

Livre blanc

Clés pour une Infrastructure de recherche numérique (IRN) innovante

Décembre 2020

Auteur·e·s :

Suzanne Talon, directrice générale de Calcul Québec (contact principal)

Marie-Jean Meurs, directrice scientifique de Calcul Québec

Victor Ionescu, directeur des systèmes d'information de Calcul Québec

Table des matières

- Construire la confiance
- Assurer l'engagement des chercheurs et chercheuses
- Offrir des services de qualité en français
- Répondre aux besoins émergents
 - Des environnements d'exploitation adaptés
 - Un nuage sécurisé
 - La gestion des données de recherche actives
 - Des logiciels de recherche « libres »

Construire la confiance

Le Québec a été un pionnier dans le domaine du calcul informatique de pointe au pays. Dès 1992, en collaboration avec l'université de Montréal et l'université d'Ottawa, l'université de Sherbrooke créait le *Centre d'application du calcul parallèle de l'université de Sherbrooke* et se dotait d'un SP2 de IBM. Cette infrastructure a permis aux chercheur·e·s de ces institutions de bien figurer au palmarès des recherches basées sur la simulation numérique.

En 1999, les universités de Montréal et de Sherbrooke, accompagnées de Polytechnique Montréal et des universités Concordia et Bishop, ont formé le *Réseau québécois de calcul de haute performance* (RQCHP) afin de faire appel à la *Fondation canadienne pour l'innovation* (FCI) pour le financement d'une puissance de calcul importante. Le RQCHP est alors devenu un leader dans l'installation de grappes Beowulf. En 2003, les universités McGill, Laval et le réseau de l'université du Québec ont formé le *Consortium Laval, UQAM, McGill et Est du Québec* (CLUMEQ). Plutôt que d'évoluer en parallèle, dès 2007 ces deux organisations ont initié des rapprochements, créant un concours d'allocations de ressources de calcul en

commun à partir de 2009 et des appels à projet pour du temps dédié d'analystes à partir de 2010.

En 2011, la collaboration interinstitutionnelle s'est consolidée dans la création de **Calcul Québec**, qui a remplacé les deux organisations précédentes. Alors qu'initialement la collaboration s'est centrée surtout sur des activités de soutien aux chercheur·e·s, en 2016, sous l'impulsion de la réduction du nombre de sites nationaux, les quatre universités québécoises hébergeant des infrastructures majeures ont décidé d'unir leurs forces afin de préparer une proposition de site québécois en co-gestion, tirant profit de l'expertise du personnel de ces quatre institutions. C'est Calcul Québec qui assure la planification, l'installation, la gestion, l'entretien et la disposition des actifs, avec une équipe distribuée dans plusieurs institutions et en collaboration avec nos partenaires institutionnels.

Cette décision de travailler en collaboration plutôt qu'en compétition illustre la volonté des universités québécoises de faire bénéficier l'ensemble de la communauté des forces de chacun et reflète l'esprit qui prévaut au comité de la recherche du *Bureau de collaboration interuniversitaire* (BCI).

En 2020, Calcul Québec devient une entité incorporée, une organisation indépendante, centrée sur les besoins des universités et des chercheur·e·s du Québec. Cette indépendance est cruciale à l'accomplissement de notre mission, puisqu'elle permet de s'abstraire des enjeux politiques. Le *Ministère de l'Économie et de l'Innovation du Québec* (MEI) salue la création de cette OBNL qui regroupe les forces vives de toute la communauté universitaire.

Mission

Procurer au milieu universitaire ainsi qu'à la communauté des utilisateurs et utilisatrices des **environnements informatiques de pointe** ainsi que de **l'expertise-conseil**, afin de contribuer à l'avancement des connaissances dans toutes les branches du savoir et à la formation de personnel hautement qualifié, capable d'exploiter efficacement les systèmes informatiques modernes.

Cette année également, Calcul Québec a été invitée à postuler au tout nouveau programme de financement AGORA des *Fonds de recherche du Québec* (FRQ), ce qui lui permettra de devenir l'une des trois premières plateformes du Scientifique en chef de la province.

Assurer l'engagement des chercheurs et chercheuses

L'un des éléments clé du modèle de *Calcul Québec* (CQ) est de rester proche de la communauté de la recherche. Concrètement, les règlements administratifs de CQ prévoient l'existence d'un conseil scientifique, qui agit à titre de comité consultatif permanent auprès du conseil d'administration et de la direction générale, composé d'au moins sept chercheur·e·s. Ces personnes proviennent à la fois des membres universitaires et des membres collaborateurs (p. ex. centres de recherche, incluant les hôpitaux de recherche, centres collégiaux de transfert technologique). CQ a également une direction scientifique, qui est impliquée au quotidien dans les décisions organisationnelles et d'orientation.

L'équipe d'analystes de CQ est répartie dans les universités et interagit avec la communauté locale, prenant le pouls des besoins. La proximité entre les analystes et les personnes administrant les systèmes, qui sont également réparties, permet une bonne adéquation entre les besoins de la recherche et les outils déployés sur les systèmes.

Offrir des services de qualité en français

Calcul Québec joue un rôle fondamental dans l'offre de service en français. Il ne s'agit pas simplement d'offrir des traductions, mais bien de **penser des contenus d'une manière différente**, ce qui est particulièrement important dans le domaine des sciences humaines et sociales.

La collaboration pancanadienne permet de généraliser cette offre aux chercheur·e·s de tout le Canada, et de positionner le pays à l'échelle internationale, en particulier en rapport à l'Europe pour qui le multilinguisme constitue un élément fondamental.

Autre point d'intérêt, le bilinguisme de la documentation est en fait un atout pour plusieurs, dont l'anglais est la langue seconde. En effet, lors de la traduction des textes, les équipes de rédaction sont amenées à remettre en cause le texte original, ce qui entraîne très souvent une amélioration de la version originale en anglais.

Répondre aux besoins émergents

Des environnements d'exploitation adaptés

Le calcul informatique de pointe (CIP) prend de l'importance dans de nombreux domaines de recherche. Cependant, les formations universitaires ne se sont pas adaptées à cette nouvelle réalité, sans compter le fait que les chercheur·e·s et les étudiant·e·s en sciences humaines et sciences de la santé ont souvent peu d'intérêt pour la programmation et les

environnements informatiques en général. **L'accès à la puissance des outils du CIP passe par le développement de plateformes intuitives et de modes d'accès simplifiés.**

Un nuage sécurisé

La recherche requiert de plus en plus de ressources de calcul performantes. Cependant, les environnements actuels ne sont pas toujours appropriés pour accueillir des données ayant des exigences de sécurité et de confidentialité, comme dans les domaines de santé et des services sociaux. En collaboration avec le Canadian Centre for Computational Genomics (C3G), CQ développe depuis plus d'un an un service d'**infrastructure d'infonuagique sécurisée**. Cette plateforme **libre et indépendante de toute infrastructure privée** s'inspire du projet de *HPC4Health* en Ontario, la plateforme qui est utilisée par la plateforme de données sur la santé pour aider à vaincre la COVID-19 en Ontario¹, un élément essentiel de la stratégie mise de l'avant pour lutter contre la pandémie dans cette province. CQ sera en position de jouer un rôle similaire dans l'environnement québécois.

La gestion des données de recherche actives

La gestion des données de recherche, afin de permettre une utilisation plus efficace des données colligées et générées par les chercheur·e·s, est en cours de structuration un peu partout au pays. La NOIRN a reçu comme mandat de structurer les efforts au niveau national et les universités sont en train de se doter de capacités pour appuyer les chercheur·e·s localement et de leur permettre de stocker des petits jeux de données dans des instances Dataverse associées à leur institution.

Les efforts actuels se concentrent cependant autour de la création d'un plan de gestion de données statique et du développement des dépôts de données pour la réutilisation future. La gestion des données actives est laissée aux chercheur·e·s, alors qu'il·elle·s ont peu de connaissances dans ce domaine. Calcul Québec pourra jouer un rôle important dans **l'accompagnement des chercheur·e·s dans leur gestion des données actives**. Ceci passera également par la création d'un outil pour **un plan de gestion de données dynamique**, évoluant avec l'avancement et le développement des projets et faisant le lien entre le calcul informatique de pointe et la gestion des données. La création de cet outil et la formation d'une équipe spécialisée permettront de :

- aider les chercheur·e·s à structurer leurs données, améliorant leur efficacité et leur capacité à travailler en équipe;
- anticiper la création des métadonnées, facilitant ainsi leur préservation et leur réutilisation dans le futur;

¹<https://news.ontario.ca/fr/release/56660/la-province-cree-une-nouvelle-plateforme-de-donnees-sur-la-sante-pour-aider-a-vaincre-la-covid-19>

- anticiper le transfert de propriété intellectuelle entre étudiant·e·s et chercheur·e principal·e lors de la fin du cycle d'études;
- éviter l'encombrement des systèmes de fichiers par des fichiers temporaires non purgés;
- éviter la duplication de jeux de données de référence téléchargés plusieurs fois par des personnes différentes dans un répertoire local plutôt que dans un dépôt central partagé.

L'embauche de personnel spécialisé ayant des connaissances à la fois en calcul informatique de pointe et en gestion de données de recherche sera essentielle.

Des logiciels de recherche « libres »

Le développement des logiciels de recherche est intimement lié à l'exploitation efficiente des ressources numériques. Ces **logiciels sont à l'interface entre les chercheur·e·s et l'infrastructure**. Il est donc primordial d'en avoir une **approche intégrée à la stratégie de développement des nouvelles infrastructures libres et indépendantes** et d'encourager leur déploiement en mode de **logiciel libre**. Dans ce contexte, rapprocher les équipes de développement des équipes de soutien aux chercheur·e·s constitue un élément de succès.

Finalement, nous aimerions mentionner que plusieurs logiciels libres jouent un rôle critique pour le bon fonctionnement de nos infrastructures et environnements (p.ex. Linux, Lustre, OpenStack, Slurm, CVMFS, EasyBuild, Jupyter, etc.). Ce sont des logiciels fondamentaux pour supporter la recherche. La NOIRN devrait avoir une **stratégie encourageant les contributions et le soutien de ces logiciels** par les équipes techniques qu'elle finance.