

2023-09-15

Nouvelles tendances en recherche - Résultats du sondage

Felipe Pérez-Jvostov, Seppo Sahrakorpi et Qian Zhang



Alliance de recherche
numérique du Canada

Digital Research
Alliance of Canada



Table des matières

1. Introduction	1
2. Résultats	2
2.1. Mégadonnées	2
2.2. Calcul et stockage à l'échelle	3
2.3. Intelligence artificielle	3
2.4. IRN intégrée	4
2.5. Équité, diversité, inclusion et accessibilité	5
2.6. Sécurité des données, politiques (ouvertes) et gouvernance	6
2.7. Technologies quantiques	7
2.8. Autres	7
2.9. Principales préoccupations	8
3. Conclusions	10



1. Introduction

Au cours des trois dernières années, l'Alliance de recherche numérique du Canada (l'Alliance) a entrepris une vaste consultation des membres de communauté de recherche pour connaître leur opinion sur l'état actuel de l'infrastructure de recherche numérique (IRN) au Canada ainsi que les difficultés et possibilités s'y rattachant. Le fruit de cette démarche a profondément influencé les priorités organisationnelles et les investissements pour la période de 2023 à 2025.

Pour planifier la période de 2025 à 2030 et cerner de futures possibilités encore non évoquées, l'Alliance a lancé un appel à contributions sur les nouvelles tendances en recherche qui pourraient changer la façon dont les chercheuses et chercheurs interagissent avec l'IRN et l'utilisent. Le sondage, qui visait les chercheuses et chercheurs de tout secteur, s'est déroulé du 7 juillet au 18 août 2023. Au total, 51 chercheuses et chercheurs (47 en anglais, et 4 en français) ont répondu aux questions suivantes :

1. *Quelles sont les tendances en recherche dans votre discipline qui, selon vous, influencent grandement votre manière d'utiliser l'infrastructure de recherche numérique (IRN)?*
2. *Dans 10 ans, comment voyez-vous l'interaction entre les chercheuses et chercheurs de votre domaine et l'IRN, en tenant compte des nouvelles tendances en recherche, des technologies futures et des avancements en matière de politiques?*
3. *Y a-t-il des préoccupations ou des problèmes techniques qui pourraient décourager les chercheuses et chercheurs du Canada dans votre discipline de s'adonner à la recherche de pointe?*
4. *Selon vous, y a-t-il quelque chose qui aura une influence sur l'IRN et la recherche en général et dont vous n'avez pas parlé lors de nos consultations antérieures ou dans vos réponses ci-dessus?*

Est présenté ci-dessous le résumé des principaux résultats, divisé en huit thèmes : mégadonnées; calcul et stockage à l'échelle; intelligence artificielle; IRN intégrée; équité, diversité, inclusion et accessibilité; sécurité des données, politiques ouvertes et gouvernance; technologies quantiques; et une section supplémentaire sur les principales préoccupations.



2. Résultats

2.1. Mégadonnées

Les chercheuses et chercheurs ont discuté de la multiplication et de la complexification rapides des données de recherche, tendance qui se maintiendra dans un avenir prévisible. Beaucoup ont indiqué que les besoins en matière de calcul et stockage augmenteront de manière exponentielle, et qu'il faudra des logiciels de recherche robustes pour analyser des jeux de données aussi volumineux. Il a aussi été grandement question de la haute disponibilité et de la haute fiabilité des instances de stockage dans le contexte des grandes analyses.

Un grand nombre de chercheuses et chercheurs ont mentionné l'impossibilité de gérer manuellement de vastes quantités de données de recherche. Il leur faudra des outils automatisés et propulsés par l'intelligence artificielle (IA) pour exécuter de telles tâches. De même, on a souvent répété que l'IA est un outil nécessaire pour faciliter l'intégration de vastes jeux de données hétérogènes et améliorer l'interopérabilité des données.

L'ouverture des données et le libre accès sont d'autres sujets récurrents. Comme la science ouverte gagne rapidement du terrain dans tous les domaines, beaucoup de chercheuses et chercheurs ont concédé qu'il faut reconnaître et appliquer plus de normes de données propres à chaque discipline. Celles et ceux qui travaillent avec des données réglementées ou sensibles (p. ex. renseignements protégés sur la santé ou renseignements personnels) ont ajouté que la future IRN doit permettre de partager, en toute sécurité et conformité, des données bien annotées ayant fait l'objet d'un processus de curation rigoureux afin d'utiliser les vastes quantités de données sur la santé du Canada de manière optimale et de relever les défis toujours plus complexes du domaine de la recherche médicale.

« [...] la communauté scientifique mondiale de la bio-imagerie doit composer avec d'immenses jeux de données. Pour les gérer et les analyser, il faut utiliser des outils automatisés et infonuagiques [...]. »

« [...] dans la génomique, les besoins relatifs au stockage et au calcul continuent de croître de manière exponentielle [...]. »

« [...] le calcul de haute performance dans le domaine de la physique des particules nécessite des solutions de stockage en réseau à l'échelle mondiale. Les instances de stockage doivent être hautement disponibles, fiables et performantes [...]. »



2.2. Calcul et stockage à l'échelle

Beaucoup de chercheuses et chercheurs ont souligné que les ressources de calcul et de stockage doivent mieux prendre en charge les mégadonnées. Le besoin de ressources actives de stockage et de calcul de haute performance (CHP) et sa croissance rapide ont aussi été fréquemment mentionnés. On estime que la demande de solutions de calcul et de stockage dans la dernière décennie est de deux à dix fois plus importante qu'auparavant.

À maintes reprises, on a fait allusion à différentes architectures de calcul, dont les processeurs, les processeurs graphiques, et les calculs nécessitant beaucoup de mémoire. Des chercheuses et chercheurs ont spécifiquement mentionné les besoins relatifs aux capacités exa-informatiques ou au calcul à haut débit, ce qui laisse croire que les utilisatrices et utilisateurs finaux de l'IRN ont rarement recours à ces termes techniques.

« [...] il y a et aura toujours une forte demande de ressources de CHP traditionnelles, les cœurs de processeurs et la mémoire étant les principaux ingrédients d'une multitude de grands problèmes fondamentaux et appliqués en sciences physiques [...]. »

« [...] les ressources de calcul nécessaires pour réaliser de la recherche de pointe seront extrêmement exigeantes [...]. »

« [...] nous avons besoin d'infrastructures de stockage de données et de sauvegarde qui peuvent gérer ces problèmes de données de nouvelle génération [...], les besoins relatifs au calcul de haute performance et à haut débit, [et] les exigences en matière de stockage et de sauvegarde à haute capacité de données [...]. »

2.3. Intelligence artificielle

L'IA et les technologies connexes (p. ex. apprentissage machine et grands modèles de langage) sont perçues comme les plus grandes tendances à venir dans l'avancement de la recherche. Bon nombre de chercheuses et chercheurs ont expliqué que les technologies d'IA seront d'importants outils et leviers dans leur domaine. Les répondantes et répondants n'ont pas mentionné l'élaboration de méthodologies ni l'entraînement de modèles comme tels, mais plutôt l'application des outils propulsés par l'IA dans le cadre de leurs recherches.



Les répondantes et répondants ont fourni des possibilités d'utilisation de l'IA dans la recherche, p. ex. pour l'analyse d'images, du son et de jeux de données (volumineux) ainsi que pour la traduction automatique, la transcription de documents manuscrits, le traitement automatique des langues, les simulations assistées ou dirigées par l'AI, et la résolution de problèmes appliquée. L'IA est souvent dite pertinente dans le champ médical et de la santé, et dans la génomique. Toujours selon les répondantes et répondants, l'IA pourrait aussi faire progresser et accélérer les avancées scientifiques dans les domaines de la physique des particules, des sciences des matériaux et de la biomécanique. Plusieurs chercheuses et chercheurs ont indiqué que l'analyse d'images et données médicales assistée par l'IA est une priorité absolue à l'intersection des technologies d'IA et des domaines de recherche.

« [...] l'utilisation de l'IA sera de plus en plus importante dans l'analyse de données et les méthodes de représentation [...]. »

« [...] bien entendu, l'apprentissage machine et l'IA seront des facteurs majeurs dans la recherche en génomique, comme dans d'autres domaines [...]. »

« [...] la quantité de données [médicales diagnostiques] produites par ces outils et l'utilisation de celles-ci avec les premières technologies d'IA permettent de résoudre des problèmes persistant depuis des décennies [...]. »

2.4. IRN intégrée

Les chercheuses et chercheurs prévoient que dans l'avenir, il sera essentiel de disposer de systèmes d'IRN intégrée (données + calcul + logiciel). Plutôt que d'accéder à chaque élément séparément, elles et ils veulent pouvoir accéder facilement à l'ensemble des ressources sur des plateformes interopérables complètes. Ont été mentionnées parmi les solutions intégrées potentielles une plateforme intégrée propre à un domaine donné, un dépôt de données ouvertes commun doté de capacités d'analyse et de calcul à grande échelle, et un logiciel-service infonuagique accessible pour l'analyse et la collaboration.

« [...] nous espérons disposer d'un écosystème de données avec une meilleure intégration des données cliniques et de recherche. [...] De façon générale, j'espère vraiment que l'Alliance exploitera le potentiel d'intégration de l'IRN à l'échelle des infrastructures, des logiciels et des données. Nous n'avons pas besoin de ressources distinctes pour ces trois éléments; il faut plutôt les regrouper en une seule plateforme propre à un domaine [...]. »



« [...] absolument, les chercheuses et chercheurs s'attendent de plus en plus à une intégration entre les systèmes [...]. »

« [...] idéalement, il faudrait potentiellement assurer l'intégration (et l'interopérabilité) complète des différentes plateformes de stockage et de gestion des différentes sources de données [...]. »

2.5. Équité, diversité, inclusion et accessibilité

Les chercheuses et chercheurs du domaine des sciences humaines ont relevé d'importantes lacunes sur le plan de l'accessibilité de l'IRN, lacunes qui devront être comblées dans l'avenir. Les inégalités dans l'affectation de fonds pour le développement de l'infrastructure et les ressources opérationnelles entre les disciplines sont considérées comme des obstacles à l'accès élargi à l'IRN et à son utilisation. D'autres répondantes et répondants ont expliqué que l'adoption d'interfaces et de logiciels conviviaux permettrait d'améliorer l'accessibilité.

Les chercheuses et chercheurs qui travaillent avec des groupes marginalisés ainsi que des membres des minorités visibles et des peuples autochtones ont discuté des questions éthiques soulevées par l'utilisation des données sur ces groupes, et souligné le nécessaire équilibre à trouver entre la souveraineté des données autochtones et le respect de la vie privée dans le cadre des bourses d'études ouvertes et du couplage de données, en particulier en ce qui a trait à l'IA et aux mégadonnées.

« [...] l'IRN au Canada ne doit exclure aucune discipline ou communauté [...]. »

« [...] j'espère que les interfaces utilisateurs deviendront plus accessibles pour les chercheuses et chercheurs qui ont des connaissances informatiques limitées [...]. »

« [...] un autre défi technique considérable est la reconnaissance de la souveraineté des données autochtones sur les plateformes actuelles [...]. Il importe de maintenir un équilibre en offrant autant de bourses d'études ouvertes et de connectivité que possible tout en assurant le respect de la vie privée des boursières et boursiers du Canada ainsi que de la souveraineté des données autochtones [...]. »



2.6. Sécurité des données, politiques (ouvertes) et gouvernance

Beaucoup de chercheuses et chercheurs ont souligné l'importance du mouvement de la science ouverte et des données liées et ouvertes ainsi que de la sécurité des données et du respect de la vie privée. Bon nombre d'entre elles et eux ont soulevé le problème de la préservation d'un écosystème « aussi ouvert que possible, mais aussi fermé que nécessaire », mettant en opposition le partage des données, la découverte, l'accès, le couplage et la réutilisation pour la stimulation de la recherche, et la protection des droits de propriété intellectuelle et le respect de la vie privée des sujets (en particulier, les renseignements protégés sur la santé). Un autre élément important mentionné est la fédération, la rétention, la curation et le stockage adéquats des données à long terme. Sur le plan des politiques, il a aussi été question de l'élaboration et l'adoption de pratiques exemplaires; sur le plan juridique, les difficultés découlant des différences entre les lois sur les données et renseignements protégés sur la santé d'un territoire à un autre sont source de préoccupation et sont considérées comme un obstacle à la recherche interjuridictionnelle.

« [...] des ententes sur le transfert de données [sont] nécessaires pour l'envoi d'images [médicales]. Le principal point d'achoppement concerne la partie des ententes sur les autorités compétentes et les droits applicables. Les deux côtés veulent que leurs propres lois et autorités figurent dans l'entente [...]. »

« [...] le mouvement des bourses d'études ouvertes et la mobilisation positive qui pourrait en découler sont au cœur de mes récents travaux [...]. »

« [...] on s'intéresse de plus en plus à la capacité de fédérer les données et de réaliser les analyses là où se trouvent les données, plutôt qu'aux enjeux de centralisation [...]. Si la fédération des données est clairement la voie à suivre, elle ne s'appliquera pas à tous les cas d'utilisation dans le domaine de la recherche sur la santé ou un domaine connexe. Il faudra probablement quand même disposer d'IRN capables d'héberger des plateformes pancanadiennes et de prendre en charge les partages des données approuvés, y compris pour les données qui ne sont pas recueillies ni gérées par des chercheuses et chercheurs ou des universités [...]. »



2.7. Technologies quantiques

Seule une poignée de chercheuses et chercheurs a indiqué que l'informatique quantique est une tendance potentielle dans son domaine de recherche. Il est largement reconnu que ce domaine a le potentiel de bouleverser le paysage de l'IRN lorsque la technologie et les applications seront plus matures et mieux comprises. Comme la phase de recherche et développement de l'informatique quantique n'en est qu'à ses balbutiements, il est compréhensible que la communauté de la recherche ne sache pas encore bien comment tirer parti de la technologie dans l'avenir.

Un cas d'utilisation potentielle est la façon dont l'informatique quantique, avec sa puissance de calcul accrue, pourrait permettre à des chercheuses et chercheurs dont les connaissances théoriques et informatiques sont limitées d'exécuter des simulations complexes. De même, on compte parmi les exemples de futures applications les simulations de modèles fondés sur la théorie quantique, p. ex. la création de modèles de physique des particules fondés sur la théorie quantique, et la reconnaissance des formes de signatures expérimentales complexes.

« [...] l'informatique quantique peut changer profondément certaines parties de la recherche en physique des particules, car la majorité du temps d'utilisation du processeur est consacrée à la simulation de processus quantiques et à la reconnaissance des formes de signatures de détecteur complexes. La disponibilité à grande échelle de processeurs quantiques pourrait entièrement transformer notre façon de travailler [...]. »

2.8. Autres

Beaucoup de chercheuses et chercheurs ont souligné des problèmes touchant à divers aspects de l'IRN; parmi ceux-ci, les enjeux liés à des infrastructures conviviales recourant à différentes technologies pour améliorer l'accessibilité, l'utilisation et la productivité ont été soulignés. D'autres ont indiqué qu'il est nécessaire de mieux intégrer les systèmes actuels dans l'avenir, mais que le soutien offert resterait minimal. Pour que le Canada et ses chercheuses et chercheurs puissent tirer le maximum des technologies actuelles et futures, il faut augmenter considérablement le financement afin de fournir des infrastructures plus robustes et mieux intégrées, sécurisées et automatisées, adaptées aux différents niveaux de compétence technologique des chercheuses et chercheurs de tout domaine. Enfin, des chercheuses et chercheurs ont affirmé que la recherche s'élargira à l'échelle mondiale, et que l'infrastructure du Canada doit donc permettre de telles collaborations tout en protégeant les droits de propriété intellectuelle et les intérêts nationaux.



« [...] au bout du compte, les nouvelles technologies sont moins importantes que le déploiement intelligent des technologies éprouvées. [...] Notre premier virage consistera à exiger à nouveau que les données et le traitement soient regroupés dans une même installation, soit le modèle de serveur, contrairement au modèle actuel, dans le cadre duquel les chercheuses et chercheurs peuvent utiliser leurs outils informatiques locaux pour gérer leurs données. Par conséquent, des centres de données bien établis avec des plateformes scientifiques adaptées aux disciplines seront nécessaires pour atteindre les objectifs de recherche [...]. »

« [...] pour faire avancer la recherche, il sera essentiel de masquer la complexité de l'écosystème de l'IRN et de laisser les chercheuses et chercheurs faire leurs travaux (sans expertise en technologie requise). Un énorme fossé sépare le strict minimum nécessaire pour l'IRN (ce que nous faisons aujourd'hui, c.-à-d. garder les systèmes fonctionnels, installer les logiciels et nous lancer) et la création d'un environnement où les chercheuses et chercheurs peuvent exceller dans leur domaine (sans obstacle technologique). »

2.9. Principales préoccupations

Les chercheuses et chercheurs ont exprimé plusieurs préoccupations relatives à l'avenir de l'IRN : sécurité et respect de la vie privée, manque de financement, dotation en personnel hautement qualifié, et enjeux liés à la diversité et aux services dans le domaine des sciences humaines et sociales.

Au sommet des préoccupations trône le risque que les restrictions relatives à la sécurité et au respect de la vie privée freinent le progrès de l'IRN et de la recherche en général. Plusieurs chercheuses et chercheurs ont donné des exemples d'incompatibilité entre les enjeux juridiques et de respect de la vie privée (p. ex. *Règlement général sur la protection des données* de l'Union européenne ou autres règlements relatifs aux renseignements protégés sur la santé) qui ont déjà bloqué ou ralenti des activités de recherche. Si l'authentification multifacteur renforce la sécurité, elle est aussi jugée lourde et peut même nuire à la collaboration internationale. Les répondantes et répondants s'inquiètent aussi du manque de fonds pour l'IRN au Canada, notamment pour le cycle d'actualisation du calcul informatique de pointe actuel ou la future croissance de l'IRN. Plusieurs d'entre elles et eux ont également dit se soucier du manque potentiel de différents types d'infrastructures dans l'avenir, soulignant la nécessité d'infrastructures, de flux de travail et d'expertise technique personnalisés dans les disciplines mal desservies. A aussi été soulevée la pénurie de personnel hautement qualifié, puisque certaines utilisations requièrent des ressources humaines supplémentaires, en particulier dans le domaine des sciences humaines et sociales.

S'ajoutent également des préoccupations quant aux coûts d'utilisation de l'IRN, à la bureaucratie du processus d'affectation des ressources, à la réutilisation des données, à



l'attention excessive portée au CHP, au manque de solutions de stockage à long terme, et au piratage et au vol de propriété intellectuelle par des rivaux étrangers et à leur influence dans le cadre des travaux pour le déploiement de la science ouverte avec des nations et établissements alliés.

« [...] il n'y a pas de soutien pour l'ensemble des chercheuses et chercheurs. Les fonds et les infrastructures sont insuffisants [...]. »

« [...] les lois et règlements – par exemple, sur la protection des renseignements personnels – sont adoptés rapidement (et de bonne foi), mais les chercheuses et chercheurs et les établissements n'ont pas l'expertise ni les ressources nécessaires pour restructurer entièrement leurs méthodes de recherche [...]. »

« [...] les ressources pour les bourses d'études et les infrastructures du domaine des sciences humaines et sociales ont toujours été plus limitées que celles consacrées aux STIM [...]. L'infrastructure des sciences humaines et sociales est plus susceptible d'être humaine et de nécessiter une expertise technique [...]. Ces projets peuvent difficilement faire concurrence avec des industries bien financées qui peuvent offrir de meilleurs salaires aux candidates et candidats compétents pour ces postes [...]. »



3. Conclusions

Les répondantes et répondants ont évoqué une vision d'une IRN du Canada plus robuste et mieux intégrée, sécurisée et automatisée, qui donnerait lieu à des collaborations de recherche au-delà des frontières tout en assurant la protection des données et des droits de leurs propriétaires. La future IRN du Canada doit tirer le maximum des avancées technologiques modernes et émergentes des domaines de l'automatisation et de l'intelligence artificielle et de l'augmentation considérable de la dotation en personnel hautement qualifié afin d'améliorer l'efficacité et la connectivité de la recherche. Il faut aussi veiller à ce que la recherche et l'infrastructure de recherche financées par des fonds publics soient fiables et que leurs retombées puissent rejaillir au-delà des travaux scientifiques.

Beaucoup de chercheuses et chercheurs ont souligné qu'il faut augmenter les ressources de calcul et de stockage, estimant que la demande de la dernière décennie est de deux à dix fois plus importante qu'auparavant. Bon nombre d'entre eux et elles ont indiqué qu'il faut disposer de différents types d'infrastructures de calcul (p. ex. processeurs, processeurs graphiques et architectures gourmandes en mémoire). Elles et ils reconnaissent que les données de recherche se multiplient et se complexifient rapidement. Beaucoup laissent entendre que dans l'avenir, les besoins relatifs au calcul et au stockage seront beaucoup plus importants qu'ils le sont actuellement, et qu'il faudra des logiciels de recherche robustes pour analyser les jeux de données volumineux. Il a aussi beaucoup été question de la haute disponibilité et de la haute fiabilité des instances de stockage dans le contexte des grandes analyses.

L'IA et les technologies connexes (p. ex. apprentissage machine et grands modèles de langage) sont perçues comme de grandes tendances à venir dans l'avancement de la recherche selon beaucoup de chercheuses et chercheurs, qui ont aussi expliqué qu'elles seront d'importants outils et leviers de recherche. Comme il sera impossible de gérer manuellement de vastes quantités de données, les chercheuses et chercheurs auront probablement besoin d'outils automatisés et propulsés par l'intelligence artificielle (IA) pour exécuter les tâches liées à la gestion des données de recherche. De même, on a souvent répété que l'IA est un outil nécessaire pour faciliter l'intégration de vastes jeux de données hétérogènes et améliorer l'interopérabilité des données.

L'ouverture des données et le libre accès sont aussi revenus souvent dans les discussions. Comme la science ouverte gagne rapidement du terrain dans tous les domaines, beaucoup de chercheuses et chercheurs ont concédé qu'il faut reconnaître et appliquer plus de normes de données propres à chaque discipline. Bon nombre d'entre elles et eux ont souligné l'importance du mouvement de la science ouverte et des données liées et ouvertes ainsi que de la sécurité des données et du respect de la vie privée. De nombreuses réponses font état du problème de



la préservation d'un écosystème « aussi ouvert que possible, mais aussi fermé que nécessaire », mettant en opposition le partage des données, la découverte, l'accès, le couplage et la réutilisation pour la stimulation de la recherche, et la protection des droits de propriété intellectuelle et le respect de la vie privée des sujets (en particulier, les renseignements protégés sur la santé).

Les chercheuses et chercheurs du domaine des sciences humaines ont relevé d'importantes lacunes relatives à l'accessibilité de l'IRN qui devront être comblées dans l'avenir. Les inégalités dans l'affectation de fonds pour le développement de l'infrastructure et les ressources opérationnelles entre les disciplines sont considérées comme des obstacles à l'accès élargi à l'IRN et à son utilisation. D'autres répondantes et répondants ont expliqué que l'adoption d'interfaces et de logiciels conviviaux permettrait d'améliorer l'accessibilité.

Les chercheuses et chercheurs qui travaillent avec des groupes marginalisés ainsi que des membres des minorités visibles et des peuples autochtones ont discuté des questions éthiques soulevées par l'utilisation des données sur ces groupes, et souligné le nécessaire équilibre à trouver entre la souveraineté des données autochtones et le respect de la vie privée dans le cadre des bourses d'études ouvertes et du couplage de données, en particulier en ce qui a trait à l'IA et aux mégadonnées.

Notons que peu de chercheuses et chercheurs ont dit que l'informatique quantique est une tendance potentielle dans leur domaine de recherche à moyen terme, ce qui indique que la phase de développement de cette technologie qui pourrait changer le paradigme ne fait que commencer.

L'une des préoccupations soulevées est la façon dont les restrictions relatives à la sécurité et au respect de la vie privée peuvent freiner la recherche. Plusieurs répondantes et répondants ont donné des exemples de situations où les politiques sur le respect de la vie privée, l'incompatibilité entre les autorités et l'authentification multifacteur ont déjà bloqué ou ralenti des activités de recherche ou des collaborations internationales. S'ajoute aux préoccupations principales le manque de fonds pour l'IRN au Canada, y compris pour le cycle d'actualisation du calcul informatique de pointe actuel ou le financement continu pour la croissance future de l'IRN.