

Séminaire jeudi le 30 avril 2015 11:00h / Seminar Thursday April 30<sup>th</sup> 2015  
11:00h

**Sujet/Subject:** Une approche de requête de base de données permettant des analyses efficaces et reproductibles des données de modèles globaux du CMIP5 /  
A database-query approach for the robust and efficient analysis of climate model outputs in the CMIP5 archive

**Langue/language** : Français

**Conférencier/Lecturer:** Frédéric Laliberté (visiting fellow, Environment Canada, Downsview, Toronto)

### **Résumé:**

Les intercomparaisons de modèles du système terrestre (MIPs) sont possible grâce à des bases de données d'échelle péta distribuées à travers le monde et qui sont continuellement mises à jours par les groupes de modélisateurs y contribuant. Pour la plupart des utilisateurs, il est difficile de découvrir au sein de ces archives les données appropriées pour leurs requêtes ainsi que d'y accéder. De plus, lorsque que ces obstacles sont surmontés, une analyse requiert souvent le téléchargement et le stockage de grandes quantités de données, ce qui a tendance à pousser les infrastructures informatiques existantes à leur limite. Dans cette présentation, je vais présenter des outils informatiques pour l'analyse efficace des MIPs tout en discutant des difficultés que j'ai rencontrées lors de l'utilisation de ces outils pour la réduction de grands jeux de données. Je vais m'attarder sur l'idée d'efficacité dans les sens humains et informatiques tout en mettant l'emphase sur la reproductibilité des résultats. Cet outil se sert des capacités de recherche avancées des archives distribuées pour limiter les quantités de données qui doivent être téléchargées. De plus, il offre un méthode structurée pour conserver les résultats des recherches, rendant ces résultats réutilisables.

Je vais passer la journée à Dorval et serai disponible pour aider quiconque le désire à installer ces outils informatiques.

### **Abstract:**

Earth System Model intercomparison projects (MIPs) are associated with globally distributed peta-scale data repositories that are continually updated by contributing modeling groups. Scientific analysis of such large archives has been hampered by technicalities of data discovery and access for the typical user. Moreover, it often

requires download and storage volumes that strain available computing resources at many smaller institutions. Here I will present an informatics tool for efficient analysis of MIPs and will discuss challenges I have encountered using these tools for the processing of large datasets from the point of view of a user without a privileged access to a data node. I will focus on efficiency in terms of both human and computing resources as well as reproducibility and timeliness. The tools rely on advanced search and data access services of the distributed archive to reduce the volume of data required to be transferred. It also provides a structured approach to storing query responses, mirroring the data structure within the archive, making these responses re-usable. Simple benchmarking will be presented and a discussion of the challenges of using such an integrated tool over the last four years will be introduced.

I will be spending the day at Dorval and will be available to help those who want to get started with the informatics tool.