

WSL ENVIDAT USER GROUP MEETING: EAWAG RESEARCH DATA

— CURRENT STATUS AND CHALLENGES —



download
PDF (869 kB)



Harald von Waldow <harald.vonwaldow@eawag.ch> – 2018-03-13

1. **technical infrastructure development:**
ERIC-internal, ERIC-public, ERIC-control
(ERIC = *Eawag Research Data Institutional Repository*)
2. **support activities:**
consulting w.r.t. "open data", DMP, general data management
3. **other tasks:**
policy development, outreach

INFRASTRUCTURE

SOFTWARE **D**EVELOPMENT

&

SYSTEM ADMINISTRATION

- Integrated **user-feedback** from public beta
- Added some necessary **features** (private data, publication package workflow)
- Connected to **11 TB** fast storage
- SOLR index in Postgres DB in **separate VM**

Improved support for publication data packages:

What are packages?

Packages are the primary units of contents in the Eawag Research Data Platform.

A package is a collection of

1. **resources** (files or links) and
2. **metadata**, such as abstract, keywords, author, ...

Packages are what users find when searching for data.

Packages have a permanent URL.

1 Create package

2 Add data

Publication data package?

YES

i

DORA-Id or DOI, if available:

http://doi.org/10.1016/j.envsoft.2017.03.002

Check

i

Open Data?

YES

i

Embargo:

i

* Title: A descriptive title

i

* URL: data.eawag.ch/dataset/<dataset>

Edit

Improved representation of the organizational structure:

The screenshot shows the profile page for the organization 'Systems Analysis, Integrated Assessment and Modelling (SIAM)' on the eawag Research Data Platform. The page has a dark teal header with a breadcrumb trail: Home / Organizations / SIAM. Below the header, the organization's name is repeated in a smaller font, followed by a 'Follow' button. To the right of the name, there are statistics for 'Followers' (0) and 'Packages' (11). The main content area features a 'Description' section with the text 'This is SIAM' and a 'Data Manager' section with a profile picture of Peter Reichert, his email address, and the number of packages uploaded (8). At the bottom, there is a list of topics associated with the organization: 'Systems Analysis, Integrated Assessment and Modelling', 'Ecological Modelling', 'Hydrological Modelling', and 'Material Fluxes in the Anthroposphere'. A 'Manage' button is located in the top right corner of the main content area.

Home / Organizations / SIAM

systems-analysis-integrated-assessment-and-modelling

Systems Analysis, Integrated Assessment and Modelling

This is SIAM [read more](#)

Followers Packages

0 11

[Follow](#)

Systems Analysis, Integrated Assessment and Modelling (SIAM)

Description
This is SIAM

Data Manager

 [Reichert, Peter](#) <Peter.Reichert@eawag.ch>
Packages uploaded: 8

- **Systems Analysis, Integrated Assessment and Modelling**
 - Ecological Modelling
 - Hydrological Modelling
 - Material Fluxes in the Anthroposphere

Improved representation of the organizational structure:

The screenshot displays the 'Systems Analysis, Integrated Assessment and Modelling' organization page on the eawag Research Data Platform. The page is divided into a left sidebar and a main content area. The sidebar contains the organization's name and a link to 'This is SIAM'. The main content area features a header with 'Edit', 'Packages', and 'Members' buttons, and a 'View' button. Below the header, the organization's details are listed in a form: Name, URL, Short Name, Description, Homepage, Data Manager, and Image URL. The URL and Short Name fields are highlighted in yellow. The Description field contains the text 'This is SIAM' and a note about Markdown formatting. The Data Manager field is a dropdown menu showing 'reichert'. The Image URL field shows a long alphanumeric string and a 'Remove' button.

systems-analysis-integrated-assessment-and-modelling

Systems Analysis, Integrated Assessment and Modelling

This is SIAM [read more](#)

Edit Packages Members View

* Name: Systems Analysis, Integrated Assessment and Modelling

* URL: data.eawag.ch/organization/systems-analysis-integrated-assessment-and-modelling Edit

Short Name: SIAM

* Description: This is SIAM

You can use [Markdown formatting](#) here

* Homepage: http://www.eawag.ch/de/abteilung/siam/

* Data Manager: reichert

Image URL: 2016-02-19-153607.500239systems-analysis-integrated-assessment Remove

- Separate **stripped-down** (no user-management and package upload) CKAN instance.
- Separate (logical) storage: Exploit **FS-level deduplication**.
- Separate Index and DB (**next step**).
- In the **DMZ**.
- Consider it **hacked**.

ERIC-CONTROL

DOI-MINTING

PUBLIC-INTERNAL-MANAGEMENT

DORA INTEGRATION

- **DOI minting works** semi-automatically (fill a JSON-template)
There will be always a manual step (correctness check)
A workflow augmented with a GUI needs to be established in the long run.
- **public-internal-management**: Not even started.
I envision an independent micro-service.
Start with CLI-scripts.

DORA integration is not yet finalized

SUPPORT CONSULTANCY & SERVICES

- Storage procurement:

We want to move 5 TB from external HDs to somewhere

=> develop IT-policy for shared storage access, develop intra-IT workflow, act as first contact for storage management.

- SNSF Data Management Plan: Individual reviews, SNSF DMP Guide, workshops.
- "Open Data" data journals, data publication strategy

- ERIC support & workshops
- ERIC system administration
- data package quality control:
calculate ~1 hr per data package
- DOI management:
 - Still quite manual
 - currently involves Zenodo or other public repository

OTHER OUTREACH & POLICY DEVELOPMENT

- Development of an Eawag data archival policy

hvw/ 2018-01-04

Eawag directive on the archival of research data

1 Preamble

Eawag recognizes the fundamental importance of research data in maintaining quality research and scientific integrity, and is committed to pursuing the highest standards. data management standards [we should say highest standards regarding what, I suggest "data management", an alternative could be "research integrity"] Eawag acknowledges that correctly and easily retrievable research data are the foundation of and integral to most researchs projects. They are necessary for the verification and defence of most research processes and results. Research data have a long-term value for research and academia, with the potential for widespread use in society.

This directive is based on and concretizes the regulations regarding the collection, documentation and archival of research data set forth in the Guidelines for Good Scientific Practice¹.

QUESTIONS?

DATA MANAGEMENT



QUESTIONS?

DATA MANAGEMENT



QUESTIONS?

DATA MANAGEMENT



QUESTIONS?

DATA MANAGEMENT

