Building a repository for freshwater quality data

Andrea Schweer
schweer@waikato.ac.nz

Information Technology and Services Division
The University of Waikato, New Zealand

Presented at the CAIRSS Community Days 2012
The data

• Freshwater quality measurement data
  • Lakes, rivers
  • Weather conditions, chemical data, biodata
• Range of collection methods
  • Automated buoys: samples every 15 minutes
  • Manual collection
• Manually processed
  • Error correction
  • Quality control information
• For a specific research project or as part of ongoing monitoring
The goals

- Publish datasets (funder mandate)
  - LERNZ project at the University of Waikato
  - Other researchers (universities / CRIs)
  - Regional government
  - Citizen science?
- Link datasets ↔ publications
  - Persistent identifiers
- Allow re-mixing of datasets
Our approach: Customised DSpace

Use what's there

- Repository interface
- Workflow

Customise where needed

- Submission steps
- Discovery
- Access
- Backend
Customised DSpace submission

Dataset submission

Choose type of dataset:
You can choose one type of dataset. Please create an individual submission for each type of data that you got.

Zooplankton

Select the variables in your dataset by dragging them from the left column to the right column or by double clicking on them. If you cannot find your variable in the list, please contact the LERNZdb administrator at and continue your submission at a later stage. All data that you have been provided so far has been saved already.

Available variables
Zooplankton presence/absence - Zo_PA

Selected variables
Zooplankton biomass - Zo_Bio
Zooplankton density - Zo_Den
Zooplankton relative abundance - Zo_Rel

Apache POI XSSF
Customised DSpace discovery & access

- Find datasets by
  - water body
  - time period
  - variables
  - quality control info
- XML file for metadata → structured MD
Looking ahead: extract datapoints

- Extract individual datapoints (rows) from dataset
- Normalise units of measurements
- Allow creation of re-mixed datasets
  - Treated just like first-generation datasets?
Open questions

- Machine interfaces
  - OAI-PMH / OAI-ORE
  - Geospatial data standards?
- Controlled vocabularies
- Datapoint backend
  - Storage
  - Querying
- Licensing
- Roles & workflows
- Advocacy
- Build it & they'll come?
LERNZ Water Quality Database

This database contains water quality data for lakes in New Zealand. It is part of the Lake Ecosystem Restoration New Zealand (LERNZ) Project run by the Centre for Biodiversity and Ecology Research at the University of Waikato.

Organisations

Select an organisation to browse water quality data supplied by this organisation.

- Regional councils [1]
- The University of Waikato [1]

Search

Enter some text in the box below to search the water quality database.

Recently Added

- Zooplankton BOP
  Unknown author (2012-04-12)
- Bay of Plenty 1990-2010
  Unknown author (2012-04-12)