Using the PICES TCODE catalog service

Igor Shevchenko

VNIRO (TINRO), Russia

S9: Applications of artificial intelligence to advance the understanding of North Pacific ecosystems

Data for AI/ML

- data architecture for AI/ML applications is different from traditional
- AI/ML works with large volumes of data represented and stored (e. g., in data lakes) not only in the relational (table) form
- with machine learning computers learn from data
- there is a strong synergy between code and used data

Data catalogs

- collections of metadata that helps users to find the data they actually need
- function in environments that provide authorization, editing, publishing, moderation, searching
- have to become core elements of modern data management

Metadata records

- features of some resource (data, information, service) that important for
    - discovery
    - understanding
    - preliminary evaluation
    - retrieving
    - use
    - management, etc.
- structured and may contain different fields
- include links to corresponding providers and lead eventually to seeking resources

Standards

- standards (represented as schemes) define syntax and semantics of metadata descriptions
- widely used
  - the ISO 19115/11139 standard for spatial resources (e.g., datasets, services, maps)
  - the Dublin Core scheme for referencing publications and reports,
  - the ISO 19110 standard for the feature cataloging

History of TCODE medadata sharing initiatives

- PICES Long Term Time Series
- Bering Sea Ecosystem Biophysical Metadatabase
- North Pacific Ecosystem Metadatabase
- PICES metadata federation project (PMFP) (Bernard A. Megrey and S. Allen Macklin, 2005) for preparing, publishing and searching metadata on marine ecosystems of the North Pacific
- PMFP was a great success in terms of number of collected records (it was funded, technical work done by paid staff) but not in terms of created infrastructure (US national FGDC standard and NSDI clearinghouse network)

Cataloging with open source technology

- at some point, the "multilingual" Geonetwork Opensource based on international standards became available
- after a positive evaluation of its features at TINRO, the Geonetwork software package was installed on a rented server as PTC
- when the US NSDI Clearinghouse Network ceased functioning, all collected metadata holdings eventually were automatically converted (with a loss of some content) and moved to the PICES TCODE metadata catalog (PTC)
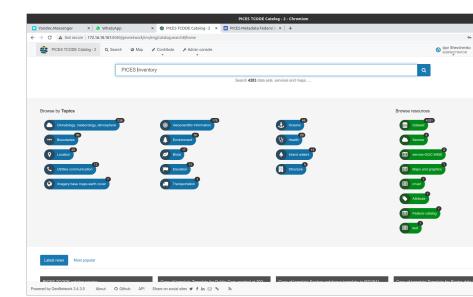
Catalog-2

- a service that allows running sub-catalogs by expert groups and individuals has been arranged
- PTC-2 (ver. 3.4, ISO 19115/19139, Dublin Core, lives on a virtual server in TINRO, `http://tcode.tinro.ru/geonetwork/`) became the primary source of collected metadata including lossy converted PMFP records from the FGDC standard
- original FGDC metadata records are available as XML files on the PMF site `https://sites.google.com/view/pices-metadata/home`
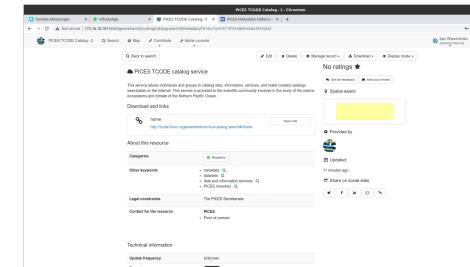
Creating and publishing metadata records

- life cycle typically goes through such states as Draft, Submitted, Approved, Retired
- a user chooses a metadata template and a group, and then fills out the chosen fields in screen forms
- records can be associated with different types of resources (files, weblinks, etc.)
- external categories (as, e.g., Datasets, Maps, PICES related datasets, etc.) may be assigned to metadata records (not parts of metadata)
- users may identify user groups and privileges (to view, to download, to edit, etc.) to metadata records and any attached data
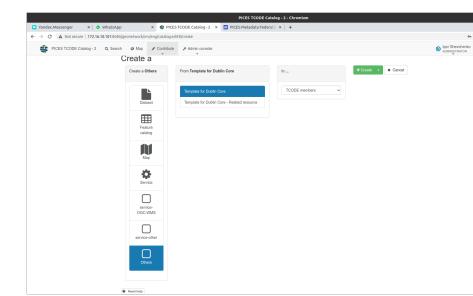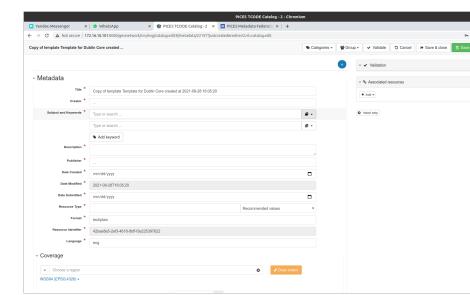
Homepage

Example of metadata record (depends on the standard)

Adding a record (depends on the standard)

Adding a record (depends on the standard)

Status quo

- PMF catalog and the catalog service are not in demand of the PICES community
- reasons are not specific to PICES
  - no adequate formal recognition of data products
  - no mandatory sharing policies
  - no an understandable and explicit legal basis regarding donors and users rights
  - no generally accepted ethical codes, etc.

Challenges

- new arising community (especially, young career scientists) who apply AI/ML could reverse the momentum by
    - adopting the catalog service as a required tool for data management of large volumes of reusable data that are usually processed
    - citing all data products properly (e.g., using links to the metadata catalog records)
    - participating in forming a data-sharing culture and establishing standards and norms (the legal and ethical codes), etc.

References

- Metadata Federation of PICES Member Countries. PICES Technical Report No. 1, Edited by Bernard A. Megrey, S. Allen Macklin, Kimberly Bahl, and P. Daniel Klawitter, 2007.
- PICES TCODE catalog - 2. URL: `http://tcode.tinro.ru/geonetwork/`
- Website: PICES TCODE catalog service. URL: `https://sites.google.com/site/picestcodegeonetwork/`

References (Contd.)

- Shevchenko, I., 2020. PICES TCODE catalog service. PICES Press, Vol. 28, No. 2, pp. 20-23. URL: `https://meetings.pices.int/publications/pices-press/volume28/PPJul2020.pdf`
- IOC/IODE Inter-sessional Working Group To propose a Strategy on Ocean Data and Information Stewardship for the UN Ocean Decade (IWG-SODIS). First Interim Report. June 2020. URL: `https://www.iode.org/index.php?option=com_oe&task=viewDocumentRecord&docID=27050`