

~~WSL ENVIDAT USER~~ **STEERING** GROUP MEETING: EAWAG RESEARCH DATA

— CURRENT STATUS AND CHALLENGES —



download
PDF (869 kB)



Harald von Waldow <harald.vonwaldow@eawag.ch> – ~~2018-03-13~~ 2018-03-15

OVERVIEW

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 DEVELOPMENT & 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

DORA-Id or DOI, if available:

Open Data?
YES

Embargo:

* Title:

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

Improved representation of the organizational structure:

The screenshot shows the profile page for '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 left sidebar contains the organization's name, a brief description, and statistics: 0 followers and 11 packages, with a green 'Follow' button. The main content area features tabs for 'Packages', 'Activity Stream', and 'About', along with a 'Manage' button. The title 'Systems Analysis, Integrated Assessment and Modelling (SIAM)' is prominently displayed with a share icon. Below this, the 'Description' section states 'This is SIAM'. The 'Data Manager' section includes a profile picture of Peter Reichert, his email address, and a note that 8 packages have been uploaded. At the bottom, a bulleted list outlines the organization's research focus areas.

Home / Organizations / SIAM

systems-analysis-integrated-assessment-and-modelling

Systems Analysis, Integrated Assessment and Modelling

This is SIAM [read more](#)

Followers: 0 Packages: 11

[Follow](#)

Packages [Activity Stream](#) [About](#) [Manage](#)

Systems Analysis, Integrated Assessment and Modelling (SIAM) [Share](#)

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
 - Mathematical Methods
 - Soil, Groundwater, Catchments
 - Systems Analysis and Water Management
 - Water, Environment and Food Security

Improved representation of the organizational structure:

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

Parent: None - top level

Delete Save Organization

TODO

- creators record (ORCID, affiliation)
- clean up resources metadata-input (automation)
- auto-generate cryptographic hash / MIME-type / size
- display content of zip-archives
- bulk download (as zip archive)
- "restricted records"
- interactive georeferencing
- multi-file upload in web-GUI
- streamline metadata-entry / complete documentation
- improved keyword parsing (comma as separator)
- complete organization display / initial datamangers

- 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**.
- Establish that quick & ugly

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**: postponed indefinitely.

DOI minting



- Currently being worked on.
- Set of tools that works
 - a) as CLI scripts (used now),
 - b) can be used as modules for webapp.
- ```
doigenerator.py, mkdcxml.py, daciapi.py, doicontrol.py
```
- Our DOI prefix is **10.25678**
- We have our first DOI (currently hosted at Zenodo): <https://doi.org/10.25678/000011>

# 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, ...
- General data management: data bases, workflow, project planning, file formats, software, ...

- 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 Practice<sup>1</sup>.

- Maintenance of Eawag research data management website:  
[opendata.eawag.ch](https://opendata.eawag.ch)



# QUESTIONS?

## DATA MANAGEMENT

