

Re-using DSpace to build a repository for freshwater quality data

Andrea Schweer, Stefan Mutter, Deniz Özkundakci
{schweer, stefanm, denizo}@waikato.ac.nz

Information Technology & Services Division
Centre for Biodiversity and Ecology Research

The University of Waikato, New Zealand



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

The screenshot shows the LERNZ Water Quality Database website. The header features a landscape image of a lake with a boat, the LERNZ logo, and the text 'Water Quality Database'. Below the header is a navigation bar with 'LERNZdb Home' and a search box. The main content area is divided into several sections: 'Browse' with a sidebar menu, 'My Account' with a login link, 'Discover' with a list of lakes, 'Search' with a search box and 'Go' button, and 'Recently Added' with two entries: 'Zooplankton BOP' and 'Bay of Plenty 1990-2010'.

Browse

Everywhere

- Organisations
- By Date Added
- Lakes
- Time Periods
- Variable

My Account

Login

Discover

Lake

- Lake Okareka (2)
- Lake Okaro (2)
- Lake Rotoehu (2)
- Lake Okataina (1)
- Lake Rerewhakaaitu (1)
- Lake Rotoiti (1)
- Lake Rotokakahi (1)
- Lake Rotomahana (1)
- Lake Rotoma (1)
- Lake Rotorua (1)
- ... View More

Time Period

- 1990-2010 (1)
- 2006-2008 (1)

Date Added

- 2012 (2)

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)

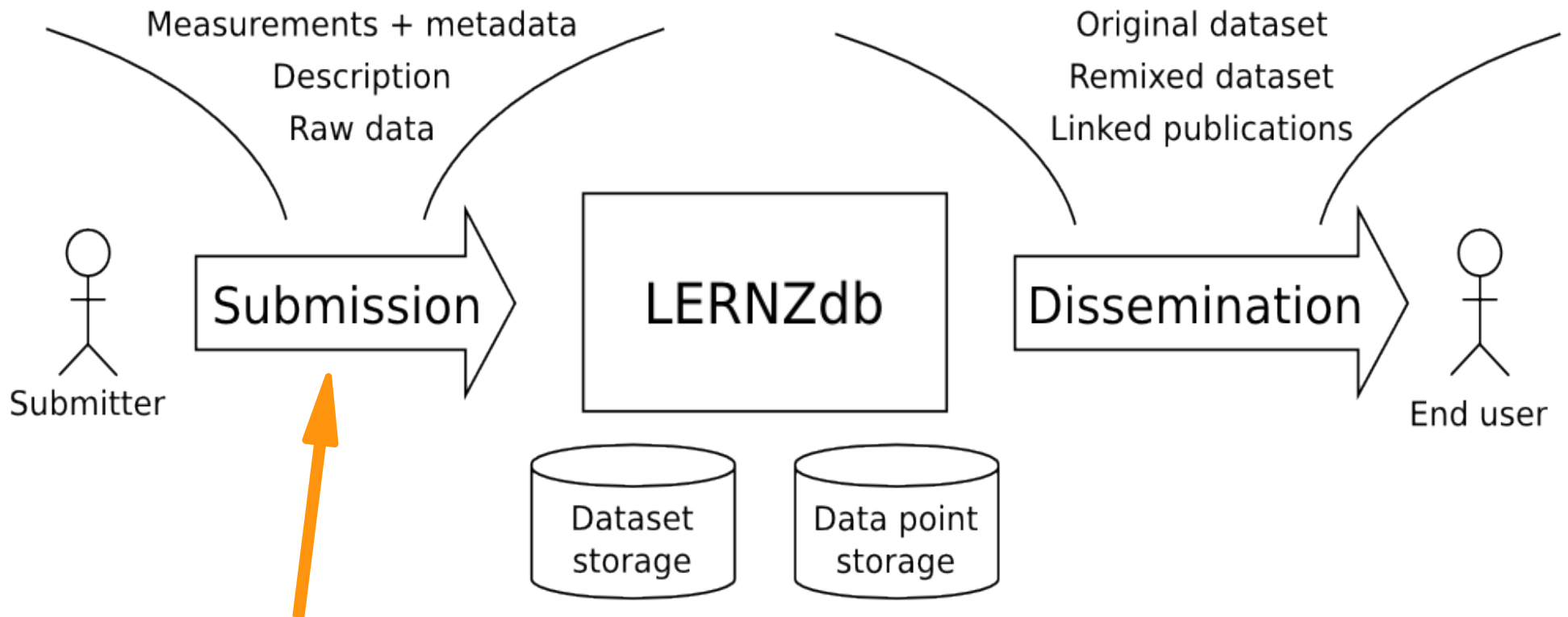
Use what's there

- Repository interface
- Workflow

Customise where needed

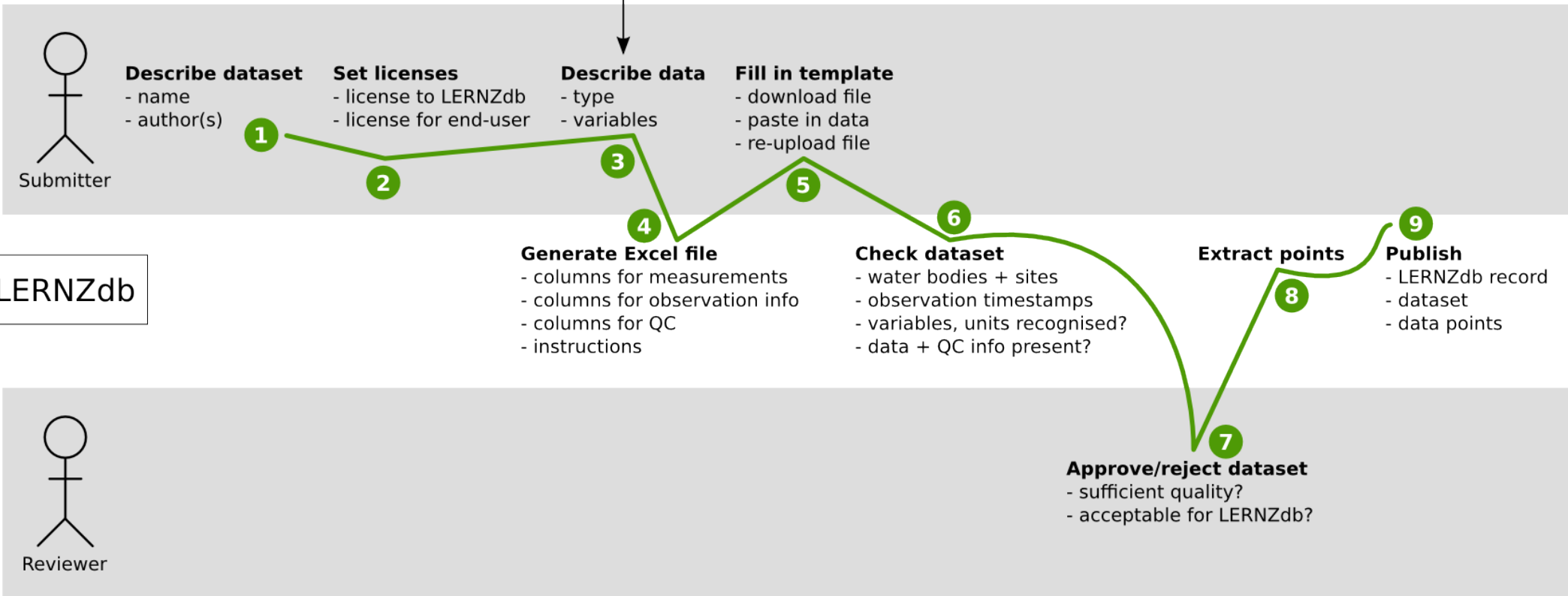
- Submission steps
- Workflow actions
- Backend
- Discovery & Access

Customised submission steps



Data vocabulary

- 9 dataset types
- 156 variables
- 44 QC types



Submitter

Describe dataset

- name
- author(s)

1

Set licenses

- license to LERNZdb
- license for end-user

2

Describe data

- type
- variables

3

Fill in template

- download file
- paste in data
- re-upload file

5

Generate Excel file

- columns for measurements
- columns for observation info
- columns for QC
- instructions

4

Check dataset

- water bodies + sites
- observation timestamps
- variables, units recognised?
- data + QC info present?

6

Extract points

Publish

- LERNZdb record
- dataset
- data points

8

9

Approve/reject dataset

- sufficient quality?
- acceptable for LERNZdb?

7

LERNZdb



Reviewer

Generating the template

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

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 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 →

Code	Description	Preferred unit
Zo_Bio	Zooplankton biomass	mg/L
Zo_Den	Zooplankton density	ind/L
Zo_Rel	Zooplankton relative abundance	%

Waterbody type ID	Waterbody name	Region	Site ID/Site name	Easting	Northing	Coordinate system	Depth from	Depth to	Depth unit	Variable	Value	Unit	Date	Time	Date and Time	Sp
-------------------	----------------	--------	-------------------	---------	----------	-------------------	------------	----------	------------	----------	-------	------	------	------	---------------	----

Summary:

There are 3 datapoints in 3 rows.

This dataset is concerned with lakes.

The category is: Nutrients.

The preliminary overall LERNZdb quality rating for your dataset is: ☆☆☆

Recommendations to improve the rating:

To get to ☆☆☆

- Please check your variables as some of them are unknown to the system. If you think, they should be added to the LERNZdb standard set, then please contact a system administrator once you have completed your submission.
- Please fill in more than 50% of the standard quality control information for your category (currently at 50%).

To get to ☆☆☆

- Check your bodies of water so that all of them can be recognised by LERNZdb.
- Please use the preferred unit as shown on the cover sheet of your template.
- Please fill in more than 80% of the standard quality control information for your category (currently at 50%).

To get to ☆☆☆

- Please fill in all of the standard quality control information for your category (currently at 50%).

Details:

Known variables in category Nutrients

- Ammonium - NH4 (found 1 time)
- Total nitrogen - TN (found 1 time)

Unknown variables

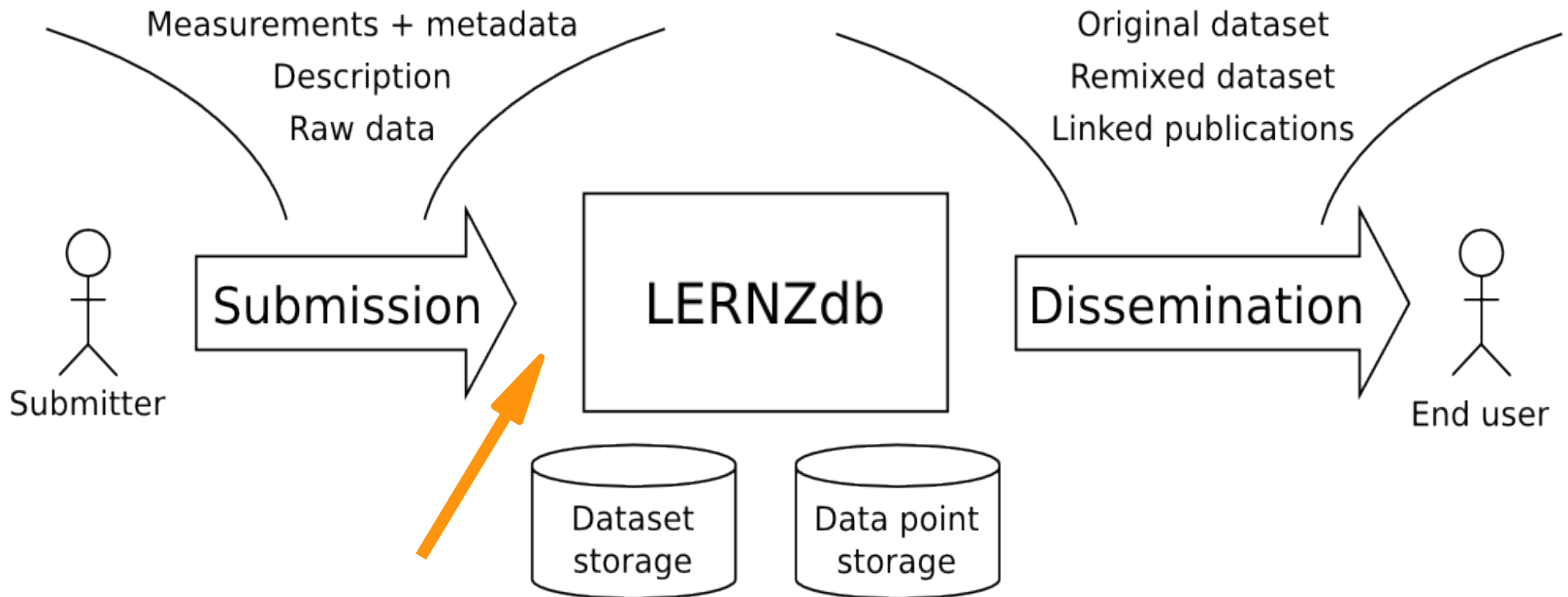
- Ammonum (found 1 time)

Timestamps

- 3 timestamps recognised.
- Earliest point in time recorded: Thu Jul 11 00:00:00 NZST 2013
- Latest point in time recorded: Thu Jul 11 00:00:00 NZST 2013

Recognised Lakes

Customised workflow actions



Extending the standard DSpace workflow

- Add extra information
 - Show rating + reasoning to reviewer
- Add extra actions
 - Allow reviewer to deactivate some of the checks
- Potentially remove some default actions

Actions you may perform on this task:

If you have reviewed the dataset and it is suitable for inclusion in LERNZdb, select "Approve".

Approve dataset

If you have reviewed the dataset and found it is **not** suitable for inclusion in LERNZdb, select "Reject". You will then be asked to enter a message indicating why the dataset is unsuitable, and whether the submitter should change something and resubmit.

Reject dataset

Select this option to change the dataset's metadata.

Edit metadata

Return the task to the pool so that another user may perform the task.

Return task to pool

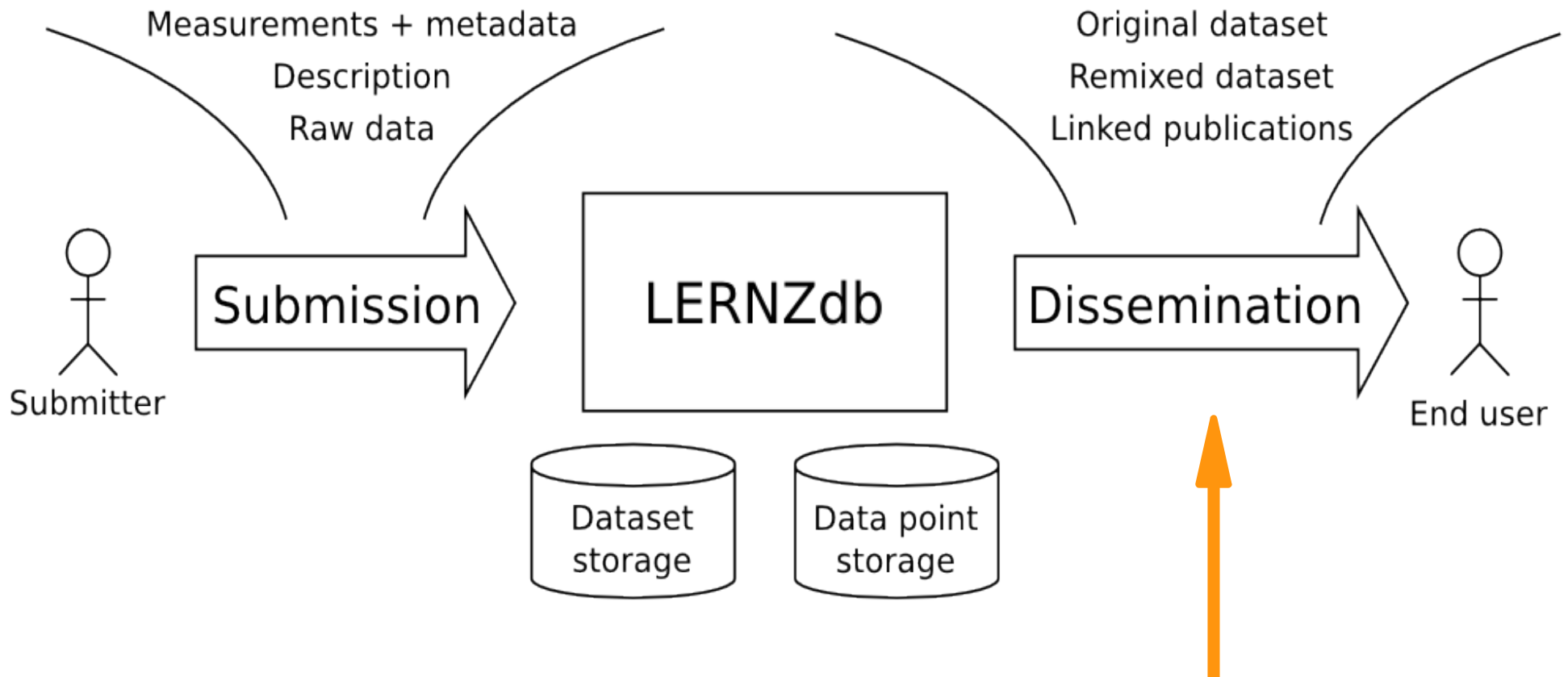
Cancel

Customised item representation

- DSpace item: metadata + bitstreams
- LERNZdb item:
 - Structured metadata in XML bitstream
 - Subset of XML bitstream copied to DSpace metadata
 - in custom namespace
 - Other bitstreams
 - Excel file
 - Raw data files (optional)

```
▼ <variables>
  ▼ <variable>
    <id/>
    <name>Ammonum</name>
    <present-for>1</present-for>
    ▼ <units>
      <unit>mg/L</unit>
    </units>
  </variable>
  ▼ <variable>
    <id>3</id>
    <name>NH4</name>
    <present-for>1</present-for>
    ▼ <units>
      <unit>mg/L</unit>
    </units>
  </variable>
  ▼ <variable>
    <id>5</id>
    <name>TN</name>
    <present-for>1</present-for>
    ▼ <units>
      <unit>mg</unit>
    </units>
  </variable>
</variables>
▼ <qc-columns>
  ▼ <qc-column>
    <id>9</id>
    <name>Analysed by</name>
    <present-for>3</present-for>
  </qc-column>
  ▼ <qc-column>
    <id>2</id>
```

Customised access & dissemination



Customised discovery & access

- Find datasets by
 - water body, time period
 - variables, quality control info
- Stored in item metadata → can just configure Discovery, DSpace browse as needed
- Item pages
 - XMLUI → already set up for XSL processing
 - can include information from metadata XML

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

- Cross-linking datasets & publications
- Datapoint backend
 - Storage
 - Querying
- “Real” controlled vocabularies
- Machine interfaces
 - OAI-PMH / OAI-ORE
 - Geospatial data standards?
- “Soft” issues

LERNZ Water Quality Database
Lake Ecosystem Restoration
New Zealand

LERNZdb Home Search

Browse
Everywhere
Organisations
By Date Added
Lakes
Time Periods
Variable

My Account
Login

Discover
Lake
Lake Okareka (2)
Lake Okaro (2)
Lake Rotoehu (2)
Lake Okataina (1)
Lake Rerewhakaaitu (1)
Lake Rotoiti (1)
Lake Rotokakahi (1)
Lake Rotomahana (1)
Lake Rotoma (1)
Lake Rotorua (1)
... View More

Time Period
1990-2010 (1)
2006-2008 (1)

Date Added
2012 (2)

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.

Go

Recently Added

Zooplankton BOP
Unknown author (2012-04-12)

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

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

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



Andrea Schweer
schweer@waikato.ac.nz

Pick and Mix CC 2.0 BY-NC-SA © 2008 by Ben Cooper
<http://www.flickr.com/photos/cycleologist/2896137842/>