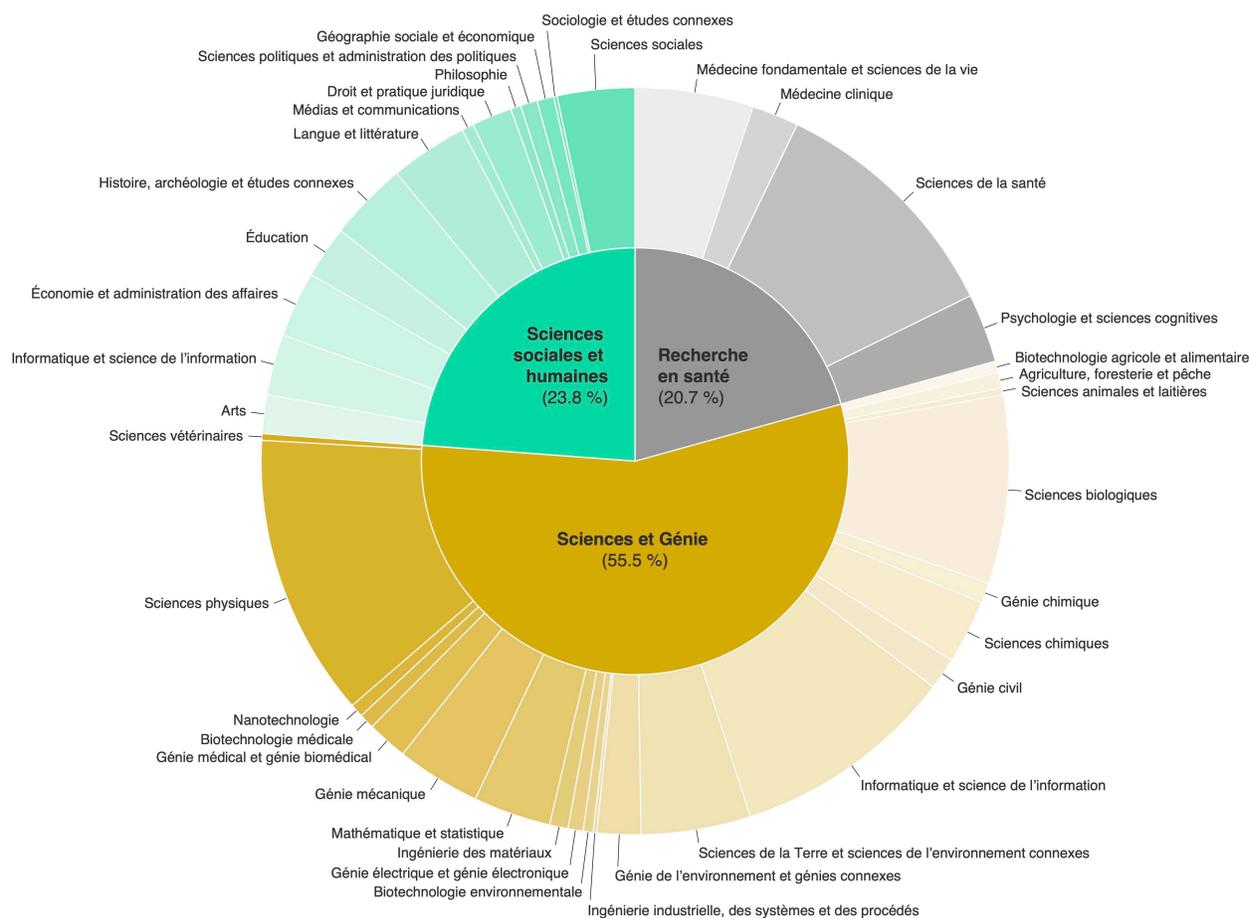


Figure 1. Répondants et répondantes par discipline universitaire



2.1. Identité sociale

Plus de la moitié des répondants (57,8 %) ont répondu qu'ils étaient de genre masculin; 34,3 %, de genre féminin, et le reste, soit moins de 2 %, se sont définis comme étant transgenres ou de genre mixte. Près de 6 % ont préféré ne pas s'identifier à un genre. La vaste majorité ne s'est pas définie comme membre d'un groupe racisé; 18 % l'ont fait et plus de 7 % ont préféré taire leur origine ethnique.

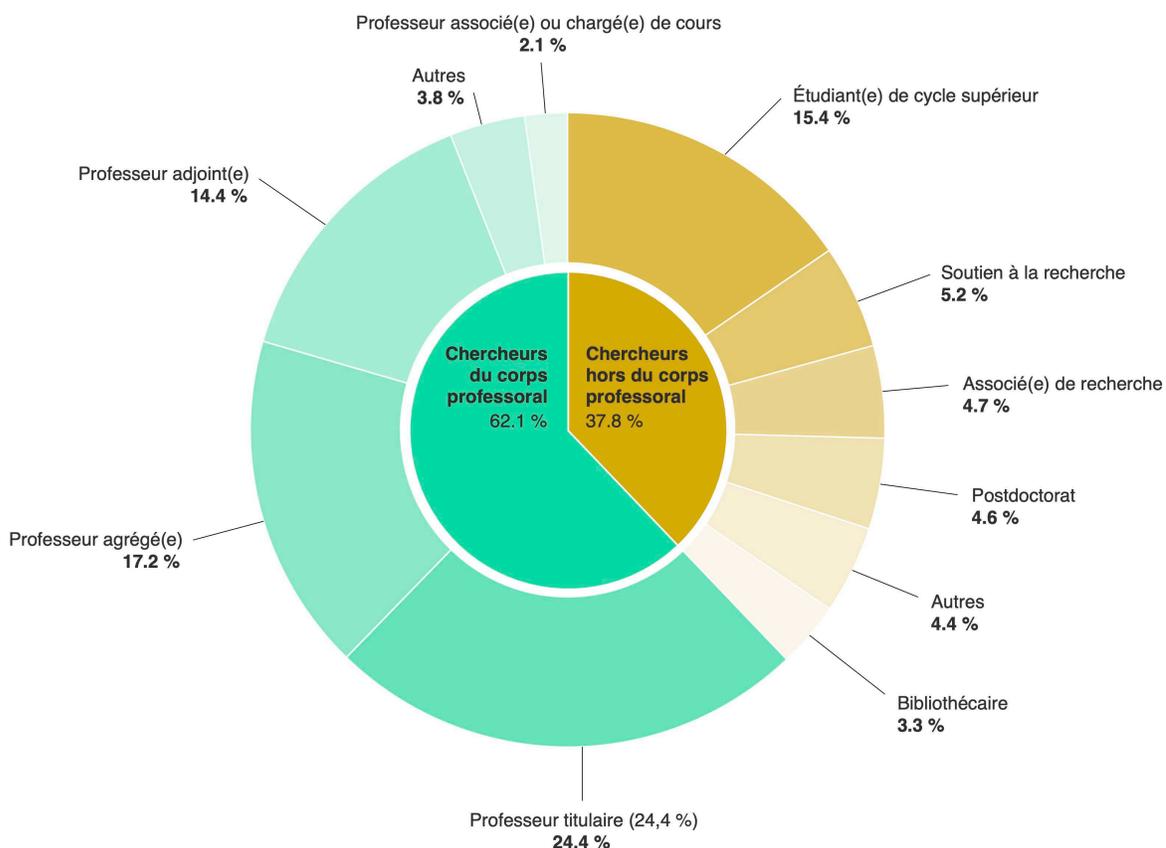


Figure 2. Répondants et répondantes par rang et poste



3. Utilisation de l'infrastructure de recherche numérique

Une série de questions portaient sur l'accès et l'utilisation de l'IRN, que ce soit via des plateformes nationales ou des fournisseurs de services ou via les établissements. Ces questions, de nature générale, couvraient divers thèmes : gestion des données, logiciels de recherche, infrastructure de calcul, stockage, formation et soutien technique.

Sur la question de savoir quelles activités étaient le plus fréquemment entreprises dans le cadre des flux de travail de recherche, trois activités, à savoir utiliser des outils *open source*, documenter la collecte, le nettoyage et l'analyse des données, et partager les données (figure 3), se retrouvaient dans les trois disciplines de la recherche en santé (RS), des sciences sociales et humaines (SSH) et des sciences et du génie (SG)². Des différences importantes sont cependant apparues. EN SG, 80,2 % des répondant(e)s écrivent le code/le logiciel, 67,3 % testent le code/le logiciel, 70,6 % utilisent les interfaces de ligne de commande et 56,5 % utilisent le calcul de haute performance; ces activités sont beaucoup moins courantes tant en SSH (respectivement 35,5 %; 24,4 %; 24,8 %; 10,6 %) qu'en RS (respectivement 55 %; 39,1 %; 40 %; 29,5 %). De même, Utiliser les documents sources numérisés (62,5 %), mobiliser les connaissances en ligne (58,4 %) et préserver les données et autres objets numériques (56,1 %) étaient des activités beaucoup plus courantes en SSH qu'en SG (respectivement 12,1 %; 15,8 %; 33,7 %) et en RS (respectivement 19,1 %; 21,2 %; 43,3 %). C'est au chapitre de la conservation des données sensibles que les répondant(e)s de la RS se sont le plus différencié(e)s, dans une proportion de 53,3 %, par rapport aux autres disciplines.

Les chercheuses et chercheurs de toutes les disciplines ont le plus souvent mentionné leur propre laboratoire ou ordinateur personnel comme étant leur principale infrastructure de calcul (figure 4). Alors que 70,9 % des répondant(e)s en SG ont aussi répondu qu'ils utilisaient l'infrastructure de la Fédération Calcul Canada, ils sont beaucoup moins en RS (35,9 %) et en SSH (18,4 %) à utiliser cette plateforme nationale. Cette plus grande utilisation des plates-formes nationales CIP à travers le SG peut être interprétée dans le contexte historique où les flux de travaux CIP ont nécessité l'utilisation de la ligne de commande et la rédaction de logiciel, activités beaucoup moins présentes en SSH et RS. Des différences entre les disciplines de recherche se sont aussi confirmées dans le recours à l'informatique des établissements, près de la moitié des répondant(e)s en SSH (47,4 %) et en RS (50,6 %) utilisant cette infrastructure comparé à seulement 23,7 % en SG.

² Par souci de clarté, nous utilisons les trois disciplines de recherche correspondant au mandat des trois organismes subventionnaires. Cette approche, qui peut être vue comme une simplification, permet néanmoins de définir les grands besoins de l'IRN au Canada.



Nombre de chercheuses et chercheurs de toutes les disciplines ont confirmé avoir accès à du personnel de soutien informatique professionnel dans leur institution d'origine (RS= 59,8 %; SG= 64,1 %; SSH= 66,9 %), ce qui souligne l'importance du personnel de soutien en place dans les établissements aux yeux des chercheuses et chercheurs de toutes les disciplines universitaires, en particulier parce que la plupart des répondant(e)s n'ont pas de personnel de soutien informatique professionnel au sein de leurs groupes de recherche (figure 5). Conséquemment, un tiers des répondant(e)s ont mentionné qu'ils ou elles n'avaient pas accès au soutien informatique dans leur établissement ou qu'ils ou elles n'étaient pas au courant de leur existence. Enfin, la moitié des personnes en SG qui ont répondu ont aussi confirmé qu'elles avaient accès à du personnel de soutien informatique professionnel par l'entremise de services nationaux. Il s'agit plutôt de l'exception qui pourrait montrer l'utilisation active qu'elles font de l'infrastructure de la Fédération Calcul Canada, un service national. La plupart des répondant(e)s ont dit ne pas être au courant de l'existence d'un tel service ou ne pas y avoir accès (figure 5).

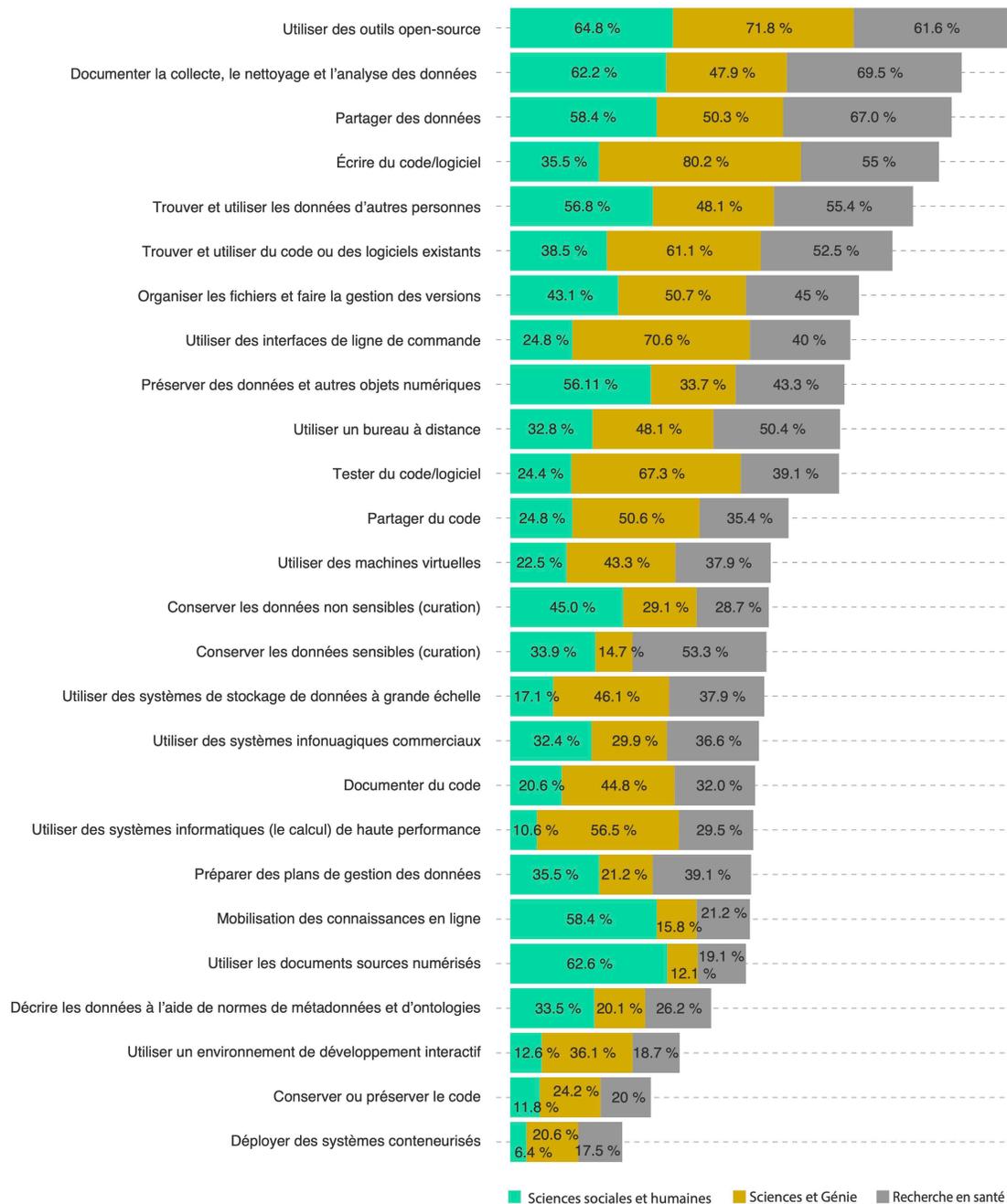


Figure 3. Question : Laquelle de ces activités fait partie de votre flux de travail de recherche? Cochez tout ce qui s'applique. Choix multiple. (Sciences sociales et humaines, n= 262; Recherche en santé, n= 240; Sciences et Génie, n= 644; total = 1146)

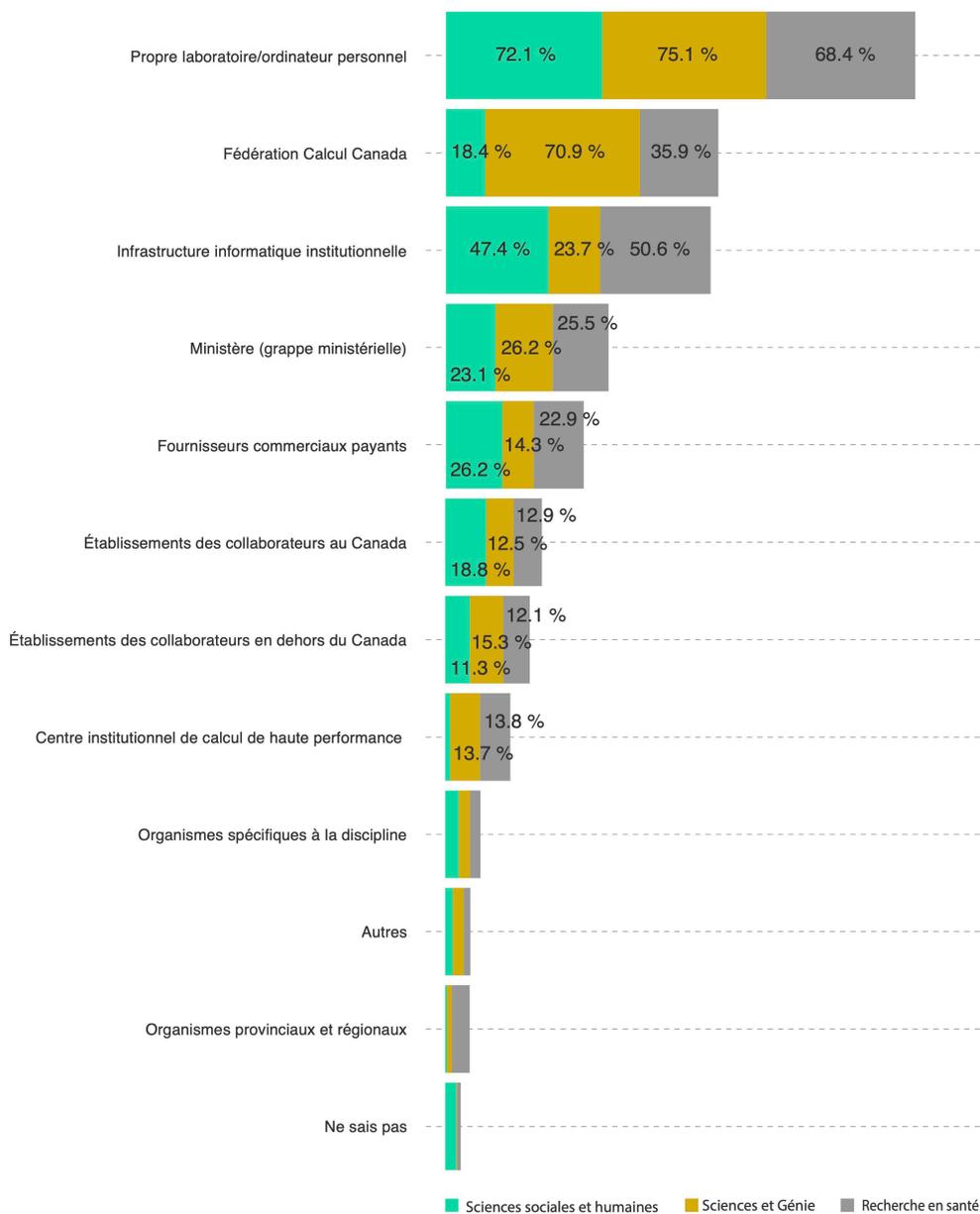


Figure 4. Question : Où se trouve l'infrastructure informatique que vous utilisez dans votre recherche? Cochez tout ce qui s'applique. Choix multiple. (Sciences sociales et humaines, n= 255; Recherche en santé, n= 231; Sciences et Génie, n= 620; total = 1106)

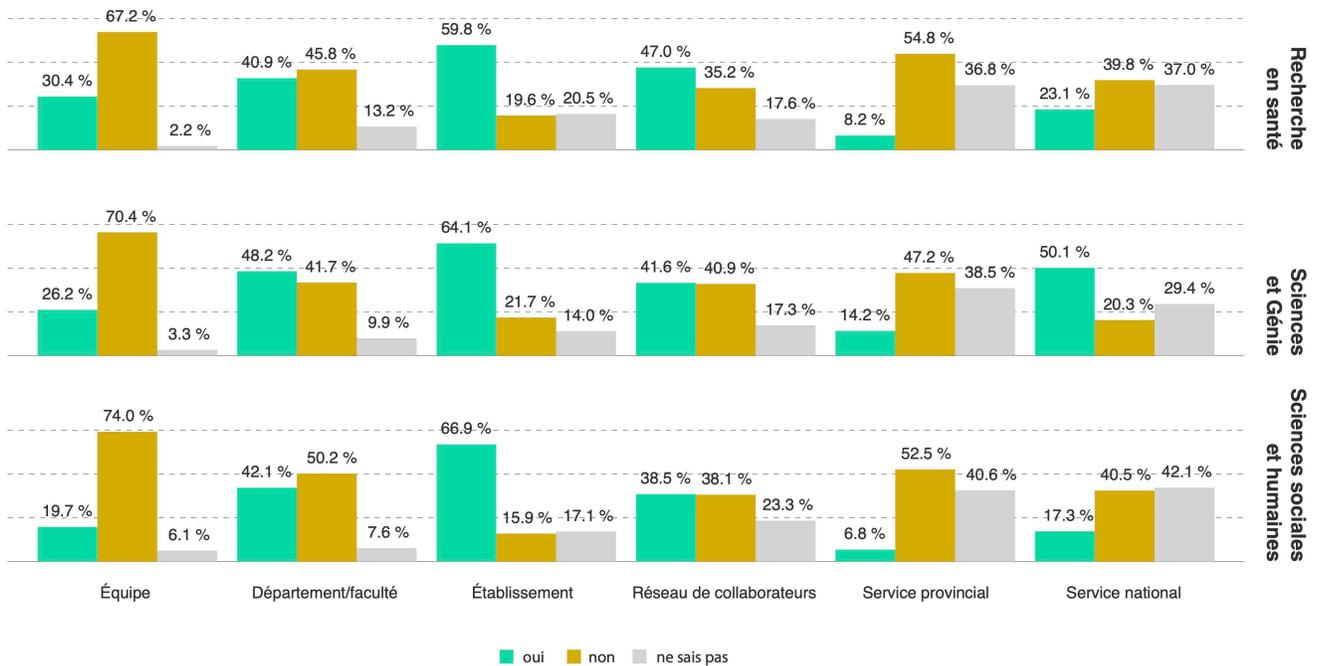


Figure 5. Question : Avez-vous accès à du personnel informatique professionnel pour soutenir vos flux de travail de recherche (p. ex. développeurs de logiciels de recherche, administrateurs système)? (Sciences sociales et humaines, n= 243; Recherche en santé, n= 222; Sciences et Génie, n= 592; total = 1057)

Quant aux moyens et mécanismes par lesquels les répondant(e)s découvrent et adoptent les outils et les ressources dans leurs recherches, il est clair que les interactions avec les pairs sont extrêmement importantes. Par exemple, les pairs/collaborateurs étaient la source d'information la plus importante pour la découverte et l'adoption des outils d'IRN pour toutes les disciplines (figure 6). Ce qui se confirme aussi dans la manière dont les chercheuses et chercheurs partagent et gèrent leurs données et leurs codes. Les chercheuses et chercheurs de toutes les disciplines ont mentionné que les pratiques propres à la discipline et les attentes des collaborateurs sont aussi, si non plus, importantes que les politiques des bailleurs de fonds et des établissements (figure 7). Malgré l'importance des interactions avec les pairs dans la dynamique au sein de la communauté en rapport avec l'IRN, il est important de souligner qu'en SG la découverte des outils et des ressources passe surtout par la Fédération Calcul Canada. Près de 66 % des répondant(e)s des SG découvrent les outils de cette manière, comparé à 33 % en RS et 21,5 % en SSH (figure 6). Ces proportions plus faibles en RS et en SSH sont probablement liées au fait, comme on l'a vu précédemment, que les chercheuses et chercheurs de ces disciplines utilisent moins les plateformes de la Fédération. Multiplier les efforts pour accroître l'accès aux plateformes nationales, comme la Fédération Calcul Canada, et leur utilisation pourrait accroître l'adoption des outils et services fournis par ces disciplines par le biais de ces plateformes, une stratégie qui pourrait aussi faire avancer la normalisation et l'interopérabilité dans ces domaines.

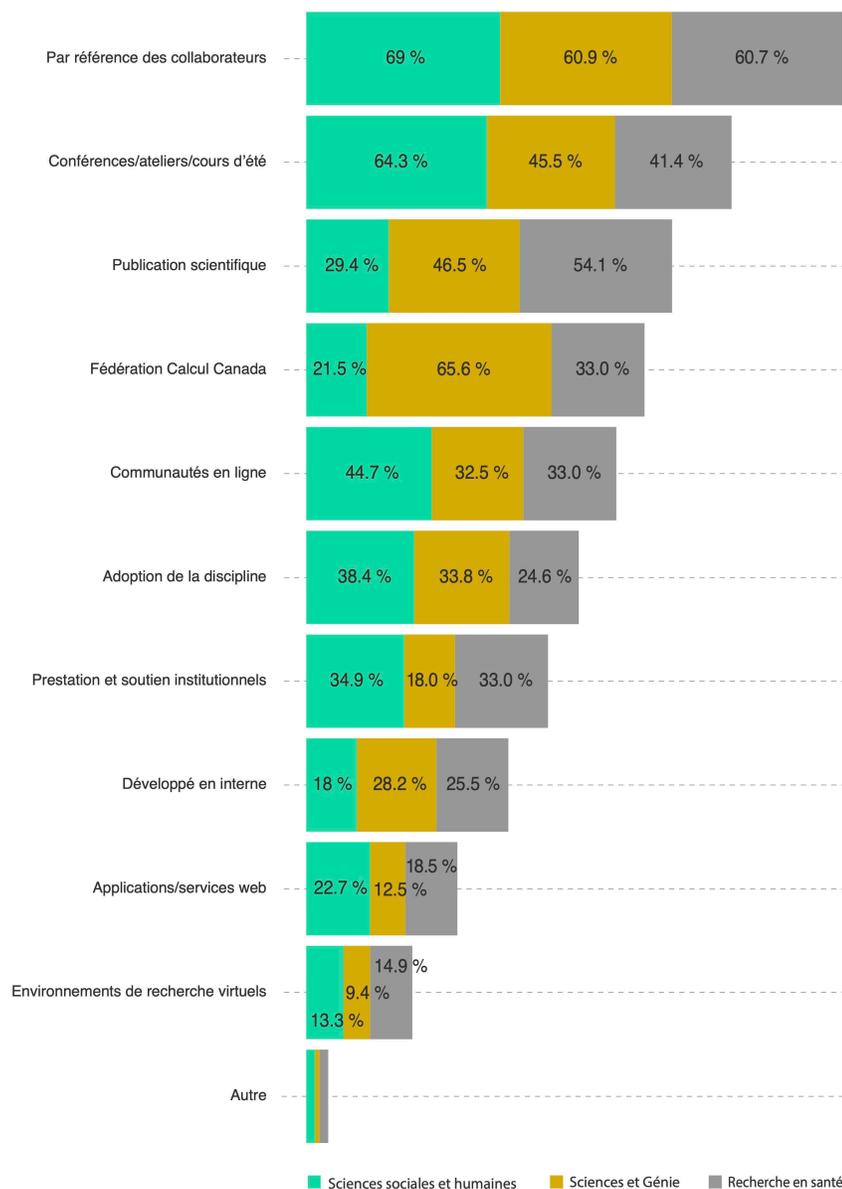


Figure 6. Question : Par quels moyens avez-vous découvert les outils et services numériques adoptés ou mis en œuvre dans votre recherche? Sélectionnez tout ce qui s'applique. *Choix multiple.* (Sciences sociales et humaines, n= 255; Recherche en santé, n= 227; Sciences et Génie, n= 615; total = 1097)

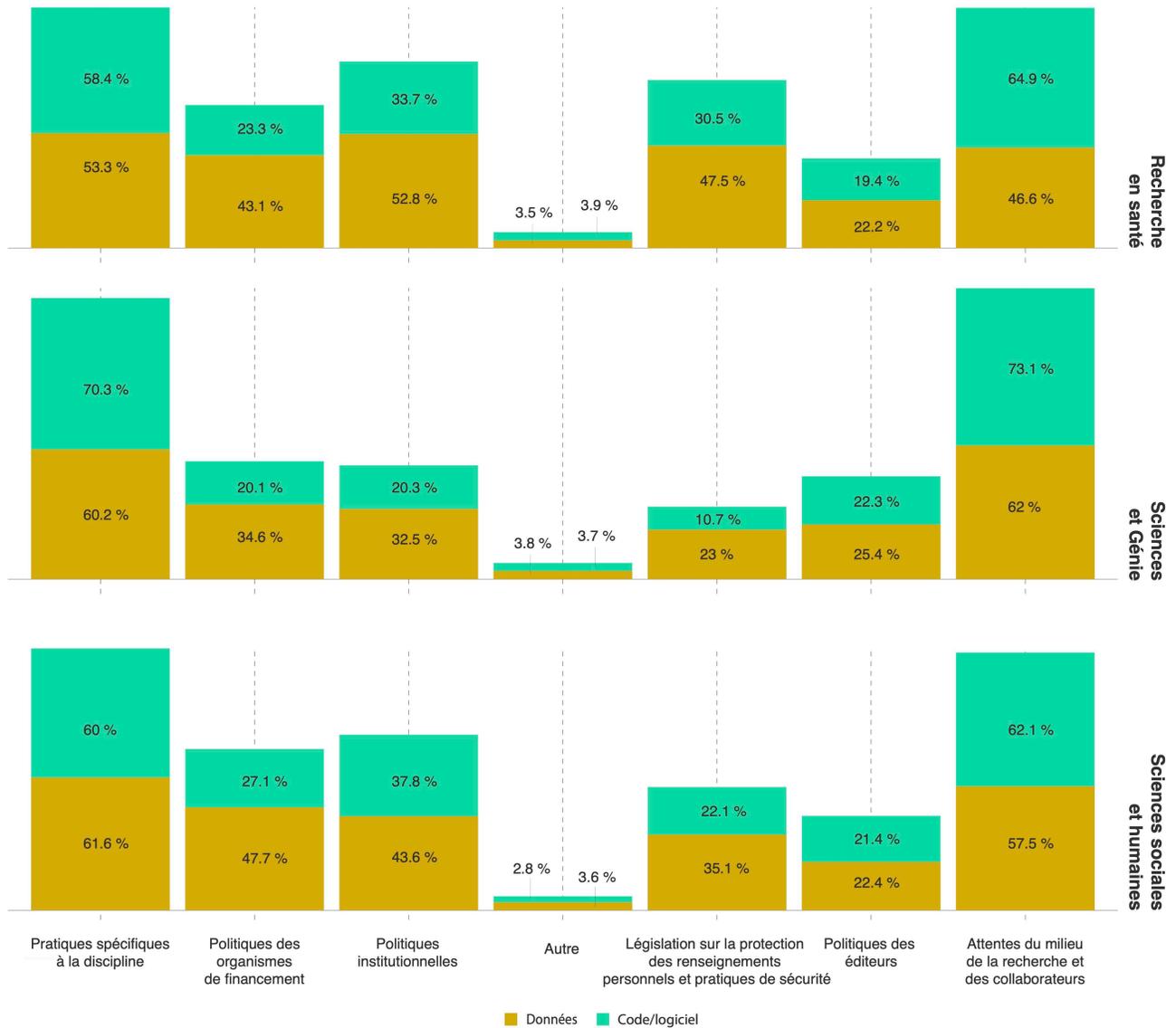


Figure 7. Question : Veuillez indiquer les trois éléments qui influencent le plus vos pratiques de gestion et de partage. (Sciences sociales et humaines, n= 245; Recherche en santé, n= 225; Sciences et Génie, n= 574; total = 1044)



4. Soutien et formation

En plus de répondre sur leur utilisation de l'IRN, les répondant(e)s devaient décrire les aspects de l'utilisation des outils et des ressources de l'IRN sur lesquels ils ou elles avaient besoin de formation et avaient l'habitude de demander ou de recevoir de la formation. L'« analyse des données (de recherche) » est l'aspect sur lequel le besoin de formation est le plus criant (figure 8), en particulier dans la RS (47,8 %) et les SSH (62,5 %). Alors que la formation sur l'analyse des données est également importante, d'autres formations plus techniques ont été priorisées par les répondant(e)s en SG (p. ex. utilisation des systèmes de calcul de haute performance, 49,6 %; pratiques exemplaires en développement de logiciel(/codage), 45,9 %; programmation parallèle, 36,7 %). Cependant, de la formation sur l'archivage et la préservation des (données) et autres objets numériques a été jugée hautement nécessaire en RS (43,6 %). L'apprentissage automatique est un aspect nécessitant de la formation tant en SG (39,2 %) qu'en RS (43,3 %). Ces divers besoins de formation reflètent la différence dans l'utilisation de l'IRN entre les disciplines, les possibilités de formation dans les programmes de premier cycle et de cycles supérieurs de même que les types de données et la culture des disciplines et des communautés.

En dépit des besoins de formation qui diffèrent entre les disciplines, les chercheuses et chercheurs reçoivent de la formation de mêmes sources. Des cours en ligne, de la formation de l'établissement et des ateliers dirigés par des pairs sont d'importants mécanismes par lesquels les chercheuses et chercheurs sont formés sur les outils de l'IRN qu'ils ou elles utilisent. (figure 9). Cependant, en phase avec les questions précédentes, plus de la moitié des chercheuses et chercheurs en SG ont dit recevoir de la formation de la Fédération Calcul Canada, contre seulement 10,9 % en SSH et 22,9 % en RS. Des congrès professionnels sont aussi un moyen de formation important en SSH (37,5 %) contre 19,2 % en SG et 26,1 % en RS.

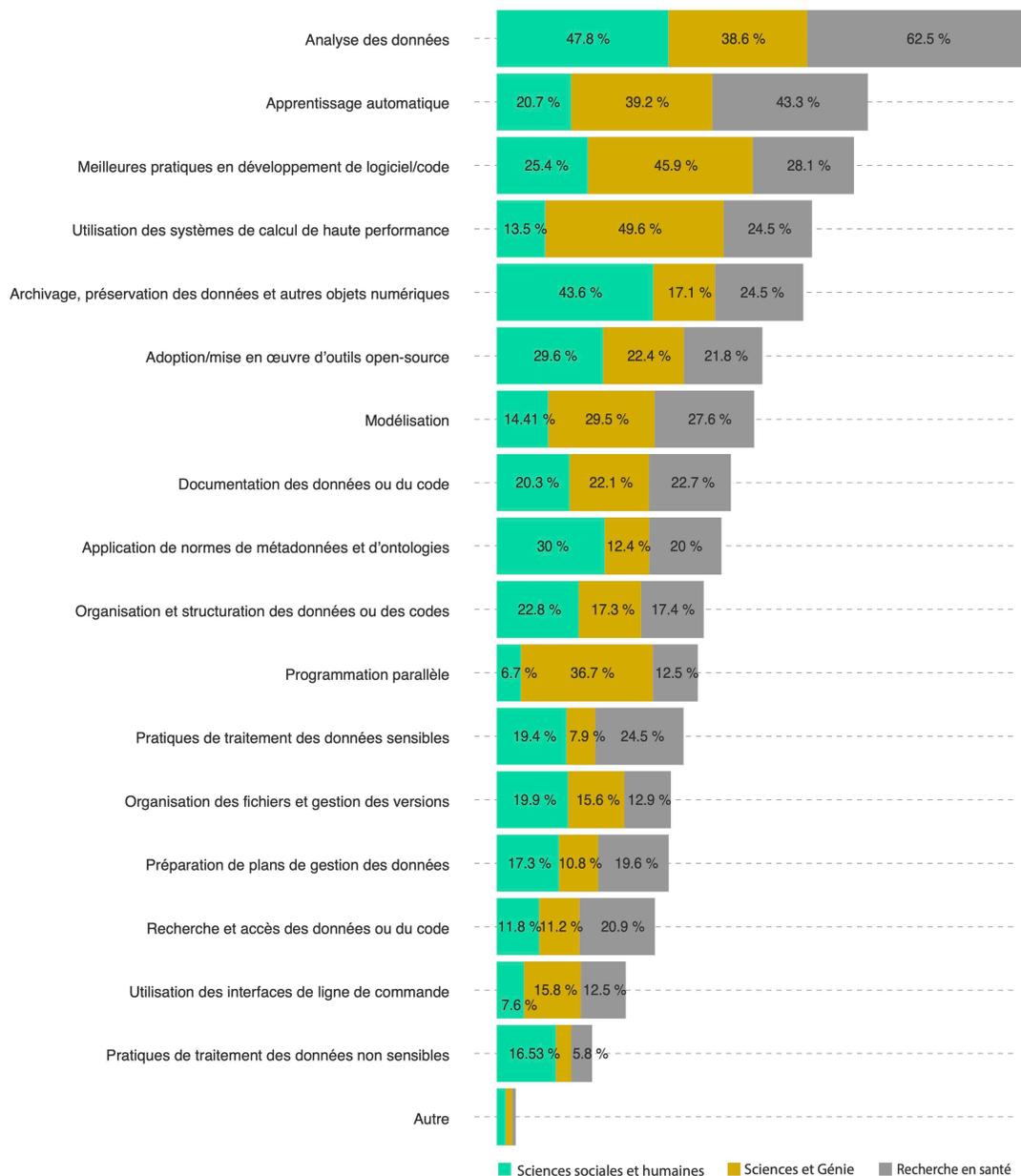


Figure 8. Question : Sélectionnez les cinq principaux domaines où vous ou les membres de votre équipe avez le plus besoin d'une formation pour améliorer ou maximiser votre utilisation de l'IRN. (Sciences sociales et humaines, n= 236; Recherche en santé, n= 224; Sciences et Génie, n= 588; total = 1048)



Figure 9. Question : Sélectionnez les trois sources d'où vous ou les membres de votre équipe recevez le plus souvent une formation. (Sciences sociales et humaines, n= 237; Recherche en santé, n= 218; Sciences et Génie, n= 587; total = 1042)



5. Besoins et priorités en matière d'IRN

5.1 Besoins

Les chercheuses et chercheurs devaient aussi répondre à la question de savoir si leurs besoins en matière d'IRN étaient satisfaits et ce que l'Alliance devrait prioriser pour répondre à ces besoins. Le niveau général de satisfaction le plus élevé au regard de l'IRN est en SG (figure 10). De la formation et du soutien manquent tout particulièrement en RS et en SSH, où les chercheuses et chercheurs se sont dit en désaccord ou fortement en désaccord dans de plus grandes proportions au sujet de la satisfaction des besoins. Cet écart entre les SG d'un côté et les SSH et la RS de l'autre est cohérent avec les questions précédentes où la plupart des chercheuses et chercheurs en SG ont affirmé recevoir de la formation et du soutien de la Fédération Calcul Canada, comparé à ceux et celles des SSH ou de la RS. En effet, les chercheuses et chercheurs des SSH et de la RS ont mentionné que les principales barrières pour ce qui est de la formation et du soutien sont le manque d'ouvertures et de formation pour du personnel hautement qualifié (PHQ). Les barrières les plus mentionnées ont été la faible documentation et le faible nombre d'ateliers et de cours. Nombre de chercheuses et chercheurs, toutes disciplines confondues, ont également mentionné des obstacles venant des établissements, où les barrières administratives et le soutien inadéquat au regard de l'utilisation de l'IRN restreignaient leur capacité d'accéder à l'IRN et de l'utiliser. Encore plus de chercheuses et chercheurs ne connaissaient pas les outils et les services qui étaient à leur disposition.

Le niveau de satisfaction au regard du financement actuel pour l'IRN a été jugé faible par toutes les disciplines. Les chercheuses et chercheurs ont mentionné que malgré les coûts importants associés au maintien en poste du personnel professionnel pour le soutien de leurs projets d'IRN, une large partie des coûts relatifs à l'IRN n'étaient pas couverts par leurs subventions de recherche. Nombre d'autres chercheuses et chercheurs ont ajouté que leurs établissements d'attache n'offraient pas de fonds pour des initiatives liées à l'IRN et que les fonds pour le personnel et l'infrastructure dépendaient du programme extrêmement concurrentiel des subventions temporaires. Selon les chercheuses et chercheurs, les salaires courants et les options de financement pour attirer et maintenir en poste du PHQ et du personnel de soutien professionnel ne sont pas suffisamment concurrentiels par rapport à ceux du secteur privé. Ces postes sont des postes temporaires n'offrant que peu ou pas de sécurité d'emploi.

Environ un tiers des chercheurs de toutes les disciplines ont souligné que leurs besoins informatiques ne sont actuellement pas satisfaits. De nombreux chercheurs évoqué le fait que les concours pour l'allocation de ressources de la Fédération Calcul Canada, basés sur des cycles de projets, étaient fastidieux et n'encourageaient pas la construction d'une infrastructure partagée et l'extensibilité à long terme. D'autres, en bon nombre, en SSH et en RS, ont cité l'immense courbe d'apprentissage et tout le savoir requis pour s'engager de manière appropriée avec l'IRN et les technologies, comme étant une barrière bloquant l'accès aux systèmes. Nombre de chercheuses et chercheurs ont décrit l'accès à l'infrastructure de la Fédération Calcul Canada



comme étant difficile, avec de longues listes d'attente et un accès non fiable. Nombre d'autres ont indiqué qu'ils ne savaient pas quels systèmes de soutien et quelles ressources d'IRN étaient à la disposition des chercheuses et chercheurs du Canada.

5.2 Priorités

Alors qu'il y a d'importantes différences dans l'utilisation de l'IRN, et donc dans les besoins relatifs à l'IRN dans toutes les disciplines, les fonds pour le personnel de soutien de recherche professionnel (SSH= 46,2 %; SG= 30,5 %; RS= 43,9 %) et la formation pour du PHQ (SSH=43 %; SG= 37,4 %; RS= 37,1 %) ont été jugés comme étant prioritaires par toutes les disciplines. L'infonuagique (SSH=32,6 %; SE= 32,5 %; HR=40,7 %) et le stockage en dépôt pour le partage de données (SSH=37,2 %; SG= 28,9 %; RS= 38,3 %) sont aussi des priorités.

Des différences dans les priorités ont aussi été observées. Par exemple, les chercheuses et chercheurs de la RS ont priorisé les ressources informatiques (p. ex. réseautage interne à haute vitesse = 48,7 %, calcul standard= 34 %); les chercheuses et chercheurs en SSH ont priorisé le financement de (plateformes) de mobilisation des connaissances en ligne (43%) et le stockage d'archives (37,6 %). Les chercheuses et chercheurs en RS ont aussi mentionné le stockage sécurisé pour les données (sensibles) comme étant une priorité (35,8 %).

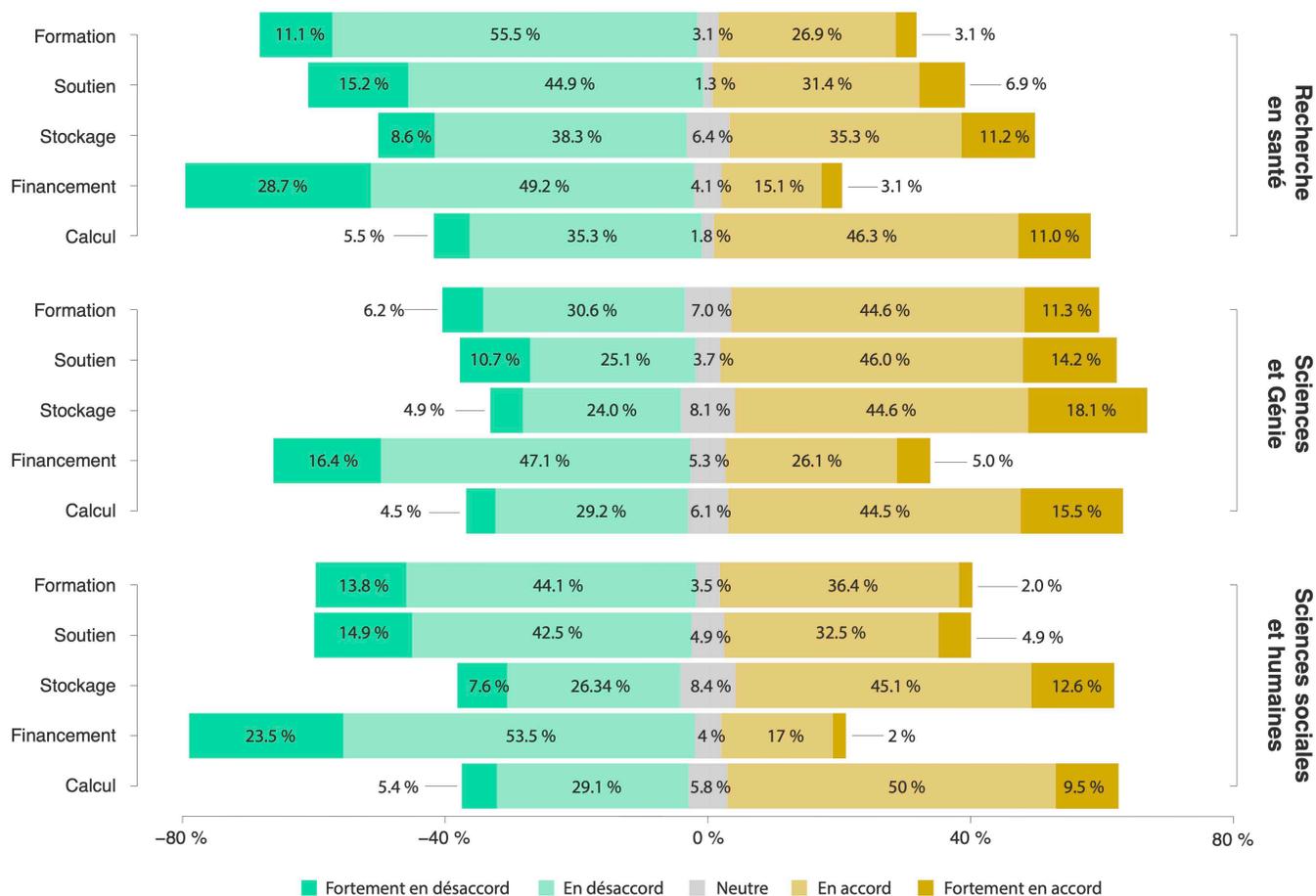


Figure 10. Question : Dans quelle mesure vos besoins actuels en matière d'infrastructure de recherche numérique sont-ils satisfaits? *Choix multiple.* (Sciences sociales et humaines, n= 224; Recherche en santé, n= 210; Sciences et Génie, n= 535; total = 969)

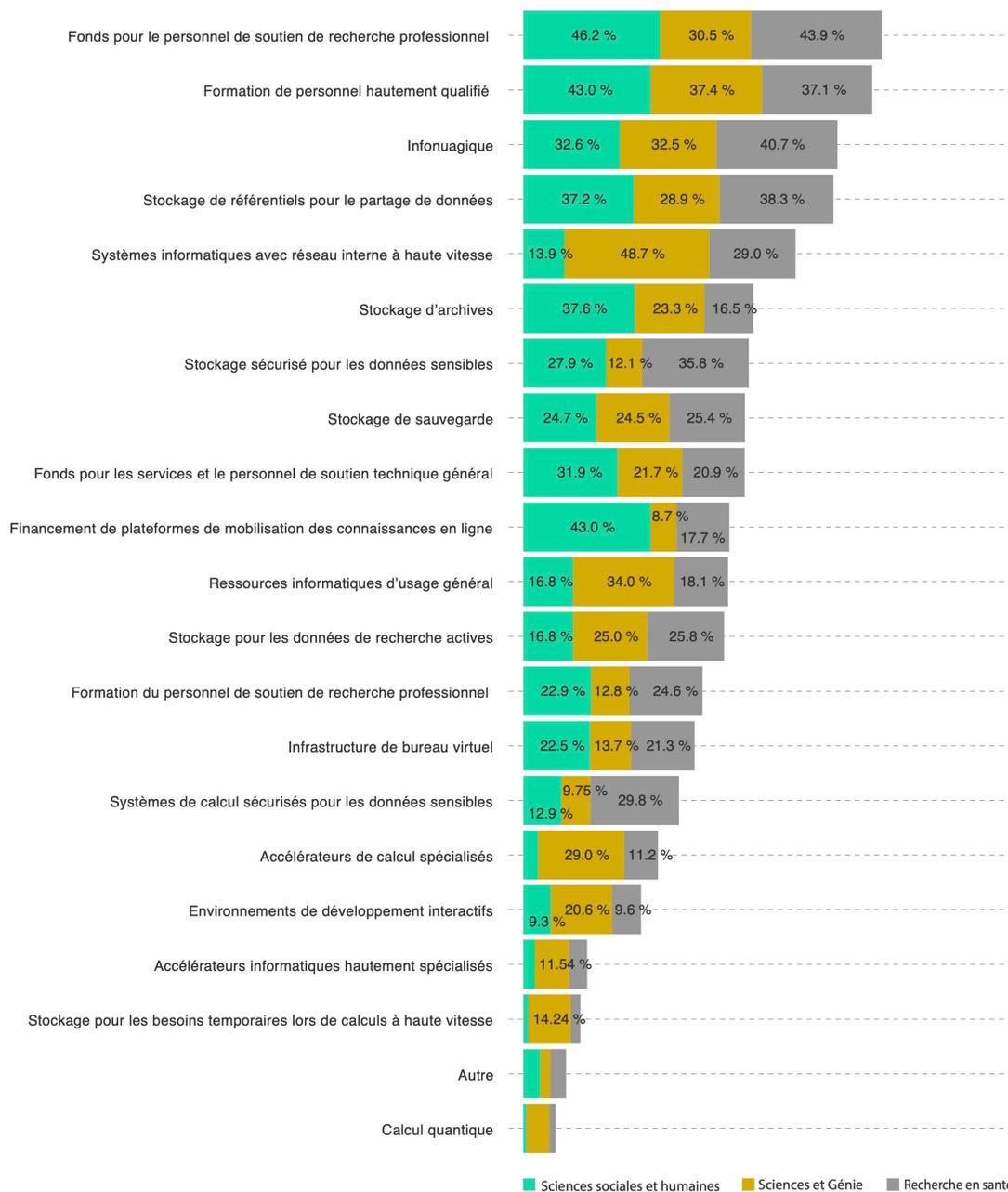


Figure 11. Question : Que doit prioriser l'Alliance pour répondre à vos besoins actuels et futurs en matière d'IRN? Sélectionnez les cinq éléments les plus importants. *Choix multiple.* (Sciences sociales et humaines, n= 279; Recherche en santé, n= 248; Sciences et Génie, n= 667; total = 1194)



6. Recommandations

On a aussi demandé aux chercheuses et chercheurs d'écrire en forme libre ce que pourrait faire l'Alliance pour améliorer l'IRN actuel au Canada. Les thèmes relevés sont les suivants (par ordre alphabétique) :

- Adopter et promouvoir l'utilisation d'équipement de logiciels gratuits et libres (*open source*).
- Améliorer l'accès à distance à l'IRN.
- Améliorer la convivialité de l'IRN (p. ex. simplicité d'utilisation).
- Améliorer la sensibilisation et la mobilisation des communautés sous-représentées/sous-utilisées et les amener sur un pied d'égalité avec les utilisatrices et utilisateurs habituels de l'IRN.
- Améliorer le soutien dans les disciplines des sciences sociales et humaines (SSH) et de la recherche en santé (RS).
- Améliorer le soutien des petits utilisateurs et utilisatrices du calcul informatique de pointe (CIP).
- Coordonner les possibilités de financement de l'IRN avec les trois conseils.
- Élargir le soutien technique adapté à chaque discipline.
- Établir un programme de financement destiné à couvrir les salaires du personnel hautement qualifié (PHQ) et du personnel de soutien professionnel.
- Étendre la formation sur les outils et les ressources d'IRN, de niveau de base au niveau avancé.
- Agir de manière transparente.
- Faire tomber les barrières administratives freinant l'accès et l'utilisation de l'IRN.
- Fournir des services professionnels pour le développement de codes, d'algorithmes et de pipelines.
- Rendre l'IRN gratuit et accessible à chaque chercheuse et chercheur au Canada, de tout établissement, de toute région ou de toute discipline de recherche.



Participation à l'enquête de l'Alliance par région

Tableau 1. Répondants et répondantes par région et établissement

Établissement ³	n
Alberta	
Université de l'Alberta	47
Université de Calgary	21
Université de Lethbridge	14
Université Athabasca	7
Université Concordia d'Edmonton	2
Collège Lakeland	1
Université MacEwan	1
Collège NorQuest	1
Colombie-Britannique	
Université de la Colombie-Britannique	88
Université de Victoria	48
Université Simon Fraser	18
TRIUMF	9
Université Thompson Rivers	3
Université du nord de la Colombie-Britannique	2
Université de Fraser Valley	2
Population Data BC	1
Université Royal Roads	1
Manitoba	
Université du Manitoba	32
Université de Winnipeg	10
Université Brandon	4
Nouveau-Brunswick	
Université de Moncton	7
Université du Nouveau-Brunswick	5
Collège communautaire du Nouveau-Brunswick	2
Université Mount Allison	1
Terre-Neuve-et-Labrador	

³ (NdT) Dans ce tableau, le nom unilingue anglais officiel de certains établissements a été librement francisé.



Université Memorial de Terre-Neuve	24
Nouvelle-Écosse	
Université Dalhousie	24
Université St. Francis Xavier	7
Université Saint Mary's	6
Université Acadia	3
Université du Cap-Breton	2
Santé Nouvelle-Écosse	1
Ontario	
Université de Toronto	185
Université York	75
Université Western	57
Université de Waterloo	42
Université McMaster	41
Université de Guelph	38
Université de Windsor	24
Université Queen	19
Université d'Ottawa	16
Université Ryerson	15
Université Lakehead	11
Université Carleton	7
Institut universitaire de technologie de l'Ontario	6
Institut Périphère de physique théorique	6
Université Wilfrid Laurier	6
Collège George Brown	5
Université Laurentienne	5
Université Brock	4
Université Trent	3
Université de l'École d'art et de design de l'Ontario	2
Collège Niagara	1
Université Northeastern	1
Santé Ontario	1
Princess Margaret Cancer Centre	1
Unity Health Toronto	1
Women's College Hospital	1
Île-du-Prince-Édouard	
Université de l'Île-du-Prince-Édouard	7



Québec	
Université McGill	127
Université de Montréal	70
Université Laval	38
Université Concordia	35
Université du Québec à Montréal	27
Université de Sherbrooke	11
Polytechnique Montréal	9
Université du Québec à Rimouski	9
École de Technologie Supérieure	7
Université du Québec à Chicoutimi	7
Université du Québec à Trois-Rivières	6
Université du Québec en Abitibi-Témiscamingue	5
Institut national de la recherche scientifique	3
Ouranos	2
CÉGEP de Sherbrooke	1
CÉGEP de Trois-Rivières /CMQ	1
Centre de recherche informatique de Montréal	1
Collège de Bois-de-Boulogne	1
Institut de recherche Robert-Sauvé en santé et en sécurité du travail	1
Collège John Abbott	1
Université du Québec en Outaouais	1
Université TÉLUQ	1
Saskatchewan	
Université de la Saskatchewan	21
Université de Regina	10
Polytechnique de la Saskatchewan	1



Annexe 2

Rapport sur les exposés de position présentés dans le cadre de l'évaluation des besoins de l'Alliance

Table des matières

1. Introduction et contexte	3
2. Sensibilisation et accessibilité	3
2.1. Sensibilisation aux infrastructures existantes	3
2.2. Accès inégal au soutien institutionnel	4
2.3. Accès inégal aux fonds	5
2.4. Bilinguisme	6
2.5. Accessibilité et facilité d'utilisation de l'infrastructure	6
3. Gouvernance et politiques	7
3.1. Reddition de comptes	7
3.2. Collaboration et relations de confiance avec les peuples autochtones	8
3.2.1. Souveraineté des données autochtones	8
3.3. Partage des renseignements permettant d'identifier une personne et des données sur la santé	9
3.3.1. Barrières juridiques	9
3.3.2. Barrières réglementaires	10
3.3.3. Données fondées sur la race	12
3.3.4. Autres modèles de gouvernance	12
3.4. Alignement sur les politiques des bailleurs de fonds et des établissements	13
3.5. Collaboration avec l'industrie	13
3.5.1. Calcul pour les partenaires de l'industrie et des universités	14
3.5.2. Influences du secteur privé	14
3.6. Mandats de la recherche ouverte	15
4. Opérations	16
4.1. Portée à définir	16



4.2.	Équité, diversité et inclusion	16
4.3.	Structure organisationnelle	17
4.3.1.	Harmonisation des politiques entre les sites hôtes	18
4.4.	Mobilisation de la communauté et collaboration avec les intervenants régionaux	18
4.5.	Collaborations à l'international	19
4.6.	Gestion de l'identité	20
4.7.	Soutien de grands projets d'infrastructure	20
4.8.	Perfectionnement professionnel et sécurité d'emploi du personnel de soutien professionnel ...	21
4.9.	Concours pour l'allocation des ressources	21
5.	Soutien.....	22
5.1.	Personnel de soutien professionnel	22
5.1.1.	Soutien national	23
5.1.2.	Soutien institutionnel.....	23
5.2.	Éducation et formation	24
5.2.1.	Prestation inégale de la formation entre les régions et les établissements	24
5.2.2.	Formations des étudiants de cycles supérieurs	25
5.2.3.	Calcul	25
5.2.4.	Logiciels de recherche (LR).....	25
5.2.5.	Gestion des données de recherche.....	26
5.2.6.	Littéracie (ou culture) des données	26
5.2.7.	Souveraineté des données autochtones	27
5.2.8.	Formation pour le renforcement de la communauté.....	27
5.3.	Partenariats	27
5.4.	Financement durable	28
6.	La gestion des données de recherche	28
6.1.	Gestion des données et science ouverte	28
6.1.1.	Interopérabilité	28
6.2.	Normes.....	30
6.3.	Identifiants pérennes	30
6.4.	Plans de gestion des données	31
6.5.	Soutien de la gestion des données de recherche à l'étape active de la recherche	32
6.5.1.	Curation des données.....	33
6.6.	Gestion des données sensibles	33
6.7.	Découverte des données	34



6.8. Numérisation	35
6.9. Soutien de la gestion des données de recherche à long terme	35
7. Logiciels de recherche	35
7.1. Stratégie durable en matière de logiciels de recherche	35
7.2. Logiciels sous-jacents du calcul informatique de pointe (CIP)	36
7.3. Plateformes de recherche	36
7.4. Logiciels sous-jacents de la gestion des données de recherche	37
8. Calcul informatique de pointe	37
8.1. Renouvellement des investissements en CIP	37
8.2. Services 24/24, 7/7	38
8.3. Communications sur le CIP	39
8.4. Infonuagique commerciale	39
8.4.1. Les risques liés à l'infonuagique commerciale	40
8.5. Besoins de calcul par discipline de recherche	41
8.5.1. Intelligence artificielle (IA) / Apprentissage automatique (AA)	41
8.5.2. Sciences sociales et humaines	41
8.5.3. Données sanitaires/biomédicales	42
8.6. Stockage	43
9. Exposés de position de référence	45



1 Introduction et contexte⁴

L'Alliance a commencé sa consultation menée dans le cadre de l'Évaluation canadienne des besoins en matière d'infrastructure de recherche numérique en sollicitant des exposés de position auprès de la grande communauté de recherche canadienne. C'était un moyen d'offrir une ardoise vierge aux autrices et auteurs pour qu'ils nous fassent part de leurs points de vue, de leurs besoins et de leurs recommandations pour un paysage d'IRN cohérent et équitable. L'appel est resté ouvert du 22 octobre au 14 décembre 2020. Un total de 107 exposés uniques a été présenté par 350 personnes de 112 organisations. La liste se trouve à la fin de l'annexe et des hyperliens renvoient aux documents originaux (voir [Exposés de position de référence](#)).

Tous les exposés de position ont été rassemblés et classés par sujet et champ d'intérêt. Une fois les exposés classés, les grandes conclusions et recommandations ont été extraites et organisées par les groupes de travail sur le Calcul informatique de pointe (CIP) et la Gestion des données de recherche (GDR) de l'Alliance. Les constatations ont ensuite été réorganisées par thème et sujet. Elles sont présentées comme des besoins par sujet. Les recommandations sont présentées comme des « Recommandation(s) de la communauté » dans la section pertinente selon le sujet. Des renvois à des exposés en particulier sont indiqués par des numéros entre parenthèses.

2 Sensibilisation et accessibilité

2.1 Sensibilisation aux infrastructures existantes

Plusieurs exposés ont souligné le fait qu'un grand nombre d'outils, de services et de ressources dont les chercheuses et chercheurs ont besoin existaient déjà (01, 13, 15, 44, 02, 62, 76); quant à savoir où les trouver et comment y accéder, la réponse variait d'une communauté à l'autre. En dehors des sciences, des technologies, du génie et des mathématiques, ce « savoir » était plutôt élémentaire (16). Même pour les chercheuses et chercheurs rompus à l'IRN, dans le contexte canadien actuel, il est assez compliqué de s'y retrouver. [trad.] « *La principale difficulté pour tout chercheur est de trouver les pièces du casse-tête nécessaires et de les assembler pour former l'image qui le guidera dans leur recherche.* » (13).

⁴ [NdT] Dans la présente annexe, le nom unilingue anglais officiel de certains établissements a été librement francisé.



Certains auteurs ont mentionné qu'accéder à l'infrastructure de la Fédération Calcul Canada peut être plus compliqué que d'établir des services infonuagiques commerciaux, en particulier pour des besoins à court terme (2; voir [Infonuagique commerciale](#)).

Recommandations de la communauté :

- Collaborer avec les entités au niveau des établissements pour mieux faire connaître les ressources nationales d'IRN (20, 54, 02, 103). Par exemple, le centre des sciences humaines numériques de l'université Ryerson voit l'Alliance comme une sorte de médiatrice au milieu d'une myriade de centres et de bibliothèques de recherche universitaires (91). Des représentant(e)s d'établissements jumelé(e)s à des employé(e)s de l'Alliance pourraient diriger les initiatives de sensibilisation (54).
- Offrir des portails conçus pour les établissements pour le soutien des besoins, notamment de sensibilisation, d'accessibilité et de surveillance (76, 56, 41).
- Élaborer et tenir à jour un répertoire ou catalogue des infrastructures existantes (01, 13, 15, 44, 02, 62, 76).
- L'université Athabasca décrit la gamme des outils, outils nationaux (p. ex. Calcul Canada, WestGrid), provinciaux (p. ex. Cybera), institutionnels, et leur passage récent à l'infonuagique des services Web d'Amazon pour l'informatique. Des outils de prise de décision quant au choix de la ressource qui convient le mieux à un projet seraient utiles pour préparer des propositions de projet et des plans de ressources (40).
- Améliorer l'accès à l'infonuagique commerciale en négociant des prix de gros et servir de rampe d'accès aux services commerciaux (2).

2.2 Accès inégal au soutien institutionnel

Comme de nombreux groupes ignorent l'existence des ressources d'IRN nationales, comme Calcul Canada ou Portage, ils demandent souvent le soutien de leurs propres établissements, notamment les centres et les bibliothèques de recherche (voir [Soutien institutionnel](#)). Plusieurs auteurs et autrices soulignent des différences importantes dans le soutien assuré par les établissements, ce qui réduit l'accès et l'utilisation des outils et des ressources disponibles et amène les chercheuses et chercheurs à organiser leur propre IRN. Cette tendance a une incidence importante sur la durabilité et la constance des données et des infrastructures (16, 18, 27). Par exemple, le fardeau imposé actuellement aux bibliothèques pour le soutien de la gestion des données de recherche au niveau institutionnel peut accroître les iniquités entre les établissements, en particulier pour les petits établissements et collèges qui n'ont peut-être pas le personnel pour soutenir leurs chercheuses et chercheurs. Les auteurs et autrices des sciences humaines qui ont besoin de soutien pour des projets numériques complexes se tourneront vers les bibliothèques même si ce soutien, par ailleurs essentiel aux sciences humaines numériques, déborde du mandat habituel de ces établissements.



Recommandations de la communauté :

- L'accès aux ressources devrait être équitable, peu importe la taille et l'emplacement des établissements (22, 53, 56, 101). L'UQAC a cité l'obligation faite à ses chercheuses et chercheurs qui veulent consulter les données hautement sensibles de Statistique Canada de se rendre dans de grands centres urbains pour accéder à certaines bases de données fournies par les Centres de données de recherche du Réseau canadien des centres de données de recherche. Ce genre de situation ne devrait pas se produire pour les solutions de l'Alliance.
- Dans un souci d'équité et d'inclusion, les petits établissements devraient être représentés dans les comités décisionnels de l'Alliance (56).

2.3 Accès inégal aux fonds

La question de l'inaccessibilité des fonds a été exposée selon trois angles d'approche : celui des disciplines, celui des établissements et celui des communautés. Les différences dans les règles d'appariement des fonds entre les provinces et les disciplines de recherche créent une certaine confusion et sont vues comme étant inéquitables (55). Les chercheuses et chercheurs des Sciences sociales et humaines (SSH) ont mentionné que le Conseil de recherches en sciences humaines ne permet pas que ses fonds servent à l'infrastructure et que le plafond de ses allocations pour le développement de logiciels est plus bas que celui des trois organismes subventionnaires, ce qui nuit de manière disproportionnelle aux chercheuses et chercheurs des SSH (18, 55). Cependant, le Conseil de recherches en sciences naturelles et en génie, qui accorde des subventions pour l'infrastructure, ne finance pas la recherche en sciences humaines (18). D'autres ont noté que la complexité des demandes à la Fondation canadienne pour l'innovation et le taux de refus élevé qui en résulte laissent à penser que le processus de soumission d'une demande est hors de portée de nombreux chercheurs et chercheuses (102).

Sous l'angle des établissements, les exposés de petites universités ont rappelé l'importance de pouvoir compter sur des services nationaux, ces derniers offrant le soutien qu'elles ne sont pas en mesure de financer et demandant des modèles de frais équitables pour y accéder (76, 101).

Le Centre de gouvernance de l'information des Premières Nations (CGIPN) répond à la nécessité de ces dernières de disposer d'un financement équitable afin de collaborer avec le système de collecte de données canadien tout en développant leurs propres structures de données pouvant s'interfacer avec le système canadien et protéger les données des Premières Nations à leurs propres conditions (98). Dans un document sur la réconciliation, Ry Moran a fait valoir la nécessité d'offrir du financement ciblé aux groupes autochtones (102).

Recommandations de la communauté :

- À la suite de la publication du rapport de la commission Vérité et Réconciliation, les trois organismes ont établi du financement ciblé pour le soutien de la recherche sur la



réconciliation, de la recherche autochtone et de la formation en recherche. Il faut prévoir plus de fonds pour identifier et réparer les préjudices communautaires et intergénérationnels en matière de recherche, notamment des fonds directement versés aux communautés pour soutenir leurs objectifs de recherche et la souveraineté de leurs données (39).

- Consolider des volets de financements équitables et agiles qui répondent aux besoins des chercheuses et chercheurs de toutes les disciplines (55).

2.4 Bilinguisme

L'aspect le plus souvent soulevé au sujet de l'équité et de l'inclusion concernait l'accès à des services et du soutien en français. Plusieurs exposés rappelaient que le français ne devrait pas être négligé et qu'il devrait faire partie intégrante de la planification des projets et de l'infrastructure dès le début (08, 49, 56, 58, 64, 65, 82, 87, 89).

Recommandations de la communauté :

- Intégrer la nécessité du bilinguisme dans les outils et les services, en ce qui concerne l'interface, la documentation (guides, modules, politiques, etc.), la formation et le soutien.
- Un exposé de l'UQAC suggère une entente de partenariat entre l'Alliance et Calcul Québec pour assurer le soutien de ses services en français (56).

2.5 Accessibilité et facilité d'utilisation de l'infrastructure

Nombre d'auteurs et d'autrices ont insisté sur l'importance que les outils soient facilement accessibles, en particulier pour le stockage, la collaboration, le partage et la préservation sur des plateformes fiables et durables pour donner des assises stables à la recherche au Canada. Cette démocratisation de l'accès aux données et au calcul et aux ressources analytiques contribuerait à l'équité, les chercheuses et chercheurs canadiens doivent être en mesure de concurrencer non seulement à l'international, mais aussi avec des géants multinationaux du secteur privé qui ont accès à une puissance de calcul quasi illimitée (84).

Recommandations de la communauté :

- Une seule interface devrait permettre de se connecter au plus grand nombre possible de fonctions administratives de l'Alliance (54).
- Offrir des plateformes de collaboration de base freinant le moins possible l'utilisation qui sont maniables et abordables et qui élargissent l'accessibilité à l'IRN (10, 15, 20, 54).
- *[trad.] « Les interfaces grâce auxquelles les utilisateurs interagissent avec les technologies de l'Alliance façonneront l'avenir de la communauté de recherche du Canada. »* Dans leur rapport, Antoniuk et Brown font valoir que bien que les outils de ligne de commande prédominent dans de nombreux contextes d'IRN traditionnels, des



interfaces utilisateurs graphiques soigneusement conçues peuvent soutenir une plus grande accessibilité aux outils et services et assurer plus d'équité à cet égard (54).

- S'appuyer sur les réalisations de la Fédération Calcul Canada, telles que l'équipe du soutien à la recherche et les autres équipes nationales, pour assurer un service plus uniforme à tous les sites via le bureau de dépannage, le wiki et les logiciels, même s'il y a encore des différences entre les sites qui posent des obstacles à l'accès et à la facilité d'utilisation (93; [voir Personnel hautement qualifié](#)).

3 Gouvernance et politiques

3.1 Reddition de comptes

Plusieurs exposés portaient sur le rôle de l'Alliance en tant qu'organisation émergente et sur les valeurs qu'elle devrait adopter. La reddition de comptes sur l'utilisation des fonds publics a été décrite comme étant essentielle pour une organisation nationale de ce genre, mais elle devrait se faire au plus bas coût possible. Les frais administratifs associés à certains éléments du modèle de financement actuel sont si élevés que de nombreux chercheurs et chercheuses de même que des organisations renonceraient à participer aux initiatives de l'Alliance, avec des conséquences néfastes sur l'innovation, la synergie et la synchronisation des efforts (54).

Recommandations de la communauté :

- Au vu de la puissance qu'exercera l'Alliance en ce qui regarde l'infrastructure de données, la gestion des données, la réutilisation et l'innovation qui y est liée, il devrait y avoir des structures et des processus de reddition de comptes sur l'utilisation des fonds publics par l'Alliance et la majorité de ses organisations membres (72).
- Il y a un risque que des sociétés influencent et contrôlent la recherche, ainsi que la prise de décisions sur l'infrastructure de recherche (voir Influences du secteur privé) (72). Par conséquent, l'Alliance devrait se doter de mécanismes de reddition de comptes clairs qui font participer le public à la gouvernance de l'IRN et aux décisions concernant la recherche et les approvisionnements (72).
- Il serait avisé de développer des indicateurs de rendement clés (IRC) pour les opérations et de les rendre publics (72, 92).
- Considérant la crise climatique actuelle, l'Alliance devrait se montrer transparente au sujet des émissions de gaz à effet de serre des installations de CIP et publier les résultats escomptés au regard de cibles de carboneutralité (19).



3.2 Collaboration et relations de confiance avec les peuples autochtones

Le Centre de gouvernance de l'information des Premières Nations a mis en lumière le fait qu'une stratégie nationale en matière d'IRN doit tenir compte des politiques, des processus et des personnes nécessaires pour établir et entretenir des relations de confiance avec les chercheuses et chercheurs et les peuples autochtones et qu'il y a lieu de prioriser la collaboration et le développement conjoint (98). La promotion et l'inclusion au sein des organes décisionnels sont des facteurs importants à considérer pour instaurer de bonnes relations (p. ex. le Conseil des chercheurs de l'Alliance); ainsi, divers niveaux d'expérience seraient représentés, non seulement le savoir technique, mais aussi la manière dont les peuples autochtones appréhendent la place des données dans leur vie, et ce qui met les données et les peuples autochtones à risque dans l'écosystème de l'IRN national actuel. Sans cette expérience, il est impossible d'admettre la présence de biais implicites dans les politiques et les systèmes (98).

Recommandations de la communauté :

- Il y aurait lieu d'établir de nouveaux modèles d'intendance conjointe des données de recherche dans le cadre d'ententes de partenariat entre les peuples autochtones et les organisations de recherche. Ces modèles doivent s'inscrire dans les relations et les modes de connaissances à l'échelle locale (36).
- L'Alliance a un rôle important à jouer dans la réconciliation. Le document présenté par Ry Moran fait valoir que [trad.] « *l'infrastructure de recherche numérique du Canada doit à tout le moins embrasser un projet complet et exhaustif de préservation et de revitalisation qui soutienne la résurgence des cultures autochtones* » (78).
- Promouvoir l'inclusion des Premières Nations, Métis, Inuits et autres peuples autochtones du Canada dans le développement de la stratégie nationale de l'IRN (98).

3.2.1 Souveraineté des données autochtones

Les peuples autochtones doivent participer au dialogue portant sur la manière de protéger leurs données tout en permettant l'interaction de ces données avec l'IRN du Canada. L'Alliance doit être la leader dans l'avancement des priorités des peuples autochtones en adhérant aux principes de l'avantage collectif, du contrôle, de la responsabilité et de l'éthique (de l'anglais *Collective Benefit, Authority to Control, Responsibility, and Ethics* ou *CARE*) et en établissant une IRN dirigée par les autochtones qui conduit au développement d'une capacité dans les communautés et à l'avancement de la souveraineté des données et de mécanismes de gouvernance uniques (01, 35).

Le travail de l'ICES (déjà connu sous le nom complet d'Institute for Clinical Evaluative Sciences) est un exemple positif de collaboration et de développement conjoint qui a conduit à un projet autochtone qui collabore avec les Premières Nations, les Métis, les Inuits et les autres peuples autochtones de l'Ontario dans le cadre de partenariats qui comprennent des ententes de



gouvernance et de partage des données qui ont permis des recherches et des analyses de données menées par les autochtones à même les données de l'ICES.

Recommandations de la communauté :

- Les politiques, le soutien et les outils de l'Alliance doivent être développés en collégialité pour que la souveraineté des données soit prise en compte dès le début (36, 39, 94, 98, 102).
- Donner aux Premières Nations la possibilité de désigner des conseillers techniques et des leaders politiques pour défendre les intérêts de leurs communautés respectives en concertation avec les experts canadiens et les leaders du monde des données.
 - *Exemple* : une table technique réunissant Premières Nations, experts autochtones et non autochtones pourrait travailler à solutionner les deux questions suivantes :
 - i. le développement et la mise en œuvre de systèmes, d'outils et de protocoles canadiens qui respectent la souveraineté des données des Premières Nations;
 - ii. le développement et la mise en œuvre de systèmes, d'outils et de protocoles des Premières Nations nécessaires à l'affirmation de la souveraineté de leurs données.
- Créer un conseil consultatif autochtone mandaté pour guider l'Alliance sur les politiques relatives à l'infrastructure de recherche numérique pour la préservation et la protection des données autochtones et de leur souveraineté (36).
- Le Réseau de recherche sur les données de santé du Canada envisage pour l'avenir des données relatives aux populations des Premières Nations, Métis et Inuits gouvernées et utilisées par des organisations autochtones pour éclairer la prise de décisions (35).

3.3 Partage de renseignements permettant d'identifier une personne et de données sur la santé

Plusieurs autrices et auteurs ont souligné les difficultés que les chercheuses et chercheurs rencontrent lorsqu'ils cherchent des solutions de stockage et de calcul pour les données sensibles (p. ex. information permettant d'identifier une personne; dossiers de santé), dont les exigences de conformité sévères aux lois provinciales et nationales (32, 86). Voici les principaux points relatifs aux barrières juridiques et réglementaires.

3.3.1 Barrières juridiques

Les architectures et les règlements courants limitent la participation des chercheuses et chercheurs canadiens à la transition qui s'opère dans le monde en faveur du partage et de la mise en corrélation des données médicales, cliniques et génomiques (61, 86). En particulier, les différences dans les législations provinciale et internationale sont des barrières à l'accès ainsi qu'à la mise en corrélation et au partage des données de recherche, cliniques et administratives



(05, 11, 32, 35, 61, 85, 86, 96). Dans son exposé, Sick Kids (hôpital pour enfants) a aussi noté l'absence d'ententes de réciprocité entre les sites de recherche qui aplaniraient les différences régionales dans les cadres de réglementation (p. ex. la *Loi sur la protection des renseignements personnels et les documents électroniques*, de compétence fédérale, et la *Loi de 2004 sur la protection des renseignements personnels sur la santé*, de l'Ontario, pour la gestion des données cliniques) (86). L'accès à l'infrastructure conforme aux différentes lois est un élément clé à considérer au regard de la collaboration avec d'autres sites, comme il est mentionné dans de nombreux autres exposés.

Recommandations de la communauté :

- Pour faciliter la recherche en santé au Canada, l'Alliance devrait soutenir l'architecture de stockage et de calcul qui est sécurisée de manière conforme aux lois sur la protection des renseignements personnels et qui est à la disposition des chercheuses et chercheurs canadiens (15).
 - Ce qui pourrait être en partie résolu par un soutien accru en faveur des options de stockage provinciales ou locales (101).
- Les autrices et auteurs d'un rapport du Réseau canadien des personnels de santé suggèrent [trad.] « *un examen juridique commandé et une série de modifications aux lois, règlements et règlements administratifs parallèlement aux examens en cours en ce qui concerne les données des patients* » (11).
- Le Réseau de la recherche sur les données de santé du Canada entreprend un examen pour établir les exigences législatives, pratiques réglementaires et choix de procédure qui diffèrent selon les paliers de gouvernement et poser les assises de la future réforme de la réglementation (35).
- Les exigences internationales imposées par la loi sont un autre élément à considérer au regard de la collaboration et de l'utilisation des infrastructures. En particulier, l'incidence du *Règlement général sur la protection des données* et des modèles qui s'en inspirent devrait être prise en compte avec les parties prenantes dans les groupes de travail et discussions des pays (35, 47).
- S'appuyer sur les normes de l'Alliance mondiale pour la génomique et la santé facilitera les collaborations nationales et internationales (12).

3.3.2 Barrières réglementaires

Les politiques de réglementation administrative sont également un défi important pour les chercheuses et chercheurs en santé, en particulier parce qu'elles limitent l'utilité des données qui peuvent être partagées (61). Un rapport d'Action Diabète Canada a constaté que chaque organisation provinciale détenant et gérant des données de santé a ses propres politiques et procédures au regard de la qualité et de la normalisation des données. Certains auteurs et autrices ont ajouté que le partage des données entre les chercheuses et chercheurs à l'intérieur



des établissements ou entre ces derniers, est rendu plus compliqué à cause de systèmes de calcul et de stockage séparés et de la nécessité d'ententes d'utilisation des données entre établissements et d'approbations éthiques au regard de la recherche prospective qui explicitement prévoient le partage étendu de données (86). Le modèle des fiducies de données a été présenté dans plusieurs exposés comme une solution de remplacement des modèles actuels d'accès aux données (04, 07, 35, 61, 85; voir [Environnements des fiducies de données](#)).

Les règlements qui empêchent les ministères gouvernementaux de rendre les dossiers administratifs disponibles pour analyse représentent un défi encore plus déterminant selon le Réseau canadien des Centres de données de recherche (96). Les autrices et auteurs des organisations de recherche en santé de l'Ouest ont insisté sur le fait que ces défis se posent à cause des structures de gouvernance contraignantes ou dépassées, qui sont censées assurer la confidentialité et la conformité aux exigences en matière de protection des renseignements personnels, y compris les comités d'éthique et les comités d'accès aux données, mais qui empêchent d'avoir accès aux données en temps opportun (04, 35). Les implications sont importantes pour la gestion des découvertes fortuites en l'absence de consentement éclairé, qui est réglementé par *l'Énoncé de politique des trois conseils : Éthique de la recherche avec des êtres humains* – EPTC 2 et le processus d'examen du comité d'éthique de la recherche, et qui ne permet pas de communiquer aux patients et aux fournisseurs de soins de santé de l'information qui pourrait sauver des vies (05, 61).

Des exemples ont été fournis. Pour les données en génomique et en santé, on a souligné la nécessité d'une approche fédérative pancanadienne extensible qui permettrait d'interroger et d'analyser des bases de données en génomique et en santé humaine à l'échelle nationale, alors que les données elles-mêmes restent contrôlées de manière sécuritaire et privée dans les hôpitaux de recherche et autres organisations au pays (12). Les autrices et les auteurs ont mentionné en particulier que les politiques, la gouvernance et l'architecture, les interfaces de programmation applicative et les ontologies relient tous les sites participants en un seul réseau cohérent : HPC4Health (Ontario) et SD4Health (Québec) sont deux nœuds de calcul qui collaborent déjà entre eux et avec le centre des sciences du génome de BC Cancer à développer des interfaces de programmation applicative conjointes qui donneraient accès aux données via CanDIG (voir [Gestion des données sensibles](#)).

Recommandations de la communauté :

- Il faut des dépôts sécurisés pour faciliter le partage de données à long terme au moyen d'un accès contrôlé aux systèmes qui sont fiables et approuvés par les dépositaires des données concernés (61).
- Il faut coordonner les protocoles de sécurité et d'intégrité des données pour respecter les ententes de partage de données et les codes d'éthique de la recherche qui s'appliquent (86).



- Il faut établir de grands principes directeurs et de gouvernance pour les comités d'éthique de la recherche des établissements en matière de partage des données et des images (84).
- Il y aurait lieu d'harmoniser les processus et le libellé du consentement entre les organisations afin de soutenir le partage des données, le cas échéant (12, 52, 39).
 - Des ontologies de consentement médical normalisées pourraient prévenir les violations de règlements (12).
 - Il faut des mécanismes plus pratiques pour obtenir des patients le consentement pour l'utilisation.
 - Exemple : [trad.] « *Le consentement (numérique), le soutien multilingue et multiculturel et les pratiques dynamiques relatives au consentement qui permettent aux patients de contrôler comment leurs données sont utilisées* » (05).

3.3.3 Données fondées sur la race

Les risques pour les patients et les communautés présentés selon les données fondées sur la race qui sont collectées et vendues dans le processus de recherche sont résumés dans un exposé de LLana James. (79)

Recommandation de la communauté :

- L'Alliance devrait s'interdire de soutenir la collecte, le couplage ou le transfert des données fondées sur la race dans son IRN jusqu'à ce que des règlements et des lois soient revus pour résoudre les risques spécifiques au recoupement des données fondées sur la race (79).

3.3.4 Autres modèles de gouvernance

Étant donné que la gouvernance des données dépasse la question de l'accès et qu'elle doit aussi concerner l'acquisition et la préparation des données, la mobilisation des intervenants et du public et le renforcement des capacités, plusieurs autrices et auteurs ont insisté sur le fait que des protocoles de stockage et d'accès des infrastructures devaient être créés conjointement avec les propriétaires des données et négociés avec eux (p. ex. les données des peuples autochtones, les données sur la santé d'une personne; 81, 85, 94, 98, 102). Des ententes et du soutien pour des modèles de gouvernance des données améliorées ont aussi été mentionnés.

Recommandations de la communauté :

- Une gamme de modèles de dépôts de données sont recommandés pour soutenir différentes communautés, allant de modèles où l'accès est accordé au cas par cas par le dépositaire des données à des modèles où la prise de décision est rapide et automatique (04).



- Adopter le cadre des « cinq objets à sécuriser » (de l'anglais *Five Safes framework*) pour le calcul de haute performance en lien avec la confidentialité et la protection des données dans les provinces et les territoires, à savoir les projets, les personnes, les installations, les données et les produits (07, 35).

3.4 Alignement sur les politiques des bailleurs de fonds et des établissements

Le CV commun canadien, la politique en matière de GDR et le calcul de haute performance sont des secteurs qui doivent être améliorés, comme le mentionnent plusieurs autrices et auteurs dans leurs exposés (08, 37, 39, 41, 54, 58, 62, 64, 92, 103).

La politique de stockage des données de la Fédération Calcul Canada qui exige le renouvellement annuel des espaces de stockage entre en conflit avec les exigences en matière de rétention des trois organismes (34) (p. ex. la politique sur le libre accès aux publications des Instituts de recherche en santé du Canada, la politique sur l'archivage des données de recherche du Conseil de recherche en sciences humaines). La récente politique des trois organismes sur la gestion des données de recherche aura également des effets sur les exigences relatives aux dépôts de données.⁵

Recommandation de la communauté :

- Établir une entente de partenariat et collaborer étroitement avec les bailleurs de fonds nationaux pour la recherche afin d'harmoniser les pratiques et les politiques (08, 37, 39, 41, 54, 58, 62, 64, 92, 103).

3.5 Collaboration avec l'industrie

Une infrastructure de recherche numérique nationale soutenant le partage de données des secteurs privé et public est un élément d'une importance capitale pour que les investissements destinés à l'innovation dans les universités se traduisent en avantages économiques pour le Canada (12, 61). Néanmoins, malgré le fait que les données commerciales sont une source courante de données dans de nombreux champs de recherche, plusieurs autrices et auteurs ont souligné que souvent ces données ne peuvent être partagées librement (37, 77, 96). D'autres ont mentionné que les éditeurs commerciaux qui agissaient dans le passé comme contrôleurs de l'accès à la recherche financée par des fonds publics ont donné une lancée aux dépôts et leurs produits de GDR (47).

⁵ Gouvernement du Canada. Politique des trois organismes sur la gestion des données de recherche https://science.gc.ca/eic/site/063.nsf/eng/h_97610.html (mars 2021).



3.5.1 Calcul pour les partenaires de l'industrie et des universités

Les projets réunissant l'industrie et les universités ne profitent pas suffisamment de l'accès aux technologies de CIP pour que la recherche procure des avantages à l'industrie, aux entreprises et à la société (77). On pourrait établir des relations avec les programmes d'accélérateurs des universités (85). Par exemple, le consortium SOSCHIP (Southern Ontario Smart Computing Innovation Platform) explique dans son exposé qu'étant donné que les priorités des sociétés ne sont pas celles de l'écosystème de la recherche, ces systèmes devraient être séparés de ceux conçus pour la recherche [trad.] « *basée sur la découverte pure* ». Néanmoins, il faut réserver des ressources pour le soutien des collaborations [trad.] « *au rythme des entreprises* ». (77)

Recommandations de la communauté :

- Les services, le soutien et les ressources de l'IRN doivent être planifiés en gardant à l'esprit les partenariats entre l'industrie et les universités.
- Les investissements en IRN devraient servir à soutenir l'infrastructure du CIP dirigée par l'industrie. Il faut des plateformes réservées au soutien de l'industrie avec un niveau de sécurité élevé, et il faut fournir des logiciels et des services pour aider à passer aux plateformes commerciales après la fin des projets, sans quoi les projets pourraient ne jamais démarrer étant donné que les résultats sont coincés dans les systèmes sans procurer de bénéfices commerciaux aux sociétés (77).

3.5.2 Influences du secteur privé

Un exposé d'un groupe diversifié de chercheuses et chercheurs de l'Ontario a soulevé des préoccupations relatives aux influences des sociétés sur la recherche et la prise de décision concernant la recherche et l'infrastructure de recherche (72). L'effet négatif de ces influences est illustré par les défis que posent l'accès et le partage des données des dispositifs médicaux et des dossiers médicaux électroniques des soins de base et des données des fournisseurs de tels dossiers du secteur privé (05, 07, 52). En particulier, les autrices et auteurs d'Action diabète Canada (5) explique que [trad.] « *les pratiques des fournisseurs de dossiers médicaux limitent l'accès aux données des patients pour la recherche en chargeant des coûts additionnels, des redevances et en causant des retards. C'est de pratique courante dans l'industrie d'imposer des frais d'accès aux données d'une clinique ou d'un patient pour utilisation secondaire, notamment pour la recherche, des études épidémiologiques ou la surveillance du système de santé* ». Ces pratiques sont exacerbées par l'absence de lois visant à imposer l'interopérabilité et à empêcher le secteur privé de bloquer le partage des données, par le péage de l'accès aux données de santé et par l'absence de surveillance des activités du secteur privé (7, 47, 52).



Recommandations de la communauté :

- Établir des structures et des processus pour garantir une reddition de comptes complète à la population et aux organisations membres [trad.] « *étant donné le pouvoir qu'exercera l'Alliance en ce qui concerne l'infrastructure des données, la gestion des données, la réutilisation et l'innovation conséquente* ». (72).
- Les représentant(e)s des entreprises commerciales/privées ne devraient pas avoir accès aux données issues de sources non commerciales aux mêmes conditions d'accès que les chercheuses et chercheurs indépendants (72).

3.6 Mandats de la recherche ouverte

L'Organisation des Nations Unies pour l'éducation, la science et la culture (UNESCO) sur la science ouverte préconise que les États membres [trad.] « se mobilisent pour faire tomber les barrières et les obstacles à la science ouverte, en particulier ceux qui se rapportent aux systèmes de récompenses et d'évaluation de la recherche et de la carrière. » Nombre d'autrices et d'auteurs ont fait valoir qu'un accès à la recherche ouverte, y compris aux données de recherche, devrait être priorisé dans les politiques, l'infrastructure et les services fournis par l'Alliance (08, 10, 33, 66). Cependant, d'autres ont aussi mentionné qu'une adoption universelle de la recherche ouverte sans égard à celui ou celle qui décide de ce qui est partagé directement entre en conflit avec la souveraineté des données des peuples autochtones, la confidentialité des données et la propriété intellectuelle. Par exemple, les spécialistes de la recherche qualitative pourraient avoir des préoccupations particulières et des appréhensions au sujet de la science ouverte/partage des données (94). Les politiques sur la propriété de données (affectant les chercheuses et chercheurs et les sujets/participant(e)s), qui a une incidence sur qui a le droit d'accorder l'accès aux données dans l'avenir et sur les contrôles qui sont requis, doivent être examinées soigneusement (94).

Plusieurs exposés ont mentionné que la politique des trois organismes sur la gestion des données de recherche est un jalon critique qui va pousser les établissements à établir des stratégies et des politiques. L'Alliance serait alors en mesure de faire avancer la reddition de comptes avec ces mandats de science ouverte en instituant des mécanismes pour surveiller et encourager la conformité des établissements et des chercheuses et chercheurs (15, 41, 62).

Recommandations de la communauté :

- Envisager des mandats et des politiques de recherche ouverte pour les partenaires de l'écosystème. Par exemple, l'exposé du groupe Dynamic Brain Circuits de l'université de la Colombie-Britannique constate que la politique de stockage des données de la Fédération Calcul Canada, qui exige le renouvellement annuel de l'espace de stockage, n'est pas en phase avec l'exigence de rétention des données pour cinq ans de la « Politique des trois organismes sur le libre accès aux publications ». (34)



- Élaborer un tableau de bord automatisé pour auditer les pratiques de recherche ouverte, y compris la publication en libre accès et le partage de données, pour promouvoir la confiance en l'infrastructure de dépôt et les pratiques des chercheuses et chercheurs (41).
- Élaborer des politiques et des messages qui font la promotion de la GDR et du partage des données et qui incitent les établissements à élaborer leurs politiques de GDR (27, 62).
- L'Alliance et les autres collaborateurs de l'écosystème universitaire devraient tenir compte de la mobilisation et de l'influence des éditeurs commerciaux dans l'adoption de principes de la recherche ouverte (47).
- Envisager de collaborer avec les organisations pour rendre leurs données repérables ou accessibles dans un environnement sûr, de sorte que leurs données puissent servir en recherche (37, 77, 96).

4 Opérations

4.1 Portée à définir

Dans son développement, l'Alliance devrait tenir compte du spectre complet du financement de l'IRN pour que les volets de financement qui existent ne soient pas perdus dans la création de l'Alliance.

Recommandations de la communauté :

- L'Alliance devrait définir clairement son rôle dans le soutien de l'IRN et qui collaborera avec les autres acteurs et actrices et de quelle manière (13).
- L'Alliance devrait tenir compte des organisations existantes qui fusionnent et définir les manques de fonds que la création de l'Alliance en tant qu'organisation peut combler (07, 13, 23, 81, 83 et autres).

4.2 Équité, diversité et inclusion

L'IRN a encore des défis à relever en matière d'EDI, pour les utilisateurs, les disciplines, les régions, les établissements et même le personnel de soutien professionnel. L'Alliance peut contribuer à concrétiser le changement par l'embauche, la formation, le leadership et les occasions (22, 46, 53, 54, 55, 33). L'EDI peut aussi servir de lentille pour l'évaluation. S'il n'y a pas amélioration sur ce point, cela signifie des occasions manquées et des lacunes à combler (41).



4.3 Structure organisationnelle

Les autrices et les auteurs ont mentionné les avantages d'une infrastructure centralisée : moins de doublons, des économies d'échelle, de la collaboration et de l'inclusion (08, 98). Exemples de réussites d'une infrastructure centralisée qu'il est possible d'exploiter : planificateur commun, pile de logiciels, authentification, documentation et systèmes de soutien, pour ne nommer que ceux-là (80). Cependant, les inconvénients d'un modèle centralisé cités sont le peu de soutien pour les utilisateurs ayant des besoins particuliers, la diminution des contacts personnels et des interactions avec le PHQ et le personnel de soutien professionnel (80). Certains auteurs et autrices ont expliqué que le modèle actuel d'un partenariat fédératif (avec les partenaires régionaux) adopté par la Fédération Calcul Canada n'était pas très adapté aux besoins des chercheuses et chercheurs, qu'il était difficile de parler d'une même voie pour que les représentations faites au gouvernement soient adéquates, et qu'il favorisait moins la reddition de comptes directe que la fomentation de divisions (14).

D'autres ont évoqué que des sites d'hébergement développent des services différents, la prise de décision se faisant localement. De plus, lorsqu'il y a coordination et que le même service est déployé dans plusieurs sites, cela se fait généralement de manière indépendante par différentes personnes, ce qui donne lieu à des décisions de mise en œuvre différentes (93). Un autre défi que pose un tel système est que le personnel du site est souvent composé d'employé(e)s des établissements locaux ou universités, et non de Calcul Canada, ce qui crée de la confusion au sujet des responsabilités (93).

En plus de financer l'infrastructure partagée, un exposé d'un laboratoire de dynamique structurale a constaté que les laboratoires ont besoin de fonds pour développer et lancer leur propre IRN, essentielle pour l'innovation (10).

Recommandations de la communauté :

- Les autrices et auteurs du groupe de recherche de Global Water Futures ont recommandé un modèle de soutien par niveaux comme moyen de répondre aux besoins de base en IRN des chercheuses et chercheurs (15). Ils ont recommandé que l'Alliance fournisse une plateforme de recherche collaborative de base à tous les chercheuses et chercheurs qui présente peu d'obstacles à l'utilisation, c'est-à-dire qui est facile à utiliser et abordable. L'exemple donné est le système de dépôt de recherche de l'université Purdue, basé sur la plateforme Hub Zero, qui offre l'accès via un modèle d'obtention des ressources à plusieurs niveaux. Les chercheuses et chercheurs reçoivent les ressources du niveau de base gratuitement, les chercheuses et chercheurs qui travaillent à des projets financés par des subventions ou autres formes d'aide reçoivent des ressources additionnelles et la capacité de stockage peut s'élargir à un coût raisonnable.



- Les autrices et les auteurs de l'Institut de physique des particules et de l'Institut canadien de physique nucléaire ont souligné que le personnel et les systèmes matériels pourraient être organisés selon les besoins de calcul et permettre aux chercheuses et chercheurs de participer à des équipes spécialisées (92). Ils ont proposé des sous-groupes, tels que curation des données, archivage et préservation, calcul de haute performance (interface de passage de messages [de l'anglais *Message Passing Interface*] ou MPI, de type « MPI »), systèmes de calcul, d'entrée et de sortie de données à grand débit et systèmes à usage général.

4.3.1 Harmonisation des politiques entre les sites hôtes

Plusieurs exposés ont mentionné la nécessité d'harmoniser les politiques des grappes de Calcul Canada et de poursuivre la simplification et l'uniformisation des opérations (67). Toutes les décisions de politique locales affectent la capacité des chercheuses et chercheurs de diriger le travail à différentes grappes. La Fondation canadienne pour l'innovation traite chaque site comme un projet distinct; l'établissement hôte de chaque site possède le matériel et exécute le processus d'approvisionnement. Certains auteurs et autrices ont demandé que la gouvernance soit plus transparente et qu'il y ait une motivation à coopérer et non à rivaliser dans le respect des différences marquantes entre les sites (93, 51). Pour l'avenir, les sites hôtes demandent [trad.] « *un financement des opérations stable et prévisible et un mandat clair de fournir des services qui sont constants dans toute la plateforme nationale* ». (67)

Recommandation de la communauté :

- Nécessité d'intégrer le développement des politiques dans les sites et avec les équipes nationales. Il n'y a pas toujours de consultation, ce qui conduit à des cycles de développement frustrants et fastidieux intégrant les infrastructures de plateforme de calcul distribué nécessaires (21).

4.4 Mobilisation de la communauté et collaboration avec les intervenants régionaux

Des établissements de recherche et dans de nombreux cas les intervenants en question ont souligné la nécessité de collaborer plus étroitement avec les intervenants régionaux et de les soutenir, en particulier les fournisseurs de services de calcul régionaux et les associations de bibliothèques (40, 46, 58, 36, 97). Par exemple, l'Atlantic Computational Excellence Network (ACENET) a insisté sur le besoin de communications soutenues et bidirectionnelles ainsi que d'une intégration étroite des employé(e)s à différents niveaux avec les groupes de travail de l'Alliance pour profiter de leur expertise (46). Dans son exposé, l'Université du Québec à Chicoutimi (UQAC) s'interroge à savoir si Calcul Québec pourrait agir comme représentant de l'Alliance au Québec, vu l'excellence reconnue de ses services. À tout le moins, les mandats de l'Alliance par rapport à ceux des entités régionales doivent être clarifiés (56).



Recommandations de la communauté :

- Le développement coopératif de l'infrastructure avec les régions a été mentionné comme un moyen de tirer parti des investissements des communautés et de livrer une infrastructure adaptable, tout en laissant une place aux différences locales. Le modèle de coopération de Dataverse Canada est un exemple de réussite (94, 89, 08, 40).
- Chaque groupe d'utilisateurs d'un composant spécifique de l'infrastructure de l'Alliance pourrait inclure les membres de l'équipe de l'Alliance qui gère ce composant. En plus de construire le soutien des pairs et des collaborations potentielles, cette approche ferait intervenir directement l'Alliance dans les services (54, 67).

4.5 Collaborations à l'international

Les collaborations internationales sont essentielles dans de nombreux champs de recherche, et l'Alliance devrait faciliter les connexions avec les homologues à l'étranger afin de faire avancer les entreprises collaboratives et les normes communes (17, 21, 12, 92). L'Alliance est bien placée pour financer les groupes de recherche dans leur participation aux efforts faits en faveur de l'interopérabilité à l'international, dans les disciplines de recherche et entre les établissements de recherche (67).

La nature internationale de nombreux champs de recherche requiert que les ressources s'intègrent dans l'infrastructure mondiale (03, 12, 17, 19, 30, 47, 71, 99), telles que le calcul parallèle qui doit se faire au moyen de collaborations internationales. Une IRN canadienne interconnectée à l'international éviterait les points de défaillance uniques et faciliterait les types de calcul décentralisé parmi les différentes IRN (p. ex. production cartographique du télescope cosmologique d'Atacama à Niagara, KwaZulu-Natal et NERSC; 71).

Recommandations de la communauté :

- Les collaborations avec des partenaires internationaux s'appuient souvent sur des avantages réciproques. Par exemple, offrir une infrastructure aux communautés internationales en échange d'un accès prioritaire aux données ou aux instruments scientifiques pour les chercheuses et chercheurs canadiens (p. ex. les communautés de l'astronomie et de la physique des particules). Il faut des mécanismes pour soutenir ces connexions, et l'Alliance devrait continuer à fournir l'accès à l'infrastructure aux collaborateurs de l'étranger.
- L'Alliance, tant du côté de la gestion que de celui des opérations, devrait consulter les bailleurs de fonds internationaux (p. ex. la National Science Foundation) et les organisations qui font tourner de gros systèmes informatiques (p. ex. le National Energy Research Scientific Computing Center et l'Organisation européenne pour la recherche nucléaire ou le CERN) (14, 92).
- L'infrastructure nationale doit permettre la collaboration internationale par l'adoption de normes communes ou de cadres stratégiques d'interopérabilité.



- Des mécanismes de coordination tels que des groupes de travail et des fédérations font en sorte que le personnel qualifié et l'infrastructure technique sont alignés.

4.6 Gestion de l'identité

L'importance de la gestion de l'identité, l'authentification à deux facteurs, et l'importance des outils de collaboration multidisciplinaire sont souvent mentionnées dans les exposés, étant donné que des outils existants, notamment Slack, GitLab, staff.wiki, ont des frontières identitaires rigoureuses entre celles à l'intérieur et celles en dehors de l'organisation. Il a été démontré qu'elles empêchent la participation de collaborateurs de l'extérieur (21).

Recommandations de la communauté :

- L'Alliance devrait collaborer avec ses établissements membres à la mise en œuvre de systèmes d'identité fédératifs, probablement sous la Fédération canadienne d'accès de CANARIE (21).
- L'Alliance devrait mettre en œuvre une authentification multifactorielle et d'autres contrôles pour protéger les données de recherche et prévenir les atteintes à la sécurité (32).

4.7 Soutien de grands projets d'infrastructure

Plusieurs exposés ont insisté sur l'importance de soutenir des projets de grande envergure par l'intervention du personnel professionnel et le financement.

Recommandations de la communauté :

- L'Alliance devrait agir comme partenaire potentiel dans le centre de données du consortium SKA (Square Kilometer Array) Canada parmi d'autres parties prenantes canadiennes (p. ex. CANARIE, Centre canadien de données astronomiques, réseau canadien avancé pour la recherche astronomique (de l'anglais *Canadian Advanced Network for Astronomical Research*) et universités canadiennes) (03).
- Collaborer au projet canadien de bibliothèque du génome humain dirigé par les Instituts de recherche en santé du Canada qui intégrera les efforts de séquençage du génome humain faits au Canada et établira un « seul point d'entrée » pour l'accès et l'analyse (61).
- Collaborer avec le CGIPN sur sa vision du réseau dirigé par les Premières Nations des centres de gouvernance régionaux de l'information servira les besoins en information des communautés des Premières Nations du pays (98).



4.8 Perfectionnement professionnel et sécurité d'emploi du personnel de soutien professionnel

Nombre d'autrices et d'auteurs ont insisté sur l'importance du personnel de soutien professionnel du pays, en particulier à cause de son rôle dans les milieux de recherche interinstitutionnels (28, 46, 67, 101). Beaucoup d'exposés reconnaissent le mérite de l'aide du personnel de soutien professionnel de la Fédération Calcul Canada, faisant remarquer toutefois que la précarité de leur emploi est un risque. Actuellement, beaucoup d'employé(e)s de la Fédération ont un contrat d'embauche à court terme alors que pour pouvoir recruter du PHQ et le maintenir en poste il faudrait offrir une stabilité à long terme (17). C'est tout particulièrement vrai dans les sciences humaines numériques où les salaires des administrateurs et développeurs de système ne sont pas des dépenses admissibles aux fins des subventions régulières (17, 54).

On a également souligné le fait que vu la vitesse à laquelle évoluent les technologies et les pratiques, le personnel de soutien professionnel doit recevoir de la formation et se perfectionner de manière permanente pour se tenir à jour pour avoir les compétences et l'expertise requises pour soutenir les chercheuses et chercheurs dans certains domaines (27; voir [Éducation et formation](#)).

Recommandations de la communauté :

- Les membres du personnel de soutien professionnel de l'Alliance devraient se former mutuellement dans leurs compétences pour éviter des points de défaillance uniques (92).
- Le personnel de soutien professionnel devrait se tenir au courant des développements en IRN afin de prévoir les besoins et suggérer les améliorations à apporter à la plateforme ou des remplacements à effectuer dès que la technologie est disponible (65).
- L'Alliance devrait soutenir le personnel de soutien professionnel au chapitre du perfectionnement continu et des possibilités de formation, des ressources d'apprentissage et des documents d'orientation dans les deux langues officielles (27).
- Développer des volets de financements précis ou des ententes pour couvrir le salaire du personnel de soutien professionnel de tous les paliers (national, régional, local) et assurer la sécurité d'emploi et la viabilité à long terme (15, 54).

4.9 Concours pour l'allocation des ressources

Nombre d'autrices et d'auteurs ont fait part de leur frustration et des difficultés que pose le processus du concours pour l'allocation des ressources (2, 15, 17, 21, 71, 92, 105). Il n'y a pas pour l'instant de connexion entre le processus de demande annuel et l'allocation des ressources et l'achat réel de ressources aux sites. Certains ont mentionné que les demandes elles-mêmes sont onéreuses et doivent être renouvelées fréquemment. D'autres ont mentionné que le calendrier du processus n'est pas en phase avec les calendriers de recherche et qu'il n'y a pas



encore de mécanisme permettant de demander du temps de calcul en dehors du cycle normal du processus du concours. Par exemple, nombre d'initiatives de calcul de haute performance demandent des ressources concentrées pendant de courtes périodes plutôt qu'un partage uniforme de ressources distribuées sur toute l'année. À l'autre bout du spectre, des autrices et des auteurs ont fait remarquer que nombre de chercheuses et chercheurs mènent des projets sur plusieurs années, ce qui demande l'accès à l'IRN pour plus d'une année avec le « volet accéléré » pour deux ans de plus après la première demande.

En plus de l'allocation réelle, des autrices et des auteurs ont insisté sur le fait que la communication entre les équipes des opérations et le comité du concours devrait être améliorée pour que les allocations répondent aux besoins de la recherche, tel qu'il est mentionné plus haut. Des autrices et des auteurs ont aussi manifesté leur frustration par rapport au processus du concours, certains projets de recherche ayant vu leur allocation subitement réduite au beau milieu des travaux à cause d'incertitudes quant à l'issue du concours annuel. Cette situation est particulièrement frustrante lorsque le projet a passé l'étape de l'examen par les pairs et s'est vu accorder des fonds des trois organismes.

Enfin, certains auteurs et autrices ont exprimé leurs préoccupations au sujet du processus du concours et les difficultés qu'il pose aux programmes de recherche qui doivent compter sur les allocations pour répondre aux besoins des projets de groupe ou d'un grand projet comportant de multiples composants de calcul distincts. Comme les allocations sont rattachées au principal chercheur, des difficultés surviennent dans les grands projets de groupe et les projets s'étendant sur plusieurs années dans lesquels il y a du remplacement.

Recommandations de la communauté :

- Aligner l'allocation des ressources sur les trois organismes (voir [Alignement sur les politiques des bailleurs de fonds et des établissements](#)).
- Accorder des allocations d'équipe sur plusieurs années.
- Créer un volet accéléré pour les chercheuses et chercheurs qui ont besoin de courtes périodes de calcul (p. ex. large, massivement parallèle ou en rafale) (02, 29, 105).

5 Soutien

5.1 Personnel de soutien professionnel

Nombre de chercheuses et chercheurs ont décrit le soutien reçu du personnel de soutien professionnel comme étant un élément de l'IRN tout aussi important que l'infrastructure matérielle, si non plus, considérant le rôle clé que joue ce personnel dans les établissements de recherche et à l'échelle nationale. [trad.] « *Les chercheuses et chercheurs se tournent en premier*



lieu vers les services informatiques, les bibliothèques, les bureaux de services de recherche, les centres de recherche/sciences humaines numériques, et d'autres unités du campus pour trouver la technologie et le soutien technologique lorsque c'est possible; en deuxième lieu vers des sociétés nationales ou des services financés par des fonds du gouvernement fédéral; et enfin, vers des produits commerciaux; quand ils sont à bout de ressources, ils abandonnent leurs projets. » (22)

5.1.1 Soutien national

Plusieurs auteurs et autrices ont aussi fait valoir qu'il fallait de l'expertise en IRN plus spécifique en certains domaines (18, 25, 67, 83). Des ressources humaines sont nécessaires pour travailler avec les groupes de recherche pour développer et exploiter notamment les instruments de code, les pipelines de données propres à leur matière. Les installations de niveau 2 de l'ATLAS canadien, par exemple, sont un modèle réussi d'intégration de personnel ayant une expertise en la matière au sein de la Fédération Calcul Canada et les sites hôtes (99, 92). L'équipe nationale de bio-informatique a mentionné les défis du recrutement qui limitent son incidence (25).

Les chercheuses et chercheurs des Sciences sociales et humaines (SSH) se sont exprimés sur le besoin de soutien additionnel sous la forme de logiciels et de calcul (55, 18). Dans les disciplines en SSH où on utilise moins souvent la technologie, il faut réapprendre à utiliser les outils à chaque projet (le soutien du système Linux a été suggéré pour l'installation et l'exploitation de serveurs Web), considérant que l'accès à du personnel de soutien professionnel spécialisé serait plus efficace et entretiendrait la « mémoire musculaire » des chercheuses et chercheurs (29).

Recommandations de la communauté :

- Il a été suggéré que l'Alliance se concentre sur les services qui maximisent l'utilisation de l'ensemble de l'écosystème de l'IRN, et non seulement ce qui se rapporte à elle, en aidant un chercheur ou une chercheuse à trouver les « pièces du casse-tête » et à les assembler pour trouver une solution (13).
- Il est proposé de soutenir la plateforme en tant que service pour la recherche en créant une équipe centrale de personnel de soutien professionnel chargée de soutenir toutes les plateformes de recherche qui s'appuient sur l'IRN (75).
- Il faut le soutien infonuagique d'une équipe nationale en vue de se doter de capacités nationales pour combler le manque d'administrateurs d'informatique et de système, de confidentialité et de sécurité, et d'expertise spécialisée en logiciels pour le codage, et réduire le dédoublement des efforts technologiques entre les projets et atténuer le risque d'iniquité (104, 38).

5.1.2 Soutien institutionnel

De nombreux exposés ont mentionné le rôle important du personnel de soutien professionnel réparti entre les établissements qui procure une aide en personne. Le rôle des bibliothèques en



particulier est jugé essentiel dans le soutien, la formation et l'orientation en IRN, notamment en ce qui a trait à la politique des trois organismes sur la gestion des données de recherche. Le personnel de soutien professionnel des bibliothèques a des liens directs avec les communautés de chercheuses et chercheurs et connaît les solutions des établissements. Cependant, les bibliothèques doivent souvent composer avec des difficultés de recrutement au regard de la capacité et du perfectionnement professionnel. Enfin, le soutien technique avancé en gestion des données de recherche s'inscrit souvent quelque part entre le mandat des bibliothèques et celui des services informatiques des établissements, un fossé que les chercheuses et chercheurs ont souvent du mal à combler (55).

Recommandations de la communauté :

- Le renforcement des capacités du personnel de soutien professionnel par des initiatives nationales comme Portage devrait se poursuivre (08, 20, 27, 28, 89, 101).
- Il est important de fournir aux établissements les fonds nécessaires pour se doter d'un ensemble de ressources humaines et soutenir leur perfectionnement professionnel pour assurer leur réussite et leur maintien en poste (8, 46, 67, 27).

5.2 Éducation et formation

On estime qu'il est impératif d'adopter différentes approches de formation afin de pouvoir répondre aux besoins de différentes communautés et à leur forme également différente de données. De plus la formation devrait être adaptée à la gamme des expériences (53).

5.2.1 Prestation inégale de la formation entre les régions et les établissements

Quelques exposés ont fait allusion aux différences importantes observées dans la prestation de la formation et les ressources de formation. Les différences dans la formation sur l'IRN entre les universités et même entre les départements en sont un exemple (46). De même, il y a des différences importantes au chapitre des possibilités de formation offertes par les consortiums régionaux (105).

Recommandation de la communauté :

- Soutenir le personnel des établissements est un moyen efficace de fournir de la formation et de la coordination à l'échelle nationale du soutien et de la formation des utilisateurs (67).



5.2.2 Formations des étudiants de cycles supérieurs

Des autrices et des auteurs ont mentionné que le peu de temps dont disposent les étudiants de cycles supérieurs pour lancer et commencer leurs recherches freine notablement le développement des compétences en matière d'IRN (15).

Recommandation de la communauté :

- Une formation plus agile pour aider les étudiants à comprendre les caractéristiques des plateformes de recherche et à repérer leurs besoins en infrastructure spécifiques à leur recherche (40).

5.2.3 Calcul

Certains ont mentionné que Calcul Canada fournit déjà les bases sur lesquelles les chercheuses et chercheurs pourraient construire leur propre système (p. ex. machines virtuelles, adresses IP, logiciels). Les connaissances restent tout de même lacunaires et il y a une bonne barrière à franchir en ce qui concerne l'utilisation de ces outils (15; voir [Accessibilité et facilité d'utilisation de l'infrastructure](#)). Il faut aussi de la formation avancée sur l'utilisation du CIP ou sur [trad.] « comment utiliser le CIP de manière intelligente » (105).

Recommandations de la communauté :

- Donner de la formation de niveaux intermédiaire et avancé sur les planificateurs, la programmation parallèle avec « MPI », l'*openACC* (NdT : standard de programmation pour le calcul parallèle) la gestion des versions et la visualisation parallèle à distance de larges ensembles de données. Formation sur les outils de diagnostic pour évaluer la performance, la communication, et la latence de lecture et d'écriture de simulations massivement parallèles (105).
- Il faudrait plus de formation adaptée aux disciplines. Par exemple, offrir des cours d'été thématiques orientés sur les disciplines et non les outils (105).
- Calcul Québec (49) décrit les outils et l'enseignement nécessaires et propose que ce travail soit confié à une équipe spécialisée qui pourrait travailler avec les chercheuses et chercheurs en vue de les guider dans la mise en application de pratiques.

5.2.4 Logiciels de recherche (LR)

Les logiciels de recherche sont un élément important de la recherche moderne, et alors que les groupes de recherche ont la capacité de recruter des développeurs, les petits groupes ayant moins de capacités comptent sur les postdoctorants et les étudiant(e)s de cycles supérieurs qui ont surtout appris par eux-mêmes.



Recommandation de la communauté :

- La stratégie de l'Alliance en matière de LR devrait s'intéresser aux pratiques exemplaires en développement de logiciels (67).

5.2.5 Gestion des données de recherche

Nombre de chercheuses et chercheurs demandent de la formation en GDR, notamment pour la curation et le stockage. Ils veulent connaître et pouvoir explorer la gamme des dépôts de données à leur disposition (02).

Recommandations de la communauté :

- Certains ont indiqué que plus de ressources de formation au regard de la préparation de plans de gestion des données étaient nécessaires pour que les chercheuses et chercheurs canadiens soient outillés pour rédiger des plans efficaces (82).
- Les chercheuses et chercheurs de l'Université Laval ont mis de l'avant le programme des champions (p. ex. <https://www.data.cam.ac.uk/intro-data-champions>) susceptible de faire avancer la mobilisation en faveur de la GDR dans les établissements (14). Ces champions des données pourraient entrer en contact avec les communautés de recherche et jouer un rôle de mentor et de formateur.

5.2.6 Littéracie (ou culture) des données

Pour réaliser les objectifs du Canada en recherche, il est primordial d'accorder une plus grande attention à l'émergence d'une culture des données (45).

Recommandations de la communauté :

- L'Alliance peut rassembler les communautés afin de les faire participer à l'avancement de l'IRN (p. ex. Réseau d'experts de Portage), y compris établir des normes et des pratiques exemplaires dans diverses communautés et champs de recherche, encourager l'adoption du partage des données, de la réutilisation et d'autres pratiques en GDR, organiser des échanges entre les chercheuses et chercheurs, les praticiens, les décideurs politiques, pour ne nommer que ceux-là (voir 46, 35, 54, 62, 71, 72, 27, 96, 40 et autres).
- Communiquer avec les organismes de financement et le Conseil des chercheurs de l'Alliance pour déterminer les disciplines pour lesquelles il y aurait lieu d'améliorer les compétences en littéracie des données.
- Évaluer s'il y a lieu de penser à des camps d'entraînement/cours d'été/subventions pour les étudiants de cycles supérieurs, chercheuses et chercheurs, dépositaires de données et assistant(e)s de recherche qui travailleront avec les chercheuses et chercheurs.
- Faire pression sur les universités pour qu'elles encouragent l'enseignement des compétences en littéracie des données dans toutes les facultés.



5.2.7 Souveraineté des données autochtones

Les Premières Nations ont besoin d'occasions et de soutien pour le renforcement de capacités, y compris de fonds pour la formation au soutien du développement de leurs propres capacités techniques, s'il y a lieu (36). De leur côté, les expert(e)s et les leaders en données du Canada tireraient avantage d'une formation sur les principes PCAP® (propriété, contrôle, accès possession) et d'une formation de sensibilisation à la culture (102, 36).

Recommandation de la communauté :

- Promouvoir une formation élargie sur les principes PCAP® et CARE et leur adoption par la communauté de recherche.

5.2.8 Formation pour le renforcement de la communauté

Des forums de partage du savoir ont été suggérés pour tirer avantage des ressources immenses de connaissances qui sont à la disposition des chercheuses et chercheurs (02, 24, 57, 83). Des modèles comme celui de FemTechNet sont vus comme d'excellentes ressources, encore plus que les sites techniques dominants tels que StackOverflow, qui désavantage les femmes (54).

Recommandations de la communauté :

- Il y aurait lieu de resserrer les liens entre les universités, les associations de discipline et les services de soutien de la Fédération Calcul Canada au moyen de formation spécialisée sur des sujets portant sur des matières spécifiques (105).
- Des occasions de formation et l'infrastructure nationale qui soutient une communauté de pratique en curation devraient être appuyées. Elles permettraient aux curatrices et aux curateurs de se réseauter, de partager leurs connaissances, leur expertise, les outils et les pratiques exemplaires (27).

5.3 Partenariats

Des modèles de soutien collaboratifs et coordonnés sont un moyen de défragmenter le paysage actuel (43, 27, 16, 15, 7, 102). Le modèle de soutien collaboratif sur lequel est fondé Dataverse Canada est un exemple de partenariat réussi entre les établissements et le soutien national qui pourrait servir de modèle pour d'autres services d'infrastructure partagés (p. ex. curation et préservation). Le réseau ontarien d'apprentissage axé sur la pratique des soins primaires (de l'anglais *Ontario Primary Care Learning Network*) est un autre exemple : gérer, nettoyer et analyser les dossiers médicaux électroniques de soins primaires par les réseaux d'apprentissage fondés sur la pratique qui établissent des partenariats entre les chercheuses et chercheurs, les clinicien(ne)s et les scientifiques des données (7).



Recommandation de la communauté :

- L'Alliance devrait favoriser les connexions et le soutien entre les établissements universitaires et les autres intervenant(e)s de l'écosystème de recherche élargi (comprenant les peuples autochtones, les dépôts des disciplines, les organismes non gouvernementaux et les groupes de recherche).

5.4 Financement durable

Plusieurs autrices et auteurs se sont dit préoccupés par le manque de fonds pour les projets existants aux prises avec des exigences opérationnelles : embauche ou maintien en poste de PHQ ou de personnel de soutien professionnel ou mises à jour de projet (20, 22, 102, 106). Les ouvertures actuelles favorisent l'embauche d'étudiant(e)s, ce qui est un avantage qui est important côté formation, mais qui ne permet pas de profiter de leur perfectionnement et de leur expérience acquise avec le temps (10, 16).

6 La gestion des données de recherche

6.1 Gestion des données et science ouverte

6.1.1 Interopérabilité

6.1.1.1 Données/métadonnées

Dans les exposés sur la science ouverte/la GDR, il a souvent été question de la nécessité de flux de travail interopérables faisant intervenir des données et des plateformes associées utilisées dans le processus de recherche. L'interopérabilité des données est un défi pour de nombreuses communautés de recherche, en particulier dans les sciences humaines, les sciences biologiques et pour les communautés qui travaillent avec des données qualitatives, ce qui fait que de nombreux projets numériques et de dépôts dans ces disciplines organisent et décrivent leurs données à leurs manières (01, 04, 52, 57, 83, 94). La production de métadonnées est souvent oubliée dans les budgets de projet. Cependant, cette façon de faire augmente les coûts pour les chercheuses et chercheurs et rend les données moins accessibles et interopérables (01). Il faut faire des analyses de rentabilité pour démontrer comment un investissement de départ dans les métadonnées rapporte des avantages en aval afin d'encourager le développement et l'adoption de normes dans ces communautés de recherche (24). Même s'il existe des normes de métadonnées et des ontologies, il faut des outils spécialisés pour saisir et produire les métadonnées et gérer les flux de travail (04, 61, 52, 24, 83).



La création automatisée de métadonnées est un secteur qui mériterait plus de soutien; tout comme l'indexation automatisée et le couplage des données (74, 83). Les processus de couplage de données avec d'autres ensembles de données peuvent être fastidieux et demander beaucoup de ressources. Des processus et des procédures plus souples sont nécessaires (7). Le couplage protégeant la confidentialité est nécessaire pour connecter les données provinciales (p. ex. en santé et en éducation) avec les données fédérales (p. ex. recensement, revenus personnels, immigration; 96). Plus de dossiers administratifs corrélables devraient être accessibles aux chercheuses et chercheurs (96; voir [Partage des données permettant d'identifier une personne et des données sur la santé](#)).

Il faut plus de soutien pour rendre les données de recherche qualitatives interopérables. Actuellement, les normes régissant les métadonnées sont d'abord conçues pour les données quantitatives et elles pourraient être extrapolées aux données qualitatives, ce qui nécessiterait des dépôts et des normes plus spécifiques (94).

6.1.2 Plateformes

Les plateformes de recherche sont importantes dans l'IRN; elles permettent l'interopérabilité des données. L'Alliance doit soutenir les demandes de financement de nouvelles plateformes (*innovations centrées sur les chercheuses et chercheurs*) et d'améliorations aux plateformes existantes (*développement d'options dirigé par les chercheuses et chercheurs*) (18). Pour connecter les plateformes, il faut un ensemble d'outils d'accès intermédiaire sûr et fiable (p. ex. interface de programmation applicative, authentification unique) capable de connecter facilement les plateformes et les infrastructures de calcul sous-jacentes (18).

Il faut coordonner les développements de la technologie entre les plateformes de recherche, les dépôts de données, les systèmes de traitement de préservation et le stockage à long terme pour rendre l'interopérabilité possible (28). Les systèmes de préservation numérique ne fonctionnent pas bien avec les systèmes de gestion de contenu, ce qui force un dédoublement des efforts d'ingestion et de curation des actifs numériques de la part des systèmes parallèles (18).

Il faut aussi des connexions entre les plateformes pour coupler les données de recherche et les autres types de données, en particulier en santé et médecine. Le Réseau de recherche sur les données de santé du Canada dirige un projet au pays visant à rapprocher les données de recherche aux dossiers cliniques, facturations provinciales, imagerie diagnostique, rapports de renvoi et données de laboratoire (5).

L'Association des bibliothèques de recherche du Canada a aussi mentionné la nécessité de rendre interopérables les dépôts de données, les dépôts des établissements, les ressources éducationnelles ouvertes et autres plateformes soutenant la recherche ouverte (08).



S'ajoute le besoin de systèmes de gestion de l'information de recherche interopérables afin d'améliorer la collecte des données et leur comparaison entre les établissements, apportant des avantages pour le développement de graphiques de recherche (08, 09, 32).

6.2 Normes

On a souvent cité la nécessité de soutenir le développement de normes pour les données et les outils. Des dépôts en biosciences, par exemple, appliquent des méthodes différentes pour présenter, décrire et extraire les objets-données (04, 52). Certaines disciplines appliquent leurs propres normes que l'IRN nationale doit pouvoir accepter afin d'assurer l'interopérabilité (p. ex. GA4GH, International Image Interoperability Framework et autres normes mentionnées) (1, 12). Pour les données de santé, les organismes provinciaux ont leurs propres procédures de contrôle de la qualité et d'uniformisation (p. ex. Institute for Clinical Evaluative Sciences en Ontario et PopData en Colombie-Britannique), qui doivent être harmonisées (p. ex. travail du Réseau de recherche sur les données sur la santé du Canada) (5). Pour plus d'information, voir [Partage des données permettant d'identifier une personne et des données sur la santé.](#)

Recommandations de la communauté :

- Il y aurait lieu de pousser les efforts de certification basée sur des normes pour les dépôts de données et l'infrastructure de préservation afin de soutenir et de connecter des infrastructures sécurisées (28).
- Il faut développer/améliorer des ontologies normalisées qui définissent des vocabulaires et harmonisent les données de différents systèmes pour établir une seule manière commune de décrire et de classer les ensembles de données (04, 12, 83).
- Il faut aussi des normes régissant la gouvernance, les politiques et l'architecture des systèmes et des établissements (12, 32, 52, 83).
- Il faut aussi des cadres de sécurité basés sur les standards de l'industrie (p. ex. applications de signature numérique, authentification, autorisation, prévention des intrusions) (12, 83).

6.3 Identifiants pérennes

L'exposé de DataCite Canada et du consortium de l'identifiant ORCID (*Open Research and Contributor ID*) décrit la nécessité de prendre en charge une masse critique d'identifiants pérennes pour atteindre leur objectif (89, 90). DataCite Canada et l'ORCID-CA fournissent en ce moment à de nombreux établissements de recherche canadiens l'accès à deux identifiants clés, ORCID et DataCite Digital Object (6).

D'autres autrices et auteurs ont suggéré que les systèmes d'identifiants courants soient basés sur des registres centraux, ce qui pose des questions d'interopérabilité, de fragmentation et de sécurité, ainsi que des coûts de maintenance. Ils suggèrent que plutôt que de maintenir les registres ou des registres de soutien, l'Alliance soutienne l'infrastructure (p. ex. outils, politiques,



logiciels, connaissances, formation) qui donne aux organismes et aux communautés la possibilité d'exploiter et de maintenir leurs propres systèmes d'ID. Ils suggèrent de plus de commencer avec un énoncé de principe et d'énumérer les technologies existantes qui pourraient être mises à profit (26).

Recommandations de la communauté :

- L'Alliance devrait soutenir le développement d'identifiants pérennes et encourager leur adoption par l'intégration dans l'ensemble de l'infrastructure gérée/financée.
- Une stratégie nationale en matière d'identifiants est nécessaire pour se renseigner sur l'adoption de nouveaux identifiants (p. ex. Research Activity Identifier (RAiD), GrantID) (6, 90).
- L'Alliance devrait intégrer tous les identifiants de ressources numériques dans son flux de travail et ses produits (26).
- Créer [trad.] « *un écosystème d'identifiants de ressources décentralisés pour la recherche qui sont plus faciles à créer et à tenir à jour, moins coûteux, plus flexibles et qui éliminent le danger de points de défaillance uniques* ». (26).

6.4 Plans de gestion des données

L'assistant PGD (plan de gestion des données) de concert avec de nombreuses autres ressources du Réseau Portage fournit aux chercheuses et chercheurs les outils pour développer un solide PGD. Ces ressources orientent leurs utilisateurs sur les pratiques exemplaires, mais il est quand même nécessaire que l'infrastructure numérique soit en place pour les adopter et pour que les chercheuses et chercheurs se conforment aux exigences de la politique des trois organismes à venir (82).

Les chercheuses et chercheurs reçoivent aussi des directives sur la manière de gérer leurs données du comité d'éthique de la recherche. Pour pouvoir les appliquer, il leur faut l'infrastructure et le soutien.

L'Assistant PGD du Réseau Portage soutient les pratiques de gestion des données bonifiées et répond aux exigences de la nouvelle politique des trois organismes sur la gestion des données de recherche. Cependant, il ne suffit pas à lui seul et il doit être appuyé par des conseils d'experts et des ressources (65, 103).

Le document présenté sur les PGD mentionne que la demande d'aide en planification avait augmenté, qu'il fallait divers modèles et exemples pour des disciplines et des communautés et du soutien multilingue (82).



Recommandations de la communauté :

- On considère que de l'orientation sur l'évaluation des PGD fait grandement défaut (40) et que l'établissement de normes nationales pourrait y remédier (32).
- Il est nécessaire que l'Alliance consolide le soutien pour les PGD et fournisse une norme nationale de référence (32).
- On pourrait mieux aligner les applications du conseil de l'éthique en recherche et les PGD (15).
- L'Alliance pourrait intégrer les PGD dans le Concours d'allocation des ressources de la Fédération Calcul Canada (82; voir [Concours pour l'allocation des ressources](#)).

6.5 Soutien de la gestion des données de recherche à l'étape active de la recherche

Il devrait y avoir une meilleure connexion entre la GDR et le CIP. Actuellement, la GDR est laissée aux chercheuses et chercheurs, qui doivent s'en occuper sur l'infrastructure de Calcul Canada alors qu'ils en connaissent peu sur le sujet. Le document de Calcul Québec (49) décrit les outils et la formation nécessaires et propose qu'une équipe spécialisée soit formée pour travailler avec les chercheuses et chercheurs et les guider dans la mise en œuvre de pratiques de GDR dans le CIP. Les autrices et auteurs ont fourni des exemples : [trad.]

- « *aider les chercheuses et chercheurs à structurer leurs données, améliorant leur efficacité et leur capacité à travailler au sein d'une équipe;*
- *prévoir la création des métadonnées, facilitant ainsi leur préservation et leur réutilisation future;*
- *prévoir le transfert de la propriété intellectuelle entre les étudiant(e)s et le principal chercheur à la fin du cycle de l'étude;*
- *éviter d'encombrer les systèmes de fichiers avec des fichiers temporaires non supprimés;*
- *éviter de copier des bases de données de référence téléchargées plusieurs fois par différentes personnes dans un répertoire local; le copier plutôt que dans un dépôt central commun ».*

Recommandations de la communauté :

- Il faut davantage de soutien au chapitre de la GDR à l'étape de recherche active, et l'Alliance pourrait fournir les outils et l'aide pour combler le fossé entre la GDR et le référentiel à long terme (15). Les secteurs demandant plus de soutien, selon les documents présentés, étaient la structuration des données, la gestion des fichiers et la gestion des versions, la création/entrée des métadonnées et la connexion des types de stockage.
- Le rapport présenté sur la curation nationale (27) mentionnait aussi que la possibilité de réutiliser les données et de reproduire les études serait maximisée si l'aide à la



curation était intégrée dans tous les processus d'aide à la recherche, y compris l'étape de recherche active.

6.5.1 Curation des données

Les bibliothèques de recherche ont souvent été mentionnées comme des joueurs clés du soutien à la curation des données dans les établissements d'enseignement supérieur, qui demandent un soutien national coordonné dans ce travail (20, 27, 36, 40, 67, 89).

Une quarantaine de dépôts de données uniques sont mentionnés dans les documents. L'élaboration de normes, ontologies et de politiques pour des données et des plateformes interopérables est à la base de ces infrastructures. Cependant, leur succès repose sur un personnel de soutien professionnel qualifié (p. ex. des curateurs, curatrices) pour développer et appliquer ces normes, ontologies et politiques aux ensembles de données, afin d'harmoniser et d'intégrer les infrastructures.

Recommandations de la communauté :

- Il faut des volets de financements spécifiques qui soutiendraient le personnel de soutien professionnel pour qu'il développe et maintienne ces dépôts et les infrastructures sous-jacentes, et en conserver les données (67).
- Le travail associé à la curation est fastidieux et le soutien est limité (p. ex. temps, formation, rémunération) (57, 94). L'Alliance est bien placée pour améliorer le soutien à la curation nationale (54).
- Les recommandations du rapport sur la curation nationale portent sur les partenariats et la collaboration nécessaires pour soutenir la curation. Par exemple, l'Alliance devrait soutenir les mécanismes de collaboration interinstitutionnelle pour que les curateurs, curatrices et autre personnel de soutien professionnel partagent les ressources, les pratiques exemplaires, etc. (27).

6.6 Gestion des données sensibles

Plusieurs autrices et auteurs ont confirmé que le manque de solutions sûres et normalisées pour la gestion des données sensibles était un défi majeur qui appelait une solution fédérative qui tienne compte des différences dans la réglementation au sujet de la confidentialité et dans normes régissant les données de santé entre les provinces, territoires et nations autochtones (voir [Partage des données permettant d'identifier une personne et des données sur la santé](#)). Un tel effort doit se traduire par une collaboration entre les « piliers » et le travail qui est en cours devrait être soutenu (p. ex. CanDIG, Réseau de recherche sur les données de santé, Réseau canadien des Centres des données de recherche) (67, 96).

Il manque une approche coordonnée de collecte de données relatives à la race, à l'origine ethnique et aux autres déterminants sociaux de la santé. Il faut que les membres des



communautés touchées participent en grand nombre à son élaboration (p. ex. cadres stratégiques, normes) (35, 79).

Plusieurs autrices et auteurs ont rappelé que les questions éthiques concernant les données sensibles – et certains ensembles de données – ne devraient être réutilisables (94). Cependant, il faut des infrastructures dont les protocoles de stockage et d'accès peuvent être coproduits et négociés avec les propriétaires de données (p. ex. données des peuples autochtones, données de santé personnelles) (94, 102).

Certaines données font l'objet d'une découverte restreinte et sont peu consultées à cause de leur taille, des ressources et des conditions technologiques. Les dépôts doivent comporter des fonctions permettant de soutenir ces données (p. ex. contrôle de l'accès, transfert sécurisé, licence; voir [Gestion de l'identité](#)), et les communautés doivent développer les métadonnées connexes, les normes de rapport et les politiques à l'appui de la découverte.

Les établissements de recherche n'ont pas accès aux ressources pour gérer les renseignements confidentiels dans l'IRN et qui sont conformes aux exigences législatives pertinentes. Par exemple, les outils pour sécuriser le transfert de données et le soutien avec les pratiques de gestion de base (p. ex. anonymisation et dépersonnalisation) (32, 40).

Recommandation de la communauté :

- Les chercheuses et chercheurs disposent de peu de ressources pour planifier l'utilisation secondaire des données au-delà de l'EPTC 2 (*Énoncé de politique des trois conseils : Éthique de la recherche avec des êtres humains*). Alors que les comités d'éthique de la recherche disposent peut-être des ressources nécessaires pour soutenir les chercheuses et chercheurs, les pratiques ne sont pas uniformes. Il faut de meilleurs outils pour soutenir les deux utilisations secondaires des données sensibles, y compris des outils de chiffrement et des environnements sécurisés pour réglementer l'accès aux données (39).

6.7 Découverte des données

On a souligné le progrès fait par Portage *via* le Dépôt fédéré des données de recherche pour soutenir la découverte de données grâce à un catalogue national et une couche de découverte, tout en reconnaissant qu'il fallait s'en occuper davantage et voir à mieux intégrer les logiciels pour la découverte des données.

Les registres des types de données qu'il est difficile de localiser doivent être améliorés : données à accès limité (37) (y compris les données de santé (04, 05, 35)), les données géospatiales et cartographiques (60), les collections de données historiques (43) et les archives Web (89).

Beaucoup de ces ressources sont régies par des organismes gouvernementaux comme Bibliothèque et Archives Canada, Secrétariat du Conseil du Trésor du Canada,



Statistique Canada, Ressources naturelles Canada, et leurs pendants provinciaux, avec lesquels l'Alliance devrait établir des partenariats.

Il faut un répertoire interopérable avec d'autres registres et catalogues (nationaux et internationaux) pour faciliter l'intégration des données de recherche canadiennes dans le Global Open Research Commons (44).

6.8 Numérisation

Il faut sans tarder documenter les renseignements historiques, le patrimoine culturel et le savoir, en particulier dans les peuples autochtones, ce qui demande une aide nationale pour la collecte et la préservation des langues et des histoires orales en conformité avec la gouvernance des nations en cause (29, 102). Le plan d'action pour le patrimoine autochtone de Bibliothèque et Archives Canada et l'initiative d'autochtonisation de la Colombie-Britannique en sont des exemples (102).

Recommandation de la communauté :

- L'infrastructure devrait permettre la collaboration avec les institutions culturelles et les peuples autochtones pour soutenir la numérisation du patrimoine culturel (16, 102).

6.9 Soutien de la gestion des données de recherche à long terme

L'exposé du groupe d'experts sur la préservation numérique de Portage et du groupe de travail sur la préservation de l'Association des bibliothèques de recherche du Canada explique qu'il n'entre pas dans le mandat de Calcul Canada de fournir des services de soutien des données de recherche au-delà du cycle de vie d'un projet de recherche financé. Cette approche est de courte vue et incompatible avec les besoins des recherches à grand volume de données (28).

7 Logiciels de recherche

7.1 Stratégie durable en matière de logiciels de recherche

L'importance des logiciels de recherche comme produit scientifique a été souvent mentionnée. Plusieurs autrices et auteurs cependant ont expliqué qu'il faut des ressources locales pour leur développement, en particulier sous la forme de formation et de soutien. De plus, ils ont souligné que trouver des collaborateurs ou des consultants peut être difficile (54) et que de compter sur des développeurs à court terme et des étudiants diplômés pose des problèmes pour la maintenance et la durabilité des logiciels (63, 67, 38).



Recommandations de la communauté :

- Nombre d'exposés suggèrent de financer le travail de développeurs au sein des groupes de recherche (17, 38, 92, autres). En particulier, ils recommandent que l'Alliance prenne le relais pour le fonds de développement de logiciels de recherche local de CANARIE. Le temps compte pour les équipes qui manqueront de fonds en 2021.
- Les autrices et auteurs de la communauté de la physique subatomique font remarquer que l'intérêt de CANARIE au regard du financement des logiciels est passé de l'impact aux mesures axées sur la valeur et devrait être revu. Il est recommandé que les fonds dans ces domaines soient concentrés sur les projets à haute priorité qui ont un grand impact international selon l'examen des pairs.
- Le développement de nouveaux logiciels devrait être encouragé par des projets pilotes plutôt que par des tentatives de prédiction de réussite de chaque outil (92).
- Le développement d'outils de source ouverte devrait être soutenu pour que soient réduits les obstacles causés par les formats de données exclusifs et les restrictions de certaines licences (10, 84).

7.2 Logiciels sous-jacents du calcul informatique de pointe (CIP)

Le rapport du département d'informatique de l'université de Victoria plaide en faveur du financement des logiciels de recherche pour soutenir les éléments logiciels qui forment l'infrastructure partagée de base et les plateformes sur lesquelles sont exécutées toutes les applications des utilisateurs du CIP. Par exemple, les sites hôtes utilisent une série de logiciels de source ouverte comme le Simple Linux Utility for Resource Management pour l'ordonnancement des tâches par lots, Lustre, un système de fichiers de nœuds partagés, et le système de fichiers *CernVM* pour la répartition des logiciels entre les sites.

Recommandation de la communauté :

- L'Alliance devrait financer un poste ou plus de développeur de logiciels réservés aux projets de logiciels de source ouverte et aux composants utilisés dans l'infrastructure de la Fédération Calcul Canada.

7.3 Plateformes de recherche

Toutes les communautés de recherche, des sciences aux sciences humaines, ont demandé que l'Alliance étende le soutien aux plateformes de recherche, appuyant le concept de plateforme en tant que service (« PaaS ») de l'infonuagique (54, 75), ce qui fait partie du mouvement de [trad.] « *démocratisation du CIP* », plus de chercheuses et chercheurs comptant sur des plateformes pour mener leur recherche (75). Pour un supplément d'information, voir [Besoins de calcul par discipline de recherche.](#)



Des autrices et auteurs ont parlé de la nécessité de services infonuagiques offrant des bacs à sable d'expérimentation (16) et des plateformes réservées au développement et à la mise à l'essai pour tester les mises à niveau ou les nouvelles expérimentations intégrées sans déranger les systèmes de production (21). Exemples cités : Kubernetes (55, 54), Git Lab, système de gestion de la relation client (de l'anglais *Customer Relationship Management* ou CRM) basé sur le Web (55) et HubZero (15). Des projets exploratoires ou la participation à des projets existants de la National Science Foundation sont suggérés à l'Alliance pour soutenir l'innovation dans ce domaine (14).

Recommandations de la communauté :

- Un dépôt national a été recommandé à plusieurs reprises pour rendre plus accessibles les codes sources et les logiciels financés par des fonds publics (56, 100) (voir 2.1 Sensibilisation aux infrastructures existantes). L'Institut national de recherche en informatique et en automatique de France est cité en exemple de même qu'une collaboration avec Software Heritage (56). Le portail CANARIE est un bon pas dans cette direction, mais les autrices et auteurs de la faculté des sciences de l'UQAM font remarquer qu'il ne soutient pas les aspects importants que sont l'hébergement ou la documentation, comme le fait GitHub (100).
- Vu le besoin de collectes de données pour recouper des sous-disciplines scientifiques (radiologie, optique, infrarouge et radioastronomie), il faut établir un portail scientifique unique pour l'astronomie permettant d'utiliser le spectre complet des données (30).

7.4 Logiciels sous-jacents de la gestion des données de recherche

La charge imposée par les tâches répétitives, y compris la curation et la découverte des données, devrait être transférée aux systèmes d'information par l'utilisation d'interfaces de programmation applicative (01, 04, 24, 74, 84, autres). Des logiciels de source ouverte fournis pour la curation, la présentation, le stockage et la découverte des données de recherche de différentes disciplines doivent être développés pour y arriver, en se basant sur les normes des organisations internationales (p. ex. GA4GH) (04, autres). Alors qu'il existe de nombreuses normes et que de nouvelles voient le jour, l'absence d'outils est un obstacle majeur (24).

8 Calcul informatique de pointe

8.1 Renouvellement des investissements en CIP

L'importance de l'infrastructure de calcul informatique de pointe est revenue dans tous les exposés des chercheuses et chercheurs universitaires. Nombre d'autrices et d'auteurs ont



expliqué que pour que les chercheuses et chercheurs canadiens restent dans la course, il fallait des investissements durables dans les moyens computationnels. De plus, ces investissements dans les dépenses en infrastructure de CIP devraient permettre de faire face à l'augmentation prévue des charges de travail des chercheuses et chercheurs canadiens (voir [Besoins de calcul par discipline de recherche](#)). Par exemple, l'exposé de l'observatoire Simons a confirmé que ses besoins annuels de traitement devraient augmenter à environ 30 000 années de base d'ici 2029 ou être 18 fois plus élevés que les cycles anticipés nécessaires pour le traitement du télescope cosmologique d'Atacama en 2021-2022 (71). De même, le sondage de Génome Canada a montré que l'infrastructure existante, bien qu'étant incroyablement utile, n'offre pas suffisamment de puissance de calcul pour traiter et analyser les données génomiques (61).

Un document venant de toutes les disciplines qui s'appuient sur des superordinateurs parallèles présente l'analyse en faveur d'une voie de renouvellement et d'expansion ambitieux pour de grands moyens de simulation parallèle dans l'écosystème de l'IRN canadien pour profiter des progrès de l'heure et étendre les ressources humaines spécialisées connexes requises pour la découverte scientifique (19).

Recommandations de la communauté :

- L'Alliance devrait commencer un programme d'investissement durable dans la capacité computationnelle (17).
- Investir dans le traitement de données en temps réel sur des ressources de calcul et d'unité de traitement graphique avec rétroaction rapide pendant les procédures expérimentales, avec connexion à des réseaux de transmission et de détection de données stables (14).

8.2 Services 24/24, 7/7

Certains ont insisté sur la nécessité de services et de soutien continus pour les plateformes nationales. En particulier, ils expliquent que certains projets comme ceux dans lesquels ils participent avec des équipes de recherche internationales, demandent l'exécution continue des tâches et donc du soutien en continu (24/24, 7/7) (21, 63, 92).

Recommandation de la communauté :

- L'Alliance doit s'assurer que les installations sont conçues et gérées pour être entièrement opérationnelles à long terme et pour que les périodes d'indisponibilité programmées ne fassent pas perdre l'accès aux ressources (92).



8.3 Communications sur le CIP

Certains systèmes d'aide à base de billets actionnés par courriel sont inadéquats; ils ne peuvent fournir l'information nécessaire à la bonne personne au bon moment; ce qui cause des retards qui aggravent la confusion sur l'identité de la personne-ressource responsable.

Recommandation de la communauté :

- Adopter des outils de communication tels qu'un système transparent de suivi des problèmes (54).

8.4 Infonuagique commerciale

Plusieurs ont rappelé l'importance des services d'infonuagique commerciale pour les chercheurs et chercheuses et que l'Alliance devrait se mesurer à ce genre de services pour des services similaires à court terme (56). Les autrices et auteurs ont commenté le besoin d'un accès flexible pour les chercheuses et chercheurs sans propositions excessivement longues et que nombre d'entre eux utilisent l'infonuagique commerciale par commodité (accès direct, facilité d'utilisation) – en particulier lorsque de tels services ne sont pas offerts par l'IRN nationale (p. ex. collaboration, allocations) ou parce qu'ils ne connaissent pas les options locales/régionales/nationales (20, 61).

D'autres ont exposé les barrières qui freinent l'accès aux services infonuagiques pour les chercheuses et chercheurs, notamment les risques relatifs à la confidentialité et à la sécurité (voir Risques, ci-dessous), les problèmes relatifs aux modèles de financement qui ne s'appliquent pas aux dépenses de fonctionnement, le manque de ressources réservées au soutien de l'utilisation de l'infonuagique (107). Les règlements actuels régissant les subventions, en particulier, ne permettent pas aux chercheuses et chercheurs d'acquitter les frais liés à l'infonuagique commerciale. [trad.] « *Une approche hybride serait le moyen le plus efficace de maximiser les avantages des investissements de l'Alliance dans le service infonuagique de Calcul Canada Cloud compte tenu des besoins infonuagiques complexes des chercheuses et chercheurs canadiens* » (50).

Enfin, les autrices et les auteurs de l'université de Victoria décrivent le risque d'iniquité dans l'offre de l'infonuagique. Certains domaines de recherche ou des petits projets dans les domaines émergents n'ont pas eu droit au même niveau de soutien que les autres domaines. De plus, ils évoquent le fait que la demande de ressources infonuagiques dépasse la capacité nationale existante, ce qui conduit les chercheuses et chercheurs à se tourner vers les solutions infonuagiques commerciales (104). Ces mêmes auteurs parlent de l'inaccessibilité de l'infrastructure en tant que service (de l'anglais « *infrastructure-as-a-service* » ou « *IaaS* ») infonuagique parce qu'elle demande trop de compétences techniques que les petits groupes de recherche ne peuvent se permettre de développer (voir [Accessibilité et facilité d'utilisation de](#)



[l'infrastructure](#)). Passer à la plateforme en tant que service (« PaaS ») abaisserait la barrière à l'entrée pour laisser les chercheuses et chercheurs se concentrer sur leurs recherches (104).

Recommandations de la communauté :

- Une approche hybride de soutien à l'accès à l'infonuagique commerciale est suggérée (50). Par exemple, à utiliser pour les demandes de pointe qui excèdent la capacité des ressources existantes (92).
- Génome Canada recommande la création d'une plateforme numérique canadienne en génomique qui allie la flexibilité du secteur privé aux ressources subventionnées et à la surveillance du secteur public (61).
- L'Alliance pourrait soutenir les établissements de recherche en achetant des ressources de l'infonuagique commerciale et en fournissant une voie d'accès aux services d'accès aux établissements et aux chercheuses et chercheurs canadiens (02).
- Développer un service à court terme qui peut se mesurer à l'infonuagique commerciale. Les éléments clés de ce service sont la flexibilité et l'accessibilité. L'Alliance pourrait peut-être offrir ces services à un prix plus compétitif.

8.4.1 Les risques liés à l'infonuagique commerciale

Soutenir l'infonuagique commerciale est difficile dans le modèle de financement actuel, qui finance l'infrastructure de recherche en couvrant les dépenses d'immobilisations plutôt que les dépenses de fonctionnement (107).

Il y a également le risque que les fournisseurs d'infonuagique commerciale soient motivés par l'appât du gain, ce qui nuit aux efforts universitaires partagés (51). Par conséquent, la plupart des barrières qui freinent l'utilisation des ressources infonuagiques viennent de préoccupations au sein des établissements au sujet de la protection des données personnelles et des coûts exorbitants liés aux grands volumes de données (86, 61). Pour plus d'information, voir [Influences du secteur privé](#).

Recommandations de la communauté :

- En s'engageant dans l'infonuagique commerciale, l'Alliance devrait régler les questions relatives à la confidentialité et au manque de soutien en développant une approche nationale standard d'adoption au moyen d'un cadre de gestion des données de recherche qui porte sur le cycle de vie complet des données et l'expertise nécessaire à chaque phase (107).
- Sécuriser les données de recherche du Canada en tirant profit de services partagés de cybersécurité (88).



8.5 Besoins de calcul par discipline de recherche

8.5.1 Intelligence artificielle (IA) / Apprentissage automatique (AA)

La stratégie pancanadienne en matière d'IA de 2017 a fait grandir les besoins computationnels en recrutant plus de chercheuses et chercheurs en IA au Canada. Dans son exposé qu'il a présenté, l'institut Vector estime que la demande d'IA pour du calcul par unités de traitement graphique (« GPU ») est comparable à la capacité entière en GPU des installations de la Fédération Calcul Canada d'aujourd'hui (48). Pendant ce temps, certains notent que l'accès à plus de calcul par unités de traitement graphique au moyen de la plateforme nationale est nécessaire pour soutenir les chercheuses et chercheurs (non affiliés à l'un des instituts d'IA) (103, 32, 61). Du côté des opérations, les calendriers à partir desquels les centres de données budgétisent les mises à niveau posent un problème, ce qui est incohérent avec le rythme des avancées technologiques, en particulier dans les domaines liés à l'IA. Il faut plus de flexibilité pour mettre les systèmes à niveau en fonction des technologies les plus récentes (48). L'accès aux unités de traitement de tenseur (de l'anglais *tensor processing units* ou TPU) sera attendu par la communauté pour soutenir la croissance de l'intelligence artificielle (32).

Recommandations de la communauté :

- L'infrastructure des provinces nécessaire pour soutenir la prise en charge des approches d'intelligence artificielle et d'apprentissage automatique dans les recherches cliniques et en santé est considérée comme prioritaire pour le développement des capacités (12, 85).
- L'institut Vector recommande à l'Alliance d'augmenter ses dépenses en infrastructure de CIP afin de pouvoir absorber l'augmentation attendue de la recherche et des besoins en IA; il est aussi suggéré que l'Alliance établisse un modèle de gouvernance paritaire des opérations et des capacités des systèmes qui réuniraient les centres de calcul régionaux et les instituts d'IA (48).

8.5.2 Sciences sociales et humaines

Le Centre de recherche interuniversitaire sur les humanités numériques fait remarquer que souvent la recherche en sciences humaines a besoin d'un hébergement sur le Web des applications spécialisées, des développements de logiciels et des services de préservation des données, et moins souvent d'un accès aux unités de traitement GPU/CPU (16).

Recommandations de la communauté :

- Mener les politiques et fournir le financement (p. ex. par des programmes pilotes) pour servir de catalyseur à l'innovation plus rapide dans l'écosystème de recherche à grand volume de données, p. ex. permettre le croisement d'ensembles de données (souvent sensibles) qui demande une puissance de calcul accrue (96).



- Fournir une plateforme technologique agile (basée sur l'infonuagique) qui répond à divers flux de travail/pipelines utilisés par les sciences humaines dans leurs travaux avec un soutien en personne sur le terrain (y compris le développement de code et les applications web), avec de la flexibilité dans l'assemblage des services (20). Cela devrait inclure l'offre d'une IRN prête à l'emploi et le soutien conséquent pour le développement et des outils de communication savante en commun, tels qu'un cluster kubernetes, une instance GitLab ou un site web avec gestion de la relation client, plutôt que de laisser à chaque équipe de recherche les frais généraux et le gaspillage associés au fait d'installer et de retoucher elle-même ces systèmes (55).
- Fournir plus de personnel de soutien assigné au calcul informatique de pointe et au développement de nouvelles applications dans les sciences humaines, y compris dans les installations de CIP régionales (18)

8.5.3 Données sanitaires/biomédicales

8.5.3.1 Approche fédérative

Un sondage mené par Génome Canada auprès de ses utilisateurs et utilisatrices a révélé une mosaïque d'infrastructures de calcul non interopérables : [trad.] « À cause des obstacles à surmonter pour accéder à une infrastructure de recherche numérique, des infrastructures informatiques locales, non interopérables, ont été établies. Notre consultation a révélé de multiples exemples de données de projet qui s'étendent sur des serveurs de labo, des groupes institutionnels de calcul de haute performance, des groupes régionaux d'universitaires et de multiples nuages privés. »

Il faut des environnements de calcul collaboratifs pour exécuter des opérations complexes sur des pétaoctets de données cliniques et génomiques sensibles (12). Étant donné les besoins de telles données sur les plans législatif et administratif, il faut une capacité accrue pour exécuter des analyses des ensembles de données distribuées qui elles-mêmes ne voyagent pas d'une compétence à une autre (et non demander que les ensembles de données soient regroupés au même endroit) (04, 35; voir [Partage des données permettant d'identifier une personne et des données sur la santé](#)).

Recommandations de la communauté :

- Développer une approche fédérative à l'échelle nationale pour rassembler l'expertise et l'accès aux données de toutes les autorités de soins de santé au Canada. Il faut se doter d'une structure de gouvernance, d'une architecture, d'ontologies, d'outils d'interopérabilité pour relier tous les sites participants en un seul et même réseau cohérent. Par exemple, on peut obtenir des données distribuées en santé et en génomique en réunissant des initiatives (HPC4Health en Ontario, SD4Health au Québec, All For One, CanDIG et la plateforme de découverte et de santé numérique (de l'anglais *Digital Health and Discovery Platform*)) avec un seul point d'accès des utilisateurs et les rendant opérationnels à



l'échelle internationale grâce à des normes établies par la communauté de pratique internationale de la discipline (alliance mondiale pour la génomique et la santé ou « GA4GH »). Parmi les modèles recommandés d'association des données de tous les dépôts, les National Institutes of Health Data Commons sont mentionnés (84).

- Élaborer une stratégie nationale visant à répondre aux besoins présents et futurs en calcul, stockage et partage des renseignements personnels de santé. Pour rester pertinente et se mesurer à l'échelle internationale, cette stratégie doit aussi couvrir le calcul et le stockage dans l'infonuagique (86).

8.5.3.2 Environnement des fiduciaires de données

Les chercheuses et chercheurs ont analysé les environnements des fiduciaires de données dans les domaines de la santé. Ils parlent de la nécessité d'une infrastructure et de personnel (gestionnaires de données et administrateurs système) pour soutenir la recherche sur des données sensibles au moyen de droits et de chiffrement plutôt que des environnements restreints ou isolés (85, 38).

Recommandations de la communauté :

- Une plateforme sécurisée d'analyse et de stockage des données sensibles est nécessaire : elle doit répondre aux exigences techniques et juridiques selon lesquelles les données doivent être stockées sur le campus ou chez l'autorité provinciale (14).
- Le fort volume de données sensibles de santé générées aujourd'hui exige des technologies de calcul et des logiciels qui dépassent ce qui est disponible dans les centres du Canada (12).

8.6 Stockage

Comme pour la demande d'investissements accrus dans l'infrastructure de calcul, nombre de chercheuses et chercheurs de toutes les disciplines ont insisté sur la nécessité de solutions de stockage supplémentaire et plus rapide. Par exemple, l'exposé du iReceptor a relevé le besoin particulier de stockage facile d'accès par les nœuds de calcul pour traiter de grands ensembles de données (13). Génome Canada a souligné le coût élevé du stockage de données à long terme. Ces coûts représenteraient un sérieux frein à l'innovation en limitant la durabilité des actifs de données et en raccourcissant leur durée de vie utile (61).

Certains ont mentionné les problèmes relatifs aux systèmes de gestion de fichiers dans les plateformes nationales et leurs effets néfastes pour les chercheuses et chercheurs. Par exemple, la redondance des données entre les grappes, telle que les ensembles de données de référence à fort volume comme l'ensemble de données protéiques non redondantes de l'outil de recherche d'alignement local de base du centre national de l'information en biotechnologie, surcharge les systèmes et nuit à l'expérience utilisateur. Cette surcharge de systèmes de fichiers de travail temporaire a pour effet de les rendre volatils (25).



Recommandations de la communauté :

- Accorder plus de ressources pour l'archivage d'une durée excédant la durée du projet initial pour permettre les explorations secondaire et tertiaire des données (p. ex. en génomique) pour que les investissements de départ continuent d'être rentables en termes de découvertes et de solutions (61, 71).
- Investir dans la capacité de stockage et de calcul à laquelle on accède via des plateformes scientifiques de domaines – la capacité existante des centres des Calcul Canada sera épuisée à très brève échéance (30).
- Établir des protocoles et des politiques pour éviter d'encombrer les systèmes de stockage temporaire actif avec des données qui devraient être éliminées (49).



9 Exposés de position de référence

1. ["Good Things Come in Small Packets": How \(Inter\)national Digital Research Infrastructure can support "Small Data" Humanities and Cultural Heritage research par Daniel Paul O'Donne](#)
2. [NDRIO White Paper: Bridging the Accessibility Gap par the University of Waterloo](#)
3. [A CANADIAN SQUARE KILOMETRE ARRAY REGIONAL CENTRE par Kristine Spekkens, Erik Rosolowsky, Séverin Gaudet, Michael Rupen and the Association of Canadian Universities for Research in Astronomy](#)
4. [A Standards-Based Digital Infrastructure for Secure Sharing of Human Biomedical Research Data par Lincoln Stein de la part de Ontario Institute for Cancer Research](#)
5. [Digital Health Research: Opportunities for Improving the Health of Canadians par Diabetes Action Canada](#)
6. [Persistent Identifiers in Canada: ORCID-CA and DataCite Canada par John Aspler de la part de ORCID Canada Consortium, DataCite Canada Consortium, and the Canadian Persistent Identifier Committee](#)
7. [Ontario Primary Care Learning Network: ORACLE; EMR data to support research, quality improvement and innovation par the Ontario Primary Care Learning Network: ORACLE](#)
8. [Future-Proofing the Canadian DRI/RDM Ecosystem through Openness and Collaboration par Canadian Association of Research Libraries](#)
9. [Research Information Management Integration for Canada par Eugene Barsky](#)
10. [Digital research infrastructure of two research laboratories in the field of structural dynamics par the Laboratory for Acoustics and Vibration Analysis at Polytechnique Montréal and the Structural Dynamics and Vibration Laboratory at McGill University.](#)
11. [Canada's Health Workforce Digital Research Infrastructure Ecosystem: Building the Foundation for Canadian Health Workforce Science & Informed Decision Making par the Canadian Health Workforce Network](#)
12. [Digital research infrastructure to support federated computing on large scale biomedical datasets par Guillaume Bourque, Michael Brudno, Steven JM Jones](#)
13. [iReceptor – A case study in the challenges/opportunities in Canadian DRI par Felix Breden and Brian Corrie](#)



14. [Considerations from researchers from Université Laval par Jacques Corbeil, Francois Laviolette and Florent Parent](#)
15. [Gaps and Opportunities for NDRIO Support of Research Data Management: Insights from the Global Water Futures Data Managers Global Water Futures Data Managers par K. Dukacz and collaborators](#)
16. [Développer une infrastructure de services numériques pour les Humanités numériques canadiennes : livre blanc pour la NOIRN par the Centre de recherche interuniversitaire sur les humanités numériques](#)
17. [Digital Research Infrastructure for Canadian Astronomy par Catherine Lovekin, Pauline Barmby, J.J. Kavelaars, Adrian Liu, Erik Rosolowsky \(U of Alberta\), Kristine Spekkens, James Wadsley](#)
18. [Gaps in Digital Research Infrastructure for Canadian Digital Humanities Researchers par Lawrence Evalyn, Elizabeth Parke, Patrick Keilty, Elspeth Brown](#)
19. [Large-Parallel Supercomputer Simulations – Frontiers in Canadian Research par Rodrigo Fernández and 16 supercomputing researchers](#)
20. [Canada's Future DRI Ecosystem for Humanities & Social Sciences \(HSS\) Canadian Society of Digital Humanities par Geoffrey Rockwell](#)
21. [Submission to NDRIO's Call for Position Papers on Canada's Future DRI Ecosystem from the Compute Canada Federation Subatomic Physics National Team par Leslie Groer de la part de the Compute Canada Federation Subatomic Physics National Team](#)
22. [All Researchers Use Digital Resources: On Campus Support, Grants, Labs, and Equity par Laura Estill de la part de St Francis Xavier University Digital Humanities Centre](#)
23. [The need for expanding spatial data capabilities: Challenges and opportunities par Dr. Ines Hessler, ACENET](#)
24. [Empowering Information Systems and Fostering Metadata Driven Data Management par Pascal Heus](#)
25. [Canada's Future Digital Research Infrastructure \(DRI\) Ecosystem: A perspective from the Bioinformatics National Team \(BNT\) Bioinformatics National Team](#)
26. [The opportunities of Decentralized Resource Identifiers in the research landscape par Carly Huitema and collaborators](#)
27. [Closing Canada's "curation gap": A national approach par the Portage Curation Expert Group](#)



28. [Digital preservation and NDRIO: a white paper par the Portage–Preservation Expert Group and the Canadian Association of Research Libraries–Digital Preservation Working Group](#)
29. [Indigenous Language Technologies and Online Resources: Algonquian Dictionaries Project and Algonquian Linguistic Atlas par Marie-Odile Junker and Delasie Torkornoo](#)
30. [Digital Research Infrastructure in Astronomy Herzberg Astronomy and Astrophysics Research Centre par Dr. JJ Kavelaars](#)
31. [The Needs of Canada's Future Digital Research Infrastructure Ecosystem par the University of Ottawa](#)
32. [Beyond Virtual or Physical Environments: Building a Research Metaverse A White Paper for NDRIO's Canadian Digital Research Needs Assessment par Paul De Decker and Stephany Peterson](#)
33. [Toward Petabyte Scale Open Neuroscience: UBC Dynamic Brain Circuits Research Excellence Cluster Experience par the UBC Dynamic Brain Circuits in Health and Disease Research Excellence Cluster](#)
34. [Population Health Digital Research Infrastructure: Building Up and Out from a Learning Collaborative Network par the Health Data Research Network Canada](#)
35. [Regional Academic Library Consortia as key partners in building sustainable, responsive Digital Research Infrastructure par the Regional Academic Library Consortia](#)
36. [Improving the Discovery of Access-Limited Data par Kevin Read, Amber Leahey, Sarah Rutley, Victoria Smith, Kelly Stathis](#)
37. [On the Need for Local Research Software Development Funding par Andrew Schoenrock](#)
38. [Strengthening Capacity to Manage, Secure and Protect Sensitive Research Data par the Portage Expert Group on Sensitive Data](#)
39. [Towards an Ideal Digital Research Infrastructure \(DRI\) par Robyn Stobbs and Vivekanandan Kumar](#)
40. [Creating an Audit and Feedback Loop to Monitor and Enhance Data Sharing in Canada par Felicity Tayler, David Moher, Kelly Cobey](#)
41. [Software support for ARC infrastructure par Ryan Taylor](#)
42. [Ensuring access to Canada's core historical data collections in the social sciences and humanities: possibilities emerging from the Canadian Census Discovery Partnership par Leanne Trimble and collaborators](#)
43. [Canadian Data Repository Inventory Peter Webster \(St Mary U\) and collaborators](#)



44. [A Proposal to Strengthen Data Literacy in the Canada Research Data Ecosystem par Ernie Boyko and Wendy Watkins](#)
45. [ACENET's Research Directorate White Paper for NDRIO par Andrew Rutenberg de la part de the ACENET Research Directorate](#)
46. [Canada's Path to a Global Open Research Commons par Karen Payne, David Castle, and Mark Leggott](#)
47. [Canada's Future DRI Ecosystem: AI Research Needs par Ron Bodkin](#)
48. [Clés pour une Infrastructure de recherche numérique \(IRN\) innovante par Calcul Québec](#)
50. [The Future of Cloud as DRI par Deborah Stacey de la part de the Linked Infrastructure for Networked Cultural Scholarship project](#)
51. [Comments for NDRIO par Mark Hahn](#)
52. [Individuals, communities and population health digital research infrastructure: incorporating primary care electronic medical record data as part of a primary health care information system par Sabrina T. Wong de la part de the Canadian Primary Care Sentinel Surveillance Network](#)
53. [CSEE/CIEE Submission on the New Digital Research Infrastructure Organization par Stephen Heard and Diane Srivastava de la part de the CSEE and CIEE](#)
54. [Interface Matters Canadian Writing Research Collaboratory par Jeffrey Antoniuk and Susan Brown](#)
55. [HQP Pathways: Engaging the Canada's Different Disciplinary Models for HQP Training and Funding to Facilitate DRI Uptake in Canada par Ray Siemens and Alyssa Arbuckle de la part de the Digital Humanities Summer Institute Directorial Group](#)
56. [L'Infrastructure Numérique en Recherche Décanat de la Recherche et de la Création par the Université du Québec à Chicoutimi](#)
57. [Animal Behaviour in the 21st Century: A White Paper on DRI for Animal Behaviour Research in Canada Animal Behaviour Research in Canada – Dr. Laura Adams \(Langara College\) and collaborators](#)
58. [Écosystème D'infrastructure De Recherche Numérique Pour Les Données De La Recherche Vision Du Sous-Comité Des Bibliothèques Du BCI](#)
59. [Federating innovation in Digital Research Infrastructure accessibility Julie faure-Lacroix, Félix-Antoine Fortin, Florent Parent, Darren Boss, Drew Leske](#)



60. [Enabling broad reuse of Canada's geospatial data and digitized cartographic materials Association of the Canadian Map Libraries and Archives par Jay Brodeur and team](#)
61. [Genomics Digital Research Infrastructure to Support Canadian Competitiveness par Genome Canada](#)
62. [Role for NDRIO in advancing open data sharing and the use of domain-specific repositories par Carolyn DuBois and collaborators](#)
63. [Steps to Success in Ensuring DRI Engages and Mobilizes Humanities and Social Science par Research Ray Siemens and Alyssa Arbuckle de la part de the Implementing New Knowledge Environments Partnership](#)
64. [Évaluation canadienne des besoins en matière de recherche numérique – Livre blanc de l'École nationale d'administration publique, l'Institut national de la recherche scientifique et l'Université TÉLUQ École nationale d'administration publique, Institut national de la recherche scientifique, Université TÉLUQ](#)
65. [University of Alberta response to NDRIO call for CDN DRI needs assessment white papers University of Alberta](#)
66. [BRISER LES SILOS DE DONNÉES ET DE SAVOIRS Un enjeu scientifique d'inclusion sociale et culturelle Université du Québec à Montréal avec la collaboration de Calcul Québec](#)
67. [Response to NDRIO Call for Position Papers on Canada's Future DRI Ecosystem McGill University](#)
68. [New Digital Research Infrastructure Organization White Paper Submission par Svetlana Komarova par the Faculty of Dentistry](#)
69. [White Paper: Open Database Support and Curation par Nicholas Probart, Gary Bader, Lincoln Stein and Michael Brudno](#)
70. [NDRIO Cartik Sharma Proposal par Cartik Sharma](#)
71. [NDRIO White Paper: Envisioning Digital Research Infrastructure for the Simons Observatory par Adam Hincks and collaborators de la part de The Simons Observatory](#)
72. [NDRIO: Limit Corporate Influence, Maximize Public Involvement and Accountability par Catherine Stinson and collaborators](#)
73. [A stronger Canada through digital infrastructures that support primary care research and practice improvement par the College of Family Physicians of Canada / Le Collège des médecins de famille du Canada](#)



74. [Standardized, linkable, analysis-ready environmental data for understanding and preventing disease and building resilience to climate change par Jeffrey R. Broo, Dany Doiron and Eleanor Setton](#)
75. [Towards Research Platform as a Service \(RpaaS\) initiatives par Carol Gauthier and collaborators](#)
76. [Brandon University White Paper Submission to NDRIO](#)
77. [SOSCIP Consortium on the Importance of Driving Industry Innovation with Canada's New Digital Research Infrastructure par the SOSCIP Consortium](#)
78. [National Research Data Infrastructure in the context of Truth and Reconciliation par Ry Moran](#)
79. [Racism's trojan horse: race-based data, health care, and risks to Black and racialized populations par LLana James](#)
80. [On Canada's Future Digital Research Infrastructure par Ross Dickson](#)
81. [A Perspective on a Canadian Digital Research Infrastructure par the McGill Centre for Integrative Neuroscience McGill Centre for Integrative Neuroscience](#)
82. [Data Management Planning in Canada par James Doiron, Maggie Neilson, and Robyn Nicholson](#)
83. [Canadian Digital Research Infrastructure Needs Assessment: An Agri-food Research Perspective from the University of Guelph par Emily Duncan, Rozita Dara, Jairo Melo, Carly Huitema, and Beverley Hale](#)
84. [A Dynamic Environment for Medical Image Computing Research in Canada par Amber Simpson and collaborators](#)
85. [Building a data science platform for better health par Laura C. Rosella and collaborators](#)
86. [PHI-Compliant Computing and Storage: a critical need for Canadian biomedical and health research par Richard F. Wintle and colleagues for The Centre for Applied Genomics/The Hospital for Sick Children](#)
87. [Bilinguisme des outils et des services: soumission conjointe des Bibliothèques de l'Université Montréal et des Bibliothèques de l'Université de Toronto Bibliothèques de l'Université de Montréal et des Bibliothèques de l'Université de Toronto](#)
88. [ORION Digital research infrastructure for Canada par the ORION Network](#)
89. [People, Platforms, and Content: The Importance of Libraries in the Emerging National DRI par the University of Toronto Libraries](#)



90. [Persistent Identifiers in Canada – Position Paper par Eugene Barsky de la part de the ORCID Canada Consortium, DataCite Canada Consortium, and the Canadian Research Data Management Community](#)
91. [DRI, University Libraries and Digital Humanities Research Centres par Jason Boyd de la part de Ryerson Centre for Digital Humanities](#)
92. [White Paper on Canada's Future DRI Ecosystem: Subatomic Physics in Canada par Randall Sobie de la part de the Institute of Particle and the Canadian Institute of Nuclear Physics](#)
93. [Successes and shortfalls of the current Canadian ARC platform and ideas to improve it further par Maxime Boissonneault](#)
94. [White Paper: Building and Supporting the Use of Digital Research Infrastructure Among Qualitative Researchers par Katherine Tamminen, Heidi Bohaker, Andrea Bundon, Denise Gastaldo, Brenda Gladstone, Cara Krmpotich, Meghan McDonough, Brett Smith](#)
95. [The Canadian Primary Care Information Network: Harnessing primary health care data to improve practice par the Canadian Primary care Information Network](#)
96. [The Canadian Research Data Centre Network \(CRDCN\) in Canada's DRI Ecosystem par Byron G Spencer de la part de CRDCN](#)
97. [The Importance of Distributed ARC HQP par Chris Loken de la part de Compute Ontario](#)
98. [The Need for First Nations Collaboration and Co-development in Digital Research Infrastructure \(DRI\) par the First Nations Information Governance Centre](#)
99. [Role of TRIUMF within the Digital Research Infrastructure Ecosystem par Jonathan A. Bagger, Reiner Kruecken, and Reda Tafirout de la part de TRIUMF](#)
100. [Livre blanc sur l'infrastructure de recherche numérique : perspective de la Faculté des sciences de l'UQAM Faculté des sciences de l'UQAM White Paper McMaster Angela Di Nello, Research & High-Performance Computing Support \(McMaster U\)](#)
101. [NDRIO White Paper par Cape Breton University](#)
102. [Clés pour une Infrastructure de recherche numérique \(IRN\) innovante par Suzanne Talon, Marie-Jean Meurs, and Victor Ionescu de la part de Calcul Québec](#)
103. [Les consortiums régionaux de bibliothèques universitaires sont des partenaires clés qui contribuent à la mise en place d'une infrastructure de recherche numérique durable et réactive par Vivian Stieda, Sonia Seck, and Cynthia Holt](#)
104. [A Cloud Ecosystem for Canada's Digital Research Infrastructure par Wency Lum and collaborators](#)



105. [DRI Needs Assessment for the Computational Fluid Dynamics \(CFD\) Research Community par the Computational Fluid Dynamics Society of Canada](#)
106. [NDRIO: Call for Position Papers on Canada's Future DRI Ecosystem par the Ontario Council of University Libraries](#)
107. [Advancing a Framework for National Commercial Cloud Adoption in Research par Jeff Gardner, Nick Rochlin, and Elizabeth Kinney](#)



Annexe 3

Assemblées générales virtuelles dans le cadre de l'évaluation des besoins de l'Alliance

Table des matières

1.	Introduction	1
2.	Sensibilisation et accessibilité	1
2.1.	Comment améliorer l'accès à l'IRN.....	1
2.2.	Goulots d'étranglement dans le financement relatif à l'IRN.....	2
2.3.	Voies de découverte de l'IRN	3
3.	Gouvernance et politiques	4
3.1.	Données de santé	4
3.2.	Souveraineté des données autochtones	5
3.3.	Harmonisation des politiques, des pratiques et des procédures de gouvernance	6
3.4.	Collaborations des secteurs public et privé.....	7
3.5.	Adoption de la recherche ouverte.....	8
4.	Opérations.....	8
4.1.	Espaces de collaboration	8
4.2.	Indicateurs de rendement clés (IRC)	9
4.3.	Transparence	9
5.	Soutien	10
5.1.	Réseau collaboratif de soutien	10
5.2.	Champion(ne)s de l'IRN	11



1. Introduction

Entre le 4 et le 7 mai, l'Alliance a fait part de ses constatations préliminaires tirées du sondage et des exposés de position dans une série de quatre assemblées générales virtuelles. Ces constatations ont été regroupées sous quatre thèmes : Sensibilisation et accessibilité, Opérations, Gouvernance et politiques, et Soutien. Les séances étaient ouvertes à tous les membres de la communauté de recherche au Canada, sans égard à leur niveau de participation à l'IRN. En tout, 477 chercheuses et chercheurs, bibliothécaires, administrateurs et administratrices et représentant(e)s des organismes de financement y ont pris part.

La rétroaction a été recueillie en ateliers où les participant(e)s ont pu exprimer leurs points de vue, soit sous forme de commentaires soit directement dans un document commun. Ce qui suit est un sommaire des principaux points qui sont ressortis de ces séances et des discussions plénières. Les commentaires, idées et points de vue exprimés en rétroaction ont été regroupés par thème et consolidés ici. Les commentaires cités textuellement sont en italiques. Ces réflexions sont le fruit des échanges qui ont eu lieu entre toutes les personnes qui ont participé aux assemblées générales.

2. Sensibilisation et accessibilité

2.1. Comment améliorer l'accès à l'IRN

Il est essentiel de disposer d'agents de liaison sur place pour orienter les chercheuses et chercheurs vers les bonnes ressources de l'IRN nationale qui sont à leur disposition. Sans la couche locale de soutien, les chercheuses et chercheurs ne seraient pas correctement intégrés et guidés dans la découverte des outils et des services de l'Alliance.

L'incidence des groupes de travail, des séminaires de formation et des séances d'information pourrait être plus grande si ces efforts étaient coordonnés avec des groupes localisés de soutien à la recherche avancée. Si les chercheuses et chercheurs, y compris les étudiants du premier cycle et des cycles supérieurs, doivent suivre des cours sur l'éthique et la sécurité, pourquoi pas sur l'IRN?

[trad.] « L'université de Toronto offre des cours sur l'IRN ouvrant droit à crédit. Elle a dû établir de nouvelles politiques pour cautionner l'intégration de l'utilisation de l'IRN dans la formation et dans l'établissement. De plus, elle a mis des ressources en place, telles que des outils et des vidéos d'orientation complétés par des ateliers offerts dans l'établissement. Un tel modèle serait utile s'il était adopté par toutes les universités. »



[trad.] « Les sites Web sont des moyens additionnels d'améliorer l'accessibilité à l'IRN, mais ils deviennent généralement utiles une fois que les chercheuses et chercheurs ont compris le paysage de l'IRN canadien. Ils sont très utiles, mais ils ne suffisent pas. »

Il est très important de découvrir les outils et les ressources mis à la disposition des chercheuses et chercheurs, mais les plateformes nationales doivent être interopérables et fournir des points d'entrée semblables et être faciles d'utilisation. Une approche générale aux plateformes nationales pourrait permettre d'éviter le dédoublement (de plateformes exécutant des fonctions identiques ou similaires) et d'améliorer l'interopérabilité.

L'IRN se concentre de plus en plus sur les données, mais il faut s'occuper aussi du réseautage (p. ex. connexion des centres de données, téléversement et téléchargement des données, flux de travail). Les chercheuses et chercheurs utilisent des ressources locales parce que le téléchargement des données peut être plus lent via un réseau national, ce qui affecte la productivité, le soutien et l'accès à l'IRN nationale. De plus, l'habitude de recourir au courriel ou au système d'aide à base de billet cause d'importants retards aux chercheuses et chercheurs.

[trad.] « Je ne veux pas être un développeur de logiciel ou un administrateur système. Tout ce que nous voulons, c'est faire de la recherche et ne pas devoir attendre huit semaines avant d'avoir l'accès. »

2.2. Goulots d'étranglement dans le financement relatif à l'IRN

Il faut que le financement des ressources de calcul pour les chercheuses et chercheurs soit stable et durable. Les allocations actuelles sont trop courtes (~ 18 mois). En sciences sociales et humaines, les restrictions sont particulièrement contraignantes pour ce qui est tant du type de dépenses admissibles que des montants accordés. De plus, on ne sait pas trop quelles dépenses liées à l'IRN sont admissibles pour quel organisme. Par exemple, la distinction entre la Fondation canadienne pour l'innovation et l'Alliance est encore ambiguë.

On croit généralement que les attentes des bailleurs de fonds ne sont pas réalistes. Par exemple, pour les subventions d'infrastructure, on exige souvent des chercheuses et chercheurs qu'ils démontrent l'extensibilité et participent aux grappes nationales de calcul. Ils se voient imposer en plus la charge d'expliquer les raisons pour lesquelles ils ne peuvent pas suivre un tel modèle.

Le financement du développement et du soutien des logiciels de recherche est précaire au Canada, à quelques exceptions près, dont celui du programme de soutien des logiciels de recherche locaux de CANARIE.



Le processus du concours pour l'allocation de ressources est fastidieux, lent, en apparence arbitraire, et mobilise une quantité considérable de ressources humaines. On pourrait recommencer à zéro et rendre le processus plus transparent, plus local et plus rapide.

2.3. Voies de découverte de l'IRN

Les chercheuses et chercheurs découvrent les outils et les ressources IRN qu'ils utilisent par le bouche à oreille, ou grâce aux recommandations et à la mise à disposition de leur institution.

Les bureaux de recherche dans les universités jouent un rôle important en aidant les chercheuses et chercheurs à repérer les possibilités en ce qui concerne l'infrastructure lorsqu'ils préparent leurs demandes de subventions. Comme les communications venant du bureau d'un établissement donné sont souvent reçues et lues par ses chercheuses et chercheurs, c'est là un excellent moyen de découvrir les outils et les ressources de l'IRN. Malheureusement, il arrive souvent que les chercheuses et chercheurs ne reçoivent pas le soutien des bureaux de recherche pour tout ce qui n'est pas demande de subventions. Des chercheuses et chercheurs se tournent alors vers les services informatiques ou bibliothécaires de leur établissement.

[trad.] « Il y aurait lieu de promouvoir l'intégration et une collaboration plus étroite entre les services informatiques, les bibliothèques et les bureaux du vice-recteur (recherche) des universités. La prestation des services est cloisonnée dans beaucoup d'établissements. »

Il faudrait définir correctement les rôles et les responsabilités des services informatiques et du soutien technique à la recherche (c'est-à-dire le CIP) des établissements. À l'université St. Francis Xavier, par exemple, les services informatiques sont fournis avec les services de CIP. Du côté de l'université de la Colombie-Britannique, les chercheuses et chercheurs reçoivent de l'aide de leurs services de calcul informatique de pointe pour accéder aux ressources de l'université et aux ressources nationales. Établir des bureaux et des services semblables dans d'autres universités améliorerait la découverte et l'accès à l'IRN.

Un consortium régional comme ACENET contribue grandement à faciliter la découverte des outils et des ressources de l'IRN, étant donné qu'il intervient dans tous les établissements de toutes tailles et de partout.



3. Gouvernance et politiques

3.1. Données de santé

En plus des solutions à long terme pour améliorer le partage des données sensibles et des données de santé entre les provinces, il faut des solutions immédiates qui fourniront les ressources et la gouvernance adaptées aux contextes actuellement plus restrictifs de nombreuses autorités sanitaires. L'alignement est un processus qui dure depuis dix ans et les chercheuses et chercheurs ont besoin de solutions qui peuvent s'adapter au paysage politique actuel.

L'infrastructure doit pouvoir héberger les données sensibles de manière sûre, mais il faut pouvoir y accéder. Des différences sur le plan législatif font obstacle à l'utilisation des données de santé, mais une interprétation trop restrictive des politiques régissant la protection des renseignements personnels pourrait en être partiellement la cause.

Il faut une approche coordonnée entre les secteurs pour que les données puissent être accessibles.

Les chercheuses et chercheurs ont besoin de soutien et de conseils de spécialistes qui possèdent une connaissance juridique des différences entre les lois régissant les données sensibles.

Les besoins en matière de CIP des chercheuses et chercheurs en santé ne sont pas satisfaits. Les ressources de Calcul Canada ne peuvent servir pour les données sensibles qui sont assujetties à des restrictions. Par conséquent, de nombreuses équipes de recherche se dotent de leur propre atelier sans prétention de GDR et de CIP.

Nombre de chercheuses et chercheurs utilisent des microdonnées de Statistique Canada par l'entremise des Centres de données de recherche (CDR) du Réseau canadien des centres de données de recherche. La puissance de calcul des CDR est limitée, les progiciels statistiques sont souvent désuets ou ne peuvent exécuter les algorithmes ou les modèles d'apprentissage automatique. Ainsi, les chercheuses et chercheurs ne sont pas en mesure d'utiliser l'infrastructure de calcul et les méthodes d'analyse des données les plus récentes. De plus, il faut se rendre dans les CDR pour effectuer sur place les analyses, ce qui est très compliqué en ces temps de restrictions sanitaires et de confinement.

Il faudrait que les chercheuses et chercheurs en santé puissent accéder aux ressources existantes et partagées. Autrement, tout se fait de manière fragmentaire, ce qui nuit à l'interopérabilité.

Il faudrait une approche fédérative d'analyse des données de santé, c'est-à-dire que plutôt que d'amener les données là se fait l'analyse, les laisser là d'où elles viennent. On pourrait ainsi déjouer certains problèmes de compétence. Cependant, la question de savoir si la chose est possible dépendra largement de la mise en place et du soutien de l'architecture.



[trad.] « L'Alliance pourrait jouer un rôle dans la mise en œuvre de pratiques et de politiques de soutien d'une architecture distribuée et aussi dans l'installation de celle-ci. »

La question du consentement est souvent négligée. Les participants pourraient ne pas avoir consenti à ce que leurs données soient utilisées à toutes fins. Le consentement devrait être intégré dans le processus parce qu'il est difficile de l'obtenir rétrospectivement. Il devrait être présent à l'esprit lors de la conception du système.

3.2. Souveraineté des données autochtones

Le système de gouvernance doit apprendre à connaître et comprendre les Premières Nations, les Inuits, les Métis et les autres peuples autochtones et établir une relation avec ces derniers. Il faut maintenir le contact et construire de véritables relations.

Les Premières Nations, les Inuits, les Métis et les autres peuples autochtones ne sont pas des sujets de recherche et ne devraient pas être vus comme tels. Ce sont eux qui aident et mènent la recherche. L'Alliance devrait adopter un modèle de mobilisation inspiré du message [trad.] « *pas de recherche sur nous sans nous* ».

Il faut d'abord construire une relation, entre les chercheuses et chercheurs et les communautés, et entre les chercheuses et chercheurs et l'infrastructure. Les peuples autochtones doivent être mobilisés et participer activement au développement des plans de l'IRN pour que leurs systèmes de gouvernance soient pris en compte et mis en œuvre, et que leurs droits ne soient pas enfreints ou supplantés. De plus, l'Alliance devrait soutenir ces relations et tendre une main rassurante, mais ces relations doivent pouvoir s'épanouir. [trad.] « *L'enrichissement en coulisses est le gage du succès.* »

Il y a des exemples de bonnes collaborations en recherche, spécialement la recherche menée par les autochtones. Le gouvernement fédéral a récemment accordé des fonds substantiels en faveur de la souveraineté des données autochtones, et les groupes autochtones établiront leurs propres stratégies en matière de données. C'est là une occasion à saisir pour l'Alliance de collaborer avec ces groupes pour qu'ils ne soient pas exclus de la planification stratégique. La manière centralisée ne peut pas fonctionner, le travail doit être réparti, bien soutenu et fidèle à l'esprit de la Vérité et Réconciliation.

Le CGIPN a joué un rôle en repensant la manière dont les données des Premières Nations devraient être gouvernées, étant donné que les modèles de consentement traditionnels se sont révélés inappropriés. Les Premières Nations ont joué un rôle de premier plan à cet égard. Le consentement de la communauté et le consentement individuel ont leur place dans la gouvernance des données.



Les peuples autochtones ont des besoins distincts, il y aurait donc lieu de permettre certaines nuances. L'accès à niveaux, par exemple, autorise la nuance et la flexibilité.

Par le passé, dans l'espace de l'IRN, la Fédération Calcul Canada et ses organisations régionales se sont concentrées sur les solutions techniques et l'infrastructure de recherche numérique. Il y a des lacunes importantes du côté des politiques. [trad.] « *La meilleure technologie ne servira pas à grand-chose s'il n'y a pas de relations et d'ententes en matière de partage des données* ».

Il serait important de prescrire la technologie et les autres services, ou d'en soutenir la prescription, qui permettraient aux peuples autochtones de mener les travaux sur la souveraineté des données. On a beaucoup parlé de l'engagement envers la souveraineté des données et de l'appui en faveur de celle-ci, mais les ressources nécessaires à la réalisation de cet engagement n'ont pas suivi les discours.

3.3. Harmonisation des politiques, des pratiques et des procédures de gouvernance

La politique des trois organismes sur la gestion des données de recherche crée un arrangement à deux niveaux pour les chercheuses et chercheurs parce qu'elle s'applique seulement au financement de ces organismes. Elle limite le contrôle que les universités exercent pour réglementer la recherche. Il faut une politique qui concerne toute la communauté de recherche.

La question de la propriété des données n'est pas claire pour les chercheuses et chercheurs. C'est un problème important. À Calcul Canada, par exemple, il est entendu que chaque site d'hébergement et chaque établissement ont leur propre politique. Calcul Canada n'a pas pour l'instant de politique claire pour changer la propriété des données sans créer de problèmes avec la propriété intellectuelle si, par exemple, le propriétaire originel ne peut plus être contacté (p. ex. qui ne répond pas aux courriels ou qui a peut-être changé d'établissement). Il faut [trad.] « *un proche parent pour les données étant donné que le principal chercheur et l'utilisateur étaient décédés et que nous ne savions pas quoi faire avec leurs données* ».

Calcul Canada s'emploie à rédiger une politique de classement des données, mais plusieurs questions et points litigieux se posent en rapport avec les types de données qui sont hébergées dans les grappes. La propriété des données et la propriété intellectuelle devraient faire l'objet d'une politique harmonisée pour tous les ensembles de données des différents domaines réglementés et dans tous les sites d'hébergement.

Avec un cadre de gestion du risque et des outils de prise de décision d'apprentissage, il serait possible d'avoir des discussions sur des données beaucoup plus ouvertes et gratuites.

Une loi de portée nationale pourrait être adoptée; et ouvrir une discussion sur le rôle que pourrait l'Alliance dans la construction d'une plateforme qui générerait automatiquement des certificats estampillés sur lesquels les organismes fédéraux pourraient s'appuyer pour accélérer le processus d'accès.



3.4. Collaborations des secteurs public et privé

La recherche financée par des fonds publics qui s'appuie sur l'infonuagique ou qui est menée en collaboration avec l'industrie ne devrait pas être préjudiciable aux chercheuses et chercheurs universitaires. Il y a une limite à la quantité totale des ressources dans les nuages privés et universitaires, donc toute relation entre les secteurs privé et public devrait alimenter le système en ressources.

De nombreux partenaires commerciaux n'ont pas besoin d'une infrastructure de calcul de plus, étant donné qu'ils ont déjà accès à des systèmes plus rapides et plus puissants. Ce qui leur fait souvent défaut, ce sont du personnel hautement qualifié (PHQ) et du personnel de soutien professionnel.

Afin de mener l'adoption de l'intelligence artificielle et des sciences des données au niveau postsecondaire et universitaire ou de l'industrie et aiguillonner des avantages commerciaux, économiques et sociaux de toutes sortes, l'Alliance devrait prendre l'initiative et fournir un soutien stable sur le long terme en faveur d'un modèle pancanadien d'exploitation de l'IRN pour les universités et l'industrie inspiré du modèle du consortium SOSCIP éprouvé au cours de la dernière décennie. Dans ce modèle, les collaborations industrie-université devraient être soutenues grâce à un accès réservé en priorité à leurs propres ressources distinctes d'IRN du CIP, qui serait entièrement complémentaire au soutien à la recherche en CIP axée sur la découverte. L'industrie se mobiliserait davantage, et les liens avec les universités et les collèges seraient renforcés, y compris la formation et l'embauche de PHQ et de personnel de soutien professionnel. L'accès des entreprises et des partenaires à une IRN fiable est essentiel tout comme la protection des données qui y résideraient.

[trad.] « Dans les dernières années, le SOSCIP s'est souvent fait demander de hausser le niveau de sécurité des données non seulement pour les projets qui portent sur la santé, mais aussi pour à peu près tous les secteurs dans lesquels nous avons travaillé. De récentes cyberattaques ont fait encore plus augmenter la demande d'une IRN de CIP cybernétique sûre si bien que les clients le demandent pour pratiquement tous les projets collaboratifs. »

Sur le plan international, il est difficile de collaborer outre-frontière lorsqu'un pays investit dans une infrastructure et des services infonuagiques et pas les autres. Le Canada devrait prêcher par l'exemple et étendre la prestation de ses services d'infonuagique.

Les systèmes devraient être conçus avec des mécanismes clairs de portage informatique pour éviter la dépendance vis-à-vis d'un fournisseur.

L'accès actuel à l'infrastructure nationale est limité aux chercheuses et chercheurs universitaires attachés à des établissements universitaires. On devrait envisager d'améliorer l'accès aux



chercheuses et chercheurs universitaires qui sont en dehors de ce secteur (p. ex. les organismes à but non lucratif).

3.5. Adoption de la recherche ouverte

Dans le contexte de la recherche ouverte, le partage des données n'est pas le produit final. Les méthodes ouvertes, les appels ouverts et la transparence en tous points sont aussi essentiels. Fournir un plan ou un modèle pourrait inciter les gens et les aider à faire ce genre de travail. Offrir une titularisation peut être une mesure incitative majeure.

[trad.] « L'absence de reconnaissance et d'approbation dans ce genre de travail est un problème réel qui freine les chercheuses et chercheurs dans leur démarche. On devrait s'en préoccuper dans les établissements; ce sont eux qui devraient donner suite aux préoccupations. »

Il n'y a pas que les mégadonnées qui sont importantes, les données de taille moyenne et petite le sont aussi.

L'Alliance devrait adopter les principes FAIR et EDI et demander qu'ils soient mis en œuvre dans tous les projets financés.

4. Opérations

4.1. Espaces de collaboration

Il faut distinguer le soutien informatique et le soutien de la recherche. Il faut aussi une collaboration plus étroite entre ces deux secteurs pour que les chercheuses et chercheurs soient aiguillés vers le personnel approprié.

Les chercheuses et chercheurs ont aussi besoin d'une voie de communication directe avec l'Alliance, étant donné qu'ils sont nombreux dans les petits établissements à ne pas avoir les ressources pour pouvoir établir des communications intermédiaires efficaces.

La formation de groupes de travail qui poursuivent un objectif précis et dont la fin du mandat est fixée dans le temps est un excellent mécanisme pour résoudre des problèmes spécifiques.

Les webinaires sont peut-être instructifs, mais ils sont impersonnels. Il est difficile d'interagir avec des écrans.



L'Alliance devrait élaborer des mécanismes de rétroaction et de communication entre la grande communauté de recherche et son conseil des chercheurs.

4.2. Indicateurs de rendement clés (IRC)

Le champ étendu des activités de l'Alliance pose un défi. L'Alliance devrait s'appliquer à réaliser correctement un certain nombre de choses et ne pas tenter de tout faire pour tout le monde étant donné que d'un point de vue opérationnel c'est probablement infaisable.

Il doit y avoir équité entre les portefeuilles et les disciplines de recherche.

Tenir compte des différences entre les disciplines : les mesures d'excellence sont différentes d'un domaine à l'autre.

Pour les indicateurs de rendement clés (IRC), l'Alliance ne devrait pas se contenter de s'en remettre aux rendements passés, elle devrait mesurer ses activités par rapport à celles des organisations internationales qui poursuivent le même mandat.

Les chercheuses et chercheurs des sciences humaines ne peuvent rivaliser avec ceux des sciences pour l'obtention de ressources si les critères d'excellence sont la taille des sorties de données, le nombre des produits de recherche ou même l'utilisation de la puissance de calcul. Les sciences humaines sortent des produits qui diffèrent par leur taille et leur nombre.

Élaborer des IRC dérivés du point de vue des chercheuses et chercheurs et non de celui de l'organisation elle-même.

4.3. Transparence

Tenir compte des émissions de carbone des ressources informatiques.

On ne sait à peu près pas qui utilise l'infrastructure informatique nationale – tout est décidé par le comité consultatif de la recherche.

C'est l'Alliance qui devrait solliciter la rétroaction des chercheuses et chercheurs et s'informer de leurs besoins. Les chercheuses et chercheurs ne devraient pas avoir à s'adresser à l'Alliance.

Il est important de comprendre que l'Alliance n'est pas la version 2.0 de Calcul Canada. Elle s'occupe d'autres activités, telles que la GDR et les LR.

[trad.] « Tous ne savent pas d'où provient le budget additionnel et qui décide où va l'argent. Les chercheuses et chercheurs veulent voir comment les décisions d'investissement dans les priorités sont prises, et qui prend ces décisions. Communiquer clairement : quelles sont les priorités, à court, moyen et long terme? Comment le conseil et la direction décident-ils ce qu'il faut prioriser? »



[trad.] « Beaucoup souhaitent de l'efficacité et de la normalisation entre les groupes. Parce que l'Alliance est financée par des fonds publics, nous ne voulons pas voir des investissements compartimentés du gouvernement fédéral viser le même but. L'utilisation des fonds doit être transparente. »

L'Alliance devrait aller vers les établissements et les inciter à participer à l'IRN national. L'absence de personnel de soutien ne devrait pas empêcher les petits établissements d'utiliser les ressources nationales.

Ce ne sont pas tous les établissements qui ont les ressources et le personnel pour soutenir le calcul de haute performance. Les établissements pourraient être disposés à collaborer afin d'accroître les services qu'ils peuvent fournir.

Des tableaux de bord interactifs accessibles au public montrant l'utilisation selon les caractéristiques des utilisateurs ou utilisatrices et des sujets pourraient donner aux chercheuses et chercheurs et au public une idée de l'allocation des ressources et accroître la confiance et la reddition de comptes.

5. Soutien

5.1. Réseau collaboratif de soutien

Pour la formation et la collaboration, il serait bénéfique que les communautés de Portage et de Calcul Canada collaborent plus étroitement.

La formation annuelle en sciences humaines numériques (du Digital Humanities Summer Institute) est une formation de pair-à-pair. L'institut participe à des projets de collaboration avec Calcul Canada. Ce modèle pourrait être repris et extrapolé à toutes les disciplines.

Les chercheuses et chercheurs consultent souvent le personnel de Calcul Canada pour des problèmes informatiques. Ce modèle surcharge le personnel de soutien qui doit faire une consultation avec chaque chercheuse et chercheur plutôt que de former les chercheuses et chercheurs sur les sujets.

Les membres du corps professoral n'ont pas le temps de participer à autant de ressources et de formations. Les étudiants des cycles supérieurs peuvent s'occuper d'établir les connexions pour eux.

Le « mois de la recherche » au début de l'année soutenu par les vice-recteurs (recherche), qui s'adresse aux étudiants des cycles supérieurs et aux chercheuses et chercheurs en début de carrière, pourrait être un bon moyen d'initiation à l'infrastructure et aux pratiques exemplaires.



Cette initiative pourrait faire partie d'un cours ouvrant droit à crédit pour les étudiants des cycles supérieurs.

Les documents du groupe d'experts national sur la formation Portage- Alliance sont largement disponibles. Il est quand même difficile de trouver quelles sont les priorités.

Les cours de Software Carpentry enseignent aussi aux chercheuses et chercheurs les éléments de base et les compétences nécessaires.

Les logiciels de recherche et la gestion de projet devraient se recouper pour aider les chercheuses et chercheurs à mettre en place de tels cadres. Le soutien local en logiciels de recherche fait une grande différence avec le progrès continu réalisé en s'occupant de l'interface du pipeline de données.

Si le personnel hautement qualifié (PHQ) compte uniquement sur des subventions, il sera toujours difficile de maintenir en poste les employés(e)s qualifié(e)s qui chercheront la stabilité, ce que les subventions ne peuvent garantir. L'idée de privilégier le financement à la subvention doit venir de la tête de l'organisation pour qu'on n'ait pas à se demander si le PHQ sera reconduit dans ses fonctions au gré des subventions; il est plus probable de voir les talents rejoindre le secteur privé qui offre la stabilité.

Le milieu universitaire ne peut rivaliser avec le secteur privé pour ce qui est des attentes salariales. On assiste à un roulement incessant et à une injustice faite à ceux et celles qui ne veulent pas quitter un projet passionnant et qui gagnent moins.

5.2. Champion(ne)s de l'IRN

Cet engagement ne devrait pas relever du bénévolat. Ces personnes devraient être prises au sérieux et se sentir soutenues. L'Alliance devrait embaucher 10 champion(ne)s et les envoyer partout au pays; ou les fonds pourraient être distribués aux établissements pour qu'elles embauchent des champion(ne)s des données. Du point de vue administratif, l'établissement a la responsabilité de garantir et de protéger ces rôles. Du soutien financier pour les personnes qui mènent des activités de soutien technique pourrait être une autre option.

La représentation des établissements est très importante.

Le terme de « champion » ou « championne » n'est peut-être pas le terme approprié. Pourquoi ne pas utiliser plutôt « ambassadeur » ou « ambassadrice » ou un terme semblable? Peut aussi étendre l'idée au-delà des « données » ou de l'IRN. Par exemple, des ambassadeurs ou ambassadrices pour les logiciels, les codes ou le calcul.

Ces rôles doivent avoir des connexions claires et solides entre le soutien local et le soutien national. Ils doivent savoir quels services il convient le mieux de fournir à l'échelle locale ou nationale.



Les champion(ne)s de l'IRN pourraient être des experts régionaux et nationaux qui sont organisés et qui offrent de la formation, des présentations et des services de consultation. Rationaliser les possibilités de formation nationales et locales pour partager les ressources, les pratiques exemplaires et pour éviter toute redondance.

Développer des communautés de pratique composées d'un mélange de chercheuses et chercheurs et de personnel de soutien de recherche en GDR, CIP et LR. Le rôle de l'Alliance pourrait être d'animer ou de coordonner les activités des communautés de pratique.

Le groupe des champion(ne)s devrait être vu comme un réseau.