

Les besoins de l'écosystème de l'infrastructure de recherche numérique canadienne

Soumis au nom de la communauté de recherche de l'Université d'Ottawa par :

Guy Levesque | glevesq4@uOttawa.ca
Vice-recteur associé, Soutien à la recherche et infrastructure
Cabinet du vice-recteur à la recherche | Université d'Ottawa

Ce document a été rédigé par :

Sean Geddes, MSc, MBA
Conseiller exécutif
Cabinet du vice-recteur à la recherche | Université d'Ottawa



Objet

L'infrastructure de recherche numérique (IRN) du Canada est un pilier du soutien à la recherche dans les domaines stratégiques désignés de l'Université d'Ottawa. Ce document est une synthèse descriptive compilée par nos chercheuses et chercheurs des problèmes, besoins et possibilités actuels en ce qui concerne les instruments, les services et le soutien en matière d'IRN, l'IRN de l'avenir au Canada, et la façon de concrétiser cet avenir avec l'aide et les conseils de la Nouvelle organisation d'infrastructure de recherche numérique (NOIRN). L'Université a tenu compte des besoins et de l'utilisation variés de toutes les disciplines de recherche et a incorporé l'information générée par la consultation de gens de multiples disciplines, dont la médecine, l'intelligence artificielle, le génie, la psychologie, les sciences sociales, les sciences de la santé, les technologies de l'information et la bibliothéconomie.

La perspective de l'Université d'Ottawa

L'Université emploie un certain nombre d'outils, de services et de ressources de l'IRN pour enrichir ses programmes de recherche et faciliter l'atteinte des objectifs. Ces composantes se divisent dans la plupart des cas en ressources internes et ressources externes.

Les ressources internes, fournies et généralement financées par l'Université sont : les grappes de serveurs sur bâti du Centre de données de recherche de l'Université; un volume de stockage (limité) de bases de données sur un réseau local à l'usage de nos groupes de recherche; un volume de stockage en nuage pratiquement illimité; une infrastructure de machines virtuelles pour faciliter l'exécution de tâches spécialisées de faible intensité qui ne nécessitent pas un serveur dédié; une plateforme Web limitée de création et de déploiement d'applications informatiques hébergée sur place et une infrastructure informatique locale haute performance, mais limitée.

Les ressources externes régulièrement utilisées sont : les services de gestion des données de recherche (assistant de planification de la gestion des données, plans de gestion des données et outils de gestion des données sensibles) du réseau Portage de l'Association des bibliothèques de recherche du Canada (ABRC), les ressources de Calcul Canada (Service d'accès rapide à des ressources de calcul de pointe, grappes de machines virtuelles), CANARIE (réseau fédérateur pancanadien), dépôts de données sur la population (p. ex. Statistique Canada, Institute for Clinical Evaluative Sciences [ICES], etc.) et SOSCIP, qui est une plateforme particulièrement utile à la collaboration entre l'Université et l'industrie.

Enjeux actuelles

À l'instar d'autres établissements universitaires du Canada dont les activités de recherche recourent aux ressources de l'IRN, l'Université d'Ottawa est vulnérable aux difficultés d'accès et d'acquisition d'instruments, de ressources et de capacités particulières en la matière.

Les ressources et les outils de Calcul Canada et de CANARIE, auxquels recourent abondamment nos équipes de recherche, ont compté pour beaucoup dans l'élaboration et l'application de programmes de recherche reposant sur l'IRN. Cette utilisation ne va pas sans mal, compte tenu d'un manque de programmes d'éducation et de formation, de difficultés d'accès aux ressources et aux serveurs distants et des problèmes d'utilisation de la plateforme d'ordonnement connexe.

La formation et l'éducation sont une source de préoccupation constante des groupes de recherche qui utilisent activement les services de Calcul Canada et de CANARIE. Ces difficultés émanent plus précisément de la méconnaissance des capacités de ces outils, du manque d'information et de formation sur

la façon de choisir l'outil le plus approprié à une tâche donnée et du manque de formation qui permette l'optimisation de ces ressources excellentes, mais peu abondantes. Il importe de souligner que, dans certains cas, les chercheuses et chercheurs ne savent tout simplement pas qu'il existe des services de formation et d'éducation conçus pour eux. D'où ce problème connexe : la méconnaissance de l'existence même des ressources, qui met d'ailleurs en lumière la nécessité de plus amples activités et services de promotion pour en favoriser l'utilisation par la communauté de la recherche.

La disponibilité des serveurs de Calcul Canada semble aussi poser régulièrement problème, notamment à cause de certaines initiatives prioritaires (p. ex. leur réquisition par les organismes de financement). On souligne également la faible quantité de processeurs graphiques pouvant répondre aux besoins de programmes de recherche qui dépendent étroitement de l'IRN. Pour contourner cette insuffisance, nous avons acquis quelques serveurs de processeurs graphiques, mais Calcul Canada a tout à gagner de l'expansion de cette capacité. Enfin, l'accès et la planification de l'accès aux ressources posent aussi certains problèmes, liés à l'utilisation de l'ordonnanceur de tâches appelé SLURM. De prime abord, ces problèmes semblent découler d'une méconnaissance de cette plateforme créée sous Linux, mais ils n'en montrent pas moins la nécessité de fournir régulièrement des ressources et des outils pour en permettre une utilisation appropriée.

Nos équipes de recherche de disciplines comme la santé, les sciences biologiques et la psychologie du comportement évoquent le problème récurrent de la relative inaccessibilité de moyens de gérer les renseignements confidentiels des participants et des patients dans le cadre de l'IRN. Comme chacun sait, les équipes doivent traiter avec la plus grande prudence les renseignements personnels, qui contiennent souvent des informations sensibles permettant de remonter aux sujets. Par ailleurs, le stockage et le transfert de données sont soumis à de rigoureuses dispositions législatives provinciales et nationales, dont la *Loi de 2004 sur la protection des renseignements personnels sur la santé* (PHIPA) et la *Loi sur la protection des renseignements personnels et les documents électroniques* (LPRPDE). La nécessité d'une gestion appropriée et d'un stockage sûr a fait de l'anonymisation des données un atout précieux et incontournable pour toute personne qui travaille avec ce type de données. Or, les méthodes ne sont pas du tout harmonisées, ce qui, dans certains cas, accroît la vulnérabilité des renseignements personnels. L'accessibilité de ressources de stockage sûr pour ce type de données serait un atout inestimable pour nos équipes de recherche.

Dans un même ordre d'idées, l'accessibilité de ressources, d'outils et de services permettant le transfert efficace et sûr de ces données vers des établissements partenaires serait aussi précieuse. De fait, la difficulté de transférer des données confidentielles à des collègues de l'extérieur en vue des objectifs de recherche est un problème de taille.

Autre composante d'intérêt pour les chercheuses et les chercheurs de l'Université d'Ottawa : les ressources nécessaires à la création et au déploiement d'applications Web, actuellement limitées à l'échelle locale. Nos équipes de recherche doivent faire largement appel à des ressources externes, notamment celles de Calcul Canada, pour ce qui est, par exemple, de machines virtuelles pour l'hébergement d'interfaces de programmation d'applications mobiles. Mais certains de nos groupes interdisciplinaires ont besoin de plus encore pour créer et raffiner les applications.

Résumé

De manière générale, les groupes de recherche de l'Université d'Ottawa peuvent faire progresser leurs programmes grâce aux vastes avantages de l'IRN actuelle, mais ils ont des problèmes manifestes et ont nettement besoin de plus de soutien pour contribuer à l'excellence de la recherche au Canada. Il leur faut notamment des outils de formation et d'éducation pour découvrir ou pour mieux utiliser toutes les

ressources de Calcul Canada et de CANARIE. Ils doivent aussi pouvoir accéder à des serveurs à distance, utiliser la plateforme d'ordonnancement, ainsi que d'autres plateformes Web permettant la création d'applications mobiles, la gestion efficace et en toute sécurité des données des participantes et des participants, et le partage sûr de ces données en vue d'une collaboration avec des groupes de recherche d'autres établissements. L'accessibilité accrue de ce genre de ressources serait un enrichissement pour l'IRN nationale et contribuerait de manière importante à la promotion de l'excellence canadienne en matière de recherche.

L'avenir de l'IRN

L'écosystème de l'IRN canadienne gagne en cohésion au fil des ans, comme en témoigne la convergence constante des centres et des grappes de données de Calcul Canada, qui permettent de constituer des piles de logiciels tout en favorisant l'utilisation d'un même logiciel par la majorité des grappes. Cette cohésion se reflète également dans la création d'une interface unique de demandes de comptes, compatible avec toutes les grappes. De manière générale, donc, elle génère une vision plus uniforme de l'écosystème de l'IRN canadienne.

L'Université d'Ottawa compte agrandir cet écosystème cohésif en y ajoutant des ressources permettant le partage dynamique et sûr de données confidentielles entre collaborateurs à l'échelle pancanadienne, l'amélioration des communications et du transfert de tâches entre grappes, le développement continu des ressources en gestion des données de recherche (GDR), l'amélioration globale du savoir-faire en matière de données, des programmes évolutifs de formation et d'éducation, et l'intégration générale et homogène des ressources de l'IRN.

L'une des difficultés, déjà évoquée, a trait à la capacité réduite, en apparence, des centres de données de Calcul Canada. La solution pourrait consister à rendre les ordonnanceurs mobiles et capables de déléguer certaines tâches programmées vers des grappes moins actives au moment voulu; autrement dit, répartir les tâches à la demande dans tout le réseau. Il serait fort utile d'augmenter en outre la capacité globale des serveurs (de processeurs graphiques, par exemple) de Calcul Canada.

La vision d'avenir de l'écosystème de l'IRN doit englober en particulier des programmes d'éducation et de formation connexes, appropriés à un cadre qui évolue très rapidement. Il faut en effet que nos équipes de recherche puissent trouver rapidement la formation nécessaire, qui soutiendra du même coup la croissance de l'IRN canadienne. Il faut aussi créer des programmes de formation adaptables à mesure de l'intégration de nouveaux outils et de l'évolution, rapide, du domaine.

Dans ce cadre, il faut aussi veiller plus activement à la création de ressources, de formations et de cadres de GDR. L'optimisation des méthodes de GDR concourt à l'accessibilité et à la sécurité des données et, partant, au savoir-faire de nos chercheuses et chercheurs. Ce type de ressources est essentiel à l'élaboration de méthodes uniformes et reproductibles dans le milieu de la recherche. Il faut par ailleurs que la NOIRN soutienne des projets de base d'élaboration de politiques de gestion des données. Sans standard national, en effet, le risque est grand d'accroître la prolifération des méthodes et l'insatisfaction à la clé.

La collaboration est un maître mot du monde universitaire et de l'écosystème de recherche. L'IRN à venir doit favoriser la collaboration et améliorer la sécurité du partage de données entre de multiples établissements du Canada (et même, d'ailleurs), tout en assurant le respect des politiques nationales. Ainsi, les chercheurs, et en particulier ceux qui travaillent avec des données particulièrement sensibles concernant les participants à leurs études, doivent pouvoir collaborer et bâtir plus activement leurs programmes de recherche (grâce, notamment, à des tests et à des activités de création multisites).

Il est à espérer enfin que grâce aux progrès de l'IRN vers cet état futur, les établissements universitaires pourront accéder toujours plus aisément à des unités de traitement à tenseurs. De fait, la grappe Graham compte actuellement 144 processeurs graphiques à cœurs Tensor T4 de NVIDIA. L'accès à ce type de moyens de diffuser l'intelligence artificielle est certes un facteur à considérer dans le développement et la croissance de l'IRN au Canada.

Comblant l'écart

Le rattrapage indispensable à l'écosystème de l'IRN canadienne passe sans conteste par un effort graduel d'envergure nationale, ainsi que par l'adhésion et la collaboration des établissements et organisations. Il peut d'ailleurs tirer parti de grandes réussites passées et récentes de nos réseaux régionaux et nationaux.

L'évolution de l'IRN au Canada passe également par l'engagement des universités à aider les spécialistes à tirer parti des ressources qui la composent actuellement. Les projets fondés sur CANARIE, par exemple, ont largement aidé le personnel et les étudiants à atteindre leurs objectifs de recherche, mais aussi à profiter de toutes les capacités de l'IRN dans sa version actuelle. Ces mesures de soutien ont permis aux chercheuses et aux chercheurs d'apprendre à utiliser les outils et les ressources existantes qui, du reste, sont particulièrement utiles aux groupes de recherche multidisciplinaire, dont les membres connaissent généralement moins bien les outils de l'IRN.

La communication est un autre moteur essentiel au rattrapage de l'écosystème de l'IRN canadienne. Il faut envisager divers moyens, notamment pour permettre une rétroaction régulière à mesure de l'évolution des ressources de l'IRN, par exemple forum annuel des parties prenantes, des communications régulières avec les établissements universitaires sur les progrès accomplis pour faire connaître les ressources créées au sein des universités (et favoriser l'adoption de méthodes efficaces de gestion dans l'écosystème), et des formations périodiques grâce auxquelles les chercheurs pourront tirer pleinement parti des capacités de l'IRN.

La sécurité de la gestion des données de recherche gagnera forcément en importance à mesure que la NOIRN ajoutera des fonctionnalités à l'IRN. Il est essentiel d'utiliser des contrôles et des outils divers pour empêcher tout risque d'incident. On pense notamment à l'authentification à multiples facteurs des usagers. Déjà, certaines activités ont été mises en œuvre pour résoudre d'éventuels problèmes de sécurité, mais ce volet doit absolument figurer en tête des priorités de l'évolution envisagée.

Le retard de l'IRN ne doit toutefois pas être comblé aux dépens de projets de recherche dont le financement est assuré grâce à divers programmes, dont le Concours d'allocation des ressources, qui améliorent directement la connaissance de l'IRN dans la communauté et en favorisent la croissance.

Enfin, compte tenu de la croissance rapide de l'IRN, le soutien des établissements est essentiel à l'exécution de nos programmes de recherche. Les universités doivent d'ailleurs déterminer les structures et les niveaux de soutien appropriés selon leurs besoins en gestion de données. Or, puisque les besoins, l'ampleur et l'intensité de la recherche varient d'un établissement à l'autre, le moyen le plus efficace d'y parvenir est de recourir à des outils comme les cadres de maturité de la GDR. Malheureusement, ces cadres sont plus ou moins utiles, à l'heure actuelle, faute d'homogénéité. À mesure de l'évolution des services de la NOIRN, il faudra adopter un même cadre à l'échelle pancanadienne pour aider les établissements à évaluer leurs services de soutien en la matière et à cerner les volets à développer davantage, en prêtant une attention particulière à la diversité des établissements au Canada.

Conclusion

Les chercheuses et les chercheurs de l'Université d'Ottawa font souvent appel à l'infrastructure de recherche numérique du Canada étant donné leurs activités très diverses. Cette interaction débouche à la fois sur des difficultés et des possibilités. Ce livre blanc vise à éclairer le tout et à contribuer à l'élaboration du plan stratégique de la NOIRN, dont le but est d'« établir une vision unique et unifiée des principales priorités relatives à l'IRN pour 2021-2024 ». Il met en lumière la nécessité d'accroître la formation et l'éducation continues sur l'IRN, d'améliorer le savoir-faire en matière de données, d'augmenter la capacité de l'IRN, d'accroître les investissements en la matière, de réévaluer les cadres de gestion des données de recherche, et de tenir compte de la diversité des établissements, des ressources nécessaires pour faciliter la collaboration et du transfert de données entre les établissements, de cadres de sécurité et de la formation connexe et, enfin, de la nécessité de faire largement connaître les ressources à la disposition des équipes de recherche.

La communauté de la recherche de l'Université d'Ottawa est heureuse d'avoir pu faire connaître son opinion à la NOIRN. Nous sommes d'ailleurs déterminés à poursuivre cette collaboration et ce partenariat productifs.