SND Doris and Sunet Drive
FAIR and sovereign data publication in a federated world

Stefan Jakobsson, Swedish National Data Service

Richard Freitag, Sunet

SND
Swedish National Data Service | University of Gothenburg - Chalmers University of Technology - Karolinska Institutet - KTH Royal Institute of Technology - Lund University - Stockholm University - Swedish University of Agricultural Sciences - Umeå University - Uppsala University

Swedish National Data Service (SND)

- A national infrastructure to support the accessibility, preservation, and reuse of research data and related materials

- Make FAIR research data accessible, nationally as well as internationally

- A consortium of nine universities with University of Gothenburg as host university

- A network of 24 additional higher education institutions and four research organizations

- Funded by the Swedish Research Council

The SND Data Organisation and Information System
https://doris.snd.gu.se
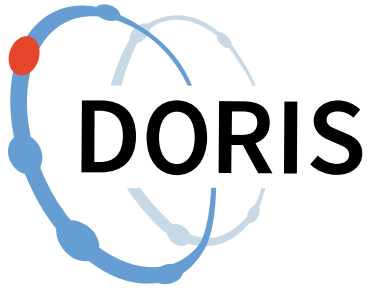
- Research data repository (≈ Zenodo) developed and operated by SND, primarily for use by Swedish universities and other Swedish research data producers

- Fairly rich metadata model inspired by DDI, DataCite and others, strong focus on controlled vocabularies

- Supports nine metadata profiles, each adapted to a research area, e.g. Social Sciences

- Assigns DataCite DOI:s to datasets and supports versioning

- Publishes to SND's data catalogue (and from there metadata to others)

DORIS
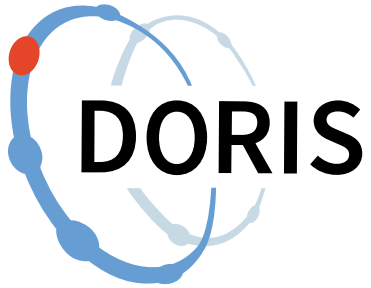
The SND Data Organisation and Information System (cont.)

- Curated repository - datasets are submitted to and reviewed by the universities' Data Access Unit (DAU) and/or SND for quality assurance, to avoid legal issues etc.

- Supports distributed data storage, enabling the research principal (university) to keep data under its governance in its own storage solution

- Datasets can be published as open access or with restricted access

- Restricted datasets can be requested via SND's data catalogue (an assessment process is conducted by the data owner before the dataset can be released)

Test Dataset

2020-2-1

Research area profile

| General | ↺ Change |

Sections ⌃

1. Files and accessibility ›

2. Citation and description ›

3. Administrative information ›

4. Collection and method ›

5. Topics and keywords ›

6. Geographic coverage ›

7. Publications and relations ›

Activities and settings ⌄

Help resources

ⓘ **Legend**

Do you have any questions or comments?

Contact SND

2. Citation and description

Citation

APA ▾

Citation suggestion is generated automatically in a number of formats..

Hegerfors, A. (2024). Test Dataset (Version 1) [Data set]. Swedish National Data Service. DOI: Available after publication.

Title ●

Enter a title that will be used in the dataset citation. It is important that the title refers to the content of the dataset, in contrast to an article title where the focus is the conclusions that the article reaches. If you are sharing data from a specific article you can clarify this as follows: "Data for: [Article's title]"

If it is difficult to translate the title of your dataset to Swedish, you can enter the English title in both fields.

Testdataset

Test Dataset

Alternative title

Enter an alternative title for your dataset, e.g. an abbreviation, the title in a third language, or another established name for the data collection.

Test

Creator/Principal investigator ●

Those who are responsible for the data material and intellectual content of the data. Creator/principal investigator can be a person(s) *or* organization(s), and only one of them needs to be specified.

Arne Hegerfors ✛ ⌄ 🗑

Save
Saved: 2024-03-08 10:27

👁 Preview catalogue entry

✈ Submit

🗑 Delete dataset

Other actions ⌄

Latest activity ⌃

✎ **Editing**
2020-05-19

View detailed activity log ›

Effects on the food-web structure and bioaccumulation patterns of organic contaminants in a climate-altered Bothnian Sea mesocosms

SND-ID: 2023-193. **Version**: 1. **DOI**: https://doi.org/10.5878/p8sx-n445

Download data

- Biological data.txt *(29.54 KB)*
- Fish.txt *(851 bytes)*
- PAR.txt *(2.52 KB)*
- POP_fish.txt *(3.99 KB)*
- POP_water.txt *(3.85 KB)*

Associated documentation

- readme-file.txt *(32.74 KB)*

Download all files

- 2023-193-1.zip *(~73.48 KB)*

Citation

Berglund, Å., Gallampois, C., Ripszam, M., Larsson, H., Figueroa, D., Grinienė, E., Byström, P., Gorokhova, E., Andersson, A., & Tysklind, M. (2024). Effects on the food-web structure and bioaccumulation patterns of organic contaminants in a climate-altered Bothnian Sea mesocosms (Version 1) [Data set]. Umeå University. Available at: https://doi.org/10.5878/p8sx-n445

Choose citation format: American Psychological Association [Copy] [Download]

Creator/Principal investigator(s)

Åsa Berglund - Umeå University, Department of Ecology and environmental science

Christine Gallampois - Umeå University, Department of Chemistry

Matyas Ripszam - Umeå University, Department of Chemistry

Henrik Larsson - Umeå University, Umeå Marine Sciences Centre (UMF)

Versions
Version 1. 2024-02-28

Homepage
The strategic research programme EcoChange

Contact for questions about the data
Åsa Berglund
asa.berglund@umu.se

Download metadata
DataCite
DDI 2.5
DDI 3.3
DCAT-AP-SE 2.0
JSON
JSON-LD
PDF
Citation (CLS)
File overview (CSV)

Sciebo RDS (Research Data Services)

- Sunet Drive will use Sciebo RDS as interface for publishing research data

- Allows researchers to export research data directly from EFSS systems like Owncloud and Nextcloud to data repositories like Zenodo and OSF

- Possible to develop custom **Connectors** for exporting to data repositories (by implementing a set of HTTP endpoints)

3 **Choose a repository.**

Choose the repository you want to publish your data to.

Doris connector

OSF
Open Science Framework

zenodo
Zenodo

DORIS RDS Connector for Sunet Drive

- SND/Sunet is developing a connector for Sunet Drive + Doris (using .NET)

- Files are not sent to Doris but published within Sunet Drive, which enables research data to reside under the governance of the university

- Metadata publication, assigning DataCite DOI, handling access to data etc. is handled by Doris/SND

- User selects files via RDS to be published as a dataset, and selects Doris as destination repository
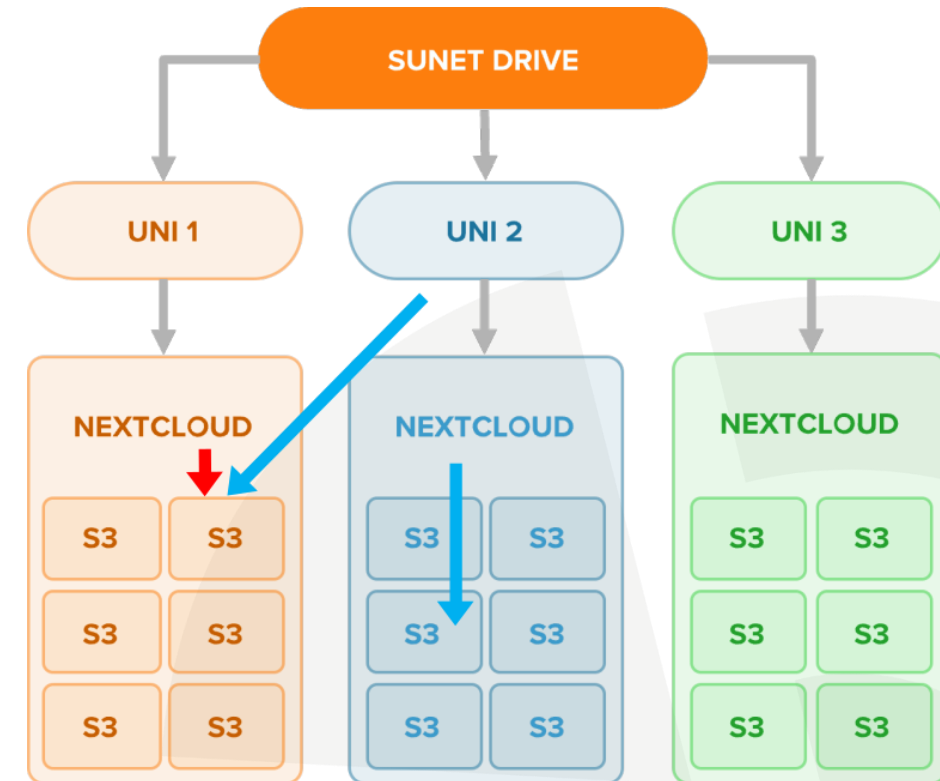
3 ACTIVE PROJECTS

/testdataset	draft
testar	

/testdataset	draft
Profile	

/DataUploadTest	draft
cat	

+ NEW PROJECT

Home

Active Projects

Published Projects

Repositories

Help

Sciebo RDS is still in its pilot phase. Please report problems to: mail@test-nextcloud.sunet.dev

Settings

1 Configuration — 2 Metadata — 3 Publish

testar

1 **Select your project folder.**

PICK   a folder to add metadata to and publish. This folder will be loaded from your sciebo account.

Currently selected: /testdataset

2 **Name your project.**

Give your project a name to help yourself to distinguish it from other projects. This name is only used internally and will not be published.

testar

Test

3 **Choose a repository.**

You need to connect a **Data Repository** to continue.
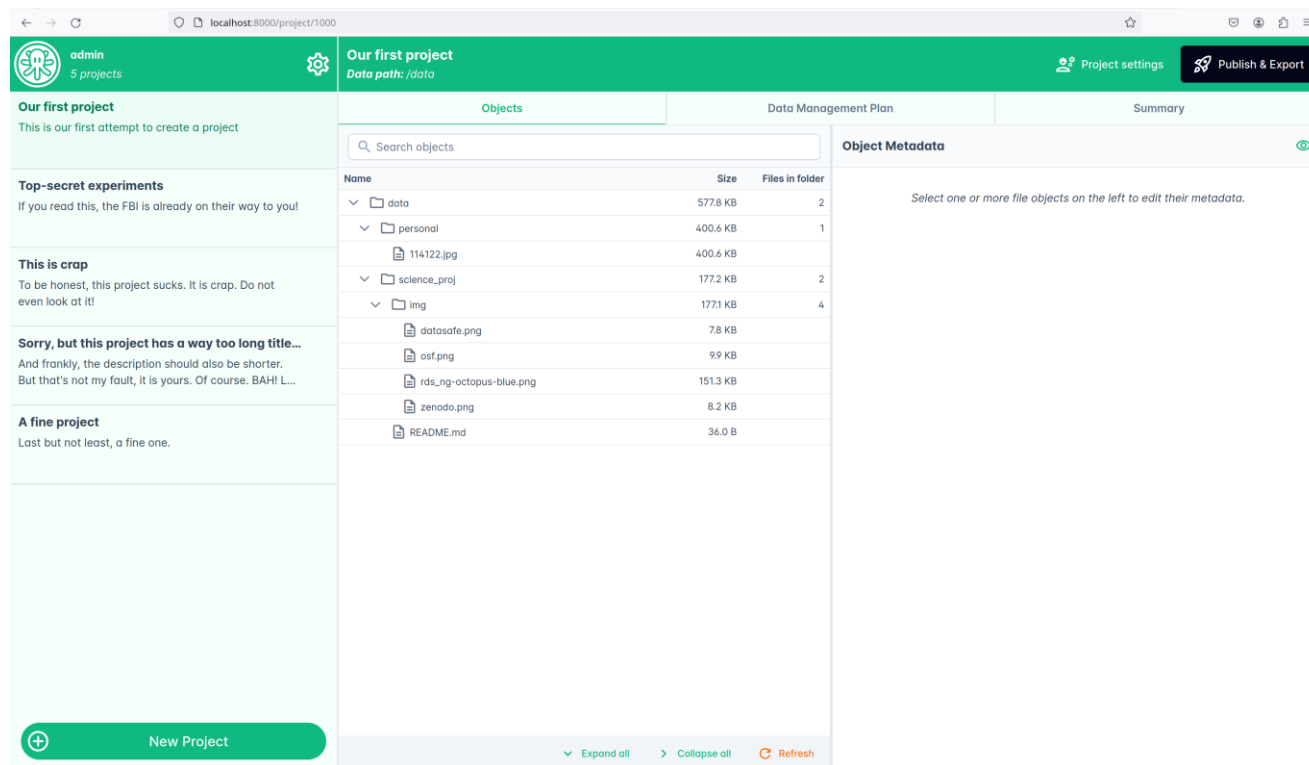
DORIS

Doris

DELETE

CONTIN

RDS-NG – Next Generation RDS

- Sunet Drive currently uses "legacy RDS"

- RDS-NG will be supported once it has been released

- Doris-connector will be ported to RDS-NG

DORIS RDS Connector (cont.)

- Files are uploaded to a dedicated research data publication area within Sunet Drive

- SHA256 checksums are calculated and stored alongside the files in a BagIt manifest

- NextCloud share link is created and protected with random password

- RO-Crate metadata describing the set of files is sent to a SND/Doris API endpoint (metadata includes checksums and share link URL)

- The user can now create a dataset in Doris and connect it to the set of selected files, enrich it with metadata, and submit for publishing

Test Dataset

2024-1

Research area profile

General | ↺ Change

Sections ⌃

1. Files and accessibility ›

2. Citation and description ›

3. Administrative information ›

4. Collection and method ›

5. Topics and keywords ›

6. Geographic coverage ›

7. Publications and relations ›

Activities and settings ⌄

Help resources

ⓘ **Legend**

Do you have any questions or comments?

Contact SND

1. Files and accessibility

Connect external files

Test Dataset - 2 files (4284 bytes) ✕ ▾

Alternate names:	• Test
Creators:	• xfors1@gu.se (forskare1.forskarsson@example.se)
	• forskare2.forskarsson@example.se
	• xfors3@gu.se (forskare3.forskarsson@example.se)
Total file count:	2
Total file size:	4284
Created at:	2024-03-08 11:47
Last modified:	2024-03-08 11:47

Data files

File name	
📄 data.csv	4.2 kB

Documentation files

File name	
📄 docs/info.txt	42 Bytes

💾 **Save** *
Saved: 2024-03-08 11:49

👁 Preview catalogue entry

➤ Submit

🗑 Delete dataset

Other actions ⌄

Latest activity ⌃

✏ **Editing**
2024-03-08

View detailed activity log ›

DORIS RDS Connector (cont.)

- University's Data Acess Unit (DAU) employees have read only access to the data publication area in order to review data

- If the dataset is accepted, suitable access level is selected (open or restricted) and the dataset is published via Doris (making it findable in SND's data catalogue)

- If dataset is published with open access, an administrator removes the password from the share link, making data available by direct download

- Datasets with restricted access can be requested via SND's data catalogue

Design considerations and challenges

- Publishing in RDS is a one-shot operation, any changes to files (e.g. due to review feedback) must be published as a new RDS project and selected again in Doris

- Complicated deployment of RDS (next generation RDS will simplify this), hard to run development environment locally

Possible future plans

- Avoid manual steps when publishing

- Add DOI and other metadata to SUNET Drive when publishing and construct complete BagIt structure

- Add support for *documentation files*, i.e. dataset files that are always openly accessible regardless of overall dataset access level