

SciLifeLab Data Management services and support

Discover services for data-driven life science

2024-05-28

Angela Fuentes Pardo, PhD

Data Steward
SciLifeLab Data Centre

Elin Kronander, PhD

Data Steward
NBIS

SciLifeLab Data Management

E-mail: data-management@scilifelab.se





“Data is one of the most valuable and long-lasting outputs of scientific research”

SciLifeLab Data Management support



- Collaborative activity between **SciLifeLab Data Centre** and the **National Bioinformatics Infrastructure Sweden (NBIS)**
- About 21 members working in different aspects of **Research Data Management (RDM)**



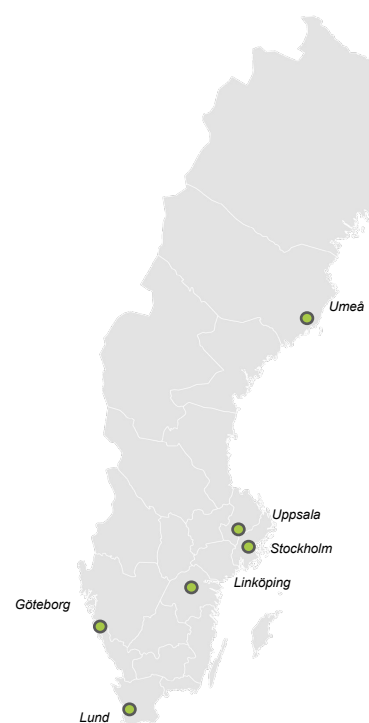
What is SciLifeLab Data Management?



Data Centre

- A multi-site, central unit within SciLifeLab established in 2016, with the responsibility for **IT-** and **data management services**
- DC develops and operates services for data publishing, data management, hosting of applications and AI models, Open Science and FAIR

<https://www.scilifelab.se/data>



Data Management team at SciLifeLab

Data Centre



Hanna Kultima
Data Manager



Anna Asklöf
Data steward



Parul Tewatia
Data steward



Katarina Öjefors Stark
Data steward



Angela
Fuentes Pardo
Data steward



Natashia
Benzian Olsson
Data steward



Joanna
Sendeka
Data steward



Elisabeth
Sundström
Project coordinator



Chris Erdmann
Head of
Open Science

What is SciLifeLab Data Management?



Data Centre

- A multi-site, central unit within SciLifeLab established in 2016, with the responsibility for **IT- and data management services**
- DC develops and operates services for data publishing, data management, hosting of applications and AI models, Open Science and FAIR

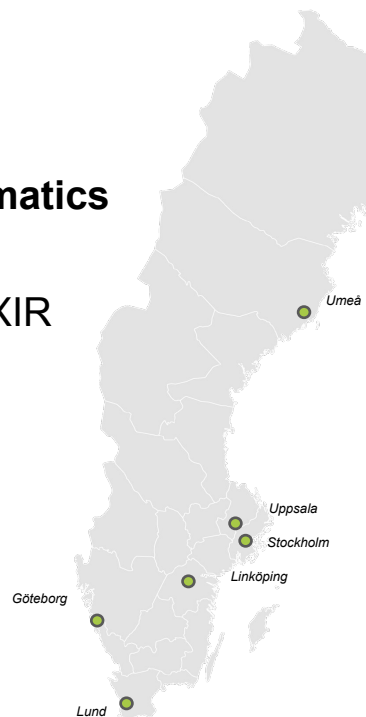
<https://www.scilifelab.se/data>



- NBIS is a consortium of all major universities in Sweden
- NBIS is the **SciLifeLab bioinformatics platform**
- NBIS is the Swedish node in ELIXIR the European infrastructure for biological data



<https://nbis.se/>



Data Management team at



Richelle Björvang
Data Steward



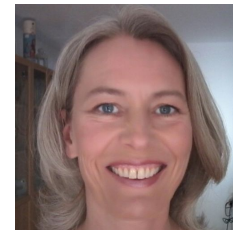
Markus Englund
Data Steward



Karin Granström
Data Steward



Erik Hedman
Data Steward



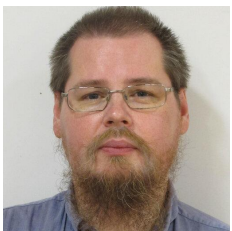
Yvonne Kallberg
Data Steward



Elin Kronander
Data Steward



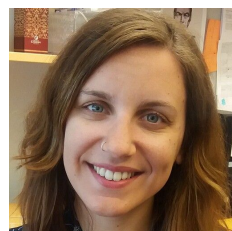
Stephan Nylinder
Data Steward



Mattias Strömberg
Data Steward



Elisavet Torstensson
Data Steward



Maria Vrettou
Data Steward



Wolmar N. Åkerström
Data Steward



Niclas Jareborg
Data Manager

SciLifeLab Data Management Support

Goals

- to **maximise the value** of life science **data**
- to promote and spread information about the **FAIR principles, Open Science** and **good data management practices**
- to **offer services, tools and support** for research data management, IT and data sharing
- to make **training** in data management, Open Science etc., easily available

We provide support to, and collaborate with both researchers and infrastructures active in life science in Sweden, during all phases of the data life cycle

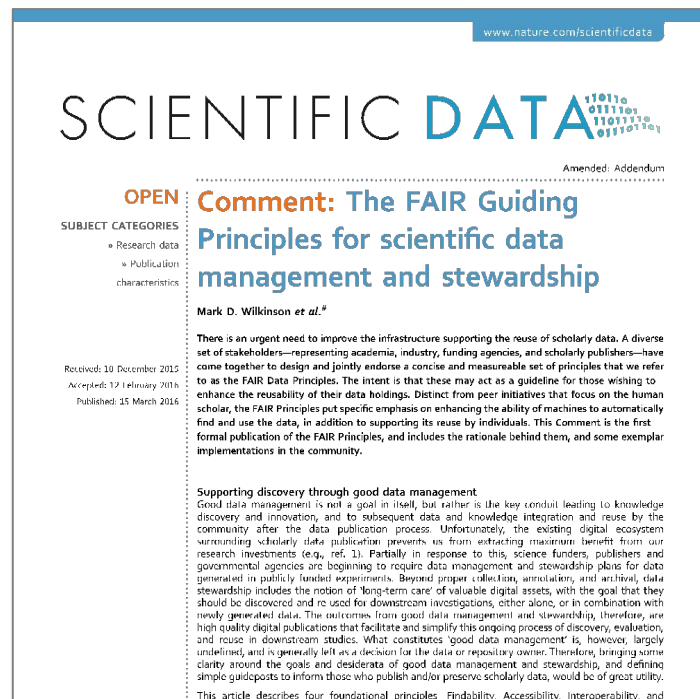


The FAIR principles

- Promote **efficient data discovery** and **reuse** by providing **guidelines to make data available** to (you) and others in the digital age (machine-actionability)

- ◆ Findable
- ◆ Accessible
- ◆ Interoperable
- ◆ Reusable

“To be useful for others, data should be FAIR for both, machines and humans”



Wilkinson et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. doi:10.1038/sdata.2016.18





Findable

- The first step in (re)using data is to find them. **Metadata** and **data** should be **easy to find for both humans and computers**. Machine-readable metadata are essential for automatic discovery of datasets and services





F

indable



A

ccessible

- The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services
- Once the user finds the required data, she/he/they need to **know how they can be accessed**, possibly including authentication and authorisation





Findable

- The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services



Accessible

- Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.



Interoperable

- The data can be **easily integrated with other data**. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing





Findable

- The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services



Accessible

- Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.



Interoperable

- The data can be easily integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing



Reusable

- The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, **metadata and data should be well-described** so that they **can be replicated** and/or **combined in different settings**



The FAIR Guiding Principles

*To be **Findable**:*

- F1.** (meta)data are assigned a globally unique and persistent identifier
- F2.** data are described with rich metadata (defined by R1 below)
- F3.** metadata clearly and explicitly include the identifier of the data it describes
- F4.** (meta)data are registered or indexed in a searchable resource

*To be **Accessible**:*

- A1.** (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1** the protocol is open, free, and universally implementable
 - A1.2** the protocol allows for an authentication and authorization procedure, where necessary
- A2.** metadata are accessible, even when the data are no longer available

*To be **Interoperable**:*

- I1.** (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2.** (meta)data use vocabularies that follow FAIR principles
- I3.** (meta)data include qualified references to other (meta)data

*To be **Reusable**:*

- R1.** meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1.** (meta)data are released with a clear and accessible data usage license
 - R1.2.** (meta)data are associated with detailed provenance
 - R1.3.** (meta)data meet domain-relevant community standards

Policy landscape of the FAIR principles

- UNESCO's recommendation on Open Science
- EU Commission
 - Open Science Policy
 - Directives
 - European Research Council
 - Horizon Europe
 - EOSC
- Swedish research bills 2016 & 2020
 - Transition to open research data implemented by 2026
 - Government assignments to KB & VR
- SUHF national roadmap for open science
- University policies
- Lund Declaration on Maximising the Benefits of Research Data
- *National guidelines for open science* (KB)

Open Science & FAIR

*“FAIR [...] open data sharing
should become the default [...]”*

*“As open as possible,
as closed as necessary”*



VR - Swedish Research Council

Vision and guiding principles for open access to research data

research process. Already existing data that have only been used in their original form and that are already managed and made accessible by another actor are not covered by this recommendation.

Metadata should also be published with open access

Both research data and data describing research data (known as metadata) should be published with open access. If there are obstacles to publishing research data, the focus should in the first instance be on making metadata openly accessible on the internet. In this way, users can find information on what research data exists, even when there are obstacles to open publication, for example lack of a suitable publication platform or technical limitations that prevent all data from being published.

Publication according to the FAIR principles

Publication of research data can be done using various digital platforms, for example via the higher education institution where the research is conducted or via other relevant national and/or international portals, infrastructures and similar organisations and platforms. The publication of research data shall always be based on the FAIR principles.

The Swedish Research Council's recommendation on data management according to FAIR

The Swedish Research Council recommends that the research data produced through research are managed according to the FAIR principles, clarified via the criteria developed by the Swedish Research Council to achieve FAIR data.

The FAIR principles should be implemented taking into account applicable legislation, and, as far as is possible and applicable, based on the technical, organisational and/or discipline-specific preconditions that apply.

The recommendations relates in the first instance to research data (and metadata) financed by public funds that can be published with open access, but the application of the FAIR principles can be made broader than this, and be used also for research data that cannot be published entirely openly. The recommendation on data management according to FAIR is overarching, and aims to create a common starting point for the implementation of FAIR data management.

*[...] The publication of research data **shall** always be based on the FAIR principles.[...]*

The Swedish Research Council's recommendation on data management according to FAIR

*The Swedish Research Council recommends that the research data produced through research are managed according to the FAIR principles, clarified via the **criteria developed by the Swedish Research Council to achieve FAIR data.** [...]*

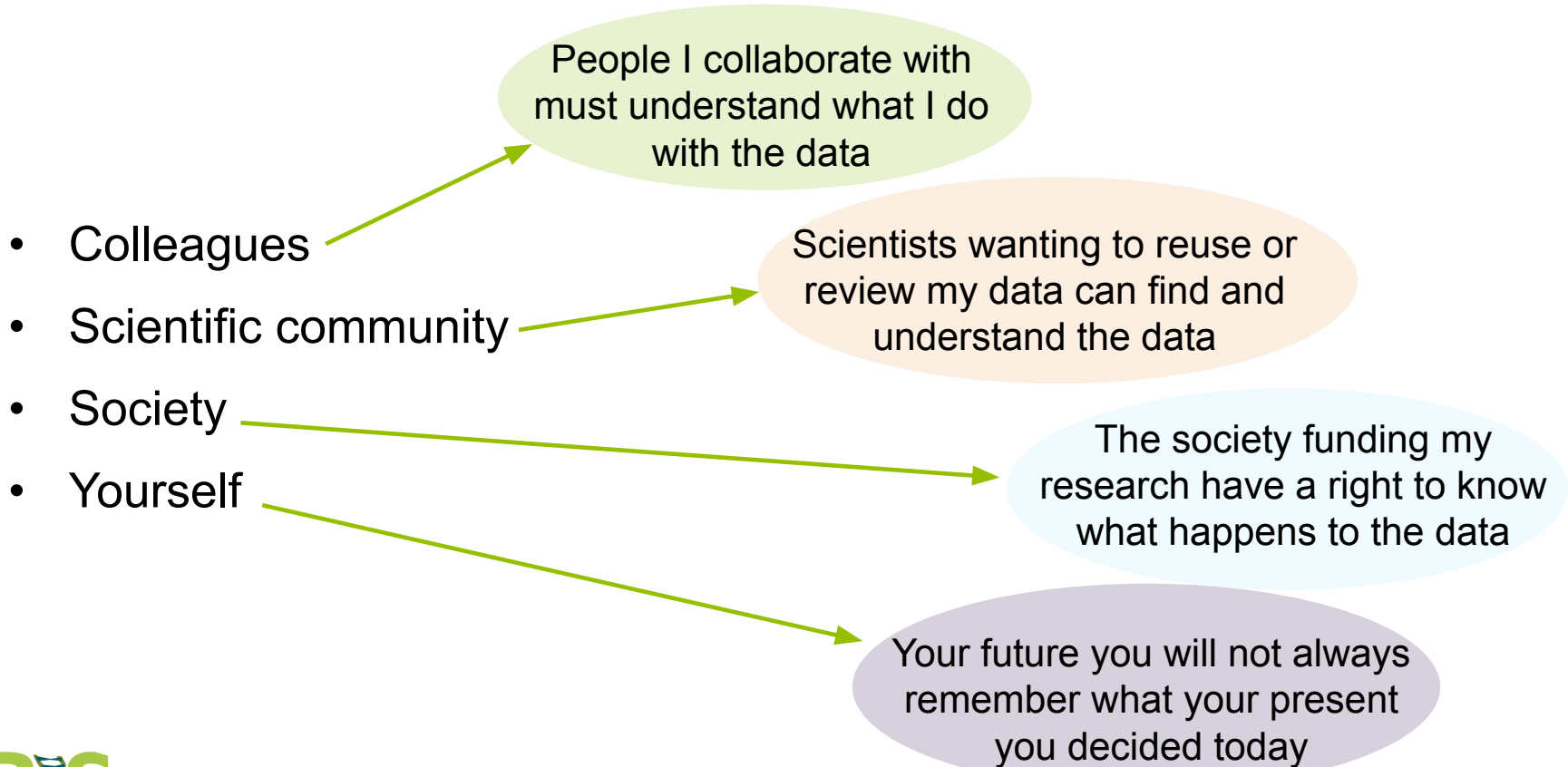


A FAIR data lifecycle

- Making data FAIR relies on **good data management practices** in all phases of research
 - Research documentation
 - Data organisation
 - Information security
 - Ethics and legislation



Data management recipients





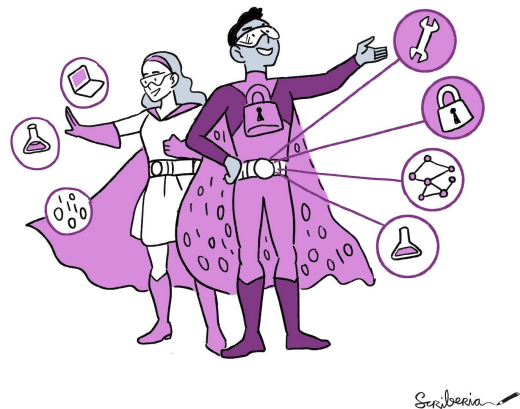
“Your primary collaborator is yourself six months from now, and your past self doesn’t answer emails”

-Rachael Ainsworth, astrophysicist at the University of Manchester, UK



How to implement FAIR in my research?

1. Make a plan
2. Submit data to a repository
3. Get trained
4. Let us help you!

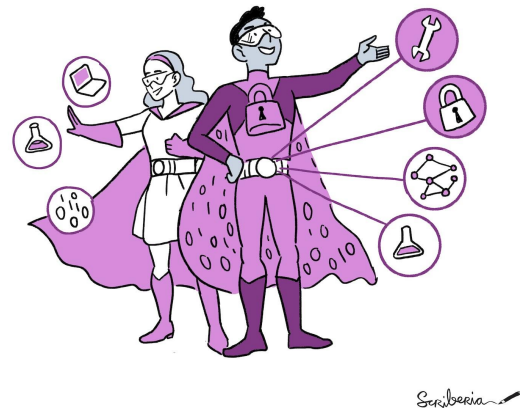


Contact us

- data-management@scilifelab.se
- <https://nbis.se/services/guidance-on-data-handling/apply>

Data management services

- Guide writing a data management plan
- Identify a suitable repository for publishing your data
- Assist during the submission process when publishing your data and code
- Advice on what needs to be done when working with sensitive human data
- Advice on describing data with proper metadata for documentation and publishing
- Data transfers, data organisation, backup, and security procedures



Contact us

- data-management@scilifelab.se
- <https://nbis.se/services/guidance-on-data-handling/apply>

Considerations

- **FAIR data \neq Open data**

Data can be OPEN without being FAIR

Data can be FAIR without being open

“As open as possible, as closed as necessary”

- **FAIR is not a binary measure**

Data can be more or less FAIR

- **There are many types of data**

Not all data is digital



RDM Guidelines

a knowledge hub for the management of life science research data in Sweden

data-guidelines.scilifelab.se/

SciLifeLab RDM Guidelines

Knowledge hub for the management of life science research data in Sweden

The purpose of these guidelines is to serve as an information resource for life science researchers in Sweden on handling research data management.

Research data management (RDM) concerns the organisation, storage, preservation, and sharing of data that is collected or analysed during a research project. Proper planning and management of research data will make project management easier and more efficient while projects are being performed. It also facilitates sharing and allows others to validate or reuse the data.

analyse, preserve, share and reuse. It is an essential part of the research process to get an overview of the management of the research data life cycle, including information on relevant resources and training material.

RDM life cycle from RDM life cycle model - Research Commission, Adapted from RDM life cycle model

Topics

Click on one of the links below to get an overview of individual research data management topics.

Get support About Contact

Home Research data life cycle Topics

Get Support

Do you need support with research data management?

We offer support to anyone involved in life science research that is affiliated with a Swedish university or research institute.

[Click here to get support](#)

Meet a Data Steward

Join SciLifeLab Data Centre & NBIS and get data management support. Each event consists of a 15 minutes mini-lecture and a 45 minutes Q&A.

Next date: Sep 19, 2023

[More information on the event page](#)

Events & Training

Upcoming conferences, webinars, workshops, and training opportunities in Sweden related to data-driven life science can be found on the SciLifeLab Data Platform.

[Go to Events & Training on the Data Platform](#)



NBIS
NATIONAL BIOINFORMATICS
INFRASTRUCTURE SWEDEN

 **SciLifeLab**

Data Centre

The purpose of these guidelines is to serve as an information resource to life science researchers in Sweden regarding Research Data Management (RDM).

Research data life cycle

Click on a section of the wheel below to get an introduction to that phase of the research data life cycle, including information on relevant resources and training material.



RDM life cycle from RDMkit licensed under [Creative Commons Attribution 4.0 International License](#).

Topics

Click on either of the links below to get an overview of individual research data management topics.

Working with human data

Data transfer

FAIR principles

Metadata

Ask us anything

Do you have a question or need support with research data management?

We offer support to anyone involved in life science research that is affiliated with a Swedish university or research institute

[Go to support page](#)

Training resources

Find resources concerning data management in form of training, guidance and tools

[Go to resources page](#)

Need more?

For more data-driven resources visit the SciLifeLab Data Platform



data-guidelines.scilifelab.se

Contact

[Home](#) / [Contact](#)

The SciLifeLab RDM guidelines are maintained by the SciLifeLab Data Centre and NBIS. We welcome suggestions, contributions, and questions related to the content of this site. This could include, for example, a suggestion for a [topic](#) that is not currently covered. Alternatively, you can request an update or correction to particular pages.

You are also more than welcome to contact us if you want support regarding data management in your project, e.g. regarding writing data management plans, publishing data, working with sensitive human data, or organising and documenting data.

You can fill in the below form to get in touch with us. If you prefer though, you can instead email us at data-management@scilifelab.se.

Contact form

Your name:

Your email:

Type of query:

Message:

data-management@scilifelab.se

[Data life cycle](#)[Your role](#)[Your domain](#)[Your tasks](#)[Tool assembly](#)[All tools and resources](#)[All training resources](#)

Are you working with data in the Life Sciences? Do you feel overwhelmed when you think about Research Data Management?

The ELIXIR Research Data Management Kit (RDMkit) is an online guide containing good data management practices applicable to research projects from the beginning to the end. Developed and managed by people who work every day with life science data, the RDMkit has guidelines, information, and pointers to help you with problems throughout the data's life cycle. RDMkit supports FAIR data — Findable, Accessible, Interoperable and Reusable — by-design, from the first steps of data management planning to the final steps of depositing data in public archives.

The RDMkit organises information into the six sections displayed below, which are interconnected but can be browsed independently.

Data life cycle

Start here to get an overview of research data management. Click on a section of the diagram below to get an introduction to that stage of the data management life cycle.



Your role

Identify your role in research data management, find data management resources relevant for your role, and learn how to use them.

Your domain

Learn about the data management problems that affect your domain or research community, and the solutions adopted to address them.

The Research Data Management toolkit for Life Sciences

- RDM best practices and guidelines
- Links to tools/resources and training material given in specific DM context
- Examples of combination of tools for RDM

<https://rdmkit.elixir-europe.org/>



Planning



RDMkit

Data Management Plan (DMP)

EDITORIAL • 13 MARCH 2018

Everyone needs a data-management plan

They sound dull, but data-management plans are essential, and funders must explain why.

By 2019, all who receive grants from us must have a data management plan

As from spring 2019, if you are awarded a grant from the Swedish Research Council you must have a plan for how the research data generated within your project shall be managed.

You must not send in your data management plan to us when you apply for a grant, but your administrating organisation will be responsible for ensuring that a data management plan is in place when you start your project or corresponding, and that the plan is maintained.

The main parts of a DMP

The **Swedish Research Council's** template is based on six central aspects identified by Science Europe

1. Description of data
2. Documentation and data quality
3. Storage and backup
4. Legal and ethical aspects
5. Accessibility and long-term storage
6. Responsibility and resources

1
Reuse and
production of
research data

1
Types of data,
amount/volume
and format

2
Data quality
assurance
procedures

2
Documentation
procedures and
standards

3
Storage,
security and
backup

4
Legal & ethical
and assurance
procedures

5
Future reuse
and long-term
access aspects

6
Responsibilities
and resources
required

Data Stewardship Wizard (DSW)



- Tool for creating DMPs
- **Interactive questionnaire** specific for life sciences.
- **Adapted templates** to national funders, e.g. **Swedish Research Council** and **H2020**.
- Login by using your university credentials (SWAMID).
- Do you **need help**? We provide support and guidance!

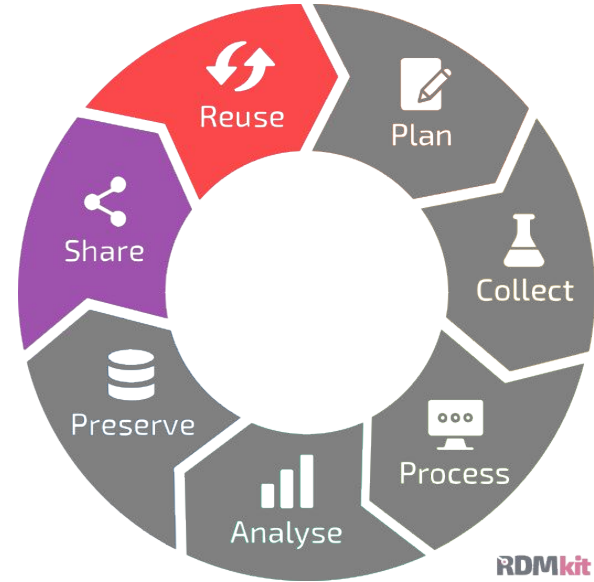
The screenshot displays the DSW interface for a project named "2023-11-30-demoproject". The top navigation bar includes tabs for "Questionnaire", "Metrics", "Preview", "Documents", and "Settings". On the right, there are user avatars and a "Share" button. Below the navigation, a "View" section shows "Current Phase" as "Before Subm" and a list of "Chapters" including Overview, Description, and various data-related sections. The main content area is titled "III. Description of data - production of new data" and contains a text box for describing data production. Below this, there are sections for specifying datasets and measurement equipment, with a "Dataset:" section where "RNA sequencing" is entered. A right-hand sidebar shows "View resolved comments" and a list of comments, including one from "Demo Researcher SciLifeLab" dated 30.11.2023, 0:21, which asks to send an email to NGI for details about instruments.

<https://dsw.scilifelab.se/>



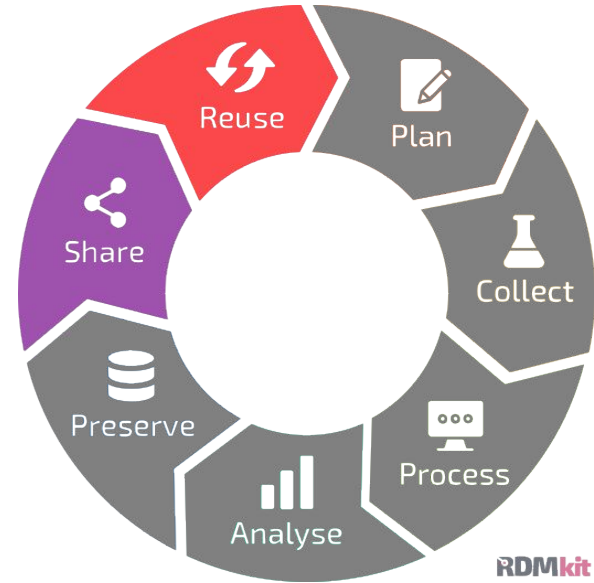
Sharing research data – why it matters

- Sharing data is a fundamental part of the research data life cycle
 - It enables reusability
- In the era of Open science and FAIR principles, data should be made available to the public
 - Scientific publications
 - Data repositories

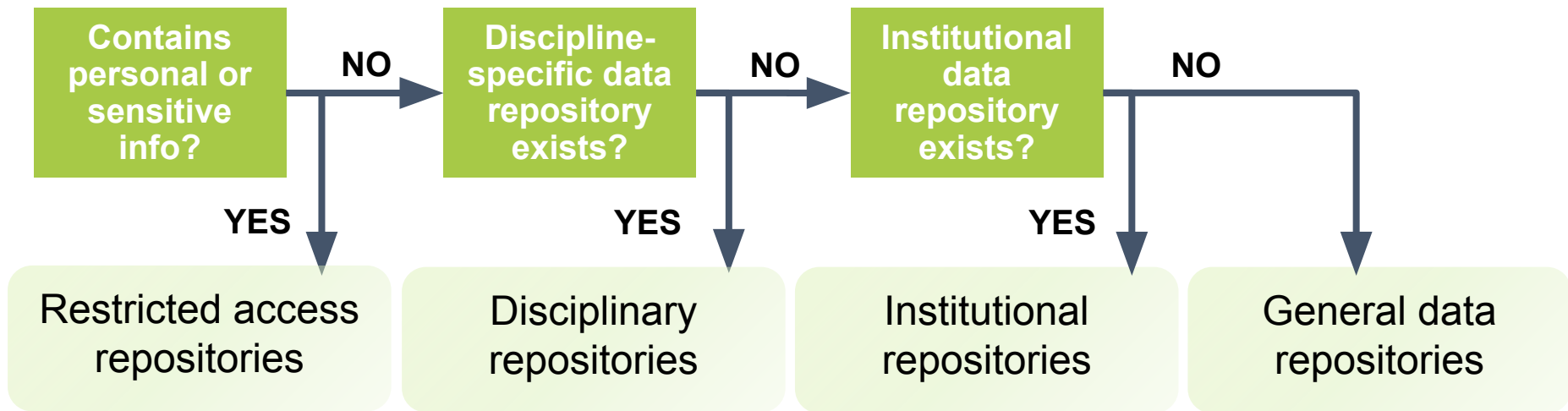


Plan for depositing data in a repository early


- What **types** of data will you generate?
- Identify **repositories** to deposit the data in
- Documentation requirements and guidelines
 - Description of study
 - Description of source samples
 - Description of technical treatment (lab methods, instruments, etc)

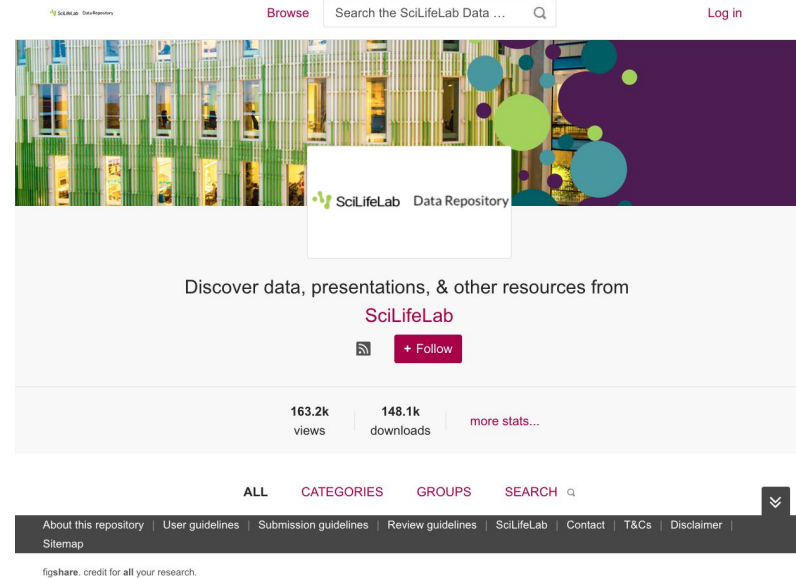


Selecting a data repository



SciLifeLab Data Repository

- An all-purpose repository. Our institutional instance of  figshare
- The repository allows for
 - Versioning
 - Embargo
 - Metadata records for data that cannot be shared openly
- Integration with e.g. ORCID, GitLab, GitHub, and Bitbucket
- Available for researchers affiliated to a Swedish university or institute working within SciLifeLab's areas of activity



FEGA Sweden

A national repository for storing and sharing personally identifiable information from Swedish biomedical research projects



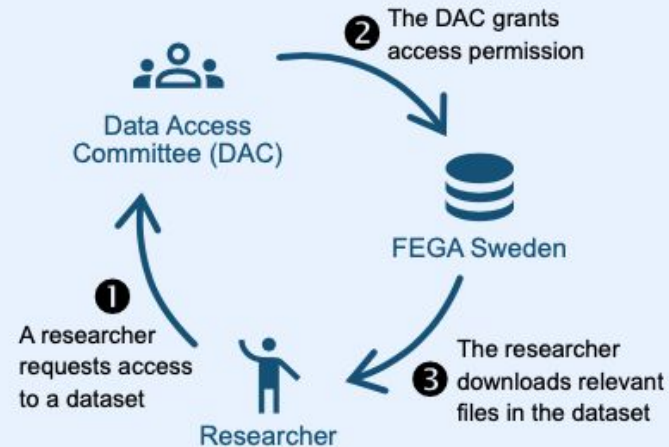
Finding data

Anonymous information about deposited datasets is searchable through the international web portal ega-archive.org.

Submitting data

- Various types of biomedical omics data are welcome.
- Sensitive data is stored with strong encryption in highly secure data centers in Sweden.
- Datasets are shared under controlled access in accordance with the research project's ethical permit.

Accessing data



Training

Introduction to Data Management Practices

Course name: Introduction to Data Management Practices

Contact: edu.intro-dm@nbis.se

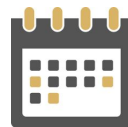
This course will introduce important aspects of Research Data Management through a series of lectures and hands-on computer exercises. The course is intended for researchers that want to take the *first* steps towards a more systematic and reproducible approach to analysing and managing research data.

Course content

Topics covered will include:

- Open Science and FAIR in practice
- Organising data, files and folders in research projects
- Describing data with metadata
- Publishing data to public data repositories
- Cleaning tabular data and metadata with OpenRefine
- Writing basic recipes for data analysis and visualisation with R
- Versioning data, documents and scripts
- Writing Data Management Plans

<https://uppsala.instructure.com/courses/79097/pages/introduction-to-data-management-practices>



Next time
October, 2024

Introduction to Data Management Practices

📅 April 23, 2024 – April 25, 2024

🌐 [NBIS Research community](#)

📄 [Registration](#)

📅 [Add to calendar](#)

Organizer:

👤 [NBIS – National Bioinformatics Infrastructure Sweden](#)

🌐 [View Organizer Website](#)

Venue:

Air&Fire, SciLifeLab Stockholm

Tomtebodavägen 23A
Solna, Sweden

[Display a menu](#)

SciLifeLab / Events / Introduction to Data Management Practices

Introduction to Data Management Practices

National Bioinformatics Infrastructure Sweden (NBIS) gives a course in research data management practices. This workshop in Data Management is a unique event to provide an introduction to practices for better management of data for PhD students, postdocs, researchers, and other employees within all Swedish universities. The workshop will introduce important aspects of research data management through a series of lectures, demonstrations, and hands-on computer exercises. The course is intended for researchers who want to take the first steps towards a more systematic and reproducible approach to analysing and managing research data.

Important dates

Application closes: 2024-03-22

Information to accepted students will be sent: 2024-04-02

Contact: edu.intro-dm@nbis.se

[Application](#) →

[Workshop Website](#) →

Format & Content

This workshop runs on-site in Solna, Stockholm. Teachers and teaching assistants will be present all week to help with exercises and

Data Management seminar series

Open Science and the SciLifeLab Open Science initiative

🕒 May 22, 10:00 – 11:00

🔗 [DDLS](#)

📅 [Add to calendar](#)

Organizer:

👤 [Data Management seminar series](#)

🌐 [View Organizer Website](#)

Venue :

Online event via Zoom

- monthly
- 15 minutes Q&A
- recorded




Next occasion: Sept, 2024

<https://www.scilifelab.se/data/scilifelab-data-management-seminar-series/>



SciLifelab Data Platform

 **SciLifeLab** Data Platform
Hub for data-driven life science research in Sweden


About Contact Cite us


Services Resources Data Highlights Jobs Funding Calls Events & Training


Request for services related to cell- and molecular biology now open!

Services

See all services



AI4Life BioImage Model Zoo
A collaborative effort to bring artificial intelligence models to the bioimaging community



SciLifeLab Serve
Platform for sharing machine learning models and data science applications, such as Shiny and Dash apps.



FAIR storage
Details about how researchers, research groups, and organisations can store their data using SciLifeLab FAIR Storage resources.

Resources

See all resources


Data sources
Sources of data that can potentially be used by researchers in Sweden. The list includes places where data and be accessed and/or submitted.


Compute resources in Sweden
Information about the compute resources (primarily for hosting and analysis) available in Sweden, as well as how to apply.


Storage resources in Sweden
Information about the storage resources available in Sweden for researchers, as well as how to apply.

What's new?

See all news

April 22, 2024
Request for descriptions of proposed national data services related to Cell and Molecular Biology

April 15, 2024
List of data sources available for researchers

April 15, 2024
New data highlight exploring a new metagenomic profiling workflow for ancient DNA

January 19, 2024
New data highlight exploring alterations to the serum proteome caused by COVID-19

Upcoming Events

See all events

Aiming for the clinic08 May

Harnessing AI Transforms Research: Smart Strategies for Advanced Productivity13 May



Contact us!

Website: data-guidelines.scilifelab.se

E-mail: data-management@scilifelab.se



Data Centre

