

# Session 2

## Cloud applications for Humanities and Social Sciences

### Chris Geroux

Chris has been using HPC since 2005 and joined ACENET in 2015, with a focus on Big Data and cloud. He has a PhD in astronomy working on computational fluid dynamics models of stars, which continued with similar work at post doctoral position at the University of Exeter.

Chris's primary focus at ACENET has been helping researchers work with the Compute Canada clouds and is a part of Compute Canada's notional cloud team. He has played a large role in much of the national public-facing cloud documentation. He is also involved in teaching workshops on parallel computing, Python, and cloud.





# Cloud Applications in HSS

*Accelerating Discovery with Digital Research*

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

**compute**canada  
*regional partner*

# Goal of Talk

By the end you will:

- Have a high level understanding of what a cloud is
- Have a rough idea of what a cloud workflow looks like
- Know some common ways that a cloud is used in HSS
- Know the difference between dynamic and static websites
- Know some important points to consider when planning a website

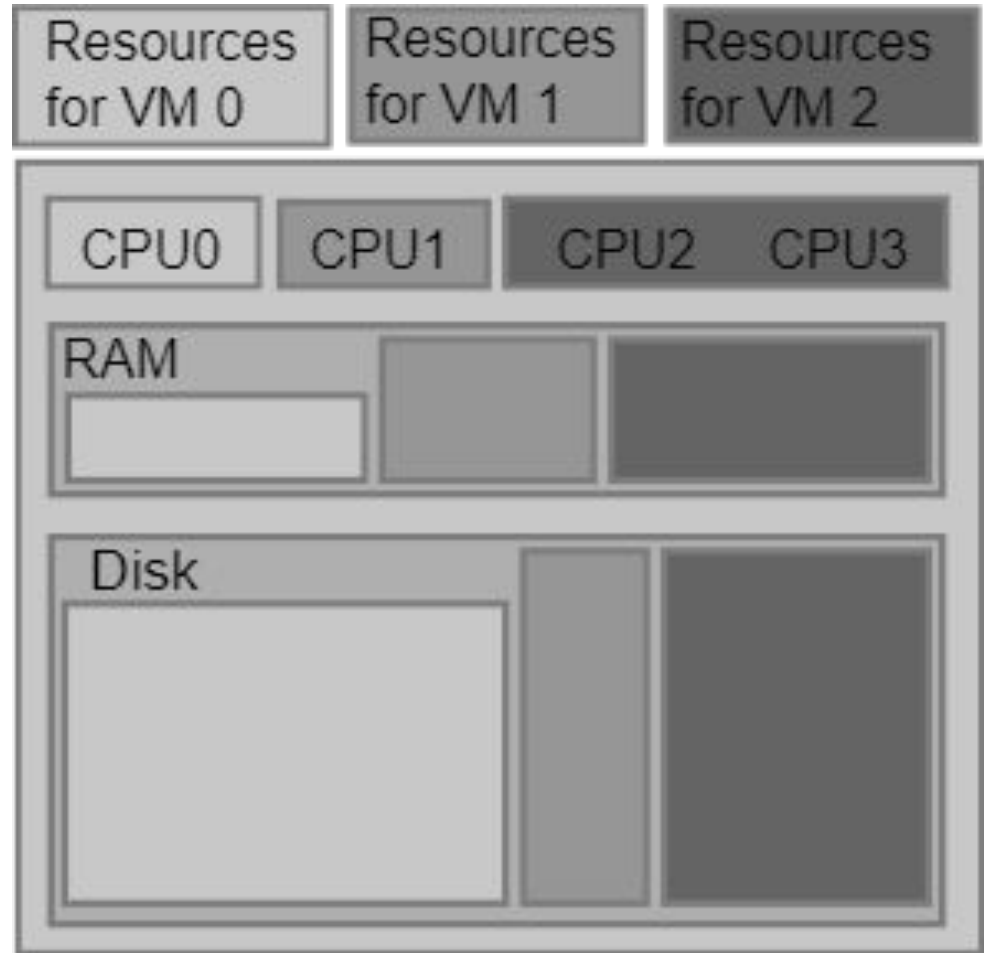
# What is a cloud?

- The word *cloud* has many different meanings
  - Cloud apps: gmail, calendars, docs etc.
  - Cloud hosting: to host your websites
  - Cloud infrastructure: cloud computers
- Physically, a cloud is a bunch of computers that can be access remotely through, apps, and remote connection tools

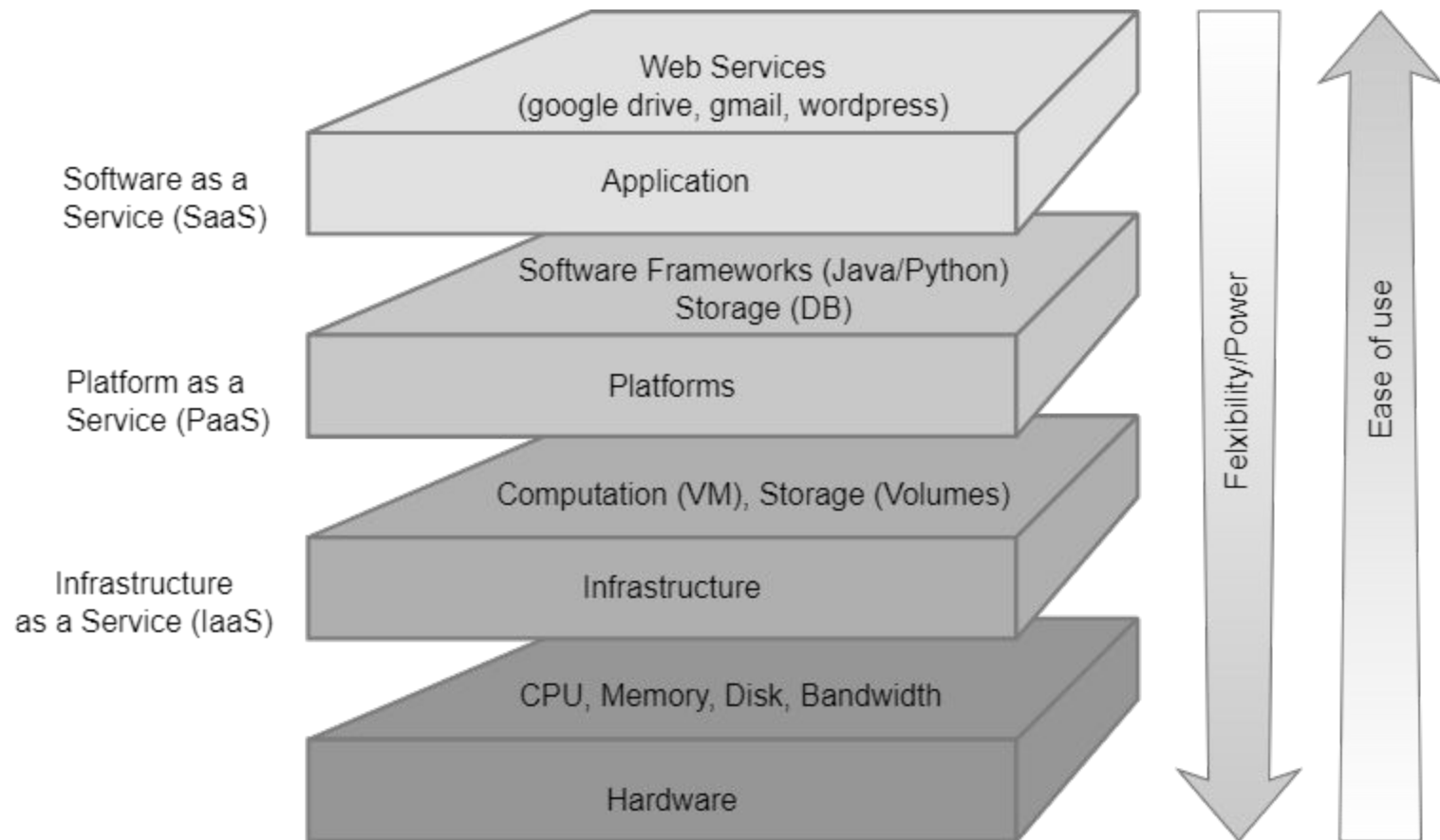


# What is a Virtual Machine (VM)?

- A simulated computer
- Runs on a real physical computer
- May use only part of a physical computer
- May use a different operating system than physical computer running VM
- Can run VMs on your laptop  
(see [virtualbox.org](http://virtualbox.org))



# Cloud Service Levels



# Compute Canada clouds are IaaS

- IaaS allows great flexibility
- Requires more effort & knowledge than PaaS or SaaS clouds
  - Manual setup and configuration of VMs
  - VM maintenance (updates/log surveillance)
  - Migration between VM operating system major versions
- Allows great flexibility:
  - Choice of Operating System
  - Choice of hardware configuration
- Compute Canada would love to offer more than IaaS (PaaS or SaaS) but currently lack required personnel & funding

# Compute Canada clouds

## Our clouds are IaaS

### Arbutus

- 484 nodes (16,008 cores)
- 5.3 PB + 12 PB storage
- 153.5 TB RAM

### Cedar Cloud

- 32 nodes (1024 cores)
- 500 TB storage
- 7.5 TB RAM

### East

- 36 nodes (576 cores)
- 100 TB storage
- 4.5 TB RAM

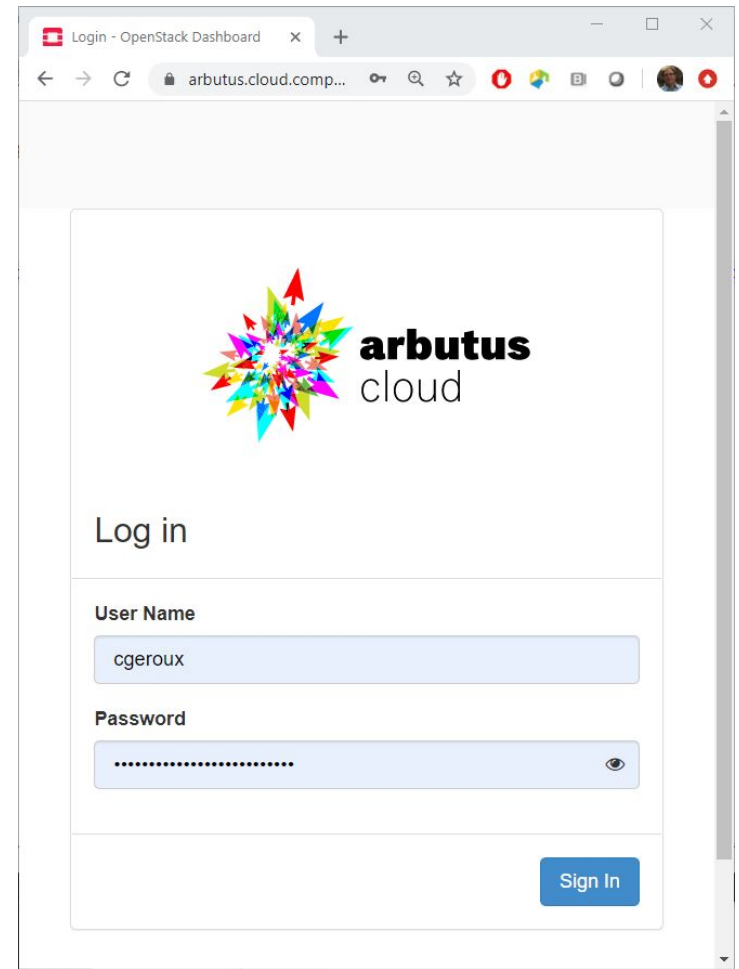
### Graham

- 24 nodes (768 cores)
- 84 TB storage
- 5.5 TB RAM



# Cloud Operating System - OpenStack

- A cloud operating system is used to manage
  - VM creation, deletion
  - storage
  - networking
  - resource quotas
  - billing
- Compute Canada clouds use the [OpenStack.org](https://openstack.org) cloud operating system
- Compute Canada clouds are free for researchers from recognized Canadian academic institutions



# OpenStack - Creating VMs

Instances - OpenStack Dashboard

arbutus cloud

def-cgeroux

Project

Comp

Launch Instance

Details \*

Source \*

Flavor \*

Networks

Network Ports

Security Groups

Key Pair

Configuration

Server Groups

Scheduler Hints

Metadata

Please provide the initial hostname for the instance, the availability zone where it will be deployed, and the instance count. Increase the Count to create multiple instances with the same settings.

Instance Name \*

Description

Availability Zone

Any Availability Zone

Count \*

1

Total Instances (36 Max)

64%

22 Current Usage

1 Added

13 Remaining

Cancel

< Back

Next >

Launch Instance

# Working with VMs

- SSH
  - Commands to configure VM
  - Commands to update VM
- SFTP
  - Transferring/Editing files
- Can have GUI desktops
  - Install desktop (e.g. **MATE**)
  - Install **VNC** to allow remote desktop connections

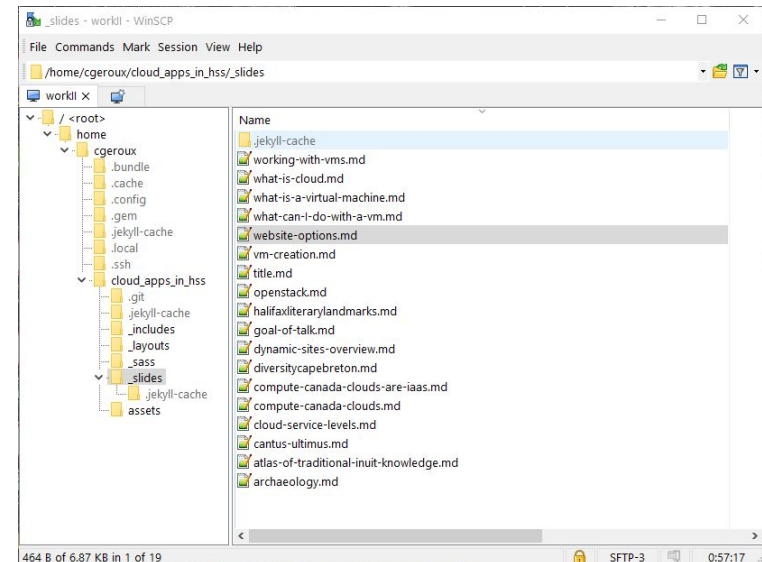
```
cgeroux@workii: ~/cloud_apps_in_hss
Destination: /var/www/html
Incremental build: disabled. Enable with --incremental
Generating...
Jekyll Feed: Generating feed for posts
done in 0.168 seconds.
Auto-regeneration: disabled. Use --watch to enable.
cgeroux@workii:~/cloud_apps_in_hss$ sudo jekyll build -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-data (=> 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 ,

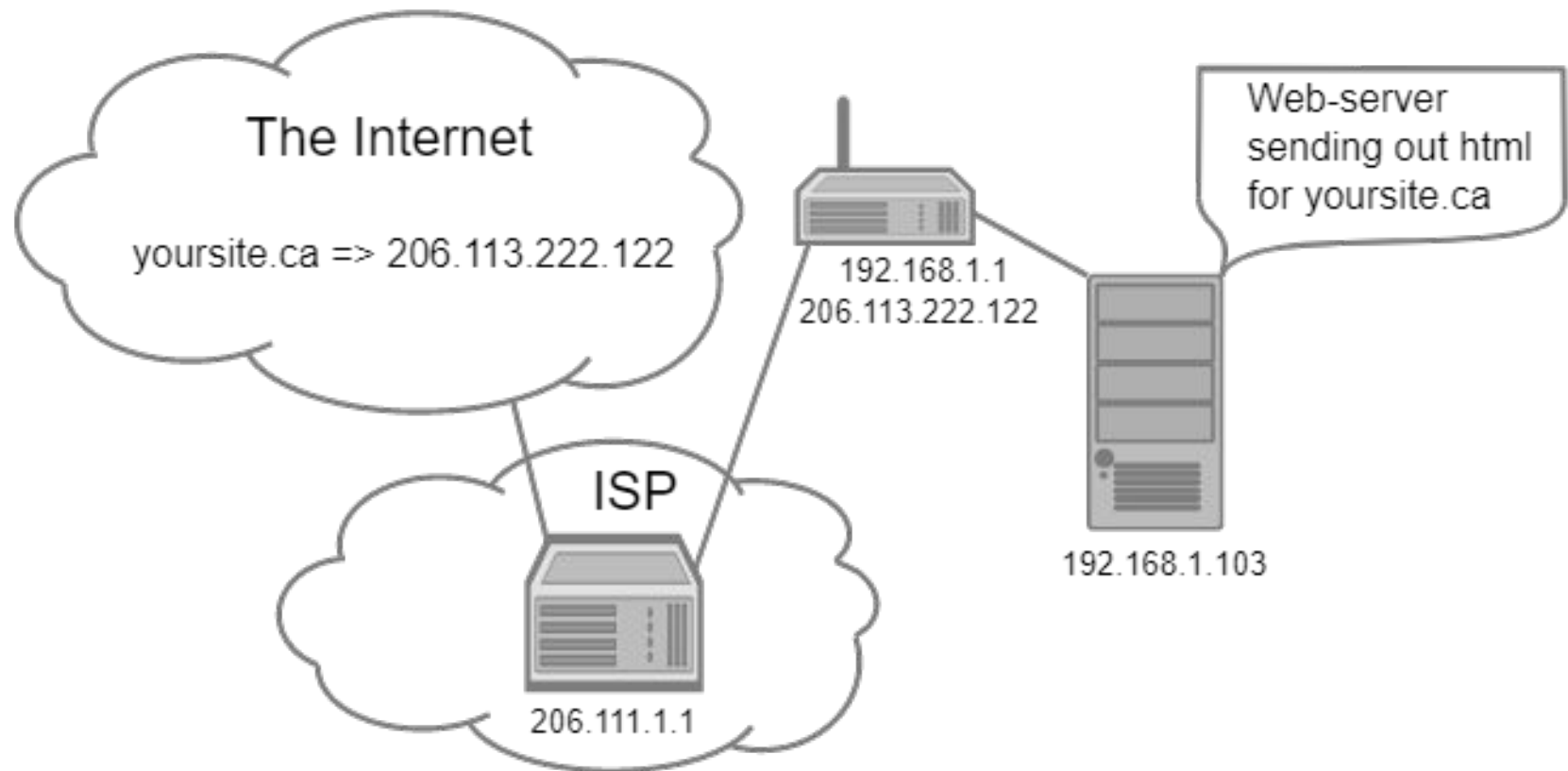


: FileZilla

# What can I do with a VM?

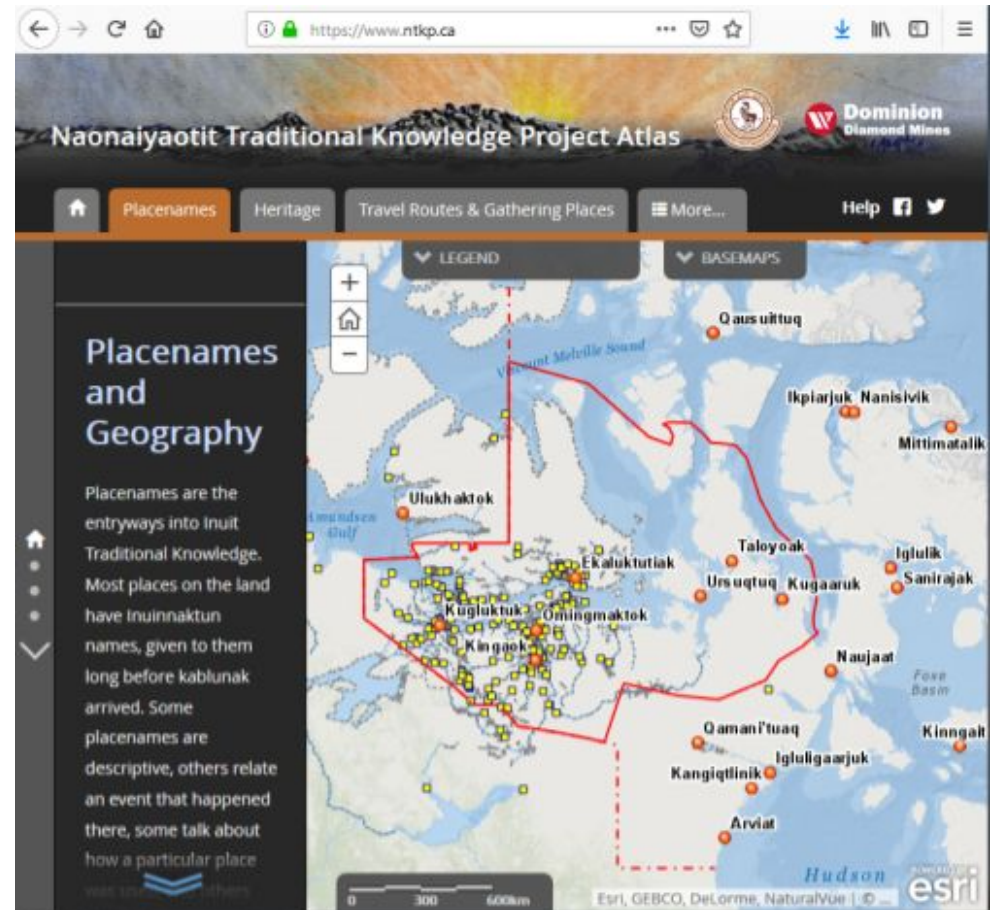
- What can I do with a computer?
  - Desktop GUI, watch movies, listen to music, browse the web (possible, but maybe your laptop would be better?)
  - Store data, host websites, run web-scrapers (probably better on a VM)
- What are researchers doing on CC clouds?
  - Scrape data from web (e.g. [curl](#), [wget](#), [Beautiful Soup](#))
    - Run programs at scheduled times (e.g. [cron](#))
  - File sharing/storage (e.g. [Globus](#), [sftp](#), [http](#))
  - Running web-servers

## What is a web-server?



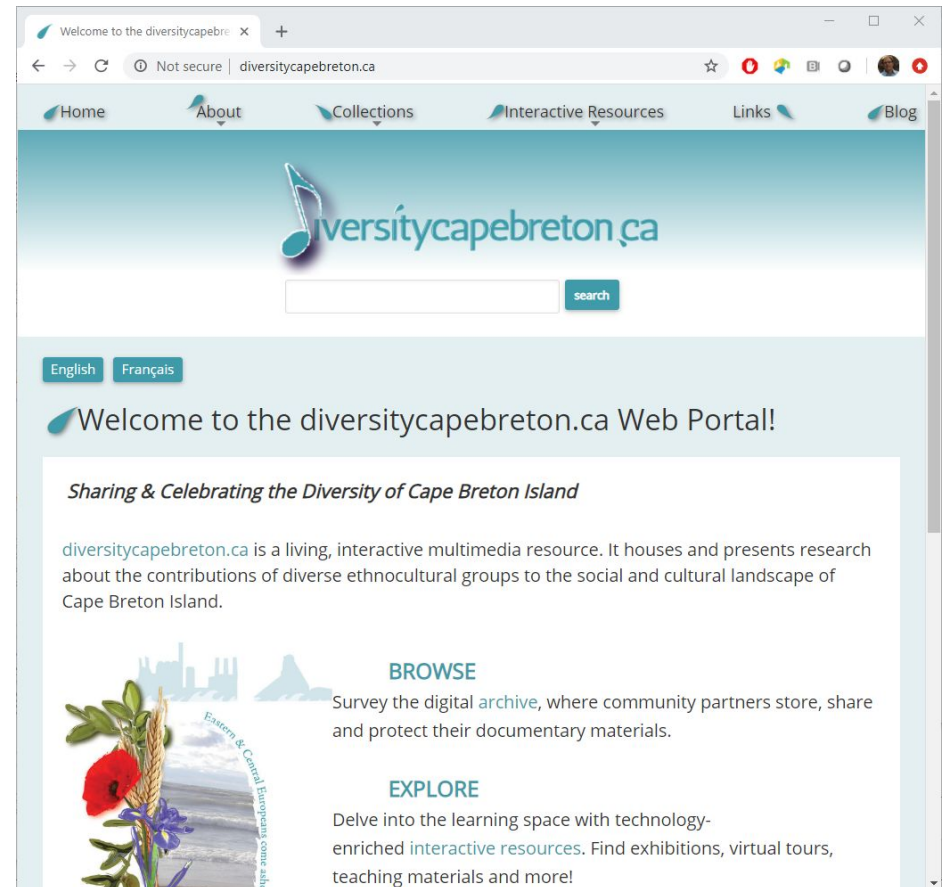
# Atlas of traditional Inuit knowledge

- Dr. Claudio Aporta
- Dalhousie University
- [ntkp.ca](https://www.ntkp.ca)
- Uses [ArcGIS](#)
- Site on CC cloud DB (hosted by [Esri](#))



# Diversity Cape Breton

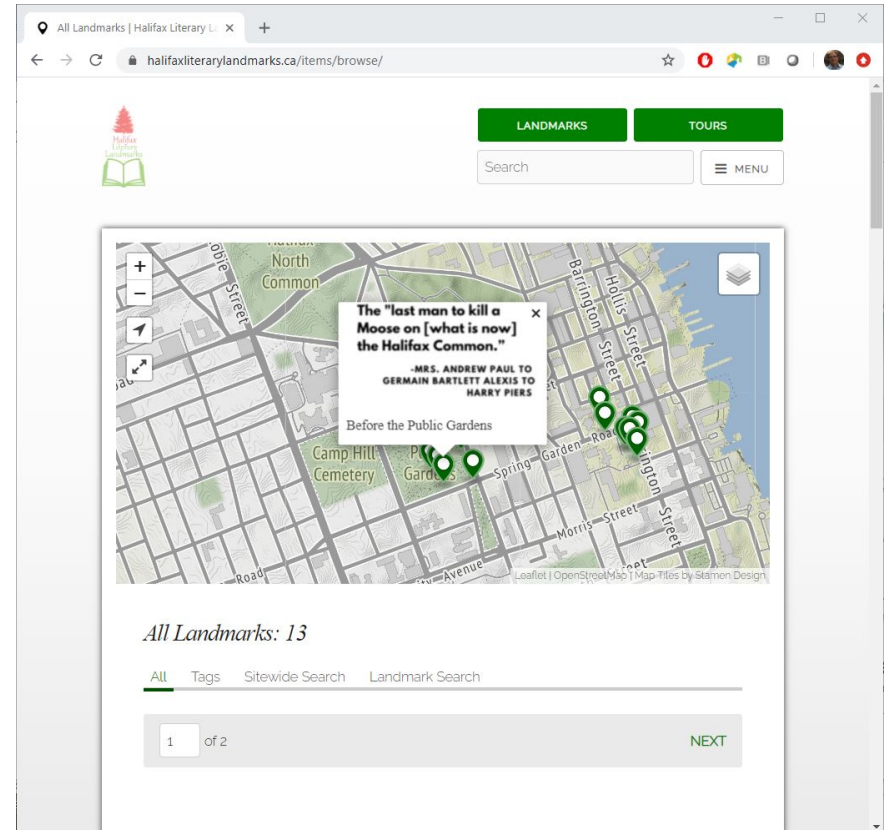
- Dr. Marcia Ostashewski
- Cape Breton University
- [diversitycapebreton.ca](http://diversitycapebreton.ca)
- Uses [Islandora.ca](http://Islandora.ca)
  - [Drupal](http://Drupal) CMS
  - [Fedora](http://Fedora) Repository





# Halifax Literary Landmarks

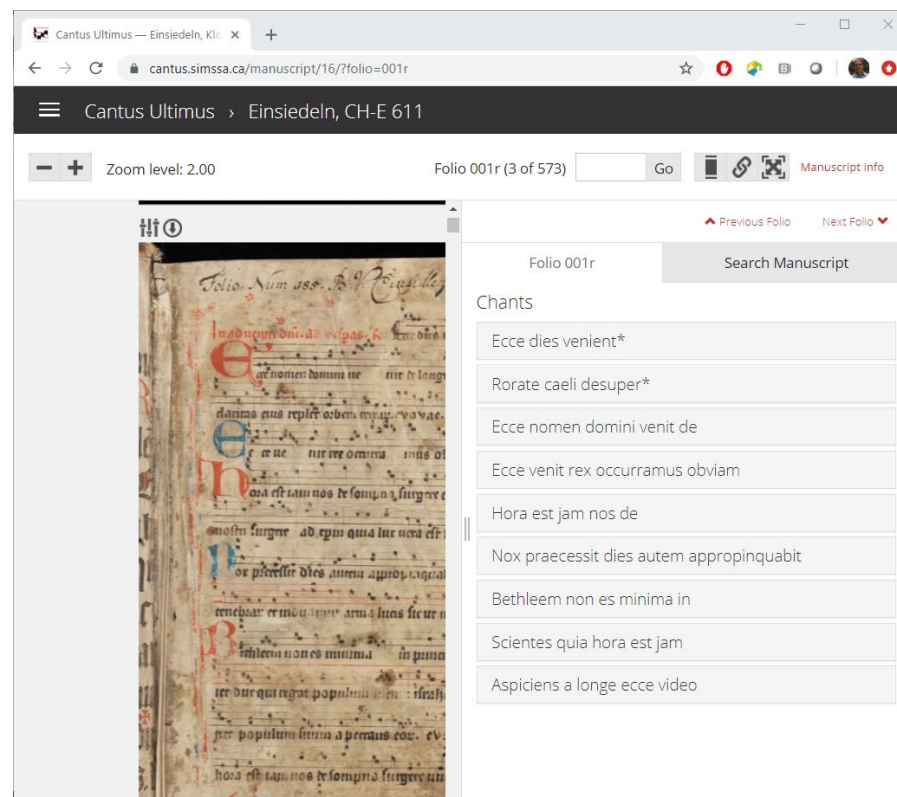
- Dr. Kate Scarth
- UPEI
- [halifaxliterarylandmarks.ca](http://halifaxliterarylandmarks.ca)
- Uses *Omeka*





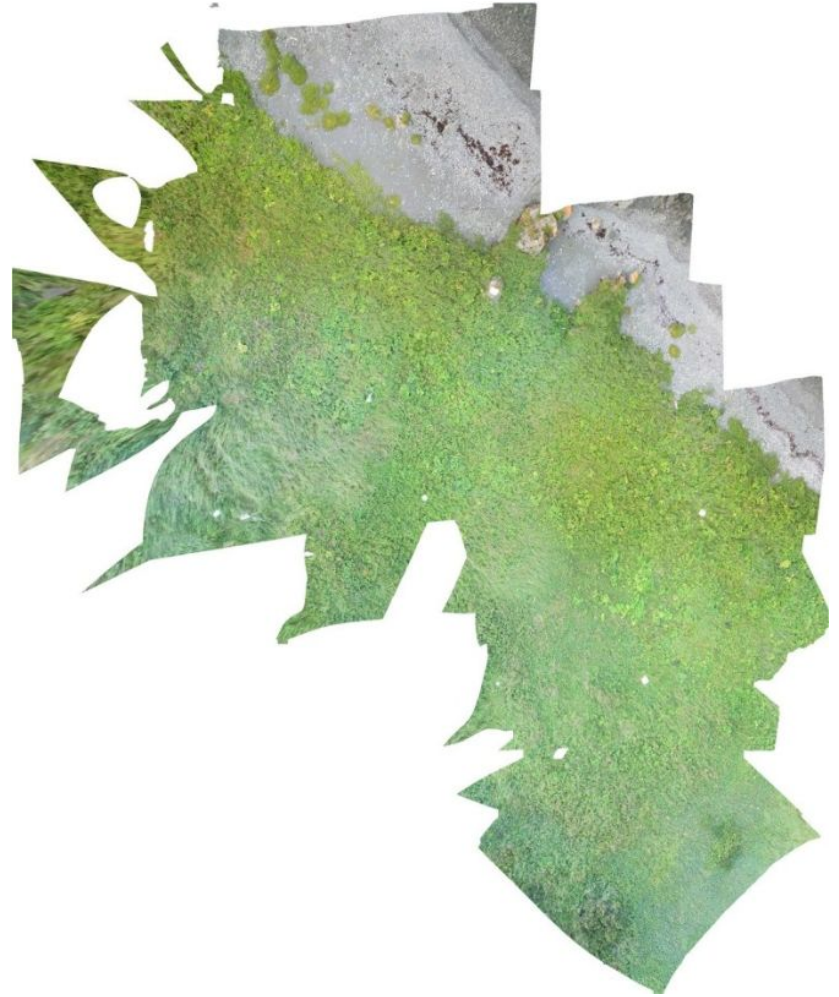
# Cantus Ultimus

- A large international team
- Local involvement:
  - Jennifer Bain
  - Barbara Swanson
- Dalhousie University
- [cantus.simssa.ca](http://cantus.simssa.ca)
- Technologies used unknown



# Archaeology with drones

- Amanda Crompton, Marc Bolli
- SMU, MUN
- VM received data stream from drone in the field. Immediate photomosaic construction which feeds back to field work.
- Used to identify anthropogenic landscapes
- No website.



# Dynamic websites

- Typically also Content Management Systems (CMSs)
- Usually written in *PHP* (Hypertext Preprocessor)
  - Web-server generates HTML on the fly
- Often require a database (e.g. *MySQL* ) to store:
  - User information
  - Page data
  - Other data required by CMS to function
- Popular CMSs
  - *WordPress, Drupal, Joomla, Omeka, MediaWiki*

## Dynamic websites - Capabilities

- Custom content per visitor e.g.
  - Admin pages to edit/configure the site
  - Shopping carts in online stores
  - Account balances
- Database can be changed in response to user input
  - Create new pages
  - Place orders
  - Paying bills
- Files can be uploaded/created by the site on the VM
- Site (and the VM) can be modified directly through the site itself
- Generally CMS are complex

# Dynamic websites - Implications

- Security becomes very important
- Visitors to the site can potentially modify it and the VM
- Logins need to be secured (*Brute-force attacks*)
- Sensitive data (e.g. passwords, card numbers) must be encrypted (*HTTPS*)
- Very important to keep CMS and all plugins up-to-date to avoid at least the publicly known exploits
  - Avoid plugins as much as possible
  - If you must use plugins ensure they are:
    - From reputable sources
    - Actively maintained

## Dynamic websites - But not me, right?

- But I am not a bank do I need to worry?
  - **Yes!** Botnets/hackers want control as many machines as possible
  - Allowing them to perform larger distributed attacks
- Examples from personal experience with server security:
  - My first Linux server
    - SSH brute-force attacks in the logs
  - Several successful SSH brute-force attacks on CC cloud VMs
  - A Drupal site I helped administer was hacked  
(and then rebuilt from backups)
- Are there alternatives to dynamic websites?

# Static websites

- Static sites are the original websites (HTML in the 1980s-1990s)
- Static files not generated by web-server
- Suitable when one set of content for all visitors is acceptable
- Can have dynamic elements using *javascript*
- Traditionally hand crafted by web-developers
- More recently there are static website generators
  - Create static files from *markdown*
  - Use either pre-made or custom themes
  - Custom themes require html/css/javascript knowledge
- Popular static website generators:
  - *Jekyll, Hexo, Hugo, Octopress, Pelican*

## 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

```
---  
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 change slides

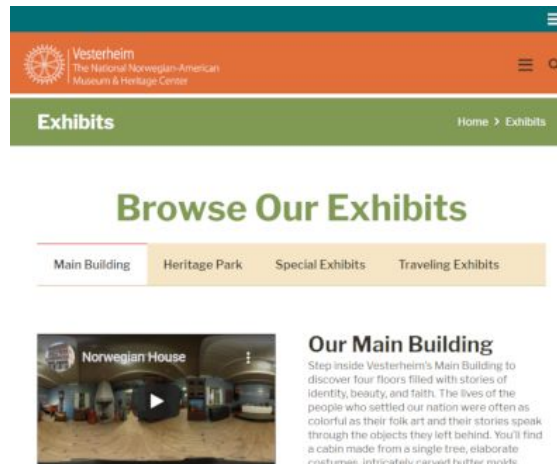


# Static sites - Jekyll examples

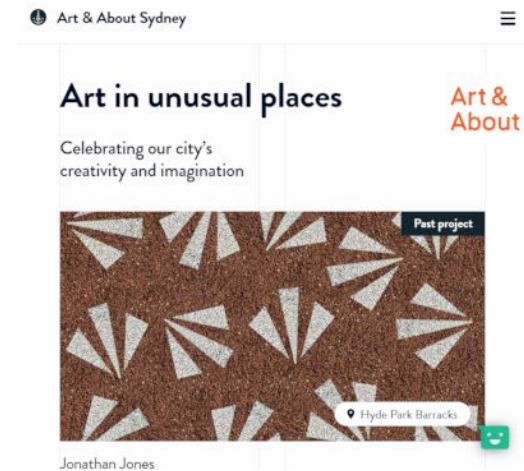
## Jekyll showcase



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



<https://vesterheim.org/>



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

# Static sites - Simple

*"Everything should be made as simple as possible, but no simpler"*  
~Albert Einstein

- If you don't need a dynamic website don't use one
- You might need a dynamic website if:
  - Visitors need to create content
  - Unique personalized information needs to be displayed
- Static websites:
  - Easy to backup; save your source folder
  - Easy to migrate; save generated HTML to your desktop
  - Easy to keep entire site under version control (e.g. [git](#))
  - Easy to set up (no SSL, no databases)

# Maintenance, Security, and Backups

- Ensure you can rebuild:
  - Keep notes on all commands used during setup
  - Backup non-reproducible data (e.g. site content)
- Eventually you will have to rebuild
  - Reinstall on next operating system major version
  - In case server is compromised.
- Keep operating system updated
- Dynamic sites
  - Keep CMS updated
  - keep all plugins updated

## Next steps

- Slides viewable here [cgeroux.github.io/cloud\\_apps\\_in\\_hss](https://cgeroux.github.io/cloud_apps_in_hss)
- Source (Markdown,HTML,CSS,Javascript) at [my github repo](#).
- Have a look at the [Cloud from A to Z](#) workshop content
  - Introduces using an OpenStack cloud
  - Creating a web server and a Jekyll site
- To get started:
  - Get a [CC account](#) + [cloud project](#)
  - Read our [cloud docs](#)
  - Our docs on [setting up an Apache2 web-server](#)
  - Look into [Jekyll](#) or [another static website generator](#)

# Housekeeping

|   |  |
|---|--|
| <p>The presentation, but not the discussion portion of this session will be recorded.</p>   | <p>Cet atelier est enregistré, à l'exception de la période de questions.</p>   |
| <p>Please stay muted unless asking a question. We will be taking on-mic questions at the end, if you have one, use the raise hand button in the reactions menu, or the participants menu (It will be in a different location depending on whether your Zoom is up to date), then you will be invited to ask your question on mic.</p>   | <p>Fermez votre microphone. Il y aura une période de questions de vive voix à la fin de l'atelier. Si vous avez une question, cliquez sur <i>Réactions</i> ou <i>Participants</i> dans la barre inférieure de la fenêtre Zoom, puis sur <i>Lever la main</i> et attendez qu'on vous invite à parler.</p>   |
| <p>During and after the session you can put questions in the slack workspace chat. The channel for this session is "session-2". We have helpers watching that channel who will be answering questions throughout the session and will direct pertinent questions towards the presenter as well. If we do not get all the Slack questions answered during the session we will continue to answer them in Slack after the zoom session ended.</p> | <p>À tout moment, vous pouvez poser une question par écrit dans l'espace de clavardage Slack, sur le canal "session-2". Vous pourrez recevoir une réponse immédiatement ou votre question sera communiquée à la personne qui dirige l'atelier. Nous resterons disponibles sur Slack après la présentation pour répondre aux questions qui n'auront pas été traitées.</p> |



compute  
canada | calcul  
canada



# Housekeeping

|  |   |
|--|---|
| <p>We will be reserving the zoom chat for technical difficulties only, it is located at the bottom of your screen near your main zoom controls, if you are having technical difficulties with Zoom please let us know in the Zoom chat.</p>  | <p>L'espace de clavardage de Zoom est réservé aux messages de nature technique. Pour rapporter un problème, cliquez sur <i>Discussion</i> dans la barre inférieure de la fenêtre,</p>   |
| <p>For this session you are invited to ask your questions in English or French.</p>  | <p>Vous pouvez poser vos questions en français ou en anglais.</p>   |
| <p>Slides for this presentation are being shared in both English and French. click on the “View Options” button at the top centre of your screen and select your preferred language from under “shared Screens” in the drop down menu. If you are having trouble switching between the slides please click the raise-your-hand button, or let us know in the Zoom chat and one of our helpers will assist you via private zoom chat.</p> | <p>Les diapositives sont présentées en français et en anglais. Pour sélectionner la langue des diapositives, cliquez sur le bouton <i>Options d’affichage</i> en haut de la fenêtre principale de Zoom et sélectionnez un des deux écrans partagés dans un sous-menu déroulant. Si vous avez des problèmes à faire afficher les diapositives dans la langue que vous voulez, cliquez sur le bouton de la main levée ou utilisez la fonction de clavardage et quelqu’un vous aidera.</p> |