

La cyber-infrastructure et le processus de recherche au Canada

Constatations de l'atelier tenu à Toronto les 17 et 18 décembre 2009

Parrainé par :
CANARIE, CDPIUC, Calcul Canada

Stephen Quesnelle
Partenaire - Sacred Cow Company
Squesnelle@sacredcowcompany.com
Janvier 2010

Rapport de l'atelier sur la cartographie de la cyber-infrastructure

Les 17 et 18 décembre 2009, CANARIE, le CDPIUC et Calcul Canada invitaient un groupe d'experts de divers domaines et d'intervenants du processus de recherche à participer à un atelier, à Toronto. L'exercice devait servir à définir, à préciser et à étayer la manière dont les chercheurs identifient puis obtiennent les ressources de la cyber-infrastructure dont ils ont besoin pour poursuivre leurs travaux, ainsi que la situation actuelle en la matière.

Participation

Quatorze personnes des quatre coins du pays et de divers milieux représentaient les groupes d'intervenants. Étaient représentés les chercheurs de diverses disciplines, les chercheurs en informatique, le secteur du calcul à haute performance (CHP), les réseaux national et provinciaux (CANARIE et RORE) et les dirigeants principaux de l'information. Étaient aussi présents cinq observateurs des groupes parrainant l'atelier. Un membre du secteur des sciences sociales a malheureusement été contraint d'annuler sa participation à la dernière minute, en raison d'un conflit.

Déroulement

L'atelier avait été conçu pour maximiser l'interactivité et faciliter la participation de chacun. Le groupe a été scindé en comités pour chaque activité. Chaque comité a rapporté ses constatations à la fin de chaque séance de façon à permettre une comparaison des similitudes et des divergences dans le travail de chacun.

Résultats

- 1.** Définitions communes pour le CHP, l'informatique répartie, l'informatique en nuage et la cyber-infrastructure.
- 2.** Carte du processus de recherche en fonction de la cyber-infrastructure pour mieux cerner les difficultés.
- 3.** Solutions à certains problèmes.
- 4.** Vision pour la cyber-infrastructure en 2015.

1. Définitions : cyber-infrastructure, CHP, informatique répartie, informatique en nuage

Pour que le groupe utilise la même terminologie, l'atelier a débuté par un exercice qui consistait à définir le calcul à haute performance (CHP), l'informatique répartie, l'informatique en nuage et la cyber-infrastructure.

Cyber-infrastructure : Milieu intégré subvenant aux besoins complets des scientifiques pour leurs recherches. Il s'agit de l'intégration des ressources de traitement et de gestion de l'information incluant les suivantes :

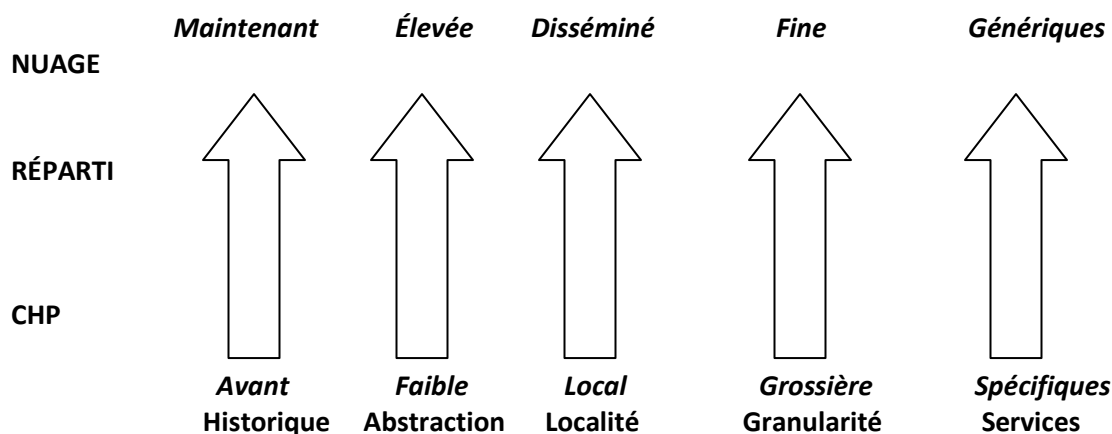
- calcul
- stockage
- réseaux
- réseaux de capteurs
- intergiciels
- personnel/expertise

Calcul à haute performance (CHP) : Tout ce qui assure une grande puissance de calcul, c'est-à-dire les calculs au-delà de ceux réalisables avec un ordinateur personnel.

Informatique répartie : Accès coordonné aux ressources réparties dans plusieurs domaines administratifs.

Informatique en nuage : Ressources communautaires virtuelles, disponibles à la carte, en informatique. Elles sont distinctes de l'environnement local et sont toutes considérées comme des services.

Un comité a brossé un tableau qui illustrent les ressources en nuage, les ressources réparties et le CHP vues sous l'angle de l'utilisateur.



Remarque : Ce graphique a été élaboré par un groupe afin de donner une idée de la direction empruntée par le développement. On ne devrait pas y voir une indication que le CHP a été supplanté par l'informatique en nuage ou le calcul réparti. À l'heure actuelle, l'informatique en nuage n'autorise pas les services de CHP.

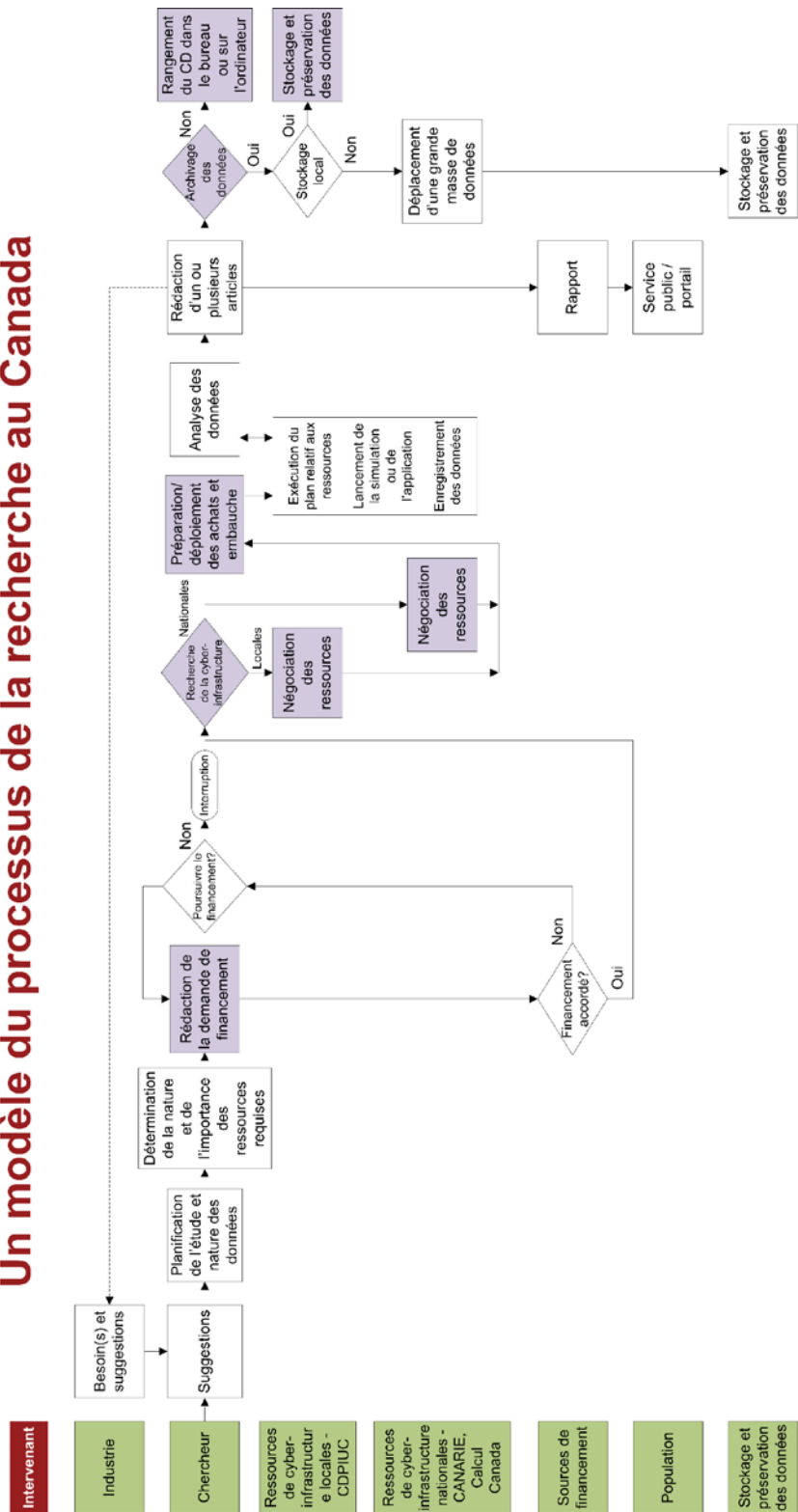
2. Cartographie du processus : identification des besoins de cyber-infrastructure et accès aux ressources par les chercheurs

Lorsqu'on trace la carte d'un processus interfonctionnel, il est possible de bien visualiser le processus dans son ensemble. Une carte bien conçue indiquera les principaux acteurs, les activités et les points où une décision doit être prise, mais aussi ceux où il y a rupture, engorgement ou incohérence.

Comment lire le diagramme d'un processus interfonctionnel

- Les intervenants sont énumérés verticalement dans la colonne de gauche.
- Des flèches indiquent le déroulement ou la séquence des activités. Bien que certaines puissent survenir en même temps, on se limite habituellement à un seul cheminement afin de ne pas embrouiller la lecture du diagramme.
- Les cases indiquent les mesures ou les activités. Les losanges correspondent aux points où une décision doit être prise.
- La personne ou le groupe responsable d'une activité est identifié dans une case sur la même ligne horizontale que l'intervenant. Par exemple, la colonne de gauche donne le nom des intervenants. En suivant cette ligne horizontalement, on verra les activités dont l'intervenant en question assume la responsabilité.

Un modèle du processus de la recherche au Canada



Constataions de l'atelier de Toronto - Décembre 2009
Parrainé par CANARIE, le CDPIUC et Calcul Canada

Remarque : Ce graphique devait faire ressortir les obstacles ou les difficultés observés au niveau de l'aide dispensée à la recherche, au Canada. Il s'agit d'un modèle de la démarche scientifique qui ne vise aucunement à refléter avec exactitude tous les aspects de la recherche poursuivie au pays.

Principales difficultés

Financement

- Le financement est fragmenté entre divers organismes dont les mécanismes manquent d'uniformité (importance variable des détails demandés, formulaires différents). Rédiger une demande de fonds détourne les chercheurs de leurs travaux.
- Même si plusieurs organismes sont sollicités, il arrive qu'un projet ne soit financé qu'en partie.
- Les organismes subventionnaires essaient de répartir les fonds entre de nombreux projets, de sorte qu'il est plus ardu de lancer les projets de plus grande envergure ou onéreux, de nombreux organismes devant être sollicités pour obtenir des fonds.
- Le financement de projets interdisciplinaires est plus difficile faute de mécanismes d'examen par des pairs pour évaluer les demandes.

Identification des ressources de la cyber-infrastructure et accès à ces dernières

- Il n'existe aucun organisme unique duquel on pourrait obtenir les ressources. Les scientifiques apprennent par expérience comment obtenir ce dont ils ont besoin (souvent trop tard dans le processus de recherche).
- Les chercheurs ignorent souvent de quelle sorte de cyber-infrastructure ils ont besoin, car leur expertise se situe rarement à ce niveau. Par conséquent, ils identifient les ressources nécessaires à leurs travaux soit au petit bonheur (en posant des questions), soit en adoptant la solution de facilité et en se limitant aux ressources les plus aisées à obtenir, qu'elles conviennent ou pas au type de recherches poursuivies.
- Il est très difficile d'obtenir des fonds à court terme pour les technologies de l'information.
- *Réseaux évolués* : Il est parfois difficile d'établir une connectivité d'une extrémité à l'autre – les pare-feu entraînent parfois des engorgements. Le problème se corse avec les projets internationaux. Les politiques relatives à l'exploitation du réseau CANARIE posent des difficultés quand le contenu doit être diffusé au grand public (à savoir quand on franchit la ligne entre le réseau CANARIE et le réseau commercial). Les chercheurs éprouvent parfois d'énormes difficultés à accéder au réseau de recherche quand ils se trouvent hors de leur institution.
- *Intergiciels* : On ne possède pas les moyens pour appuyer le développement d'une plateforme d'intergiciels nationale et il est très difficile de coordonner les activités de sorte que les équipes de recherche mettront au point leur propre plateforme, une méthode qui s'avère aussi très inefficace. Aucun groupe canadien ne se penche sur les intergiciels et il est difficile de financer cet aspect (bien que CANARIE procure certains fonds). C'est pourquoi le Canada tire de l'arrière sur les autres pays.
- *Télédéttection* : Les organismes de financement exigent habituellement des propositions et des données différentes dans des formats variés.
- *Visualisation* : Ce domaine peine à s'implanter au Canada. Le manque d'expertise est à la racine du mal. Il faut inclure la visualisation à la cyber-infrastructure.
- *Calcul* : Le Canada doit rester concurrentiel avec les autres pays s'il veut attirer et garder les chercheurs. Une capacité insuffisante poussera le chercheur à réduire l'ampleur de ses travaux en fonction de la capacité existante, ce qui diminuera l'utilité de ses recherches. Beaucoup de chercheurs préféreraient concourir au maintien d'une installation qu'acquérir, installer et exploiter leur propre infrastructure.

- *Outils de collaboration* : Outils pour les vidéoconférences, le clavardage etc. afin d'aider les chercheurs à travailler ensemble. Un réseau d'accès réparti devrait-il faire partie de cet aspect et devrait-on l'implanter sous la tutelle des DPI? En font partie des éléments comme eduoam.

Personnel/experts pour exploiter et gérer la cyber-infrastructure

- Les équipes de recherche possèdent rarement l'expertise nécessaire pour exploiter l'équipement et optimiser les logiciels. Les organismes subventionnaires ne financent pour la plupart que les immobilisations.
- Le manque de personnel hautement qualifié en mesure d'appuyer les chercheurs signifie que ceux-ci doivent s'occuper eux-mêmes des TI ou confier ce travail à leurs étudiants.
- Les méthodes locales pour engager le personnel qui assistera les chercheurs ne permettent pas toujours à ces derniers d'embaucher la personne idéale pour la tâche.

Archivage et stockage adéquats des données

- Les chercheurs ne disposent pas de bons mécanismes pour stocker, préserver et partager leurs données. Il est capital de bien stocker et préserver les données si l'on veut les valider et les comparer. Dans certains cas, les résultats sont mal archivés et préservés (« on les grave sur un CD qui est gardé dans le tiroir »). Parfois, on égare des données, ce qui équivaut à dilapider les fonds investis dans la recherche.

3. Solutions possibles à quelques problèmes

Le groupe a réfléchi aux solutions qu'on pourrait apporter à certains problèmes, a identifié ceux qui pourraient y remédier et a donné une priorité à la réalisation d'une solution.

PROBLÈME, RESPONSABLE et PRIORITÉ	SOLUTIONS POSSIBLES
<p>Le problème du « dernier mille » : des conduites à haut débit acheminent les données à la périphérie du campus, où des pare-feu, des fibres moins performantes et des engorgements ne permettent pas aux chercheurs de déplacer et de partager aussi facilement les données.</p> <p>Responsable : CDPIUC Priorité : élevée</p>	<ul style="list-style-type: none"> • Fournir aux DPI une liste d'exigences et une architecture de référence pour les aider à mettre en place des réseaux qui soutiendront à la fois l'enseignement et la haute performance • Amener chaque université à œuvrer davantage avec les RORE • Faire en sorte que les universités soutiennent davantage les centres hors campus • Sensibiliser davantage les chercheurs au processus pour qu'ils sachent ce qui est disponible
<p>Absence de grandes plateformes de développement bien intégrées</p> <p>Responsables : CANARIE, Calcul Canada Priorité : moyenne à élevée</p>	<ul style="list-style-type: none"> • CANARIE devrait financer une sorte de Programme de plateformes sur réseau (PPR) plus général pour permettre aux intergiciels d'évoluer • Donner accès aux ressources et aux installations pour tester les technologies novatrices mises au point dans le cadre du PPR • On a besoin d'un fonds de développement pour mettre en place les conditions propices à la production, y compris un mécanisme pour élaborer des plateformes de développement intégrées
<p>Le financement est fragmenté et son inefficacité empêche les chercheurs de se consacrer à la recherche</p> <p>Responsable : organismes subventionnaires Priorité : élevée</p>	<ul style="list-style-type: none"> • Fournir aux chercheurs UN seul formulaire de demande et réunir les organismes subventionnaires pour qu'ils l'examinent ensemble et octroient des fonds • Aider les chercheurs à fournir l'information appropriée dans leurs propositions • Faire évaluer les projets interdisciplinaires par des pairs • Créer un mécanisme pour qu'on passe moins de temps sur les ententes • Réduire les besoins d'administration générale et de comptabilité, car ces activités prennent trop de temps au début et à la fin des projets
<p>Les grandes organisations manquent de souplesse et ne peuvent se plier aux demandes particulières</p> <p>Responsable : Calcul Canada</p>	<ul style="list-style-type: none"> • Inciter les investisseurs à se conformer aux demandes réglementaires et aux demandes spéciales en matière de sécurité • Assurer la conformité des demandes traitées en lot • Les systèmes/ordinateurs ne se conforment pas tous à

<p>Priorité : moyenne à élevée</p>	<p>l'HIPPA</p>
<p>Les chercheurs connaissent mal les ressources de la cyber-infrastructure à leur disposition. On rate des occasions</p> <p>Responsables : universités, CDPIUC Priorité : élevée</p>	<ul style="list-style-type: none"> • Nécessité de signaler aux chercheurs quelles ressources de la cyber-infrastructure sont à leur disposition et lesquelles conviennent le mieux à leurs travaux • Former les DPI pour qu'ils procurent les services requis ou aiguillent les chercheurs dans la bonne voie; nécessité d'une architecture de référence • Distribuer une brochure sur les ressources aux étudiants des cycles supérieurs • Donner la liste des ressources de la cyber-infrastructure sur la première page des anti-virus pour s'assurer qu'on en prenne connaissance • Utiliser un site Web du genre Facebook ou un portail de la cyber-infrastructure pour énumérer les ressources • Regrouper CANARIE, le CDPIUC et Calcul Canada et mettre en place une entreprise de communications qui servirait de guichet unique pour la cyber-infrastructure • Créer un organisme national qui examinerait tous les projets et les composants connexes de la cyber-infrastructure pour s'assurer que celle-ci soit rationnelle et efficace, et qu'on la dote d'une stratégie pour stocker et préserver les résultats des projets • Organiser des présentations d'une demi-journée pour les chercheurs; assister à leurs colloques. Sensibiliser le gouvernement, les organismes subventionnaires, les institutions, les chercheurs et les étudiants des cycles supérieurs • S'associer aux bibliothèques
<p>On stocke et gère mal les données. On ne partage pas les données dont une partie est perdue complètement.</p> <p>Responsable : tous Priorité : élevée</p>	<ul style="list-style-type: none"> • Créer un dépôt de données avec des normes et une architecture autorisant le partage • Inclure un plan de gestion des données à la proposition • S'associer à la Bibliothèque nationale ou à Archives Canada • Sensibiliser les chercheurs à l'importance de cet aspect

4. L'avenir – qu'envisage-t-on pour la recherche et la cyber-infrastructure en 2015?

Le groupe a recommandé que les organismes qui parrainaient l'exercice rédigent un court document (3 ou 4 pages) sur ce qu'ils envisagent pour la cyber-infrastructure canadienne dans l'avenir.

Après réflexion, les participants estiment que les éléments que voici devrait faire partie de cette vision pour 2015, car ils intègrent une vision nationale cohérente ainsi que la cyber-infrastructure stable et durable, indispensable à la recherche.

Structure organisationnelle

- Consolidation des organisations et création d'une nouvelle entité appelée « Recherche Canada » qui servirait de guichet unique aux chercheurs pour leurs besoins de cyber-infrastructure (fonds, soutien, ressources)
- Organisation qui permettrait d'engager du personnel pour identifier les besoins des projets de recherche au niveau de la cyber-infrastructure et posséderait les compétences pour accéder à ces ressources
- Établissement de centres d'excellence pour la recherche sur la cyber-infrastructure sur les plans de l'informatique, de la technologie, des intergiciels et de l'archivage des données
- Organisme stable et durable qui permettrait à la fois la recherche sur la cyber-infrastructure et offrirait des services connexes – affectation dynamique de ressources en fonction des besoins et capacité de réagir à la nécessité de nouveaux programmes et applications
- Organisation capable de tirer parti des possibilités dans les domaines des TI vertes et de l'atténuation de l'empreinte carbone
- Organisation qui aidera les chercheurs à faire le pont avec l'industrie

Financement simplifié

- Paysage simplifié de la recherche avec guichet unique pour les besoins de cyber-infrastructure – financement, soutien, ressources, outils intégrés et technologie
- Processus prévoyant un seul processus pour les demandes de financement et l'évaluation des projets; ce processus inclurait tous les besoins de cyber-infrastructure y compris l'expertise en la matière

Soutien sous forme de :

- Services regroupés qui aideront les chercheurs à présenter leurs demandes de financement et à mener à bien leurs projets
- Mécanismes pour partager, stocker et préserver facilement les données
- Accès faciles aux pratiques exemplaires
- Technologies de collaboration et visualisation coopérative pour partager un environnement sur de grandes distances