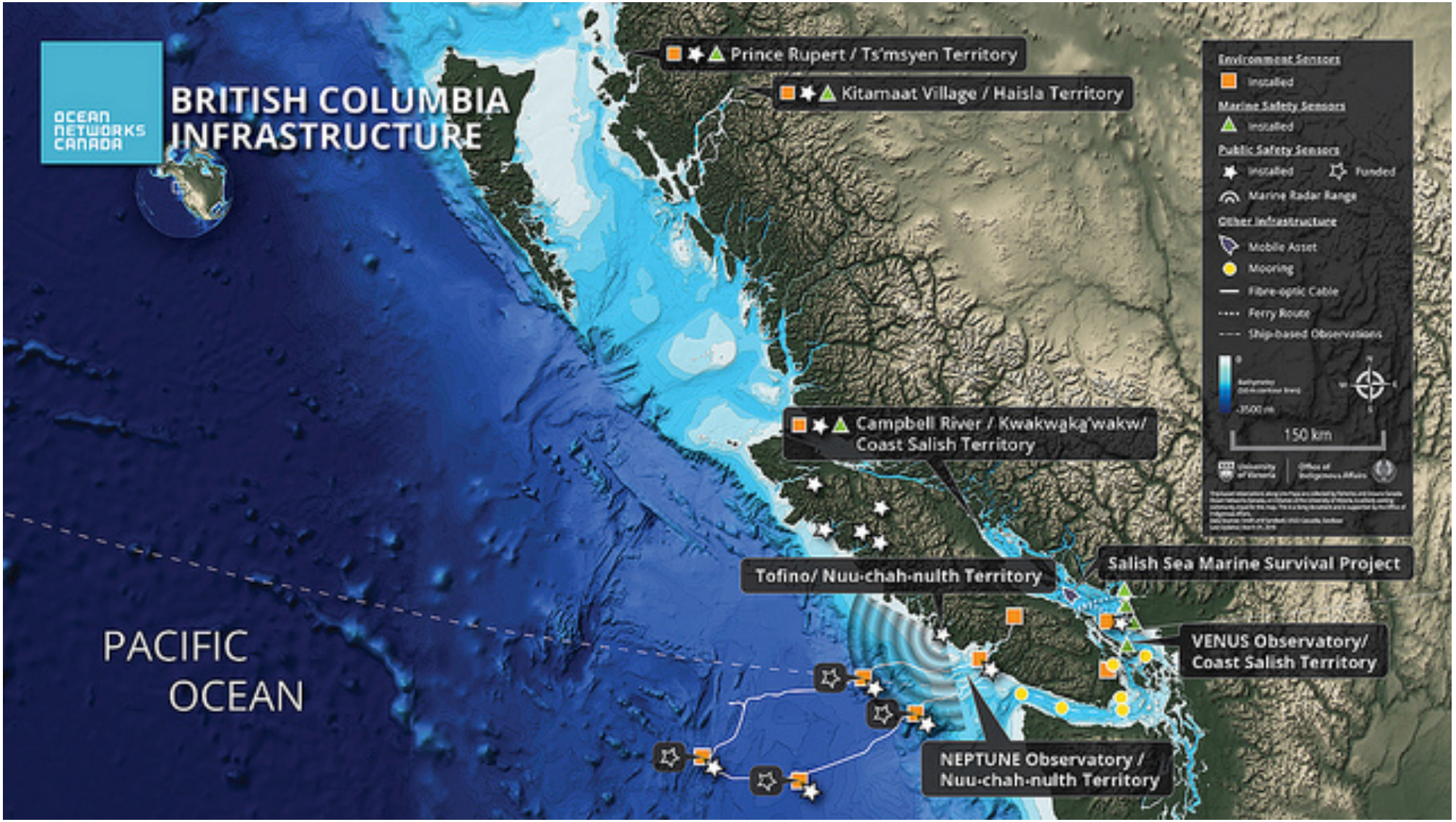
OCEAN NETWORKS CANADA

Making Identifiers Necessary to Track Evolving Data (MINTED) – A Brief Overview

Reyna Jenkyns

Ottawa, ON, 2019-01-24


OCEAN NETWORKS CANADA
BRITISH COLUMBIA INFRASTRUCTURE
Prince Rupert / Ts'msyen Territory
Kitamaat Village / Haisla Territory
Campbell River / Kwakwaka'wakw/ Coast Salish Territory
Tofino/ Nuu-chah-nulth Territory
Salish Sea Marine Survival Project
VENUS Observatory/ Coast Salish Territory
NEPTUNE Observatory / Nuu-chah-nulth Territory
PACIFIC OCEAN
Environment Sensors
Installed
Marine Safety Sensors
Installed
Public Safety Sensors
Installed
Funded
Marine Radar Range
Other Infrastructure
Mobile Asset
Mooring
Fibre-optic Cable
Ferry Route
Ship-based Observations
150 km

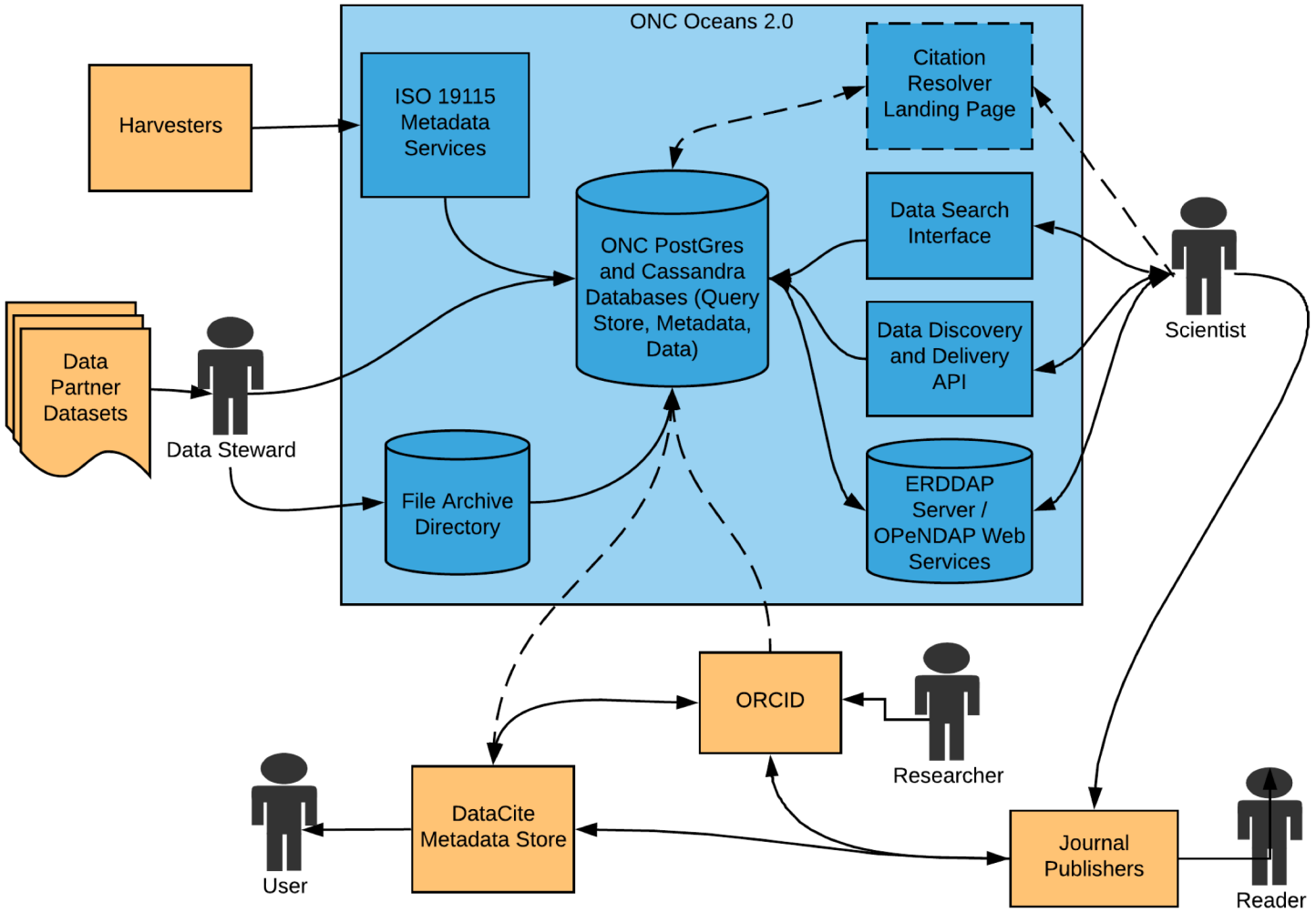
MINTED: Making Identifiers Necessary to Track Evolving Data

- Implement **dynamic data citations**, applying 14 recommendations from RDA Data Citations WG output which covers versioning, query store, resolver landing page, technology migration resilience
- Improve **provenance, versioning, and ISO 19115 metadata records** as they relate to data citation framework
- Utilize **DataCite Canada** membership for access to services for registering datasets
- Introduce ORCIDs for dataset contributors and user accounts, leveraging **ORCID-CA** frameworks and advice
- Deliver citation text provision service and a citation resolver services to **National Data Services Framework**
- As a member of the **World Data System** (WDS), adhere to the new **CoreTrustSeal data repository certification** requirement for Data Discovery and Identification (R13), such that users can discover and refer to data in a persistent way through proper citation
- Participate in **FREYA ambassadorship program**
- Consult with **RDA Data Versioning, Provenance Patterns** and new **Data Granularity Working Groups** for relevant expertise

MINTED: Making Identifiers Necessary to Track Evolving Data

- ONC data are very **dynamic** - continually accumulating data streams, data reprocessing and data product code versioning
- Highly **heterogeneous** data – fixed and mobile platforms, instrument types, data formats and processing levels, real-time vs autonomous
- many building blocks already exist (but more to go):
  - local identifiers and metadata for individual data queries,
  - software versioning,
  - metadata history tables,
  - reprocessing records,
  - archived file metadata (timestamping, history of changes –due to manual fixes or re-generation of derived data products, etc)
  - parser modification history
  - data agreement attributions (using ISO 19115:2014 terms) and restriction framework for third party data partners

MINTED: Making Identifiers Necessary to Track Evolving Data



System architecture description: The ONC Oceans 2.0 system (in blue), and third party sources and applications (in orange). Dotted lines indicate aspects that need to be added, while all ONC components would be modified. The ONC components can be directly controlled via the project, with expected modifications to include a new data model and tables within the database, additional web services, integration of third party APIs, and data citation features.

MINTED: FAIR Intersections

F1. (meta)data are assigned a globally unique and persistent identifier.

- Datasets would be assigned a DataCite DOI, in combination with a local query PID.
- Individual contributor attributions are associated with an ORCID

F3. metadata clearly and explicitly include the identifier of the data it describes.

- ISO 19115 metadata records that accompany the dataset would refer to the dataset DOI

A1. (meta)data are retrievable by their identifier using a standardized communications protocol.

- RDA WG recommendations detail the desired resolving features for a landing page and machine-readable capabilities.

A2. metadata are accessible, even when the data are no longer available

- if underlying data is no longer accessible in the same version, the resolver would provide explanations and redirect to the current version.

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

- ISO 19115 metadata records and non-proprietary formats (e.g., NetCDF) included

R1.2. (meta)data are associated with detailed provenance.

- Provenance that details the lineage of the data will be used to determine dataset versioning, which would be provided on the dataset resolving landing page and embedded within the ISO 19115 lineage section

OCEAN NETWORKS CANADA

THANK YOU!

Ocean Networks Canada is funded by the Canada Foundation for Innovation, Government of Canada, University of Victoria, Government of British Columbia, CANARIE, and IBM Canada.

@ocean_networks OceanNetworksCanada visit: **oceannetworks.ca**