

# Session 2

## L'infonuagique pour la recherche en sciences humaines et sociales

Chris Geroux

Chris utilise le calcul haute performance depuis 2005 et s'est joint à ACENET en 2015 où il se spécialise en infonuagique et en données massives. Il détient un doctorat en astronomie pour son travail sur la modélisation numérique de la dynamique des fluides des étoiles, ce qui l'a conduit à un poste de niveau postdoctoral avec l'université d'Exeter. Chris fait partie de l'équipe nationale des services infonuagiques de la fédération Calcul Canada et il soutient les chercheuses et chercheurs dans leur utilisation des nuages de la fédération, notamment par la rédaction de documentation disponible au public. Il a aussi dirigé plusieurs ateliers sur le calcul parallèle, le langage Python et l'infonuagique.





# L'infonuagique pour la recherche en sciences humaines et sociales

*Accélérer la découverte avec la recherche numérique*

*info@ace-net.ca  
www.ace-net.ca*

**compute | calcul**  
canada | canada  
*partenaire régional*

# Objectifs

À la fin de cette présentation,

- vous saurez ce qu'est un nuage;
- vous aurez un aperçu de comment travailler dans un nuage;
- vous verrez comment les nuages sont utilisés en sciences humaines et sociales;
- vous saurez la différence entre un site Web dynamique et un site Web statique;
- vous connaîtrez des points importants à considérer dans la planification de votre site Web.

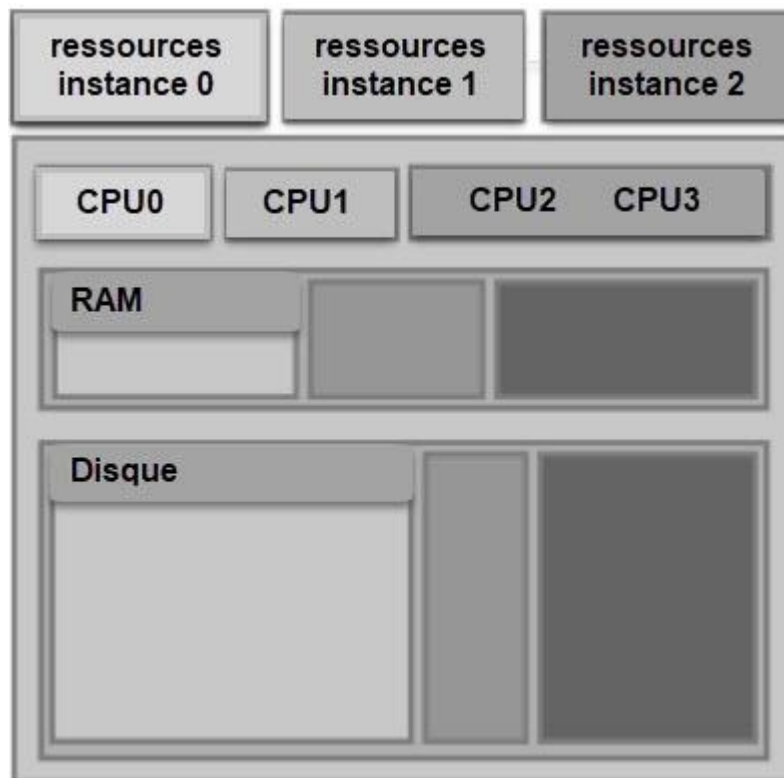
# Qu'est-ce qu'un nuage?

- Dépendant du contexte, on parle  
d'applications infonuagiques, par exemple Gmail, calendriers, documents, etc.,  
de l'hébergement d'un site Web,  
d'une infrastructure infonuagique.
- Un nuage est physiquement constitué d'un groupe d'ordinateurs auquel il est possible d'accéder à distance via des applications et des outils de connexion.

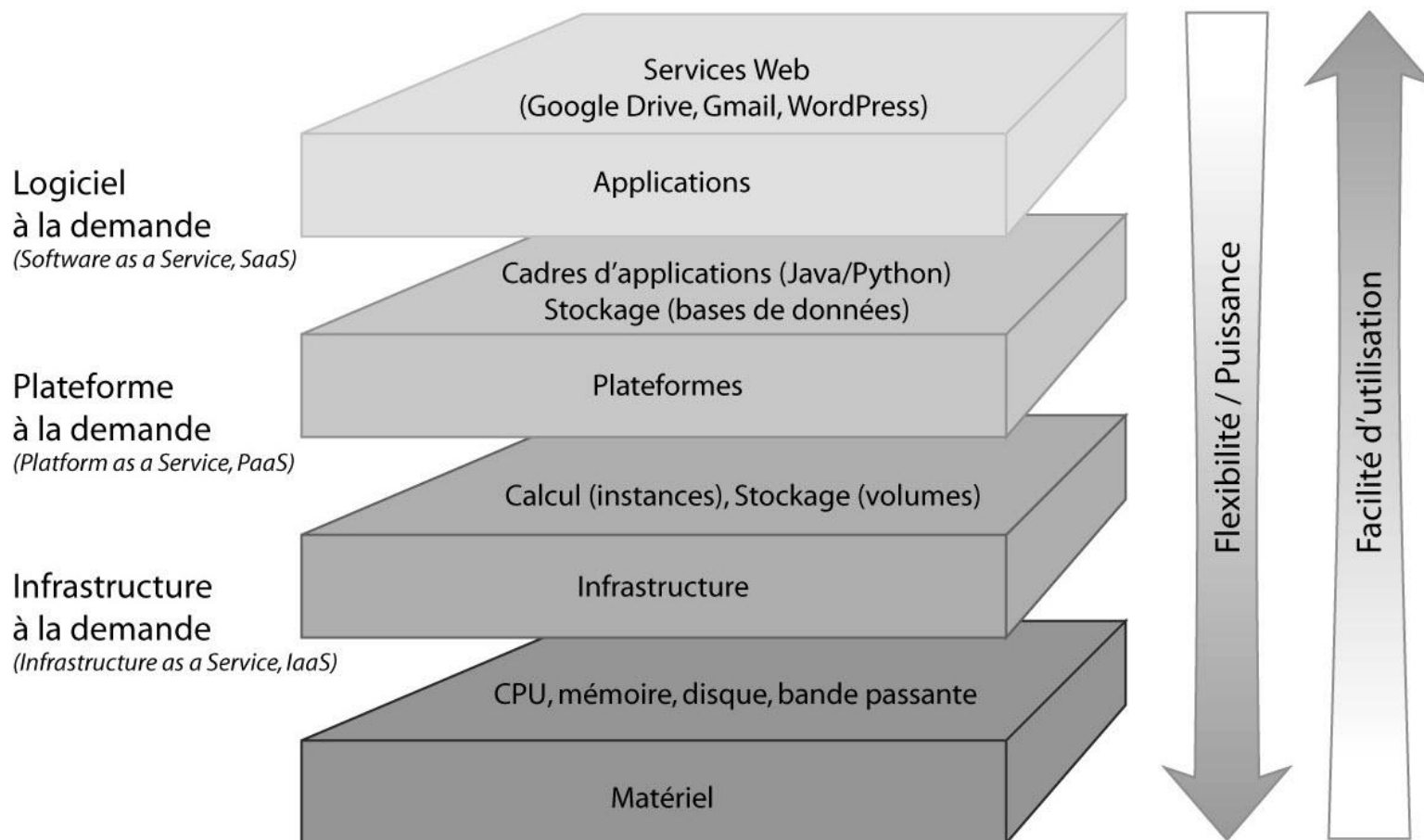


## Qu'est-ce qu'une instance?

- un ordinateur simulé;
- en opération sur un ordinateur physique;
- peut utiliser partiellement l'ordinateur physique;
- son système d'exploitation peut être différent de celui de l'ordinateur physique;
- une instance peut être en opération sur votre ordinateur local (voir [virtualbox.org](https://www.virtualbox.org)).



# Niveaux de service



## Nos nuages sont sur le modèle IaaS

- Un nuage IaaS permet une grande flexibilité.
- Par contre, ce modèle demande plus d'efforts et de connaissances que PaaS et SaaS :
  - les instances sont installées et configurées manuellement;
  - elles nécessitent plus de maintenance, notamment pour les mises à jour et le suivi des journaux (*logs*);
  - les instances doivent souvent être recrées à chaque changement de version majeure de leur système d'exploitation.
- Grande flexibilité :
  - dans le choix du système d'exploitation;
  - dans le choix de la configuration matérielle.
- La fédération Calcul Canada aimerait offrir plus que IaaS (PaaS et SaaS), mais ne dispose pas suffisamment de ressources humaines et financières.

# Nuages IaaS de la Fédération Calcul Canada

## Arbutus

- 484 nœuds (16 008 cœurs)
- 5.3 Po + 12 Po de stockage
- 153.5 To de mémoire vive (RAM)

## Nuage East

- 36 nœuds (576 cœurs)
- 100 To de stockage
- 4.5 To de mémoire vive (RAM)

## Nuage Cedar

- 32 nœuds (1 024 cœurs)
- 500 To de stockage
- 7.5 To de mémoire vive (RAM)

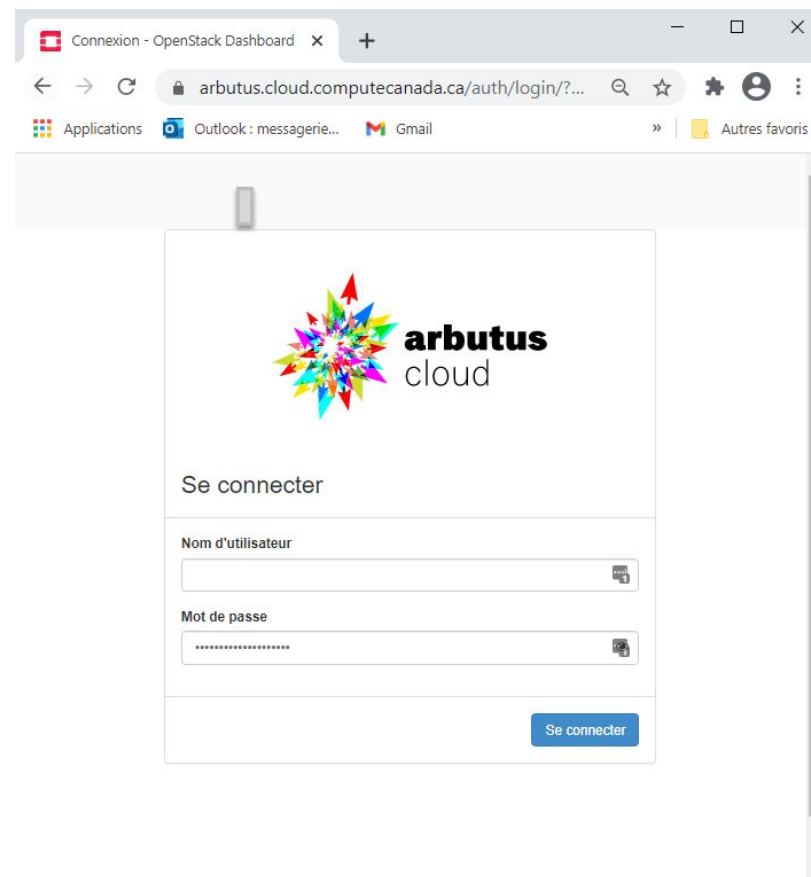
## Nuage Graham

- 24 nœuds (768 cœurs)
- 84 To de stockage
- 5.5 To de mémoire vive (RAM)



# Système d'exploitation OpenStack

- Son rôle est de gérer
  - la création et la suppression des instances,
  - le stockage,
  - la réseautique,
  - les quotas imposés sur les ressources,
  - la facturation.
- Site Web [OpenStack.org](https://openstack.org).
- Les équipes de recherche des établissements universitaires canadiens peuvent utiliser les nuages de la fédération gratuitement.



# Créer une instance sous OpenStack

Lancer Instance

Détails \*  
Source \*  
Gabarit \*  
Réseaux  
Ports réseaux  
Groupes de sécurité  
Paire de clés  
Configuration  
Groupes de serveurs  
Scheduler Hints  
Métadonnées

Veuillez fournir le nom d'hôte initial de l'instance, la zone de disponibilité où elle sera déployée ainsi que le nombre d'instances. Augmenter le nombre pour créer plusieurs instances avec les mêmes paramètres.

Nom de l'instance \*

Description

Zone de disponibilité

Nombre \*

Total des instances  
(2 Max)

100%

■ 1 Utilisation actuelle  
■ 1 Ajouté  
■ 0 Restant

✕ Annuler

< Retour

Suivant >

Lancer Instance

# Travailler avec une instance

- SSH
  - configuration des instances
  - mise à jour des instances
- SFTP
  - transfert et édition des fichiers
- Interface graphique
  - en installant un bureau (comme **MATE**)
  - en installant **VNC** pour permettre des connexions au bureau distant

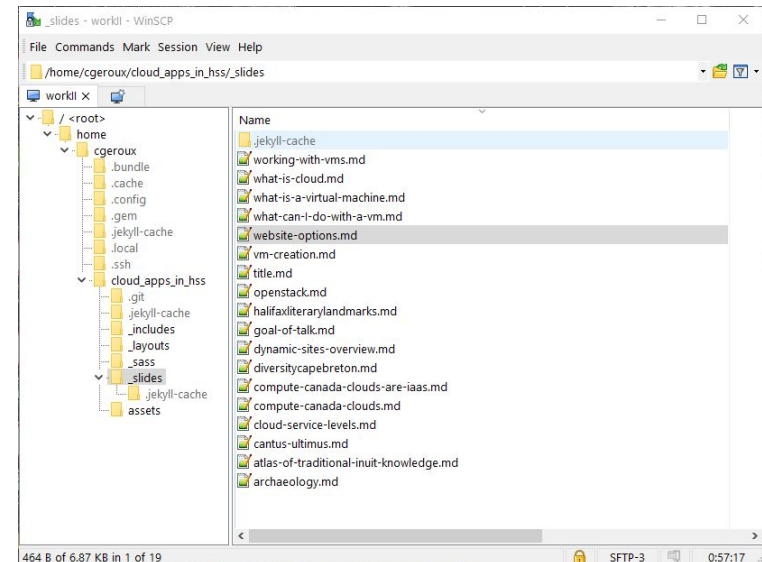
```
cgeroux@workii: ~/cloud_apps_in_hss
Destination: /var/www/html
Incremental build: disabled. Enable with --incremental
Generating...
Jekyll Feed: Generating feed for posts
done in 0.168 seconds.
Auto-regeneration: disabled. Use --watch to enable.
cgeroux@workii:~/cloud_apps_in_hss$ sudo jekyll build -d /var/www/html/
Warning: the running version of Bundler (2.1.2) is older than the version that created the lockfile (2.1.4).
We suggest you to upgrade to the version that created the lockfile by running 'gem install bundler:2.1.4'.
The dependency tzinfo (~> 1.2) will be unused by any of the platforms Bundler is installing for. Bundler is
installing for ruby but the dependency is only for x86-mingw32, x64-mingw32, x86-mswin32, java. To add those
platforms to the bundle, run 'bundle lock --add-platform x86-mingw32 x64-mingw32 x86-mswin32 java'.
The dependency tzinfo-data (~> 0) will be unused by any of the platforms Bundler is installing for. Bundler
is installing for ruby but the dependency is only for x86-mingw32, x64-mingw32, x86-mswin32, java. To add th
ose platforms to the bundle, run 'bundle lock --add-platform x86-mingw32 x64-mingw32 x86-mswin32 java'.
The dependency wdm (~> 0.1.1) will be unused by any of the platforms Bundler is installing for. Bundler is i
nstalling for ruby but the dependency is only for x86-mingw32, x64-mingw32, x86-mswin32. To add those platfo
rms to the bundle, run 'bundle lock --add-platform x86-mingw32 x64-mingw32 x86-mswin32'.
Configuration File: /home/cgeroux/cloud_apps_in_hss/_config.yml
Source: /home/cgeroux/cloud_apps_in_hss
Destination: /var/www/html
Incremental build: disabled. Enable with --incremental
Generating...
Jekyll Feed: Generating feed for posts
done in 0.258 seconds.
Auto-regeneration: disabled. Use --watch to enable.
cgeroux@workii:~/cloud_apps_in_hss$
```



: PuTTY, MobaXterm ,



: Terminal



: WinSCP, MobaXterm ,

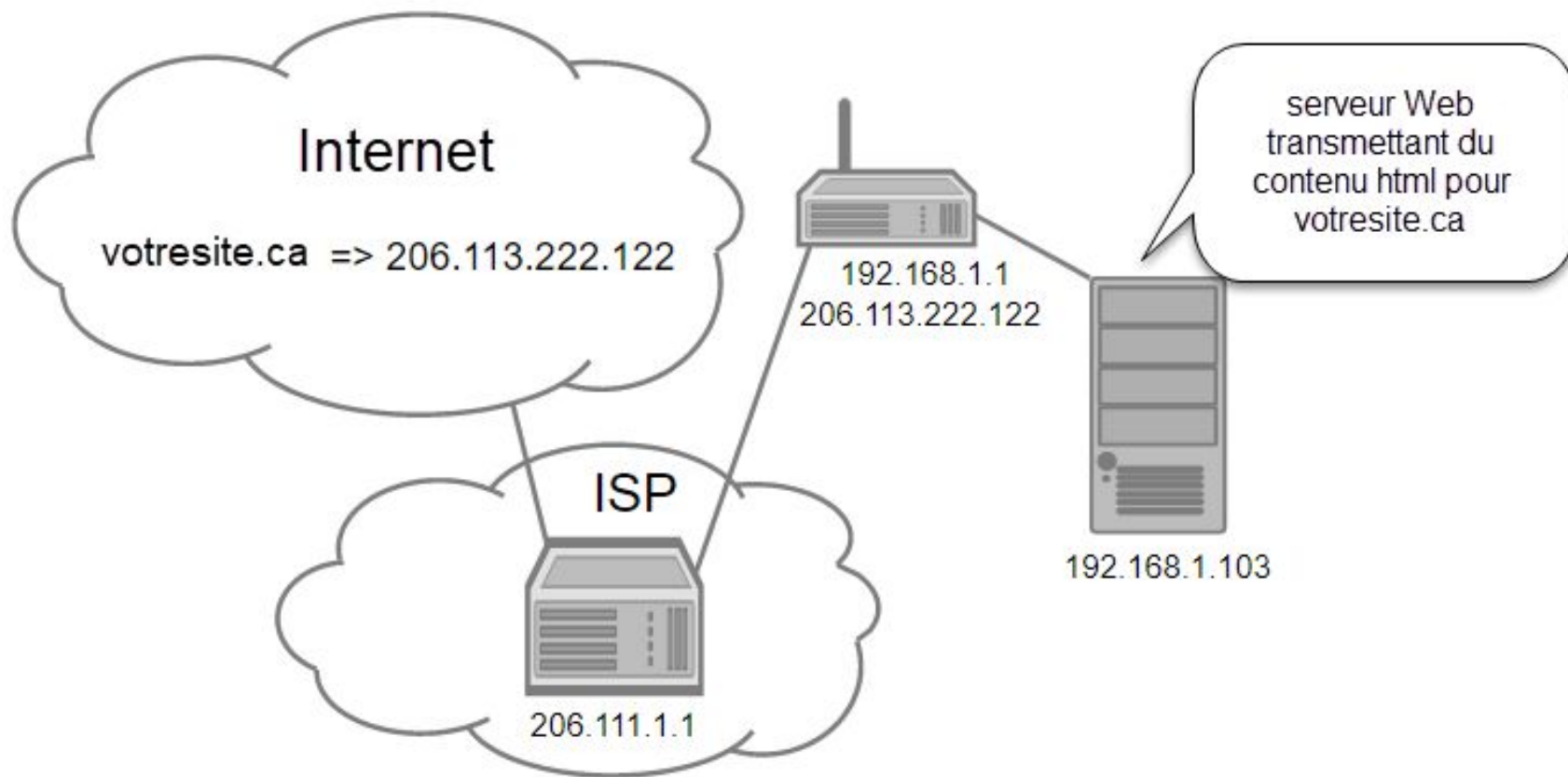


: FileZilla

# Quoi faire avec une instance

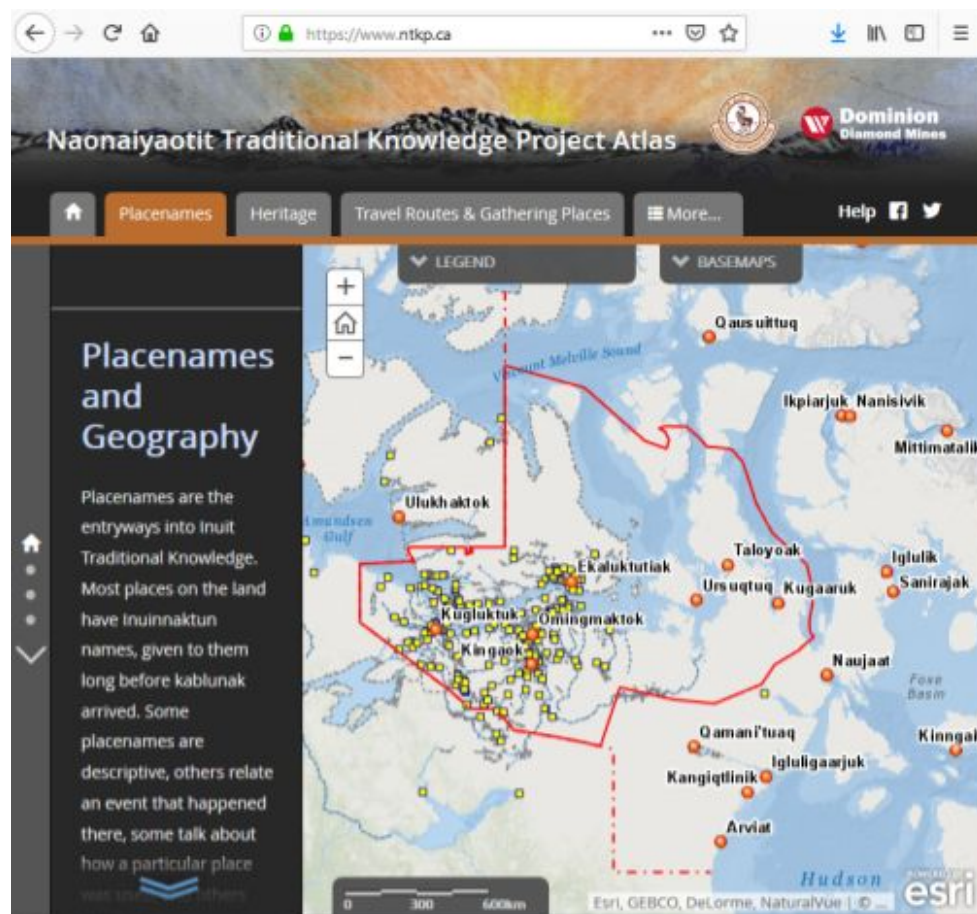
- Quoi faire avec un ordinateur
  - Par l'interface d'un bureau, visionner des films, écouter de la musique, naviguer sur le Web  
(oui mais, votre ordinateur peut faire beaucoup plus)
  - Stocker des données, héberger des sites Web, moissonner le Web (*Web scraping*)  
(probablement plus efficace avec une instance)
- Comment les nuages de la fédération Calcul Canada sont-ils utilisés en recherche?
  - Moissonnage du Web (avec par exemple [curl](#), [wget](#), [Beautiful Soup](#))
    - Exécution d'applications selon un horaire établi (avec par exemple [cron](#))
  - Stockage/partage de fichiers (avec par exemple [Globus](#), [sftp](#), [http](#))
  - Hébergement de serveurs Web

## Qu'est-ce qu'un serveur Web?



# Atlas du savoir traditionnel Inuit

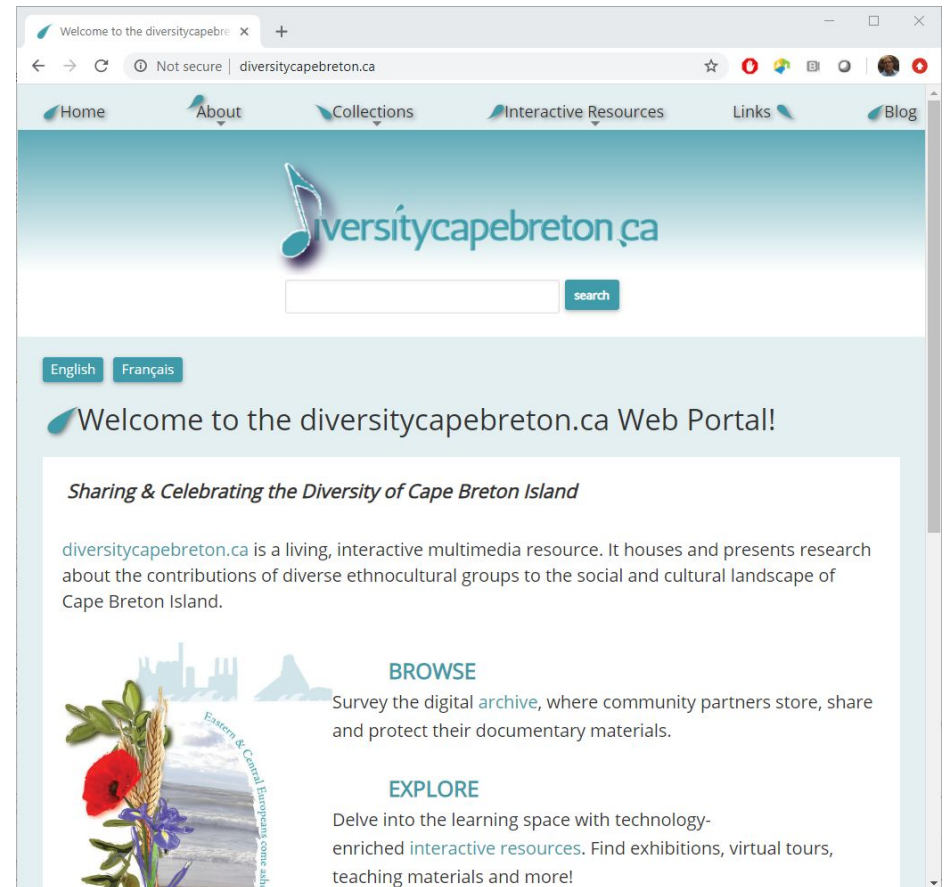
- Claudio Aporta, Ph. D.
- Université Dalhousie
- [ntkp.ca](https://www.ntkp.ca)
- Utilise *ArcGIS*
- Site sur un nuage de la fédération Calcul Canada
- Base de données hébergée par *Esri*





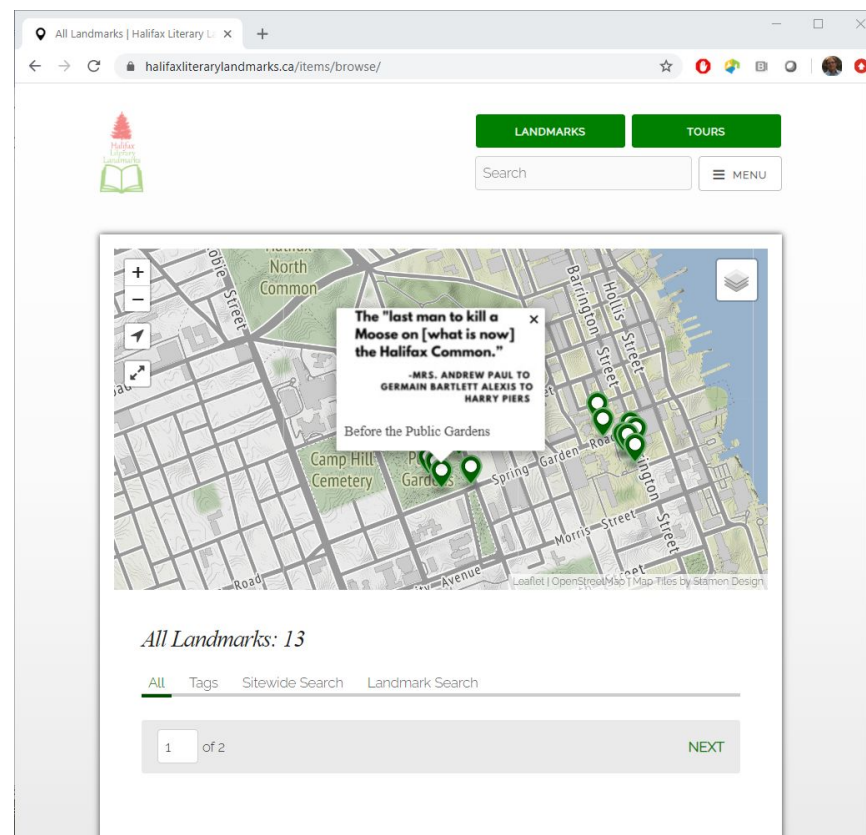
# Diversity Cape Breton

- Marcia Ostashewski, Ph. D.
- Université Cape Breton
- [diversitycapebreton.ca](http://diversitycapebreton.ca)
- Utilise [Islandora.ca](http://Islandora.ca)
  - système de gestion de contenu [Drupal](http://Drupal)
  - bibliothèque [Fedora](http://Fedora)



# Halifax dans la littérature

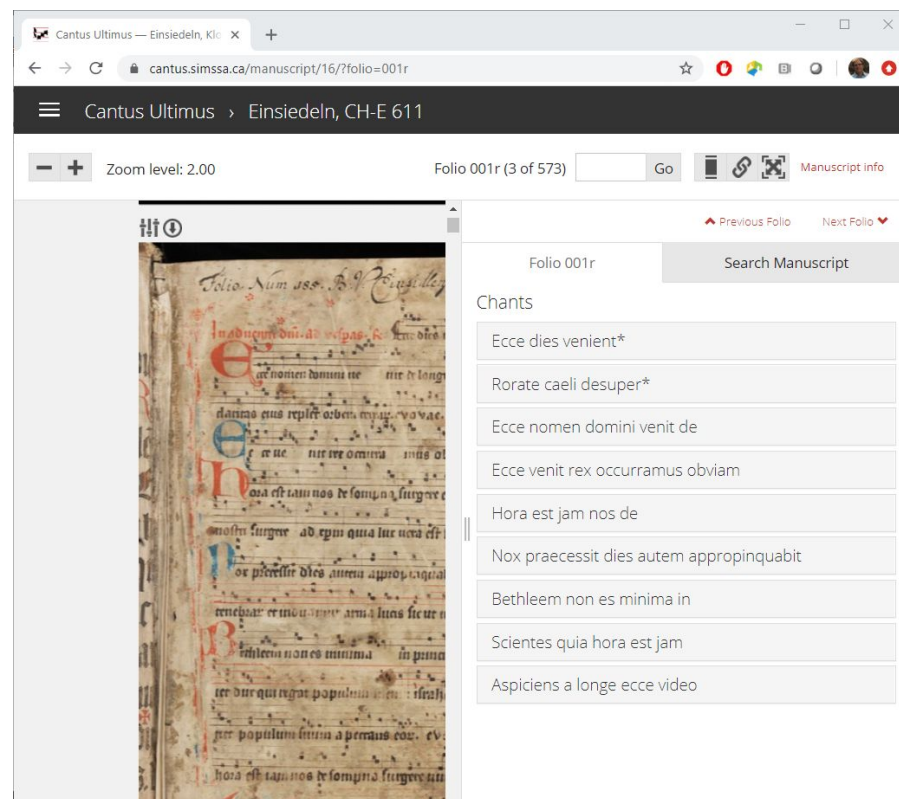
- Kate Scarth, Ph. D.
- Université de l'Île-du-Prince-Édouard
- [halifaxliterarylandmarks.ca](http://halifaxliterarylandmarks.ca)
- Utilise *Omeka*





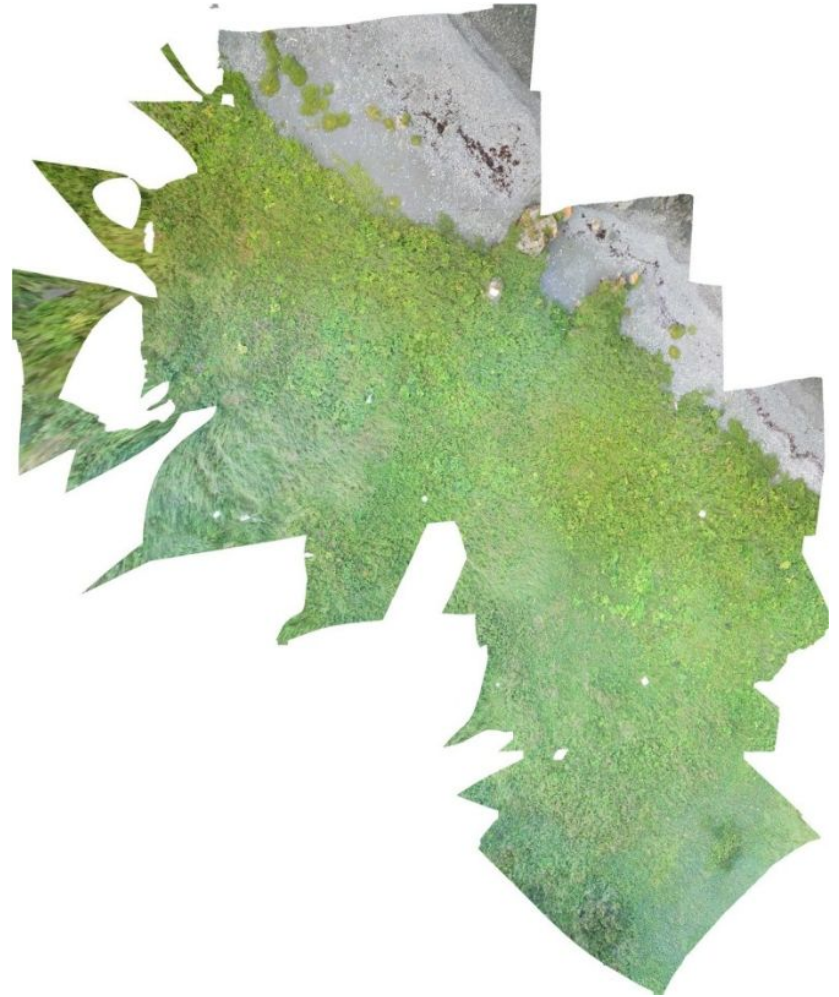
# Cantus Ultimus

- Grande équipe internationale
- Participantes canadiennes :
  - Jennifer Bain
  - Barbara Swanson
- Université Dalhousie
- [cantus.simssa.ca](http://cantus.simssa.ca)
- Technologies inconnues



# Drones et archéologie

- Amanda Crompton et Marc Bolli
- Université Saint Mary's, Université Memorial
- Transmission de données à une instance par un drone; création d'une composition photomosaïque immédiatement retournée à l'équipe sur le terrain
- Identification des paysages anthropiques
- Pas de site Web



# Sites Web dynamiques

- Sont aussi typiquement des systèmes de gestion de contenu (CMS pour *Content Management System*)
- Habituellement rédigés en *PHP* (Hypertext Preprocessor)
  - le serveur Web génère le HTML sur demande.
- Nécessitent souvent une base de données (par exemple *MySQL* ) pour stocker
  - les renseignements sur les utilisateurs;
  - les données sur les pages;
  - les autres données nécessaires au bon fonctionnement du système de gestion du contenu.
- CMS les plus connus :
  - *WordPress, Drupal, Joomla, Omeka, MediaWiki*

## Sites Web dynamiques : Fonctionnalités

- Le contenu est personnalisé en fonction de la personne qui visite le site, par exemple :
  - pages d'administration pour configurer et modifier le site;
  - paniers d'achat virtuels;
  - soldes de comptes.
- La base de données peut être modifiée selon les entrées faites par les personnes qui utilisent le site :
  - création de nouvelles pages;
  - prise de commandes;
  - paiement de factures.
- Des fichiers peuvent être téléversés ou créés sur l'instance, via le site.
- L'instance et le site peuvent être modifiés directement à partir du site même.
- Les systèmes de gestion de contenu sont généralement complexes.

# Sites Web dynamiques : Conséquences

- La sécurité prend plus d'importance.
- Les personnes qui visitent le site peuvent le modifier, ainsi que l'instance.
- Les connexions doivent être sécurisées (*attaques par force brute*).
- Les données sensibles (mots de passe, numéros de cartes, etc.) doivent être cryptées (*HTTPS*).
- Il est très important que le système de gestion de contenu et les modules d'extension soient à jour pour parer aux types d'attaques connus :
  - évitez les modules d'extension;
  - si vous devez les utiliser, assurez-vous
    - qu'ils proviennent d'une source sûre,
    - qu'ils sont maintenus activement.

## Sites Web dynamiques : ça ne peut pas m'arriver à moi...

- Je ne suis pas une banque, alors est-ce que je dois m'inquiéter?
  - **OUI!** Les pirates cherchent à prendre le contrôle d'un maximum d'ordinateurs afin de créer un réseau d'ordinateurs zombies;
  - ils peuvent ainsi lancer de plus grandes attaques distribuées.
- J'ai personnellement vécu quelques mauvaises expériences :
  - la journalisation des activités sur mon premier serveur Linux rapportait des attaques par force brute via SSH;
  - plusieurs attaques par force brute via SSH ont perturbé des instances sur les nuages de la fédération Calcul Canada;
  - un site Drupal auquel je participais a été piraté et a dû être reconstruit à partir des copies de sauvegarde.
- Y a-t-il d'autres options qu'un site Web dynamique?

## Sites Web statiques

- Les premiers sites Web étaient de ce type (HTML des années 1980-1990).
- Les fichiers statiques ne sont pas générés par un serveur Web.
- Ce type de site convient quand un contenu unique est possible pour tous.
- Des éléments dynamiques peuvent être introduits avec *javascript*.
- Ces sites étaient développés par des spécialistes en conception Web.
- Il existe aujourd'hui des générateurs de sites statiques :
  - création de fichiers statiques avec le langage *Markdown*;
  - thèmes existants ou personnalisés;
  - connaître html/css/javascript est requis pour personnaliser les thèmes.
- Générateurs bien connus : *Jekyll*, *Hexo*, *Hugo*, *Octopress*, *Pelican*.



# Sites Web statiques : Un exemple avec Jekyll

La version anglaise de cette présentation est un site Web statique généré par *Jekyll* avec un thème personnalisé, deux modèles *HTML*, du *CSS*, et un peu de *javascript* pour changer de diapositive avec les flèches.

```
---
layout: slide
title: "Static sites - These slides are a static website"
prevslide: static-sites-overview
nextslide: static-sites-simpler
---

* Uses the [Jekyll](https://jekyllrb.com/) static website generator,
with a
  custom theme
* Two [HTML](https://www.w3schools.com/html/) templates
* Some [CSS](https://www.w3schools.com/css/)
* And a little [javascript](https://www.w3schools.com/js/) so arrow
keys
```

## Static sites - These slides are a static website

- Uses the *Jekyll* static website generator, with a custom theme
  - Two *HTML* templates
  - Some *CSS*
  - And a little *javascript* so arrow keys change slides

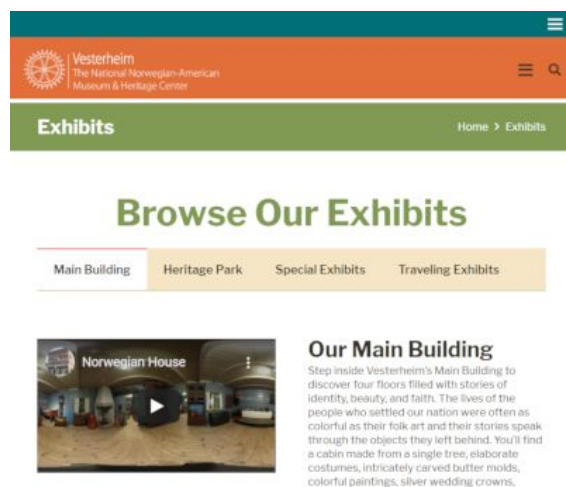


# Sites Web statiques : Exemples avec Jekyll

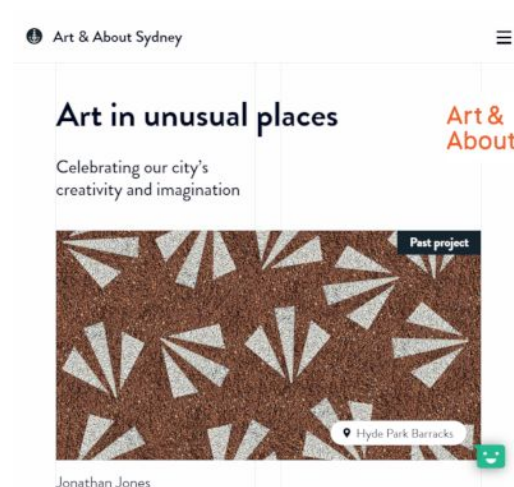
## Jekyll showcase



<https://www.foia.gov/>



<https://vesterheim.org/>



<https://www.artandabout.com.au/>

# Sites Web statiques : la simplicité

*Tout doit être rendu aussi simple que possible, mais pas plus simple.*

~Albert Einstein

- Utilisez un site dynamique si c'est absolument nécessaire.
- Vous pourriez avoir besoin d'un site dynamique si
  - les personnes qui le consultent doivent créer du contenu;
  - des renseignements personnalisés et uniques doivent y être affichés.
- Un site statique est
  - facile à sauvegarder; conservez votre fichier source,
  - facile à migrer; conservez le HTML généré sur votre bureau,
  - facile à configurer (pas de SSL, pas de base de données).
- Le contrôle des versions se faire pour le site en entier.

# Maintenance, sécurité et sauvegarde

- Assurez-vous de pouvoir reconstruire votre site :
  - notez les détails sur toutes les commandes utilisées pour la configuration;
  - gardez au moins une copie des données non reproductibles (par exemple, le contenu du site).
- Vous aurez éventuellement à reconstruire votre site :
  - utilisez la version majeure du plus récent système d'exploitation;
  - si le serveur est compromis.
- Gardez à jour le système d'exploitation.
- Site dynamique :
  - gardez à jour le système de gestion de contenu
  - gardez à jour tous les modules d'extension

## Prochaines étapes

- Les diapositives se trouvent dans [cgeroux.github.io/cloud\\_apps\\_in\\_hss](https://cgeroux.github.io/cloud_apps_in_hss).
- Le fichier source (Markdown, HTML, CSS, Javascript) se trouve dans [github.com/cgeroux/cloud\\_apps\\_in\\_hss](https://github.com/cgeroux/cloud_apps_in_hss).
- Prenez connaissance du contenu de l'atelier *Cloud from A to Z* pour
  - une introduction sur comment utiliser un nuage OpenStack;
  - créer un serveur Web et un site Jekyll.
- Pour commencer dès maintenant
  - demandez un [compte avec Calcul Canada](#) et un [projet infonuagique](#);
  - parcourez [notre documentation](#);
  - apprenez [comment configurer un serveur Web Apache2](#);
  - initiez-vous à *Jekyll* ou à *un autre générateur*.