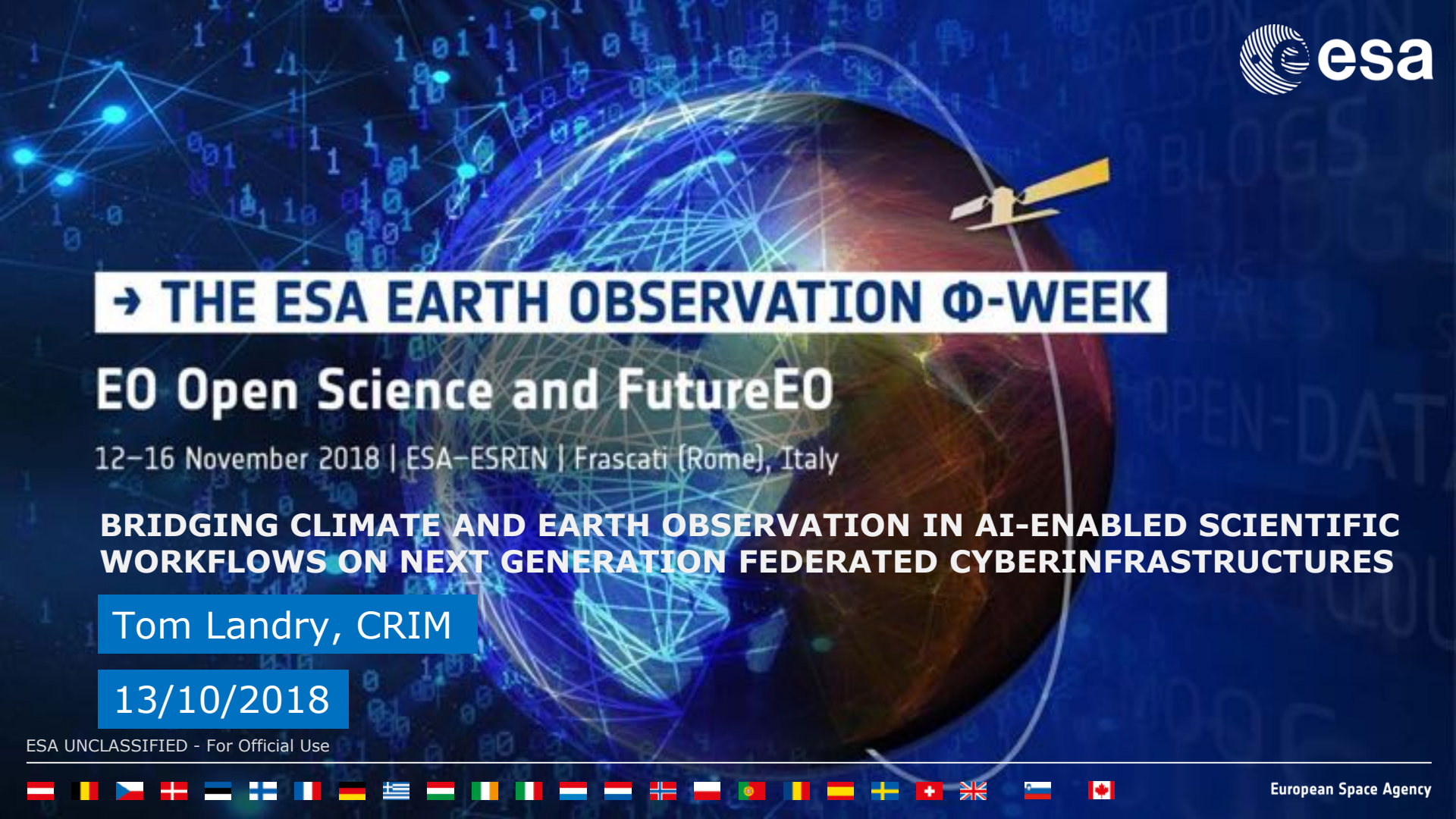


BRIDGING CLIMATE AND EARTH OBSERVATION IN AI-ENABLED SCIENTIFIC WORKFLOWS ON NEXT GENERATION FEDERATED CYBERINFRASTRUCTURES

Tom Landry, CRIM

13/10/2018

Computer Research Institute of Montreal

Computer Research Institute of Montreal (CRIM) is a not-for-profit applied research center.

- In operation for more than **30 years**
- **56 employees** (15 Ph.D., 16 Masters)
- **80 to 100 projects**, ~50 papers a year

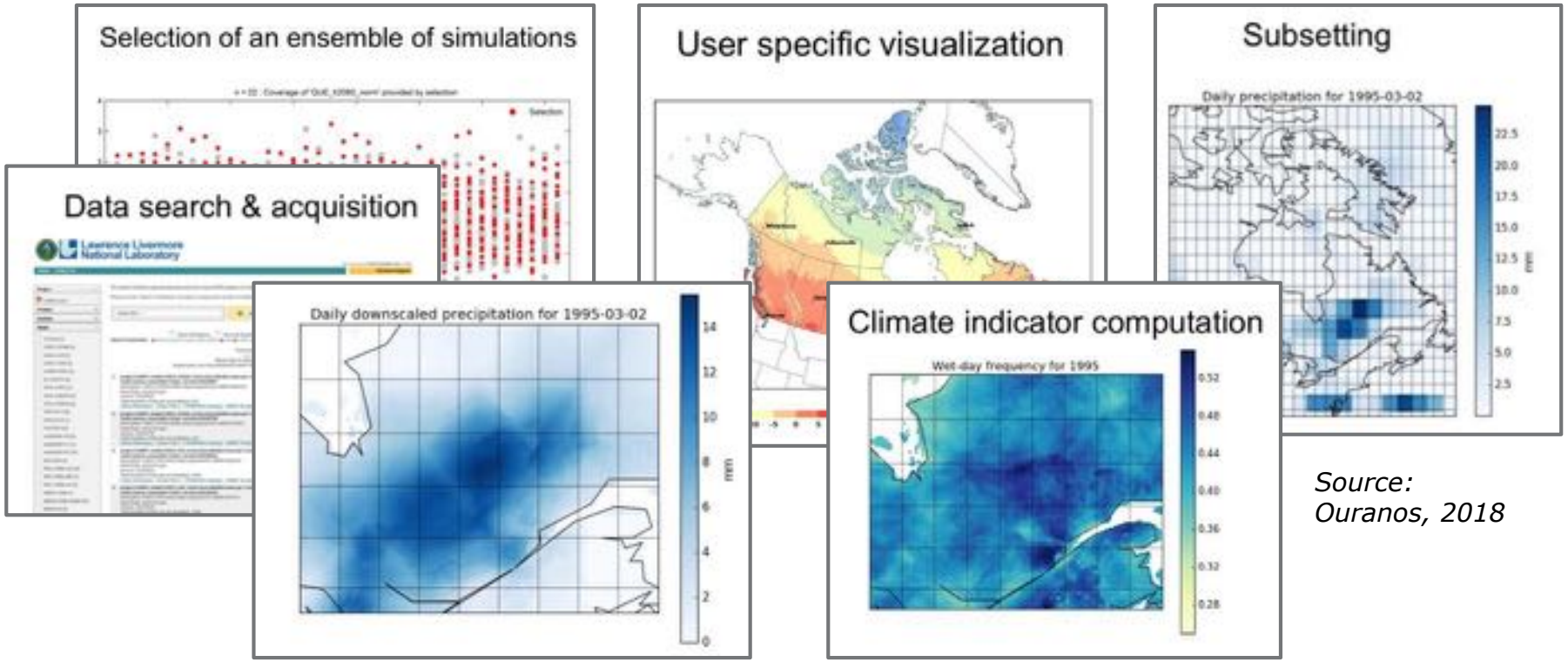
Introduction

- **Vision**: bridge Earth observation (EO) and climate data
  - EO: Past and present, state of a changing and dynamic Earth
  - Climate models: possible future Earths, uses EO as input and validation
  - Different science, technology, terms, scales and users. And meteorology?
- Execute in federated **cyberinfrastructures**
  - Data and computation scattered around the world
  - Need for interoperability, security, billing & quoting
- Enable advances of **Artificial Intelligence** (AI)
  - Natural Language Processing (NLP)
  - Machine Learning (ML) as a service
- Built on workflows
  - Computational, operational, scientific

With financial support from:

Économie et Innovation Québec

Climate processes and workflows



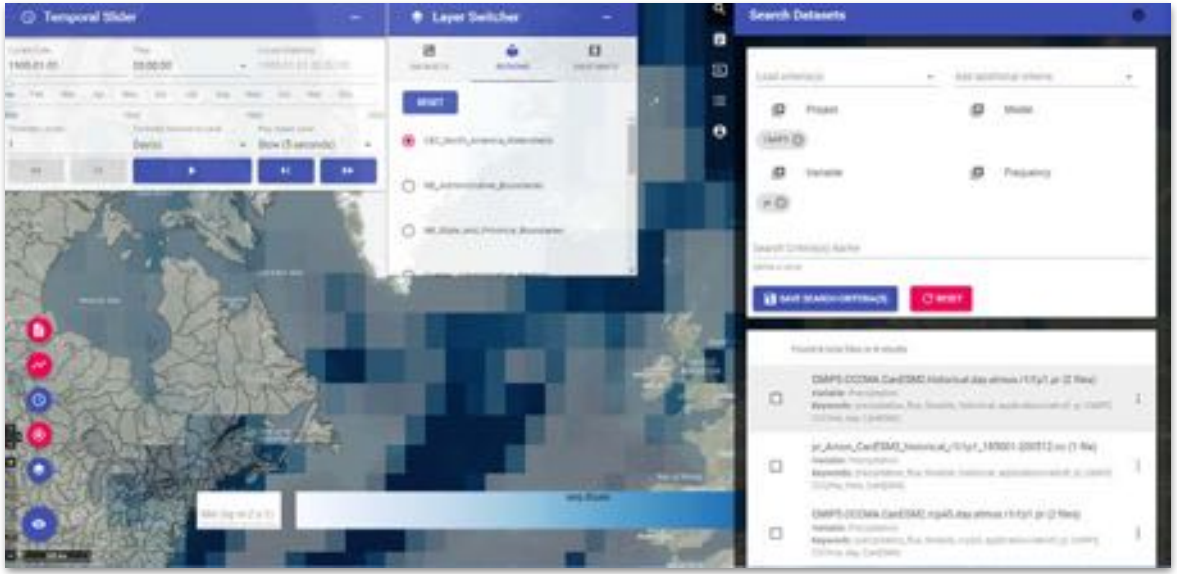
Source: Ouranos, 2018

2016-2018: CANARIE funds PAVICS

CANARIE operates national backbone network of Canada's National Research and Education Network (NREN).

CANARIE funded **PAVICS**, a research platform that streamlines climate scientists *workflows* and provide tools to analyze climate data.

PAVICS is meant to speed up the analysis of climate data and the creation of **climate scenarios** for impact studies



2018-2020: platform reuse and maintenance

GeoImageNet: Collaborative platform for VHR images annotation with extended taxonomies. Provides training sets, trained model repository and ML benchmark services. For now, Pleiade and WorldView images (30-50cm) over Canada.

PAVICS-Hydro: Integrates suite of pre-build hydrological models and North American watersheds. Calibrate and run hydrological models to simulate hydrographs and indices. Extensions to run on national HPC facilities and through notebooks.

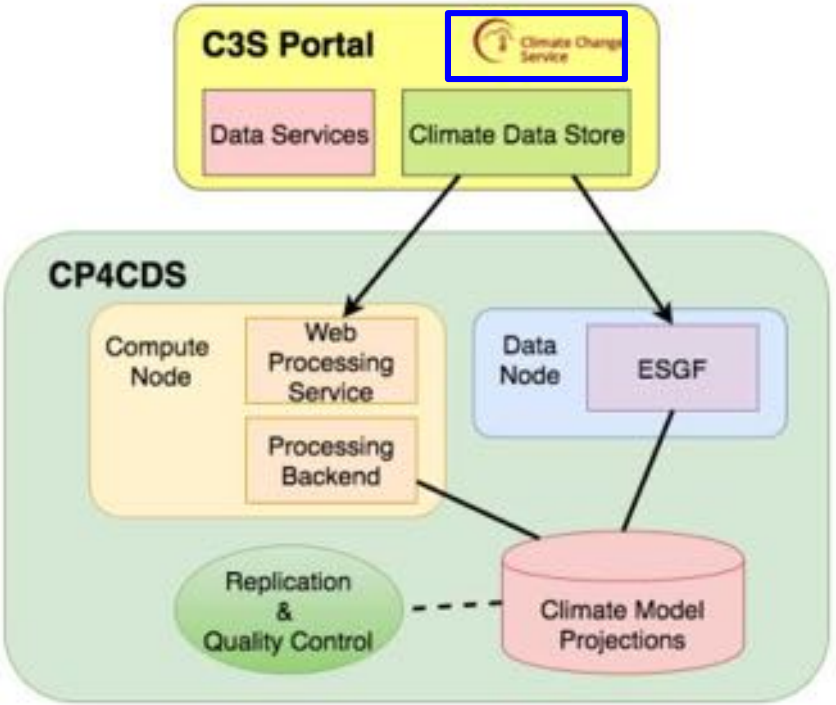
2013-now: Birdhouse

Birdhouse is a software framework and collection of Earth Systems Web Processing Services (WPS).

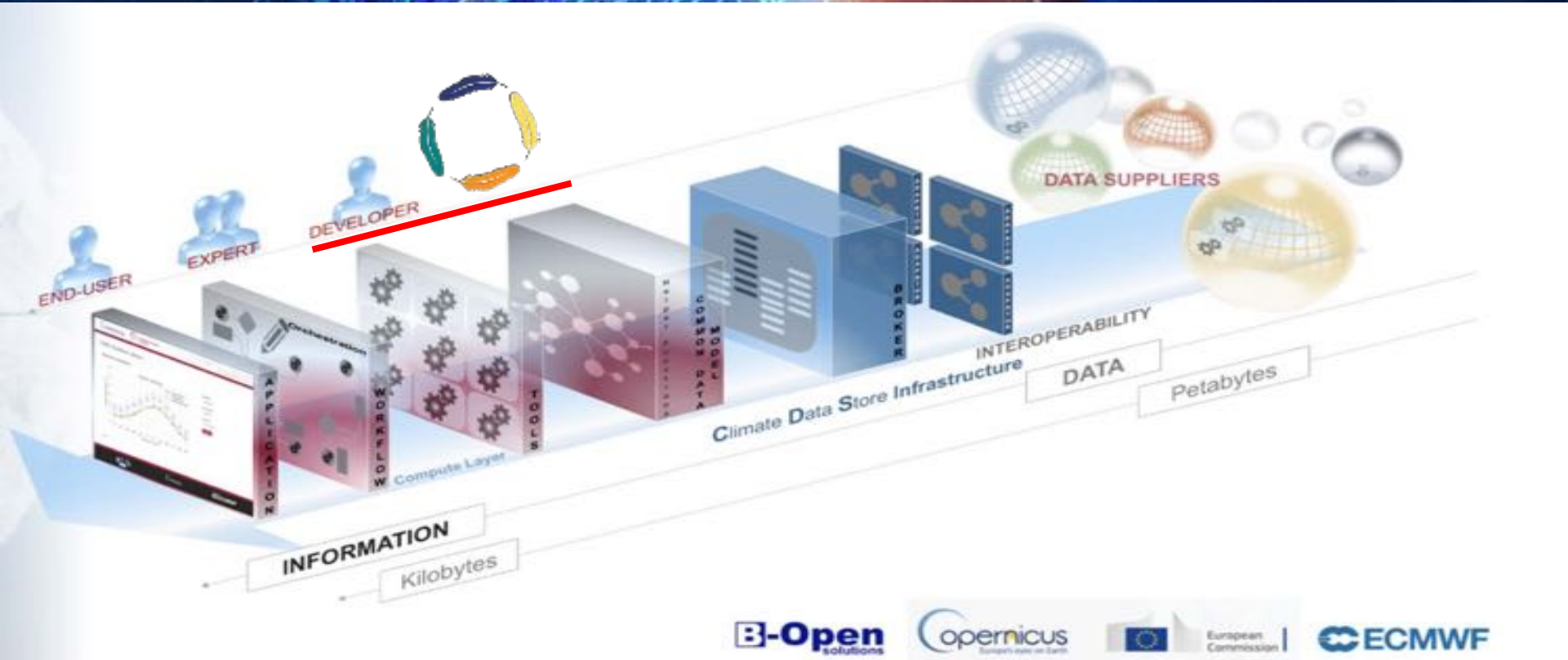
Seamless use of processes developed by national meteorological offices, climate service providers, academics, research centers and private industry.

Considers juridical conventions of the **United Nations,** the **Sendai** Framework for Disaster Risk Reduction, in line with the **Paris Agreement,** provides processes to support sustainable development goals (**SDG**).

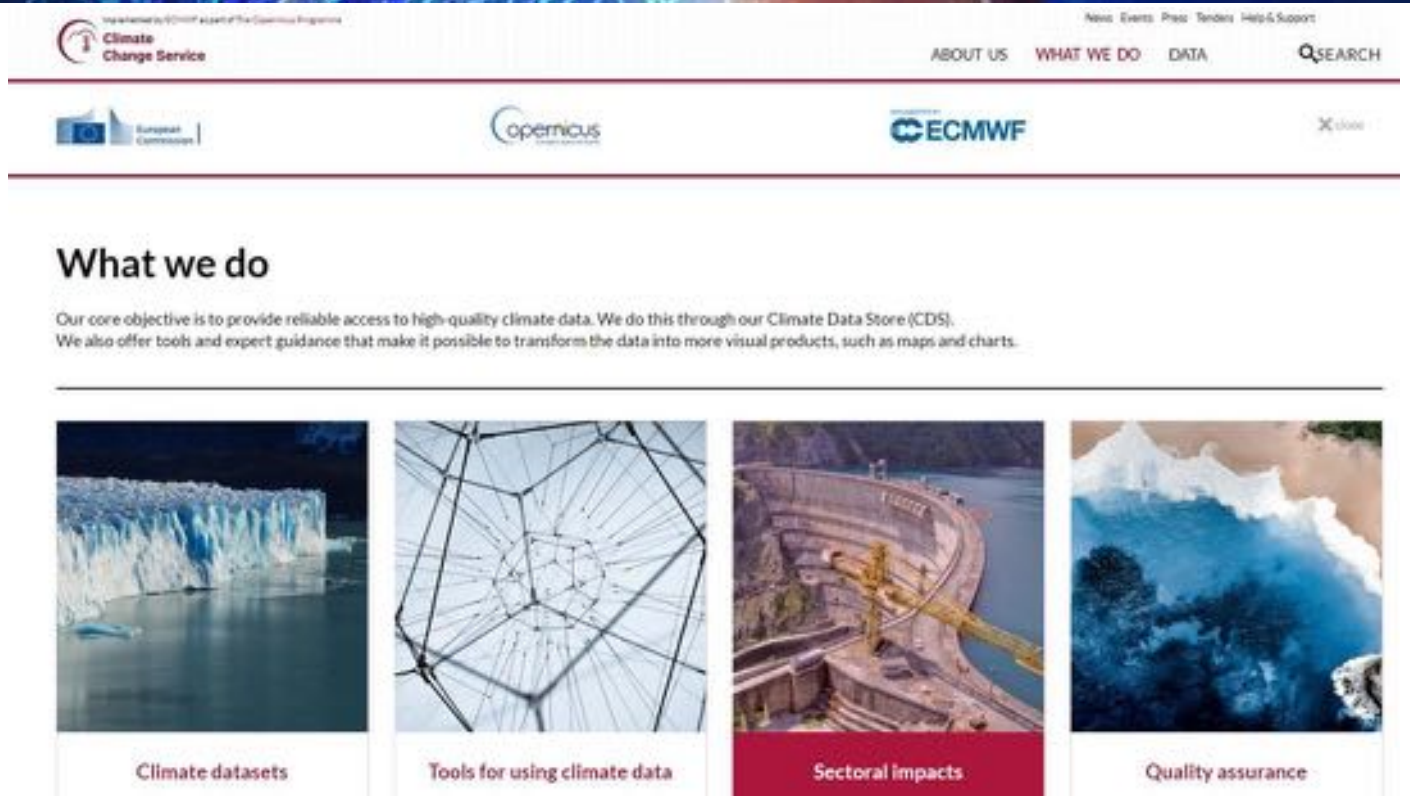
DKRZ
DEUTSCHES KLIMARECHENZENTRUM



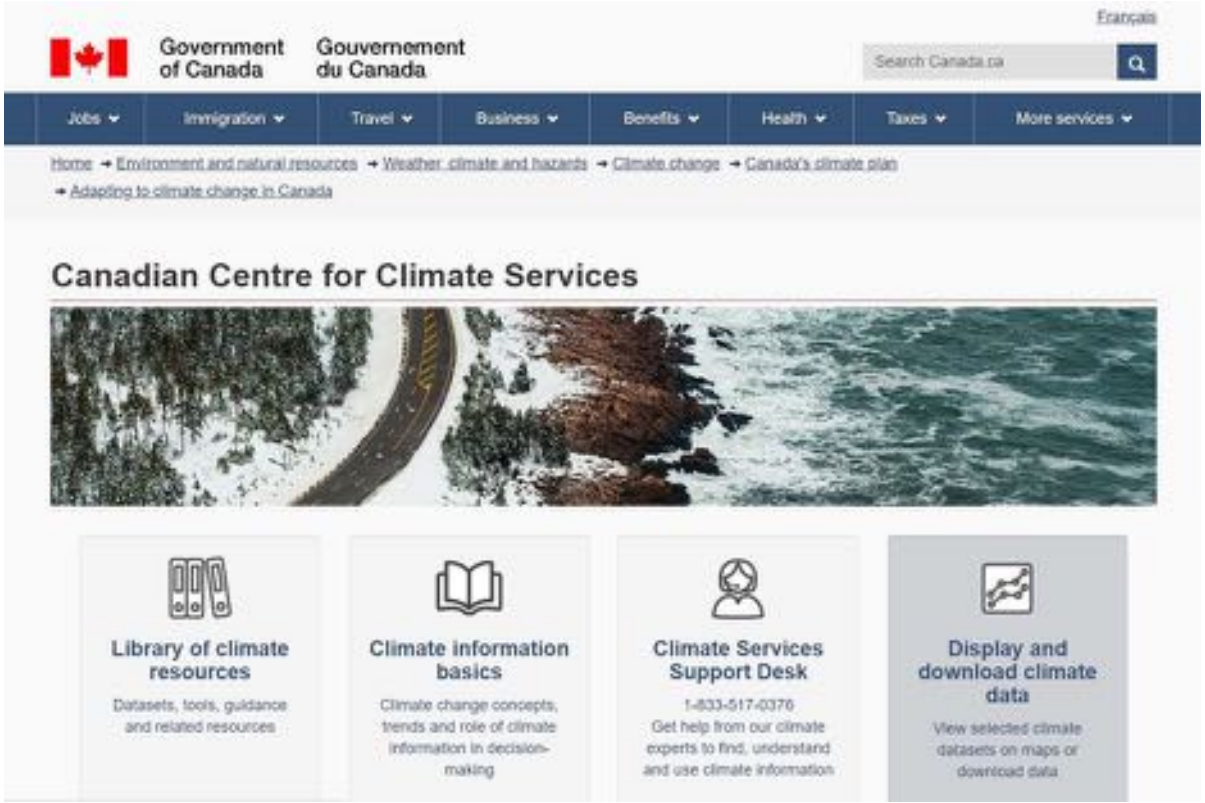
C3S Climate Data Store Toolbox



Copernicus Climate Change Service (C3S)



Canadian Center for Climate Services (CCCS)



2018-2019-2028: Climate data portal and modules

Environment a shared jurisdiction between federal, provincial & territorial. Established **climate services** consortia (Quebec, British-Columbia and Prairies) already partners of CCCS. Now fostering creation of new ones.

CCCS welcomes collaboration with European countries & Copernicus sharing needs for **user-driven** climate services for population, including **northern**, **Indigenous** or **remote** communities.

CRIM overseeing collaborative development of an interactive, web-based, climate information **portal** with **sectorial** modules (due 2019). Reuses PAVICS and Birdhouse.

Canada

OURANOS

PACIFIC CLIMATE IMPACTS CONSORTIUM

CRIM

Habitat Seven

Prairie Climate Centre

2019-2022-2027: Cyberinfrastructure challenge

DACCS: Data Analytics for Canadian Climate Services

Advancing a workflow-based Science Gateway, interoperable with Earth System Grid Federation (ESGF) for CMIP6 support. Several new services:

- Sea ice from observations and model simulations
- Ensemble diagnostics
- Downscaling, regridding
- Climatic niches
- CO2 and methane concentrations measurement
- Climate extremes and cyclone tracking
- Coastal vulnerability analysis
- EO data cubes
- Machine Learning toolboxes

2019-2022: Canada EO Data cube

- No EO data cube in Canada so far. Federal entities (Natural Resources Canada, Agriculture Canada) and industry currently working on it.
- Remote sensing data is crucial for monitoring, reporting, and verifying climate change mitigation policies.
- Some remote sensing products are part of the Essential Climate Variables (ECV), enabling production of long-term Climate Data Records.

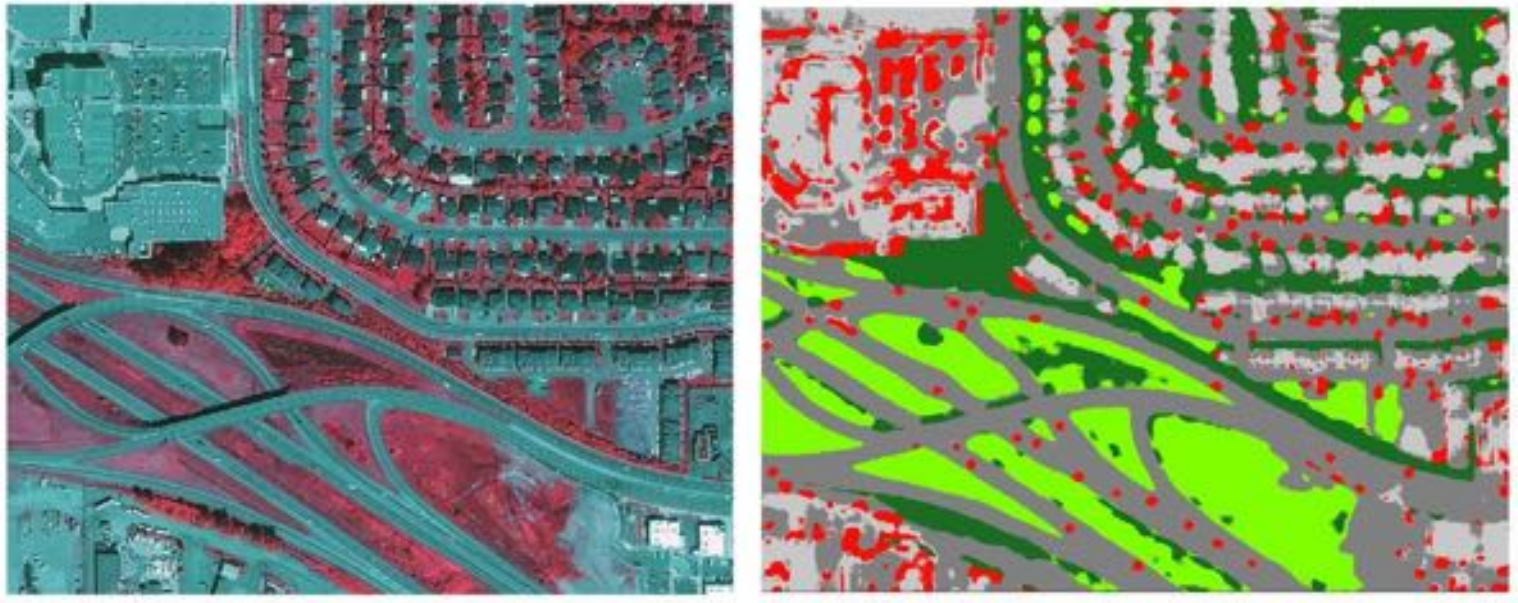
Project objectives:

1. Implement a Canadian EO Data Cube
2. Deploy and share ML workflows
3. Open to the research community (Notebooks?)
4. Implement EO services
5. Help in maintaining long term ECV series

OGC: innovation, interoperability and standards

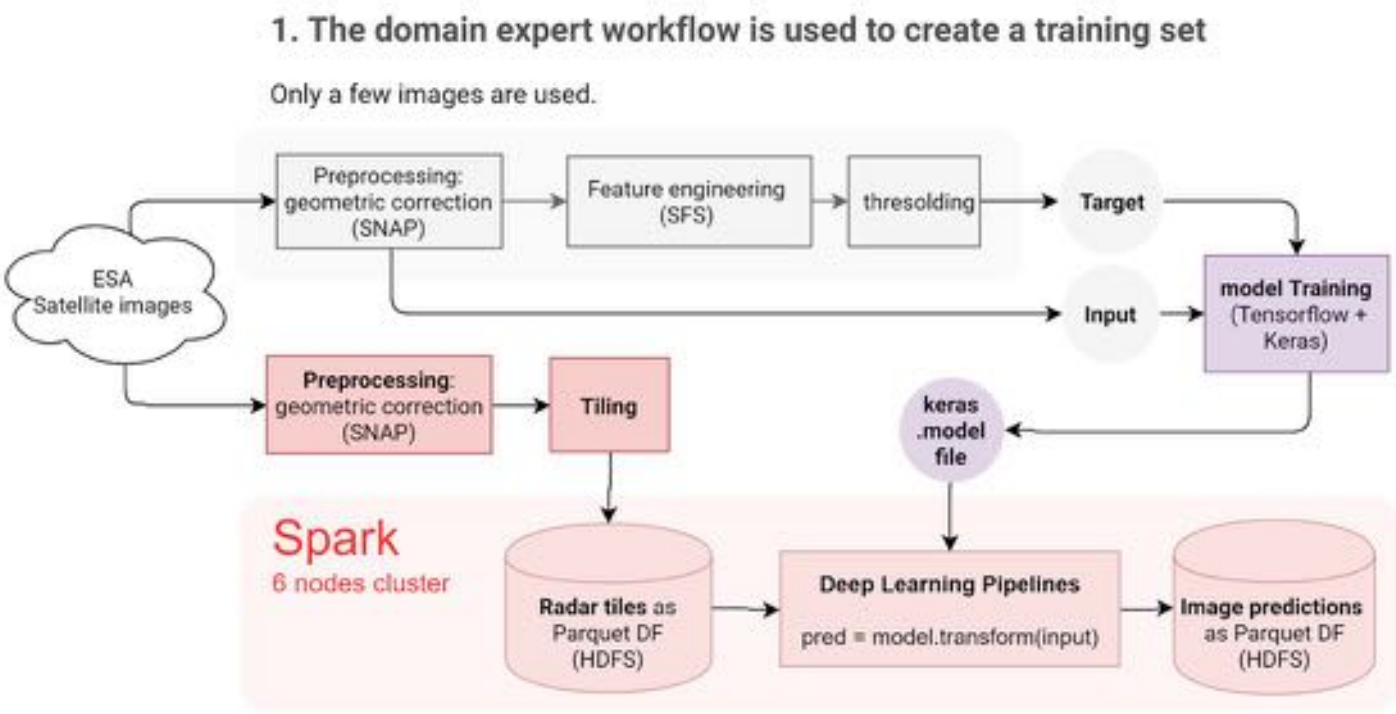
- Open standards for global geospatial community through consensus process.
- Canadian Forum on Geospatial Standards to provide open collaborative environment for all level of governments, Indigenous Organizations, academia, and industry. Currently 5 co-chairs, supported by OGC.
- OGC's *testbeds* design, develop, and test candidate interface and encodings.
  - Advance ESA's **Thematic Exploitation Platforms** (TEP) and support for **Mission Exploitation Platforms** (MEP).
  - Also advance federated clouds, workflows, security, satellite swaths, LIDAR, **machine learning**, etc.

Application 1: VHR imagery semantic segmentation



Pleiade VHR imagery of the city of Vancouver and its associated semantic segmentation, produced by a Deep Learning detector delivered by CRIM in OGC Testbed-14, ML task. *Image courtesy of Effigis Geo Solutions. Includes material © CNES 2018 (year of production), Distribution Airbus DS Geo SA / Airbus DS Geo Inc., all rights reserved.*

Application 2: Transfer learning for flood detection



2. A convolutionnal neural network is trained to replicate the target water segmentation

The model is trained with Keras on single machine (GPU recommeded).

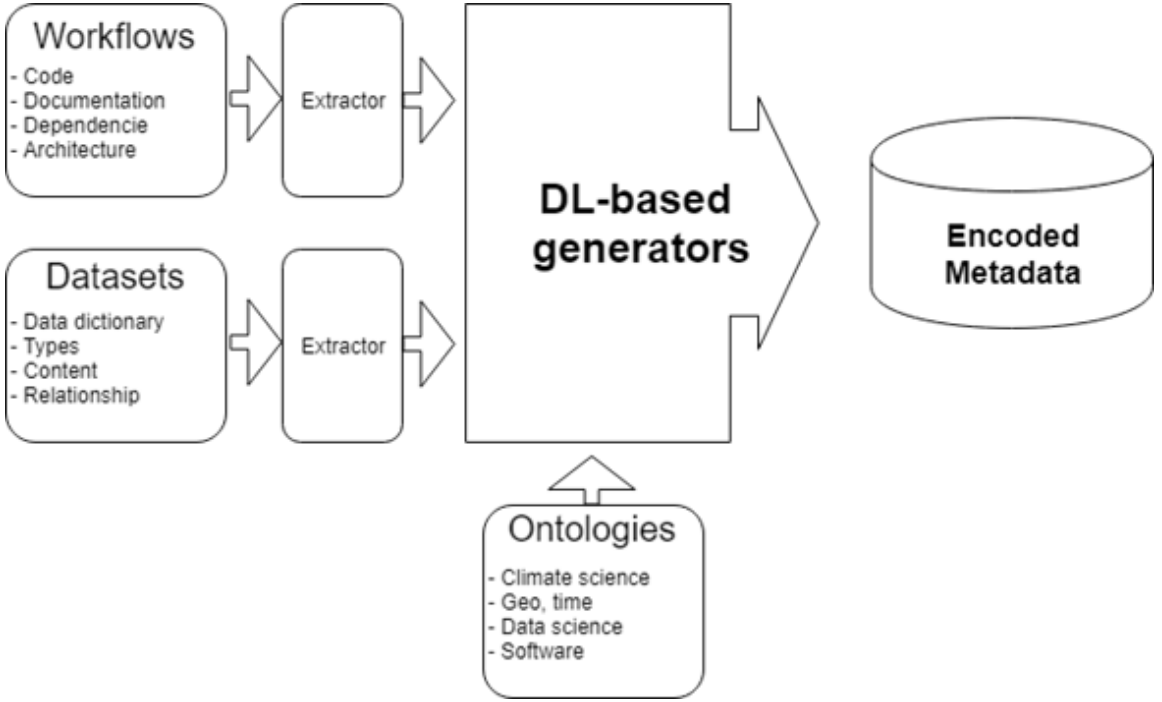
3. Batch segmentations are performed on the spark cluster

For now, geometric corrections and tiling are still performed out of the cluster.

Source: Rajotte & Al. "Apache Spark MLlib applied to geospatial imagery for flood indication" presented at ApacheCon 2018 and described in OGC Testbed-14, ML task

Application 3: future NLP modules in DACCS

- Convert and encode resources into queryable form of metadata.
  - **Deep learning** to vectorise knowledge and add it to metadata repository
- User-oriented query alignment to enrich and transcode user query.
  - **Active learning** to drive the alignment between perceived intent of user query and the appropriate resources



Source: Behrer, H. & Ménard, P.A. - CRIM 2018

Conclusion

- Extraordinary **momentum**!
  - Multiple projects, reinforcing each other
  - Sustainable, far-reaching, open and ambitious
  - Academia, industry, governmental stakeholders at provincial, national and international
  - Great conditions for AI, EO and climate science in Canada
  - Perfect fit for Copernicus
- What do we need?
  - **Funding** for projects, but also travel
  - Access to large-scale **infrastructure**
  - **Data**, volume but also variety
  - Personnel: Experts, interns, researchers, developers
  - Increased **coordination** and collaboration
    - Technical and scientific
    - Administrative and political