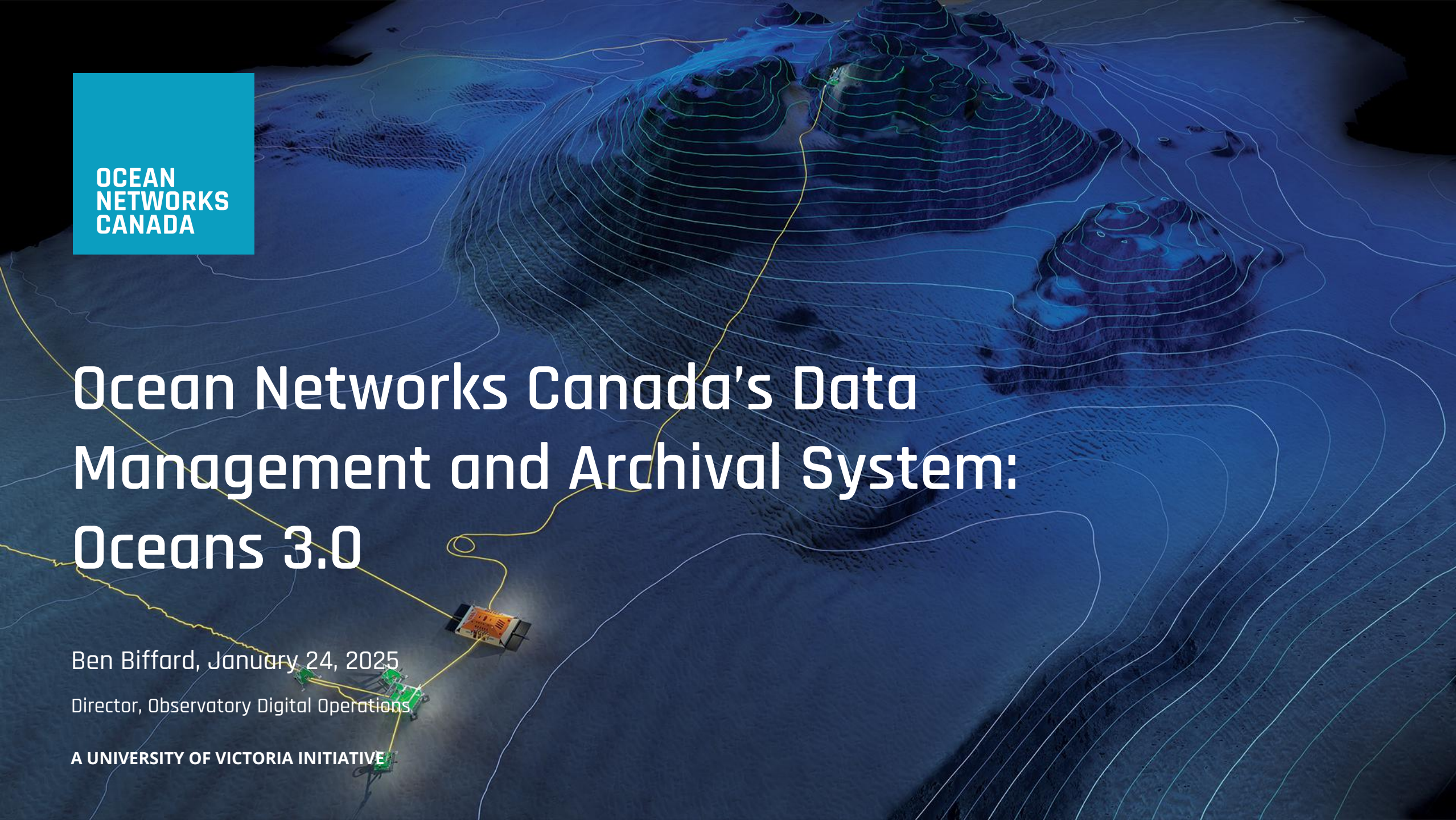


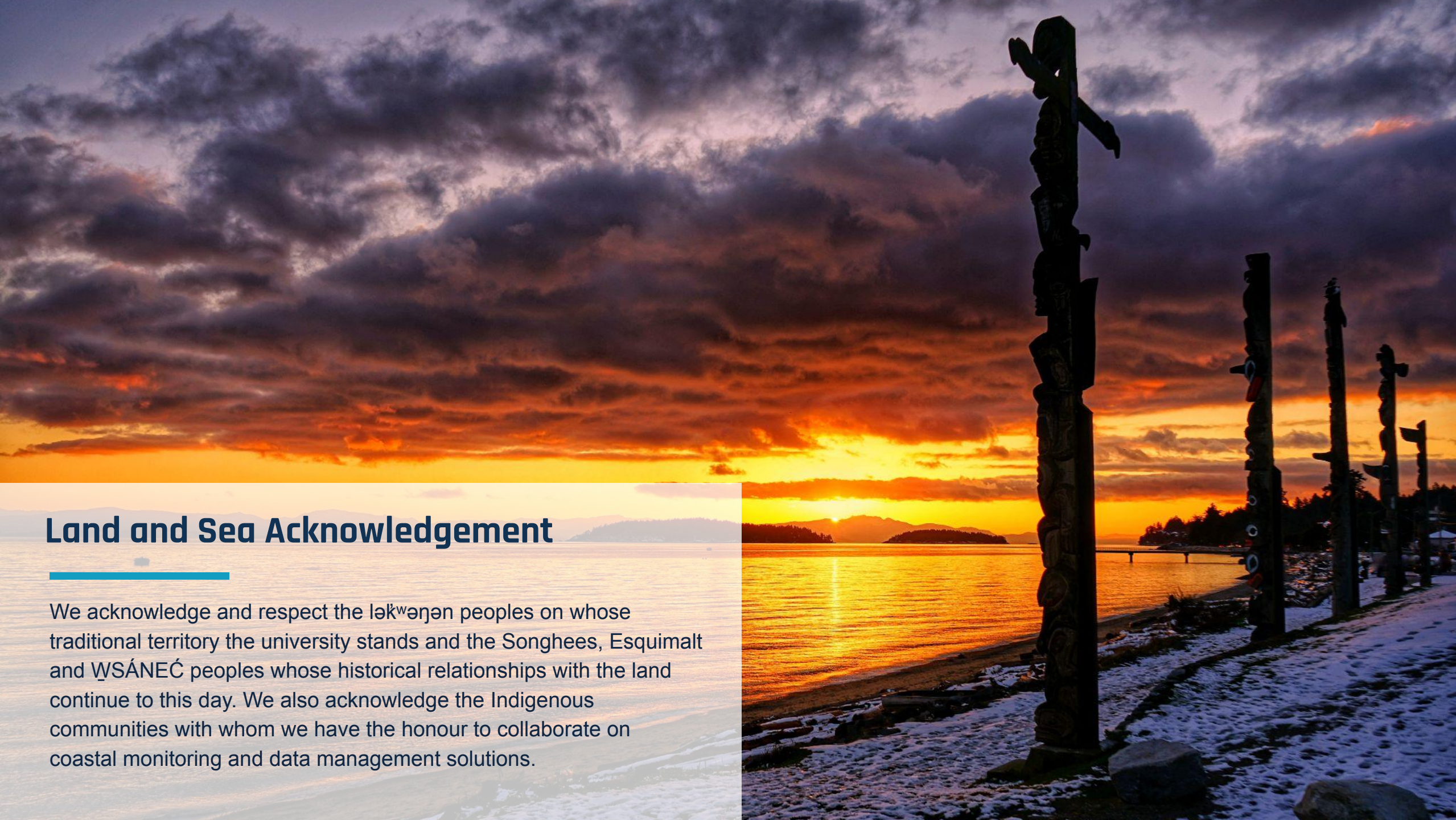
OCEAN NETWORKS CANADA

Ocean Networks Canada's Data Management and Archival System: Oceans 3.0

Ben Biffard, January 24, 2025

Director, Observatory Digital Operations

A UNIVERSITY OF VICTORIA INITIATIVE



Land and Sea Acknowledgement

We acknowledge and respect the lək̓ʷəŋən peoples on whose traditional territory the university stands and the Songhees, Esquimalt and W̱SÁNEĆ peoples whose historical relationships with the land continue to this day. We also acknowledge the Indigenous communities with whom we have the honour to collaborate on coastal monitoring and data management solutions.

ONC Strategic Vision

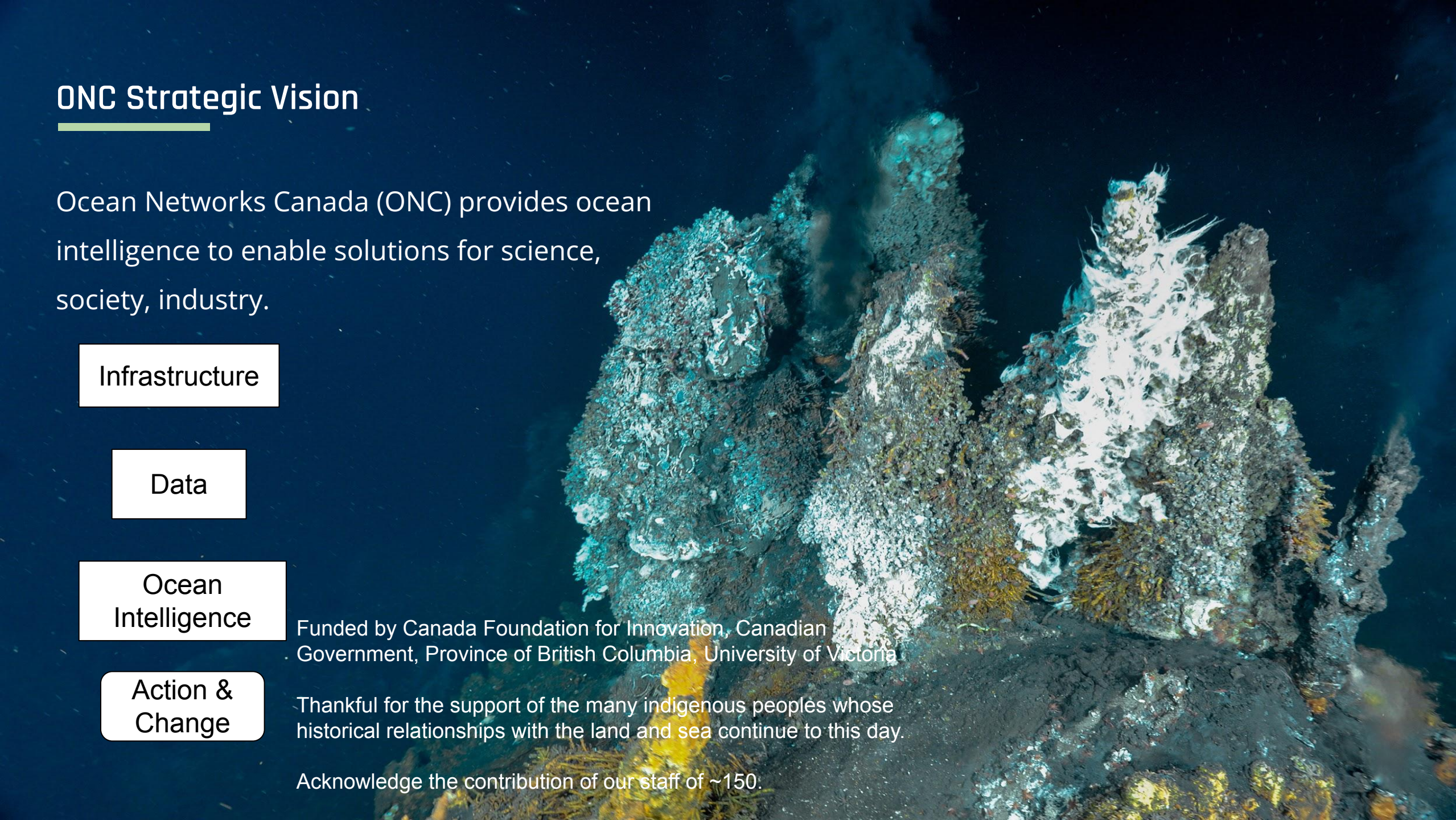
Ocean Networks Canada (ONC) provides ocean intelligence to enable solutions for science, society, industry.

Infrastructure

Data

Ocean Intelligence

Action & Change



Funded by Canada Foundation for Innovation, Canadian Government, Province of British Columbia, University of Victoria

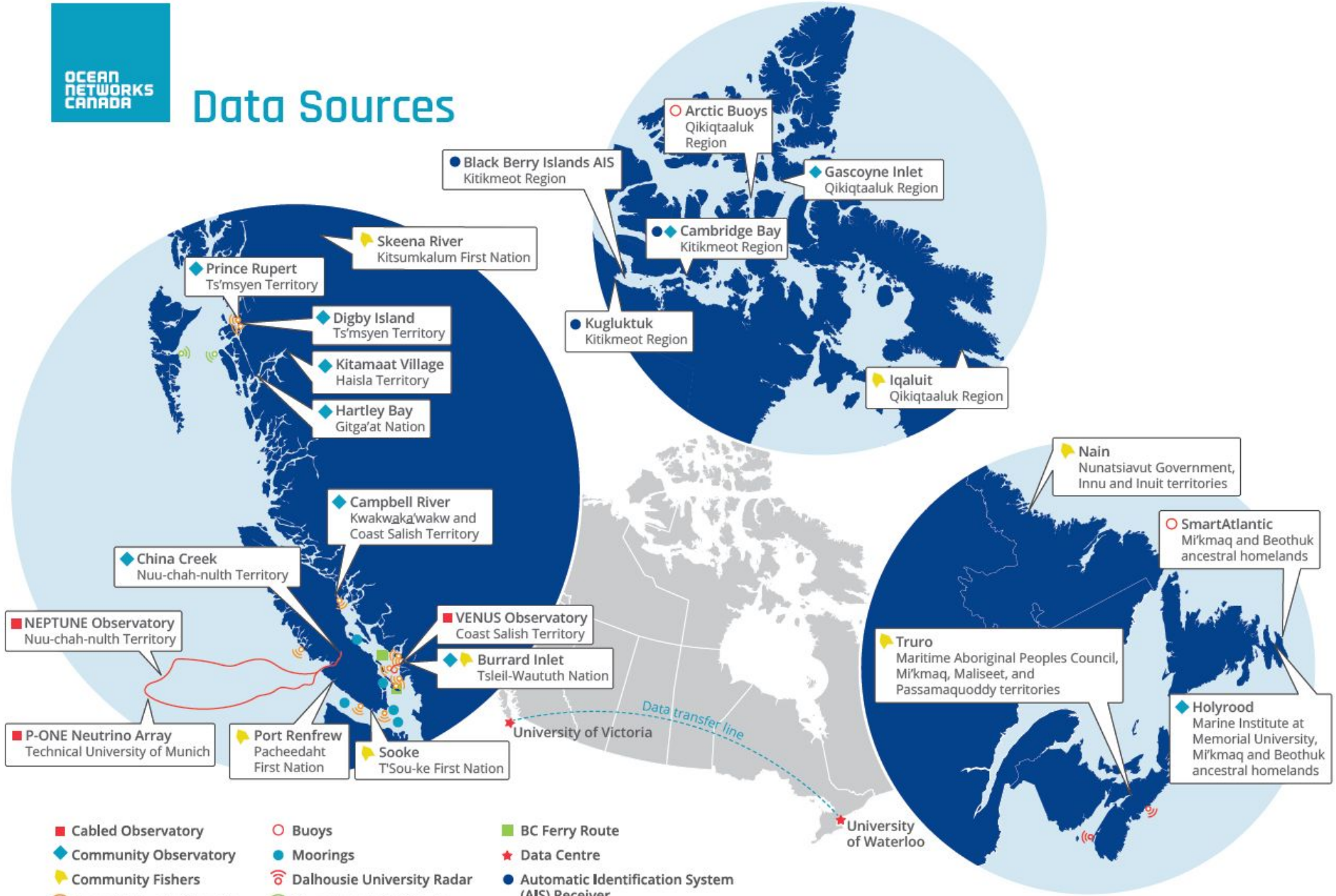
Thankful for the support of the many indigenous peoples whose historical relationships with the land and sea continue to this day.

Acknowledge the contribution of our staff of ~150.


Arctic Ocean
OCEANS 3.0
DATA MANAGEMENT PORTAL
Ocean Health Monitoring
Scientific Discovery
Tsunami Preparedness
Climate Change Mitigation
Maritime Safety
Arctic Observing
Neutrino Detection
Blue Economy
Forensics
Marine Mammal Avoidance
Ocean Soundscapes
Underwater Imagery
Ocean Sciences Education
Inform Incident Response
Earthquake Early Warning
Tectonic Plates Monitoring
Community-Based Monitoring
Pacific Ocean
Atlantic Ocean
Southern Ocean
DATA SOURCES
Major Observatory
Coastal Community Observatory
Coastal Observatory
Geo-Seismic Sensor (ONC)
Geo-Seismic Sensor (Natural Resources Canada & ONC)
Neutrino Array
Community Fishers Mobile Assets
Subsea Fibre Optic Cable
Mooring/Buoy
Data Center
Mobile Asset
AIS Reciever
RADAR (Fisheries and Oceans Canada)
RADAR (ONC)
RADAR (Dalhousie University)

OCEAN NETWORKS CANADA

Data Sources



- Cabled Observatory
- Community Observatory
- Community Fishers
- Ocean Networks Canada Radar
- Buoys
- Moorings
- Dalhousie University Radar
- Department of Fisheries and Oceans Radar
- BC Ferry Route
- Data Centre
- Automatic Identification System (AIS) Receiver

(Excluding Earthquake Early Warning)


OCEAN NETWORKS CANADA
EARTHQUAKE EARLY WARNING STATIONS
Port Hardy
Newcastle Ridge
Tahsis
VANCOUVER ISLAND
Cumberland
BRITISH COLUMBIA
Port Alberni
Vancouver
Tofino
Bamfield
Clayoquot Slope
Barkley Canyon
Cascadia Basin
Explorer Ridge
EXPLORER PLATE
Sovanco Fault Zone
Nootka Fault Zone
NORTH AMERICAN PLATE
Cascadia Subduction Zone
JUAN DE FUCA PLATE
Juan de Fuca Ridge
PACIFIC PLATE
WASHINGTON
Node
Shore Station
Ocean Networks Canada (ONC) Geo-seismic Sensors
Natural Resources Canada + ONC Geo-seismic Sensors
Fibre-optic Cable
Fault Line
Subduction Zone
Spreading Centre
Bathymetry (m)
0
-3500
N
W
E
S
50 km
AN INITIATIVE OF
University of Victoria
BRITISH COLUMBIA - CANADA
WASHINGTON - USA
Data Sources: Plate margins estimated from fig. 2 in Hyndman, 1995, Can. J. Earth Sci.; Digital Elevation Model derived from Smith and Sandwell Global Altimetry, Natural Resources Canada topography, Canadian Hydrographic Service (CHS) bathymetry.
This product has been produced by the Univeristy of Victoria based on CHS data, pursuant to CHS Direct User License No. 2019-1007-1260-ON. The incorporation of data sourced from CHS in this product shall not be construed as constituting an endorsement by CHS of this product. This Product does not meet the requirements of the Charts and Nautical Publications Regulations, 1995 under the Canada Shipping Act, 2001. Official charts and publications, corrected and up-to-date, must be used to meet the requirements of those regulations.
Last Updated: 16 January 2025

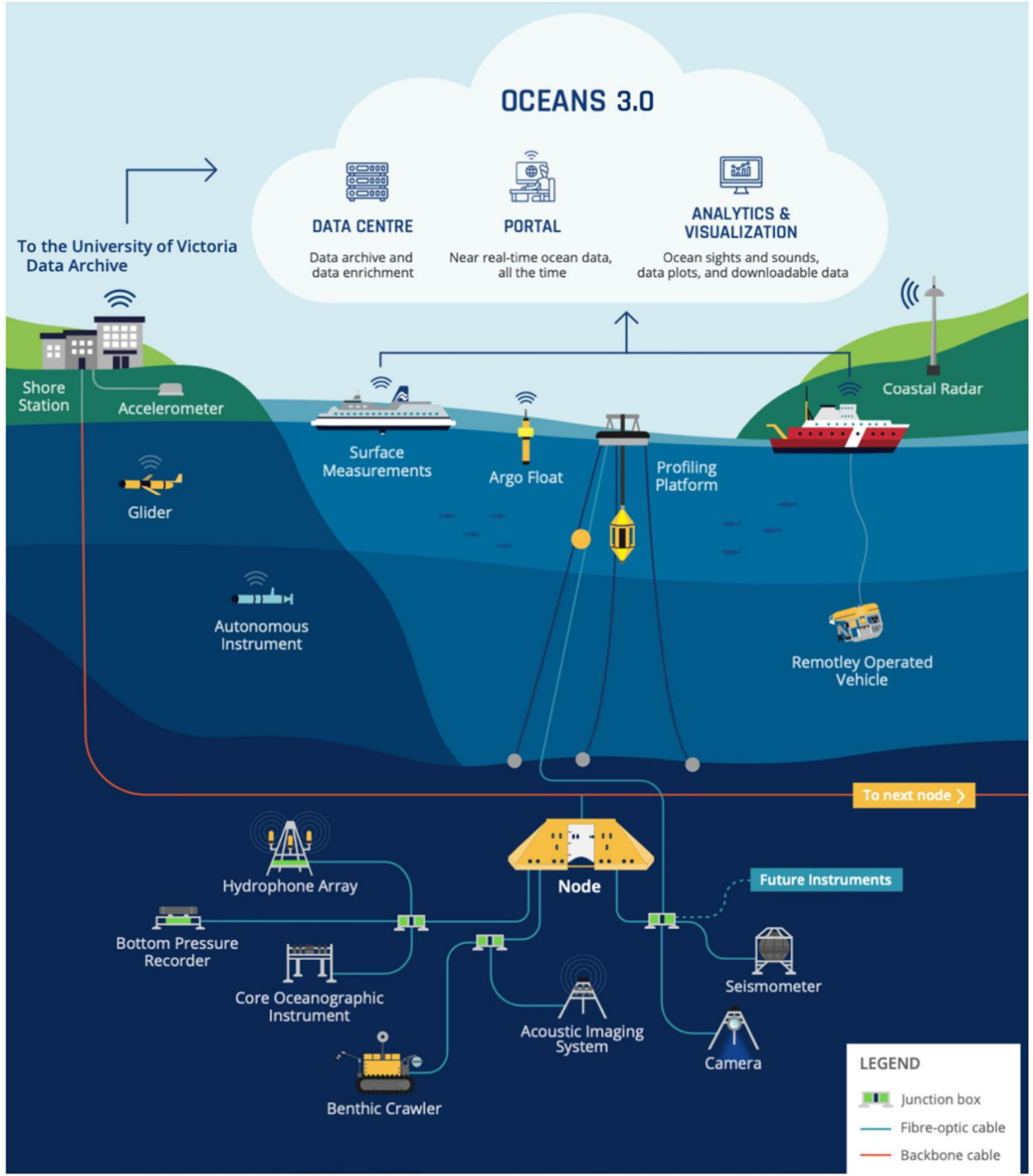
Oceans 3.0 runs the *Network* in Ocean Networks Canada

Began 20 years ago powering world's first large-scale internet-connected ocean observatory.

Originally called DMAS: Data Management and Archival System, but it does more than that now.

Enables continuous, near real-time ocean observation, data collection, and communications.

Data collected from cabled, autonomous, and mobile instruments in deep-sea, coastal, and land-based locations.



Oceans 3.0 in Numbers

12500+ Active sensors producing data;

977 Instruments producing data daily;

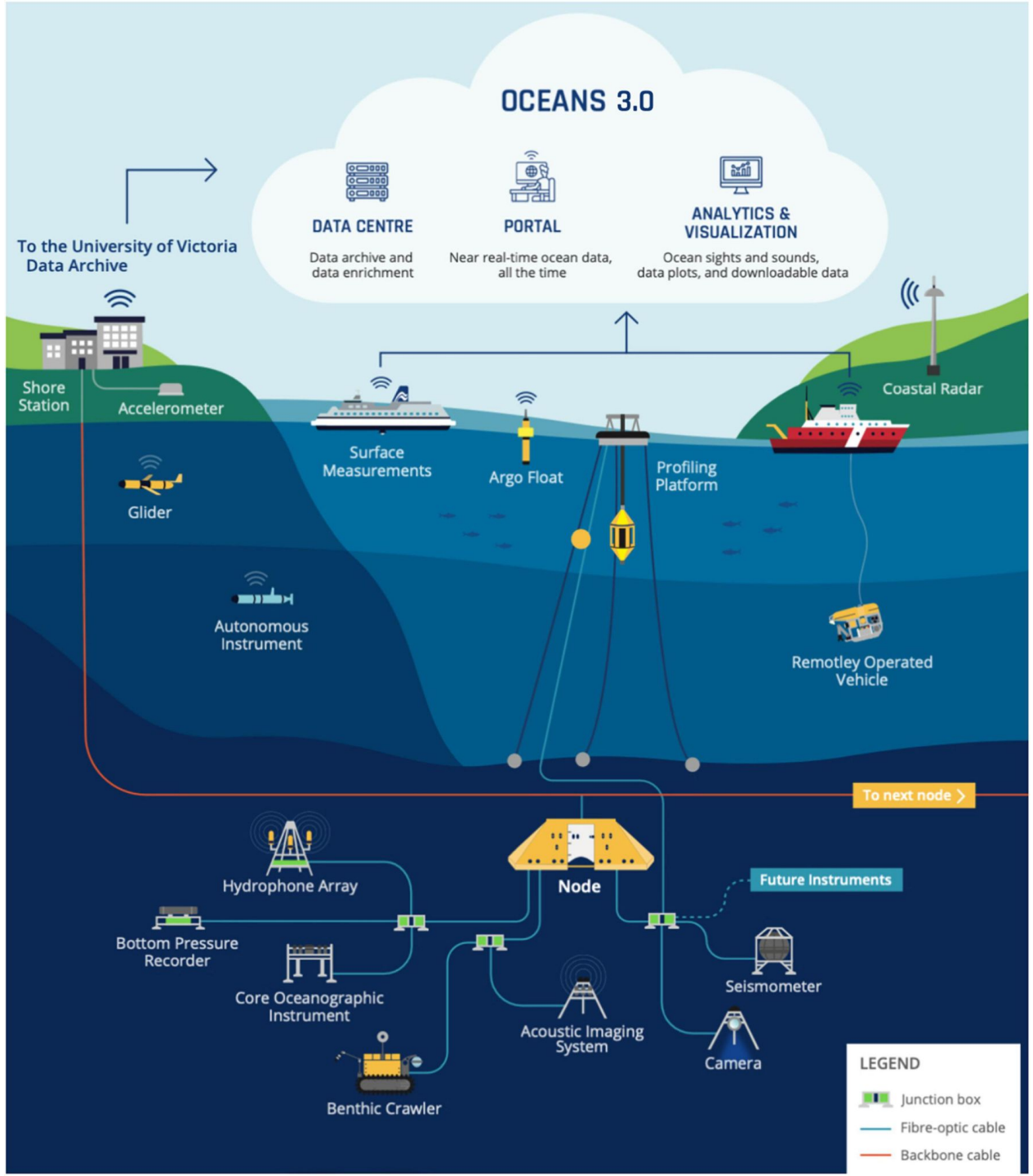
284 Unique file-based data products*;

9921 Average pre-generated plots produced daily;

9654 Average daily data requests;

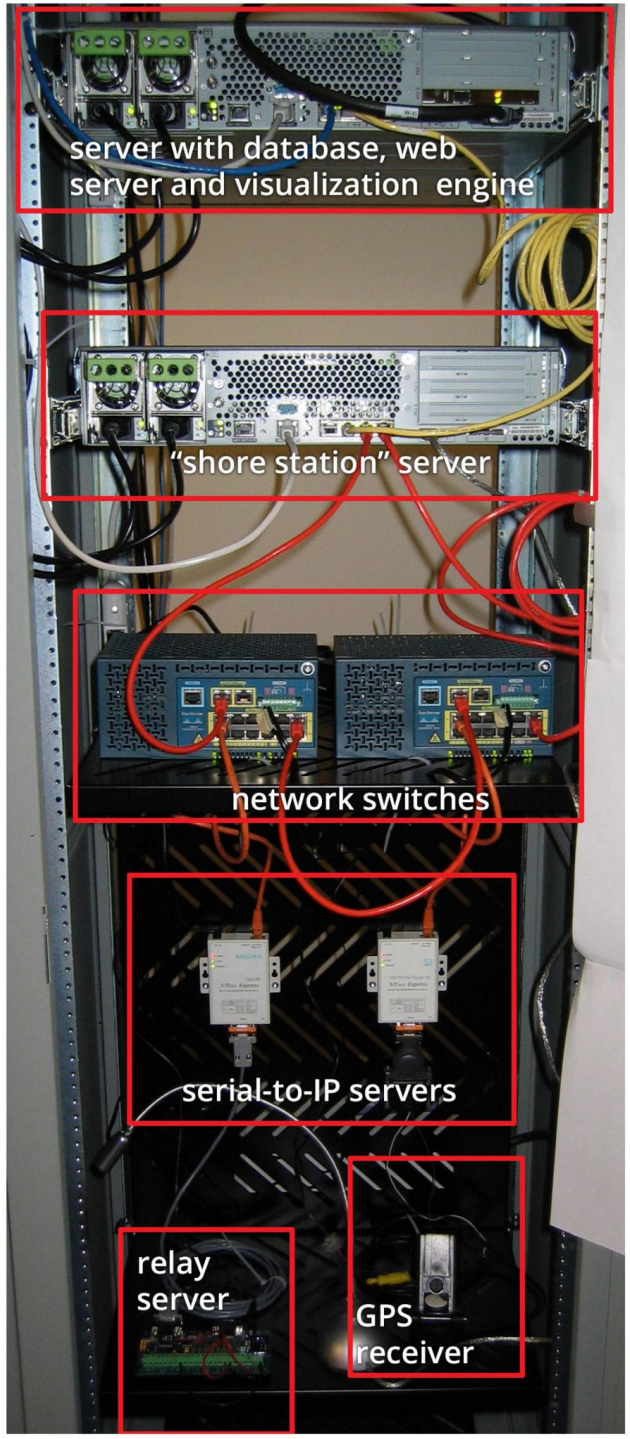
300 GB Average volume of uncompressed data archived per day;

1.74 PB Total uncompressed volume of archived data.

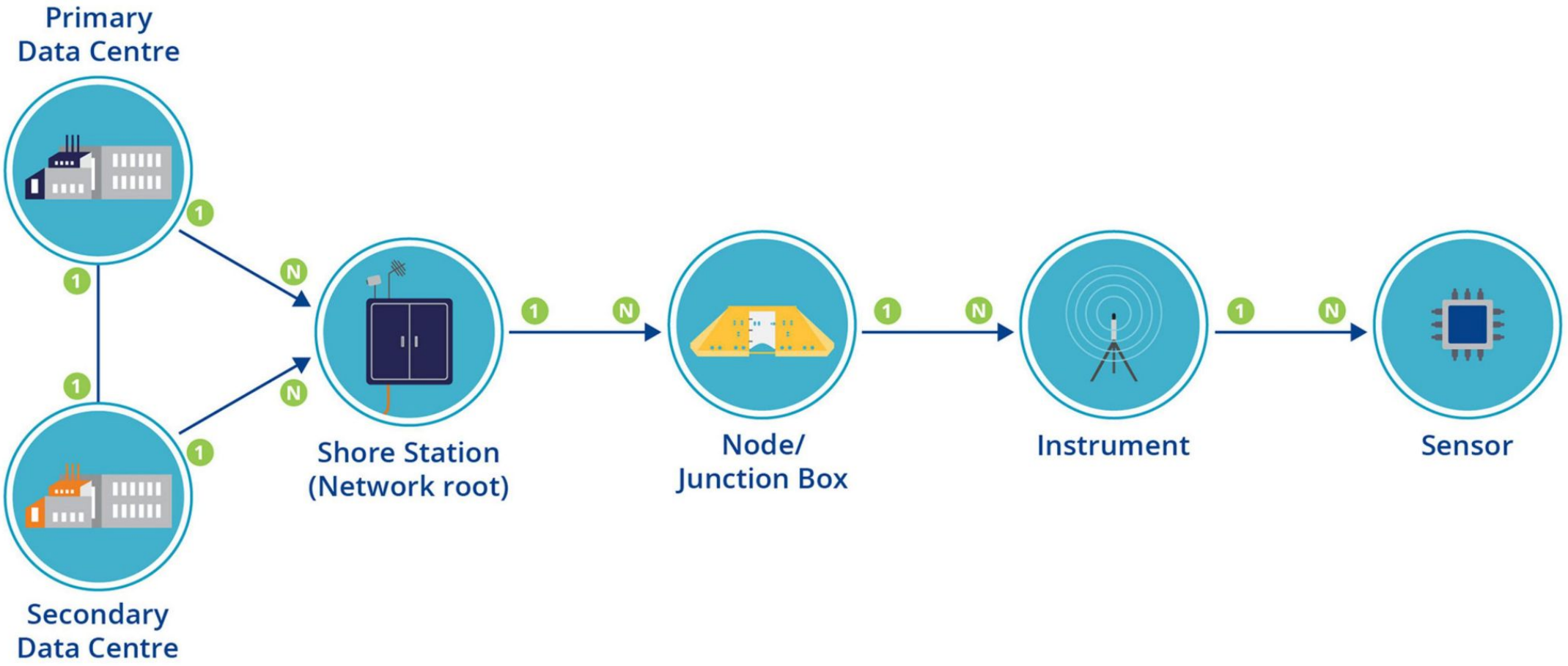


Prototype DMAS Hardware

1. Server for running data repository (database, web server, visualization software); the “data centre”
2. Shore station server with drivers that implement communication protocols for each instrument (configure, control, parsing and pre-processing)
3. Network equipment for tree topology of infrastructure, including switches
4. Serial-to-IP servers for interfacing with instruments
5. Relay server to control instrument power



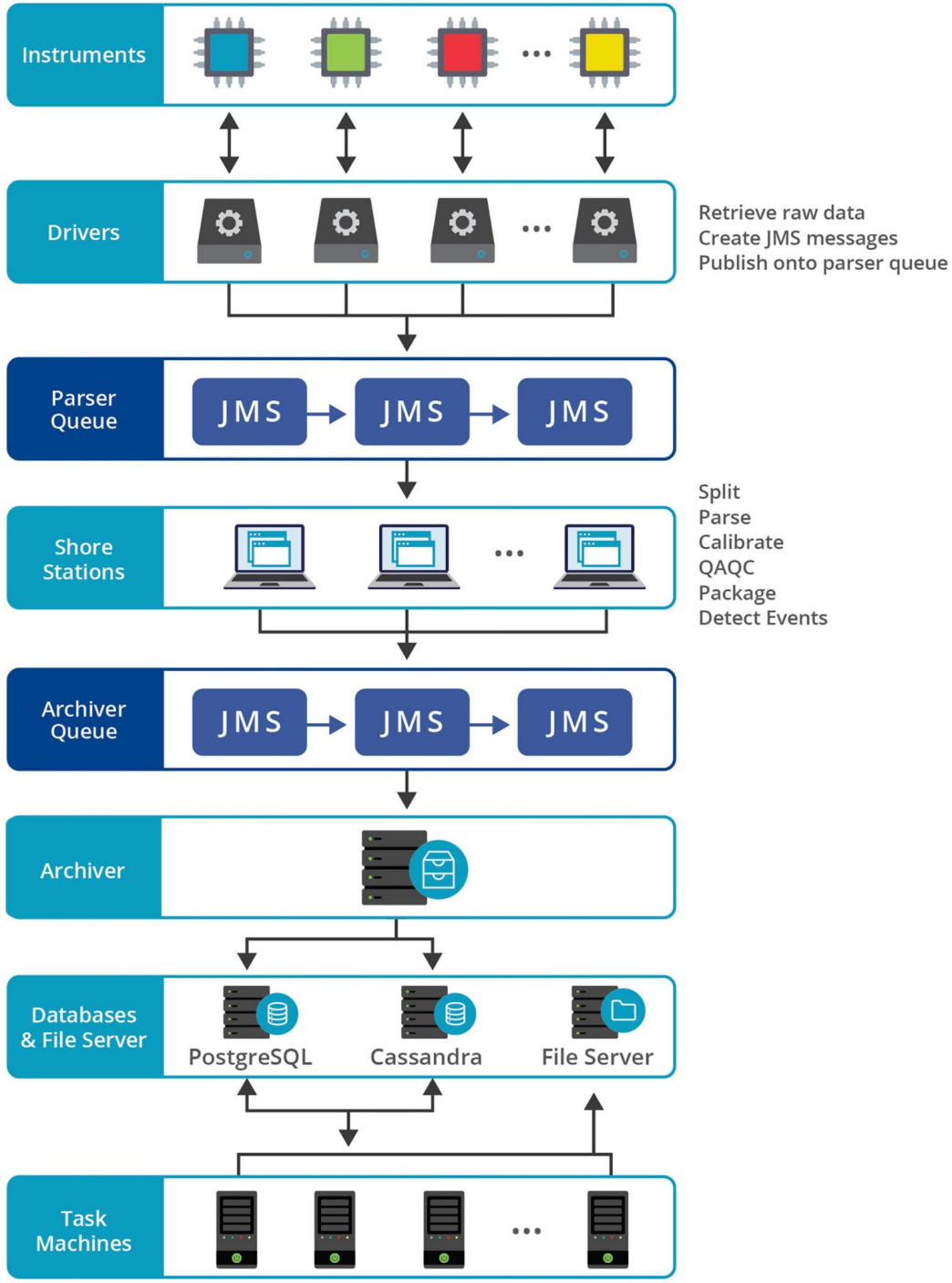
Infrastructure Hierarchy from Data Center to Sensor



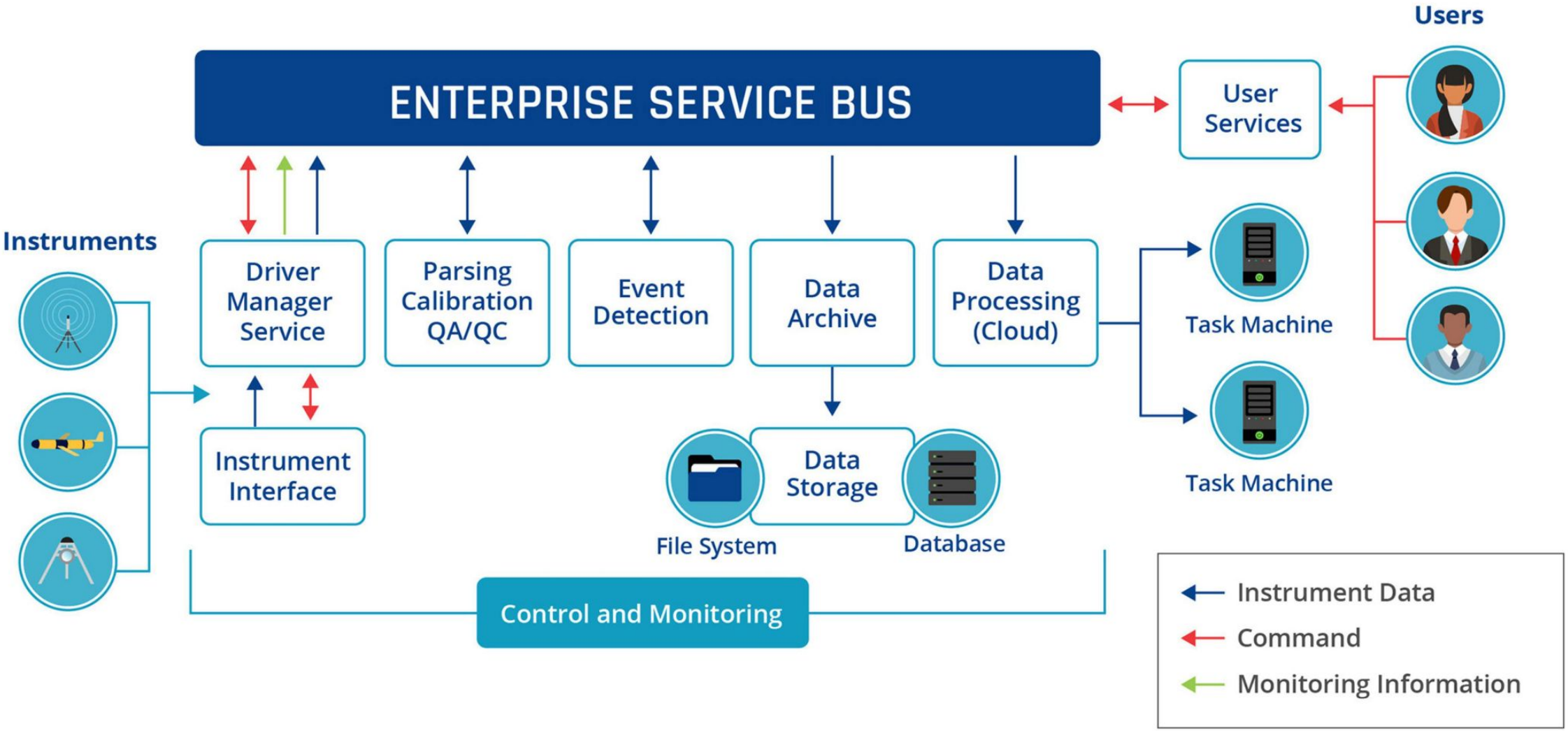
Data Acquisition Process

1. Acquiring data readings
2. Publishing onto parser queue
3. Processing by shore station
4. Publishing onto archiver queue
5. Archival of raw and processed data
6. Task machines for value-added processing and data products
7. Available on the Oceans 3.0 data portal, 100% live!

   Latency ranges from milliseconds (driver acquisition) to minutes (file acquisition).



System Architecture: How the Network works



Technologies

Metadata database: PostgreSQL

Raw data holding tank: Cassandra

File archive: “AD” (NetAPP and Synology)

Value add data:

- Scalar data: Cassandra
- Complex data:
  - Generate on-the-fly
  - Pre-generate and store in the file archive
- Interoperability: ERDDAP
  - Scalar data: custom cassandra driver
  - Complex data: NetCDF store

ERDDAP
Easier access to scientific data

ckan

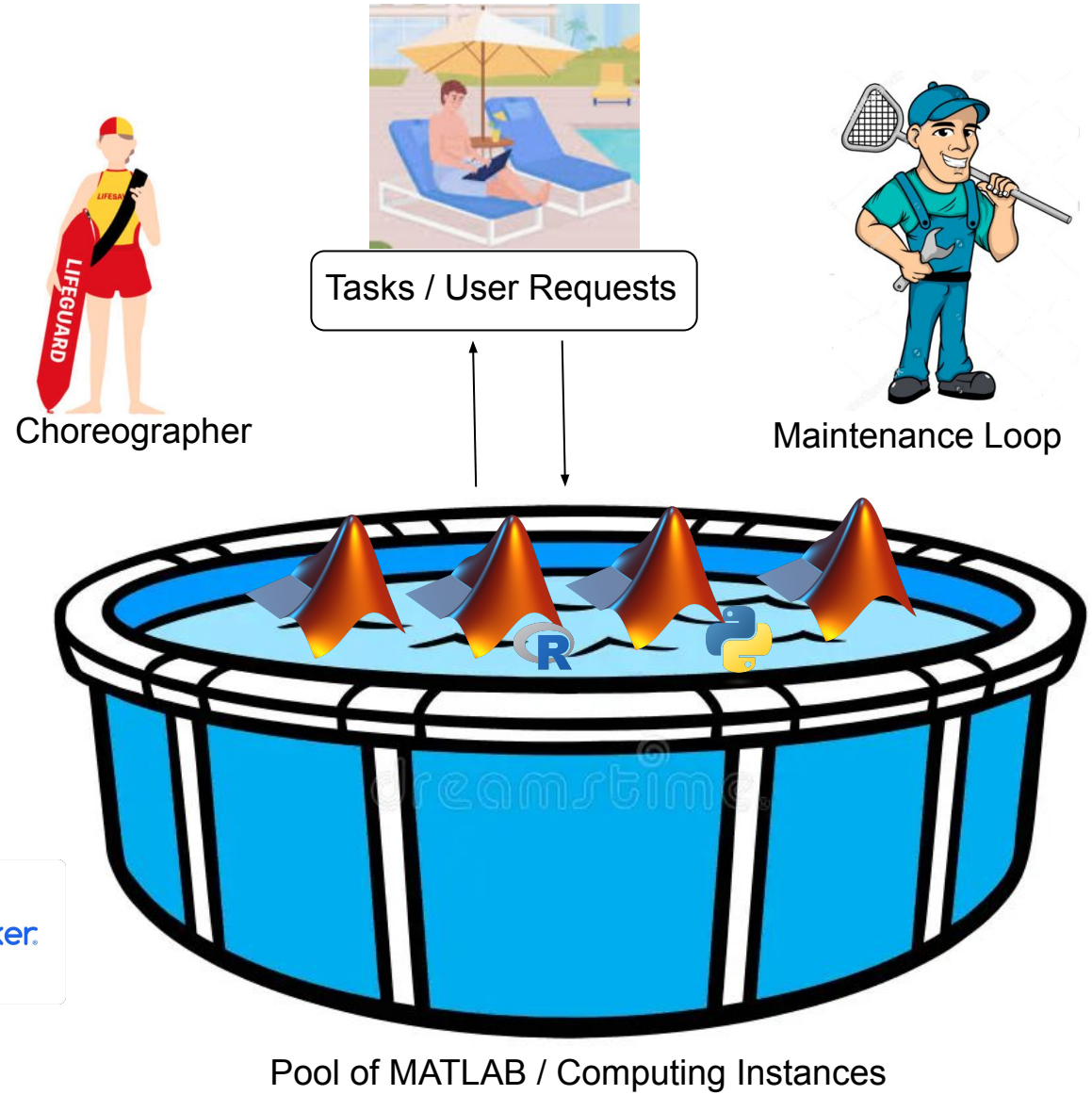
File Formats and Data Products

- Anything goes as long as it has appropriate metadata and documentation
- Support manufacturer's original formats
- Maintain a list of preferred formats: FLAC for audio, MP4/264 for video
- Raw data collected in a flexible raw log hex format
- Support compression
- Minimal requirements: raw format, manufacturer's original format, accessible data format and a visualization (plot). Over 300 types of instruments.

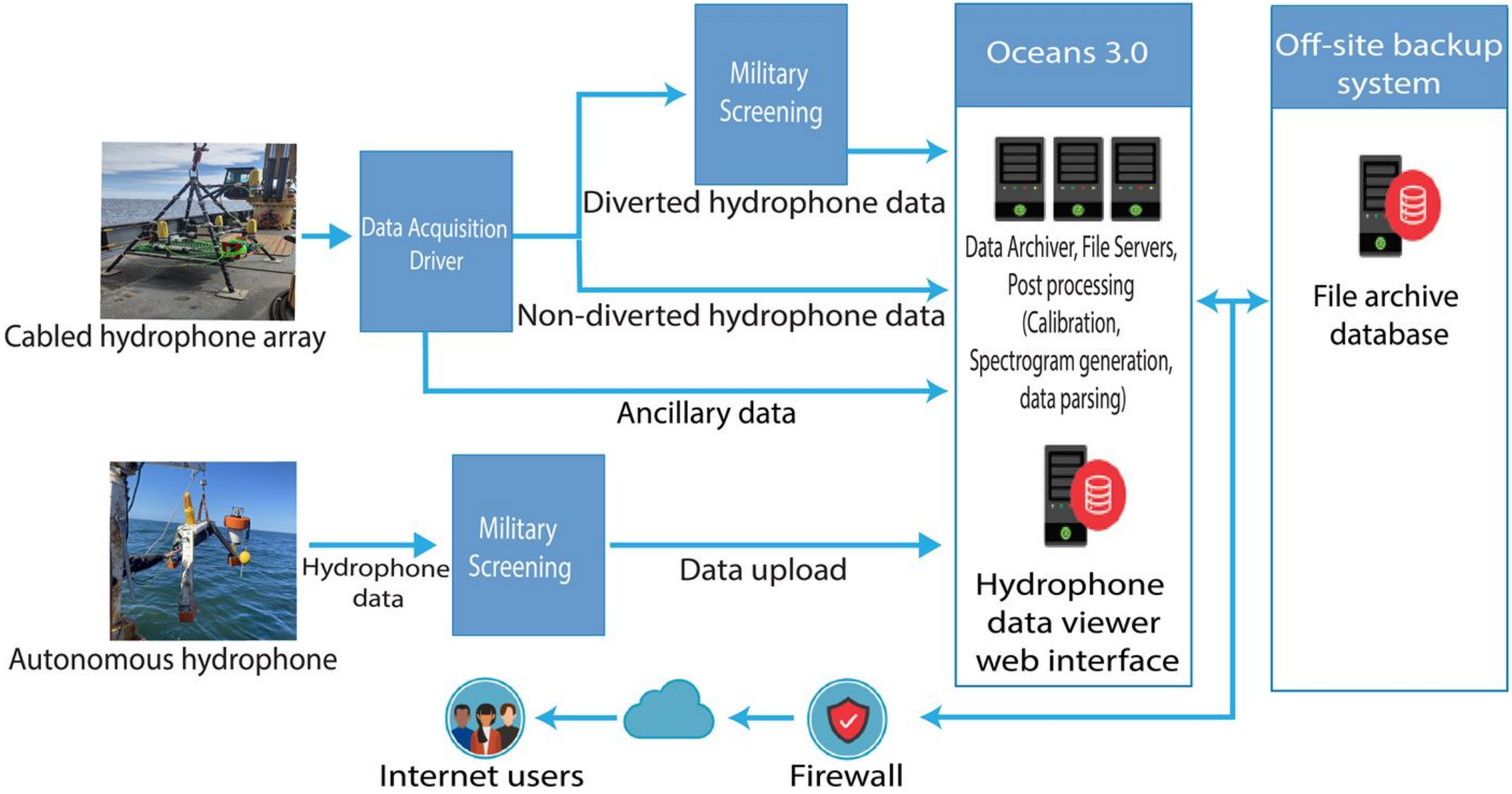
- > 400 different data product formats available
  - Examples: NetCDF, CSV, MAT, HDF5, FLAC Audio, MP4 Video
- Generated within Oceans 3.0 by Java, MATLAB or Python
- On-the-fly generation is easier to maintain, allows us to support option sets
- Support citations and persistent identifiers, allowing reproducibility
- Encourage P.I.'s to contribute code
- In-progress: citable code and complete provenance

Challenges: generating data products on-the-fly: *MATLAB-as-a-service* Low Latency System

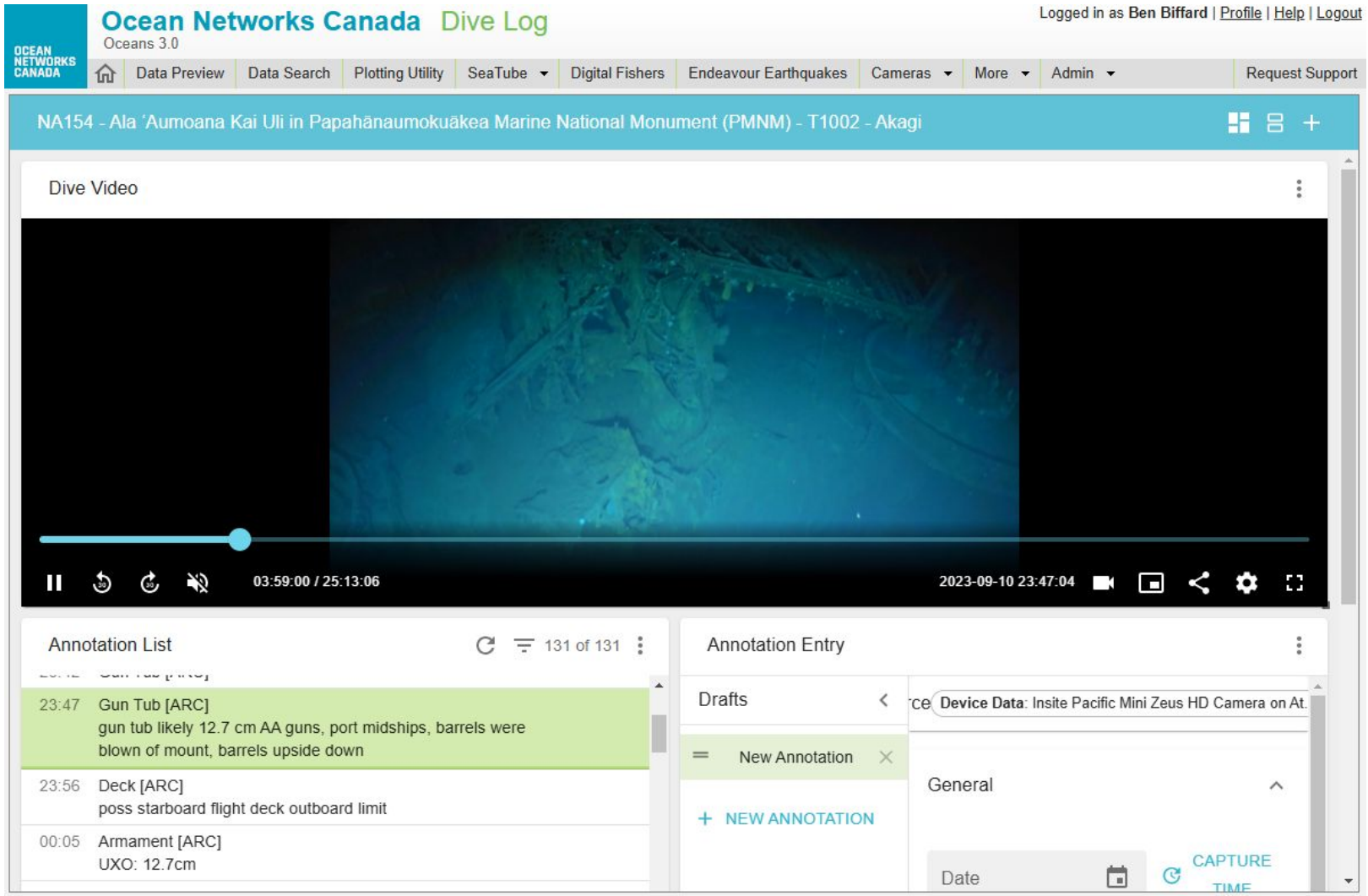
- Eliminates start-up latency and resource use. Instances can be re-used hundreds of times!
- On-demand latency can be as low as 1 second for a simple plot or small data file.
- Average data search request completion time and computation resource utilization reduced by >60%. Even greater gains for nightly processing tasks such as preview generation and automated testing.
- Fully integrated with task management features such as error handling, reporting, task scheduling and cancellation.
- Extensible to Docker images (in-progress).

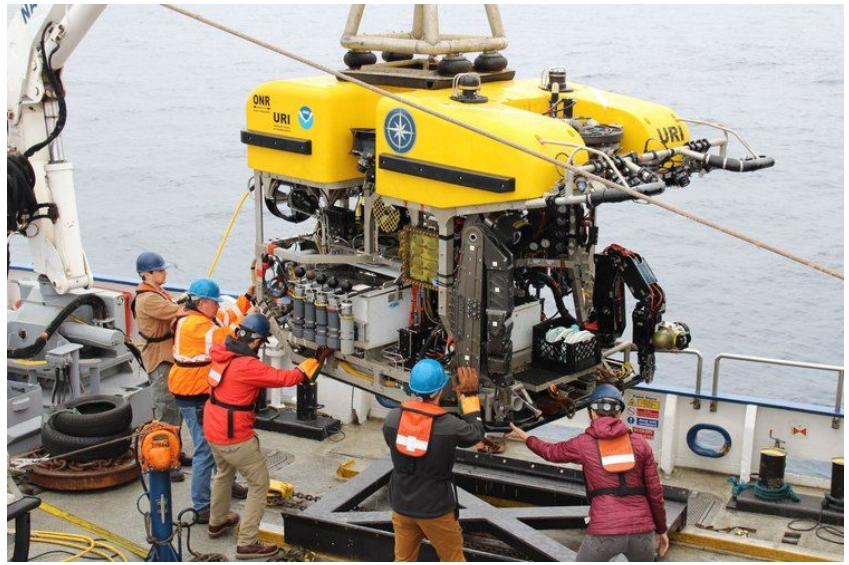


Challenges: Handling Sensitive Data (Hydrophones, Seismometers, etc)

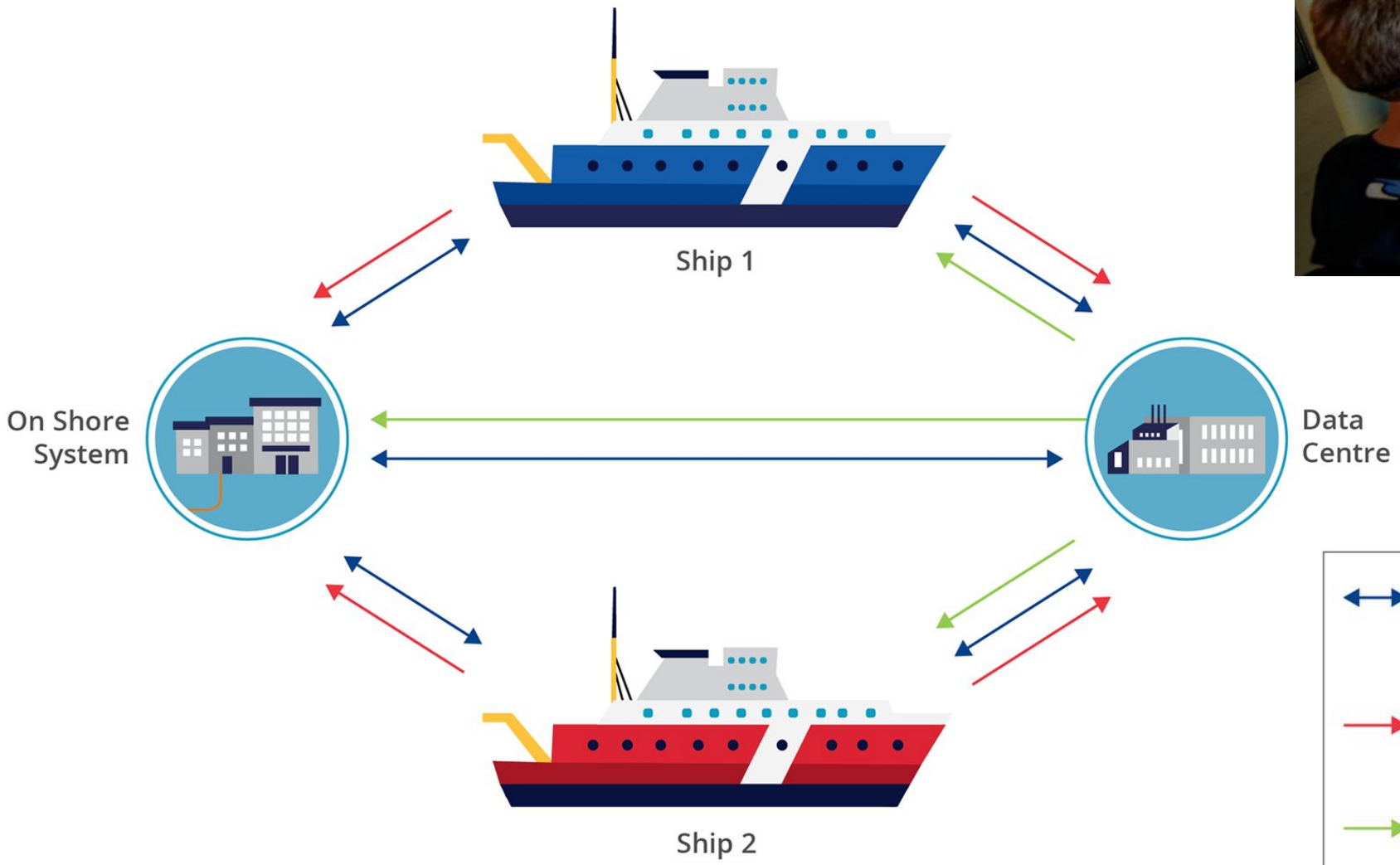


Challenges: Multi-point Synchronized Live Data and Annotations





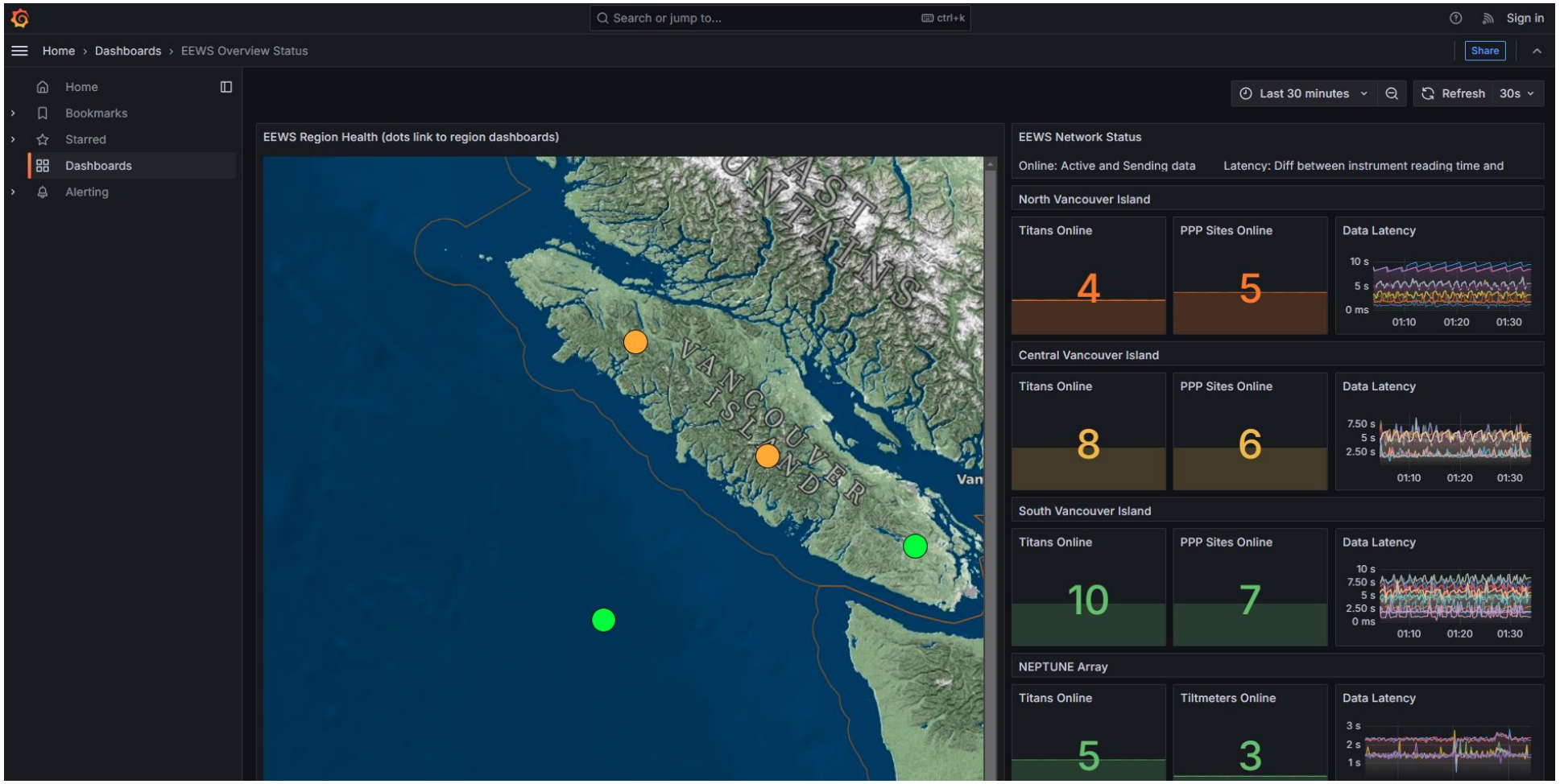
Challenges: Multi-point Synchronized Live Annotations



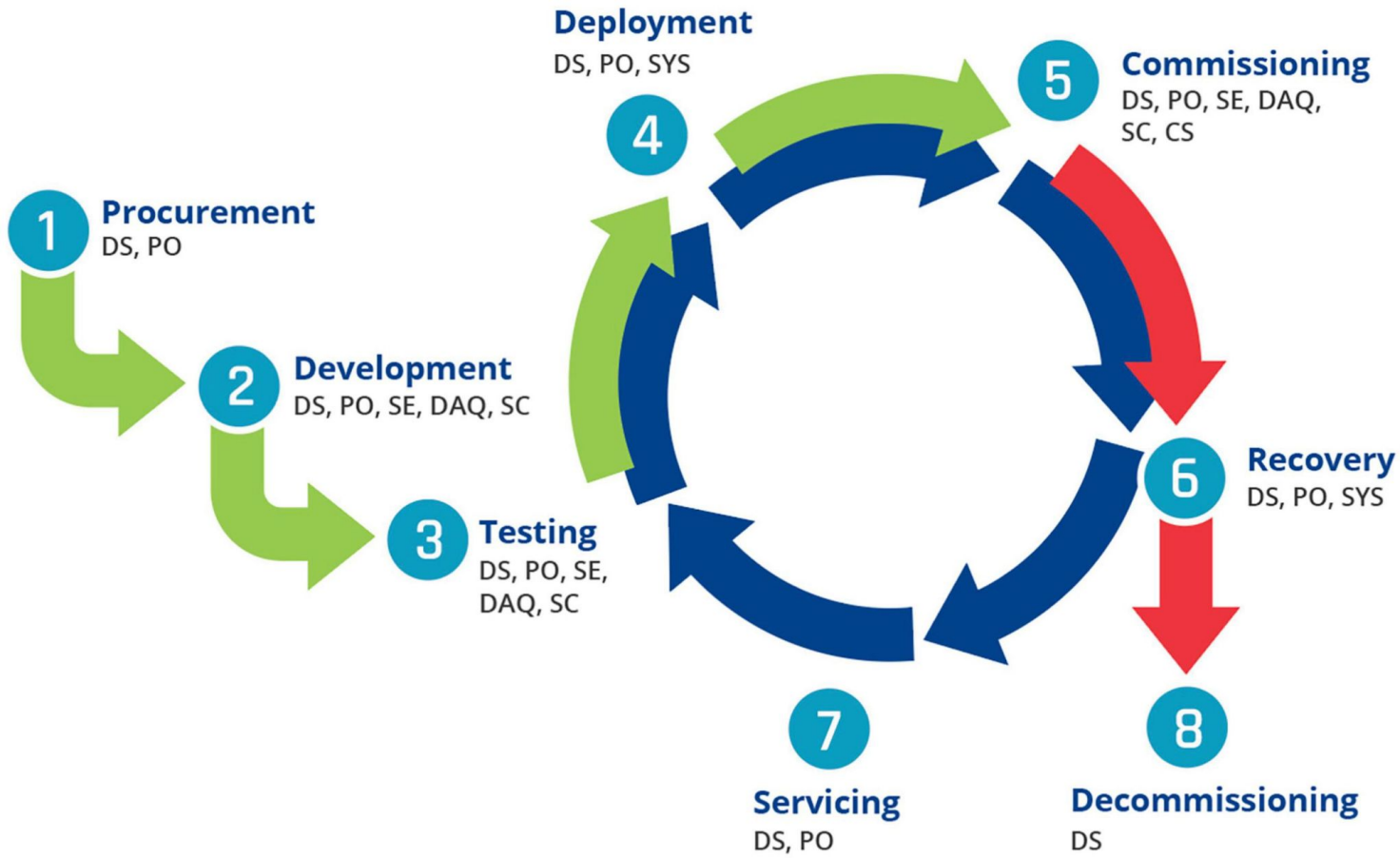
- Two-way flow of user information, dive & annotation data, and taxonomy and button set data
- Device and sensor data flowing from ship to shore system and data centre
- Expedition data flowing from data centre to ships and on-shore system

Lessons Learned: Monitoring Tools

- 24/7 monitoring and alerting with on-call staff
- We use Grafana:

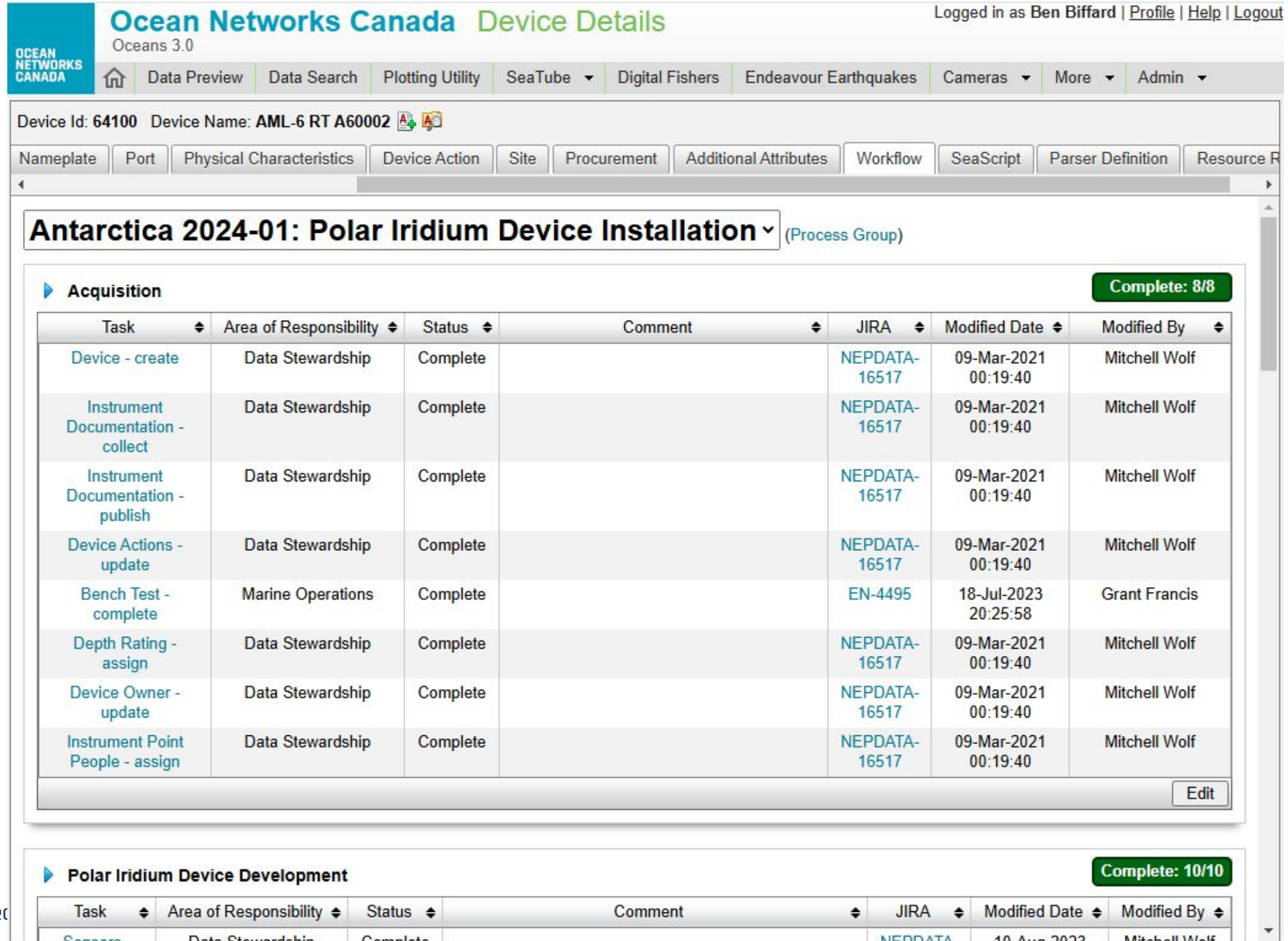


Lessons learned: instrument life cycle + quality control/assurance



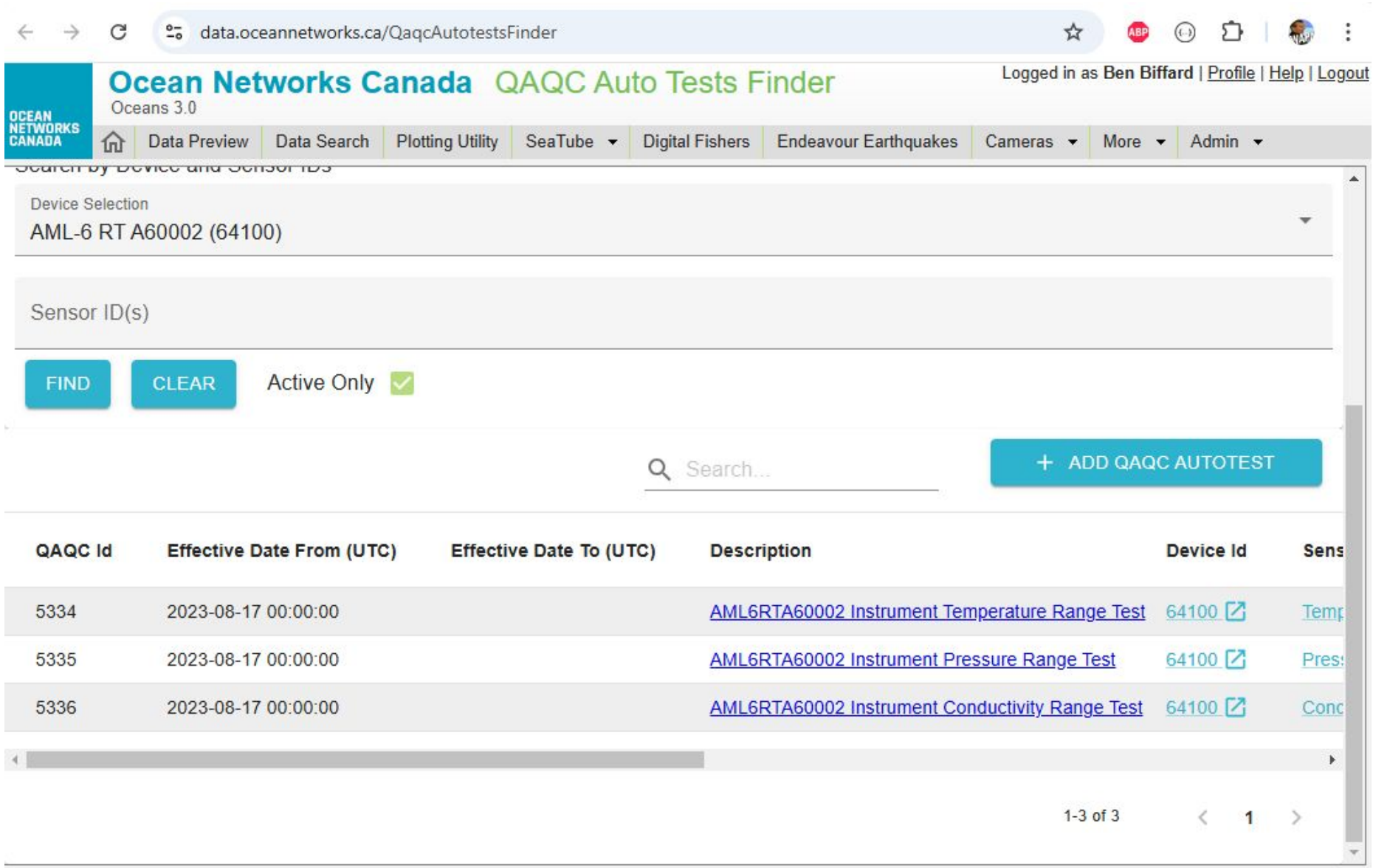
Workflow Tool

Oceans 3.0 supports the full instrument lifecycle



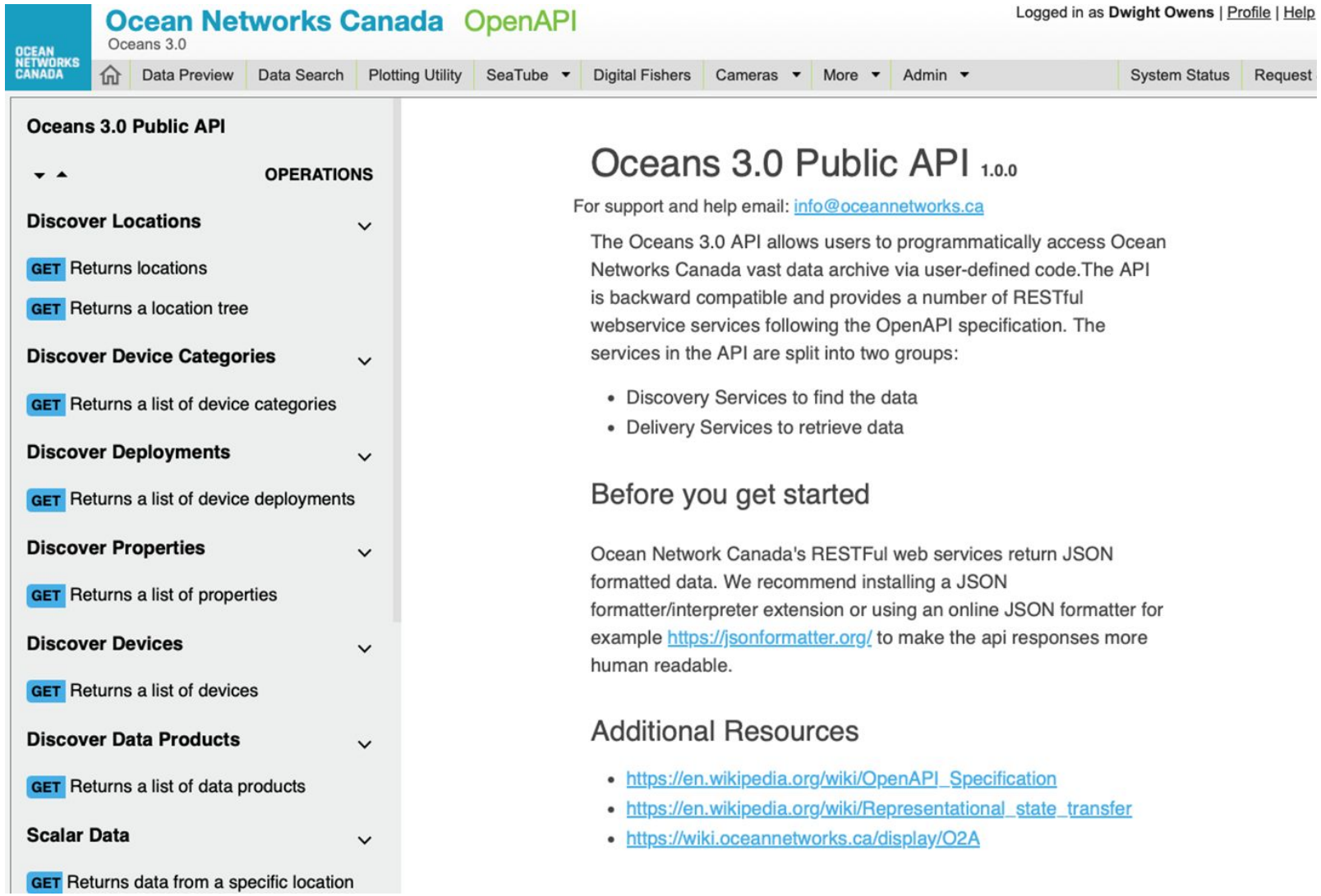
QAQC

QAQC tests and flags fully integrated



Data Distribution

- Foundational infrastructure needs to be designed and scalable to facilitate all phases of data acquisition, storage and distribution.
- Distribution includes video players, scrolling image viewers, data preview, search, plot and download.
  Best of all: **API →**
- Putting it all together is a game changer, enabling solutions such as earthquake early warning, tsunami warning, observations of climate change, biodiversity, etc.



For More Information

Frontiers in Marine Science publications:

- Canada's Internet-Connected Ocean

  https://www.frontiersin.org/articles/10.3389/fmars.2021.805134/full

- The Oceans 2.0/3.0 Data Management and Archival System

  https://www.frontiersin.org/articles/10.3389/fmars.2022.806452/full

https://www.oceannetworks.ca/

https://data.oceannetworks.ca/

bbiffard@oceannetworks.ca

OCEAN NETWORKS CANADA

Thank you
Questions?

University of Victoria