

SciLifeLab Data Management support and services

Introduction to Data Management Practices course

NBIS DM Team

data-management@scilifelab.se

Data Centre and NBIS



Help **maximize impact** of SciLifeLab and other data

Promoting FAIR, open science, and good data management practises throughout the data lifecycle

Provide services and resources for data management, IT and data sharing

Make bioinformatics support and training easily accessible

RDM Support and services



Support

Consultations
Hands-on support
Drop-ins



Guidelines

SciLifeLab RDM Guidelines
RDMkit



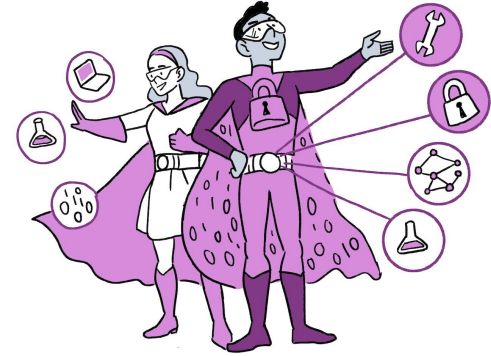
Services



Tools



Training



Scilifelab

Contact us

- data-guidelines.scilifelab.se
- data-management@scilifelab.se

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 to life science researchers in Sweden regarding 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 as well as reuse the data.

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 as well as reuse the data.

RDM life cycle from RDMit licensed under Creative Commons Attribution 4.0 International License.

Click on either 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.

[Get 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.

Research Data Management

Research Data Management (RDM) concerns the organisation, storage, preservation, and sharing of data that is collected and 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 as well as reuse the data.

Research data life cycle

The research data life cycle can be divided into several phases as seen in the wheel below: **plan, collect, process, analyse, preserve, share and reuse**. 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](#)

Get Support

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



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:

Request support



Message:

data-management@scilifelab.se

Planning

During this phase you plan for which data is needed to answer your research question, and how that data is going to be managed during as well as after the project.

It is wise to write a data management plan (DMP), using either a tool provided by your university or [DS Wizard](#) provided by SciLifeLab. The main benefits of writing a DMP are to document how the data is managed and identify possible gaps in the current data management practices. Think of the DMP as a checklist to go through when planning for a new project.

High quality science is often only possible if the infrastructure resources you intend to use get involved already in the planning phase of a project. Consultation and advice regarding data management planning, data generation and data analysis are offered by [NBIS](#) and [SciLifeLab units](#). Many institutes / universities have established [research data offices](#) which provide guidance, policies, and support regarding data management questions. There might also be some [policies](#) regarding research data from funders or other stakeholders, that you need to be aware of and adhere to.

Also, some resources have specific application periods and thus needs to be contacted well in advance. If your project includes sensitive (human) data, note that there are ethical and legal issues that you have to consider, such as apply for an ethics approval, establish necessary agreements and report the data processing to your [Data Protection Officer](#). See the page on [Research involving human data](#) for more information.

Resources

Please find below resources concerning the research data life cycle phase plan in form of training, guidance and/or tools.

Training resources

- [BAS Online](#)
- [DMP module](#)
- [Introduction to data management practices](#)
- [Metadata module](#)
- [PhD course in Open Science and Reproducible Research](#)

Guiding resources

- [AIDA Data Sharing Policy](#)
- [How to write a Data Management Plan](#)
- [RDMkit on Data Management Planning](#)
- [RDMkit on Data organisation](#)
- [RDMkit on Metadata](#)
- [SciLifeLab Data Stewardship Wizard - A short introduction](#)
- [SND on Research material with personal data](#)

Tools

- [ELSI checklist in Data Stewardship Wizard](#)
- [SciLifeLab Data Stewardship Wizard \(DSW\)](#)

Table of contents:

- [Resources](#)

data-guidelines.scilifelab.se



- The Research Data Management toolkit for Life Sciences

[Home](#)[About](#)[Contribute](#)[Contact](#)[GitHub](#)[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 you, and information to help you manage it.

Your domain

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

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

“The ELIXIR Research Data Management Kit (RDMkit) has been designed to guide life scientists in their efforts to better manage their research data following the FAIR Principles. It is based on the various steps of the data lifecycle, although not all the steps will be relevant to everyone.”

“RDMkit is recommended in the [Horizon Europe Program Guide](#) as the "resource for Data Management guidelines and good practices for the Life Sciences."

- The Research Data Management toolkit for Life Sciences


The Research Data Management toolkit for Life Sciences

Best practices and guidelines to help you make your data FAIR (Findable, Accessible, Interoperable and Reusable)

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


What can we help you find?

Browse all topics by




Data life cycle

Start here to get an overview of research data management based on stages in the data life cycle.




Your role

Identify your role in research data management, find data management resources relevant for you, and information to help you progress in your career path.




Your domain

Learn about data management tasks that affect domain or research community, and the solution to address them.




Your tasks

Find guidelines and solutions for tackling common data management tasks.




Tool assembly

Find concrete combinations of tools and resources assembled into an ecosystem for research data management.




National resources

Find pointers to country-specific information or national research data management practices.




All tools and resources

Browse the RDMkit's catalogue of tools and resources for research data management.



All training resources

Browse all training resources mentioned in RDMkit pages.


[Data management](#) [About](#) [Contribute](#) [GitHub](#)

Data quality

How do you ensure the quality of research data?

Description

Data quality is a term that can be understood in many ways. In enterprise context, it often refers to master data management as defined by the ISO 8000 standards. In science, the quality of data is closely linked to the suitability of the data for (re)use for a particular purpose and it is a key attribute of research data. Data quality affects the reliability of research results and it is a key factor increasing the reusability of data for secondary research. Data quality control can take place at any stage during the research data lifecycle. That said, you should ensure that the necessary procedures are defined during data management planning.

Considerations

Quality control is most typically performed during data collection but it should not be neglected in later stages of research data lifecycle. The type of data as well as instruments and processes adopted for data collection in your research will determine the quality assurance measures you can take. Examples of such measures are:

- setup data management working group (DMWG) that includes people who generate data, analyse data and data managers;
- for data collection: DMWG to plan and define data dictionary (including validation rules) before collecting data;
- for metadata collection: DMWG to plan and define metadata data templates;
- use electronic data capture systems;
- automated quality monitoring through tools, pipelines, dashboards;
- training of study participants and researchers, surveyors or other staff involved;
- adopting standards;
- instrument calibrations;
- repeated samples;
- post collection data curation;
- data peer-review.

Certain areas such as clinical studies, or those involving Next Generation Sequencing have commonly working methods to ensure data quality. Consider familiarizing yourself with data quality standards or established working practices in your field of study.

There are many frameworks proposed in the literature to define and evaluate overall data quality, such as:

Data management

- Data life cycle
- Your role
- Your domain
- Your tasks
- Compliance monitoring
- Costs of data management
- Data analysis
- Data brokering
- Data management coordination
- Data management plan
- Data organisation
- Data security
- Data sensitivity
- Data provenance
- Data publication
- Data quality**
- Data storage
- Data transfer
- Documentation and metadata
- Existing data
- GDPR compliance
- Identifiers

Your tasks

On this page

- How do you ensure the quality of research data?
- More information
- Relevant tools and resources

Data Services

Explore the SciLifeLab Data Platform

The SciLifeLab Data Platform is a new technical hosting environment for data-centric tools, databases, and support for data-driven life science. The website will function as a hub for the life science community. The platform, funded by the SciLifeLab & Wallenberg National Program for Data-Driven Life Science (DDLS), will serve all life science researchers and data-producing facilities in Sweden.

Make Your Research Data-Centric With The Data Platform →

Interact with our IT and Software Services

The SciLifeLab Data Centre provides services for IT and data management. We also work to facilitate the communication between SciLifeLab platforms, their users, and the research community, and develop databases and tools. The purpose is to maximize the scientific impact of SciLifeLab generated data and provide a good helpful IT environment for the SciLifeLab platforms and their users.

Get Started - Use Our Services →

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

Open Science

SciLifeLab Data Centre promotes Open Science and data sharing according to the FAIR principles

Share Your Data Open And Fair →

AI model and compute application sharing

SciLifeLab Serve is a service for publishing apps and serving machine learning models. Serve is currently under development (in beta)

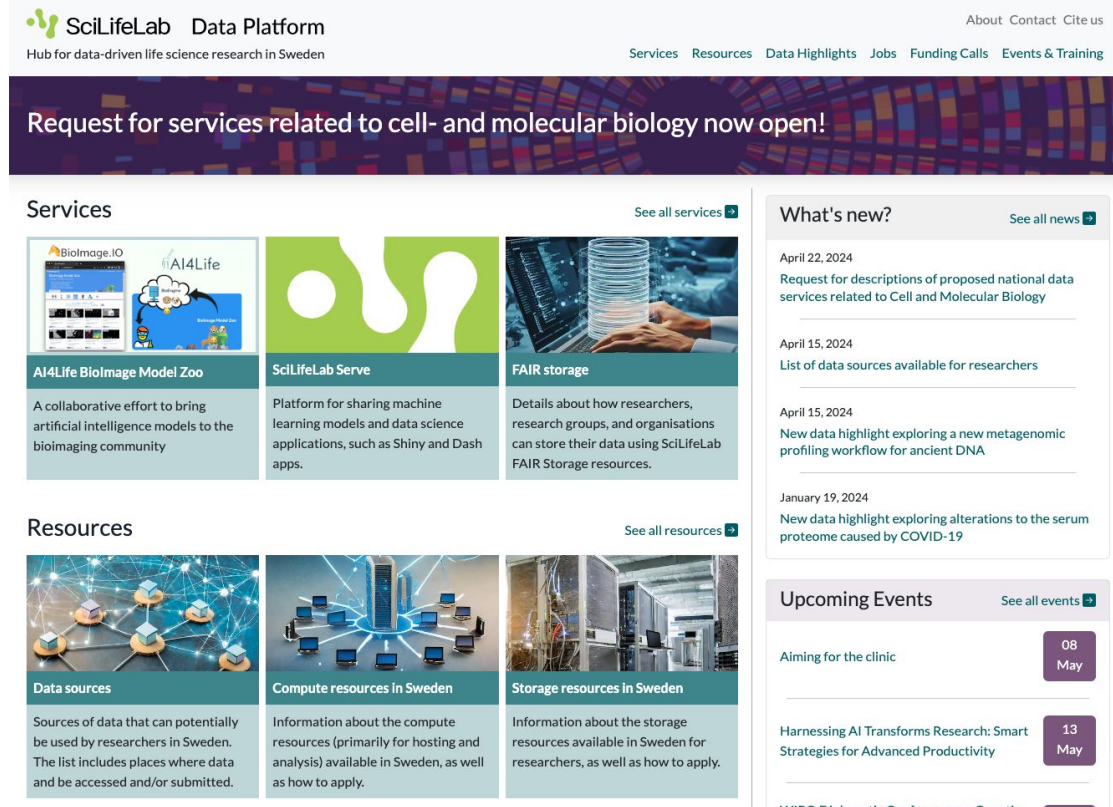
Engage With Serve →

Pandemic data

SciLifeLab Data Centre maintains the Swedish Pathogens Portal. The aim of the portal is to provide information, guidelines, tools and services to support researchers in utilising Swedish and European infrastructures for data sharing

Use The Pathogens Portal →

SciLifeLab Data Platform



The screenshot shows the SciLifeLab Data Platform website. At the top, the SciLifeLab logo is on the left, and navigation links 'About', 'Contact', and 'Cite us' are on the right. Below the logo, the text 'SciLifeLab Data Platform' and 'Hub for data-driven life science research in Sweden' is displayed. A secondary navigation bar includes 'Services', 'Resources', 'Data Highlights', 'Jobs', 'Funding Calls', and 'Events & Training'. A large purple banner with a grid pattern contains the text 'Request for services related to cell- and molecular biology now open!'. The main content area is divided into two columns. The left column has a 'Services' section with three cards: 'AI4Life BiImage Model Zoo' (describing a collaborative effort for AI models), 'SciLifeLab Serve' (describing a platform for machine learning and data science), and 'FAIR storage' (describing details for researchers to store data). The right column has a 'What's new?' section with three news items dated April 22, 2024, April 15, 2024, and January 19, 2024, covering national data services, data sources, and COVID-19 proteome alterations. Below these is an 'Upcoming Events' section with two events: 'Aiming for the clinic' on May 08 and 'Harnessing AI Transforms Research: Smart Strategies for Advanced Productivity' on May 13. A 'Resources' section at the bottom left has three cards: 'Data sources', 'Compute resources in Sweden', and 'Storage resources in Sweden'. Each card includes a representative image and a brief description.

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

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 BiImage 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 can 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 clinic **08 May**
- Harnessing AI Transforms Research: Smart Strategies for Advanced Productivity **13 May**
- WIBO Diplomatic Conference on Genetic

2 parts:

- A website (data.scilifelab.se)
- An underlying technical environment for hosting

Aims to provide:

- Access to & hosting for appropriate services
- Promotion for data-driven life science research
- Resources related to community building and support
- Guidance for researchers

IT and Software services

IT and Software Services

We have a number of services available, and some things are in development.

Systems for operations

ORDER PORTAL

A web based system for unit communication with prospective users. We can provide you with an instance configured according to your own specific requirements. Try out a demo instance at orderportal.scilifelab.se, contact us or Per Kraulis directly.

PUBLICATIONS DATABASE

The Data Centre built and maintains the [ScilifeLab Publications](#) database where the results of research assisted by ScilifeLab units can be found. A second database was set up for publications of researchers and fellows affiliated with ScilifeLab.

DATA MANAGEMENT PLANS

We encourage ScilifeLab units to support users with data management plans. We provide templates and a software system to manage such plans, the ScilifeLab instance of the [Data Stewardship Wizard](#).

TOOL SUPPORTING REPRODUCIBLE METHODS

ScilifeLab Data Centre recently purchased a Premium license for [protocols.io](#). Current accounts using your ScilifeLab email addresses have automatically been upgraded to a Premium account. For other accounts, to upgrade and take advantage of the ScilifeLab license, please email dataportal@scilifelab.se.

SURVEY TOOLS

The online survey tool [Typeform](#) is offered to the units and administration. It allows to collect questionnaire responses from users/researchers. Email us for access.

DATA DELIVERY SYSTEM

The [Data Delivery System \(DDS\)](#) is a cloud-based system for the delivery of data from ScilifeLab platforms to their users.

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

Systems for operations

Collaboration and
communication systems



Data Stewardship Wizard (DSW)

- Tool for creating Data Management Plans

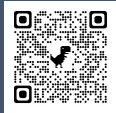
- **Interactive questionnaire** specific for life sciences.
- **Adapted templates** to national funders e.g. **Swedish Research Council** and **H2020**.
- Login by using your university account (SWAMID).
- Do you **need help**? We provide support and guidance!


Learn more:


<https://dsw.scilifelab.se>





youtu.be/HY2DVnNGkAs
(short DSW introduction)





 **SciLifeLab DSW**


 Projects

 Guide - Write a DMP


 Guide - NBIS Support Checklist


 Short intro DSW

 DSW workshop

 About project templates

Log In


 Email

 Password

[Forgot your password?](#)

[Log In](#)

Or connect with

 Life Science RI (university)



Data Management Plans

DSW - Data Stewardship Wizard

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

The screenshot displays the SciLifeLab DSW interface for a project titled '2023-11-30-demoproject'. The top navigation bar includes links for 'Questionnaire', 'Metrics', 'Preview', 'Documents', and 'Settings'. A 'View' dropdown is located on the left, and 'TODOs', 'Comments', and 'Version history' are on the right. The main content area is divided into three sections: 'Current Phase' (set to 'Before Subm'), 'Chapters', and a detailed description of data production. The 'Chapters' list includes 'I. Overview', 'II. Description...', 'III. Description...', 'IV. Docume...', 'V. Storage ...', 'VI. Legal an...', and 'VII. Accessib...'. The 'III. Description of data - production of new data' section contains a green checkmark and the text 'Please specify what datasets you will acquire using measurement equipment'. Below this, a 'Dataset:' section shows 'RNA sequencing' as an answer. A 'Comments' sidebar on the right shows a comment from 'Demo Researcher SciLifeLab' dated '30. 11. 2023, 0:21' with the text 'Send email to NGI to ask for details about instruments'.

2023-11-30-demoproject

Questionnaire Metrics Preview Documents Settings

View TODOs 1 Comments 1 Version history

Current Phase
Before Subm

Chapters

- I. Overview 3
- II. Description... ✓
- III. Description... 5
- IV. Docume... 4
- V. Storage ... 7
- VI. Legal an... 10
- VII. Accessib... 5

III. Description of data - production of new data

We will make sure that we know what data will be produced during the project. We also need to make sure that we have adequate storage space to deal with it, and that all responsibilities are cared for.

✓ III.1 Please specify what datasets you will acquire using measurement equipment

You can use any name for the dataset, make sure that it identifies the data set to yourself.

Desirable: Before Submitting the DMP

✓ III.1.a.1 Dataset:

Desirable: Before Submitting the DMP

RNA sequencing

Clear answer

Answered 5 months ago by Demo Researcher SciLifeLab.

✓ III.1.a.2 Who will do the measurements? And where?

Are there easily accessible specialized service providers for data capture, e.g. SciLifeLab facilities such as

View resolved comments

Comments 1 Editor notes

Demo Researcher SciLifeLab
30. 11. 2023, 0:21

Send email to NGI to ask for details about instruments

Reply...

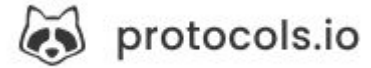
Create a new comment...



- Life Science specific guidance
- 'Bite-sized' questions
- Most questions with answer options

Protocols.io

An online platform for the creation, management, and sharing of research protocols or methods



Create

Every new protocol starts out private. You can then choose to keep it private, share it with others, or to make it public.



Run

Run protocols as checklists. Once your experiment is complete, the changes you made are recorded in your File Manager. You can keep your experiments private or share them with others.



Publish

Receive a DOI after you publish a protocol. After publication, your protocol is archived with CLOCKSS to ensure longterm preservation of the knowledge.

Sharing and reusing data

Explore the SciLifeLab Data Platform

The SciLifeLab Data Platform is a new technical hosting environment for data-centric tools, databases, and support for data-driven life science. The website will function as a hub for the life science community. The platform, funded by the SciLifeLab & Wallenberg National Program for Data-Driven Life Science (DDLS), will serve all life science researchers and data-producing facilities in Sweden.

Make Your Research Data-Centric With The Data Platform →

Interact with our IT and Software Services

The SciLifeLab Data Centre provides services for IT and data management. We also work to facilitate the communication between SciLifeLab platforms, their users, and the research community, and develop databases and tools. The purpose is to maximize the scientific impact of SciLifeLab generated data and provide a good helpful IT environment for the SciLifeLab platforms and their users.

Get Started - Use Our Services →

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

Open Science

SciLifeLab Data Centre promotes Open Science and data sharing according to the FAIR principles

Share Your Data Open And Fair →

AI model and compute application sharing

SciLifeLab Serve is a service for publishing apps and serving machine learning models. Serve is currently under development (in beta)

Engage With Serve →

Pandemic data

SciLifeLab Data Centre maintains the Swedish Pathogens Portal. The aim of the portal is to provide information, guidelines, tools and services to support researchers in utilising Swedish and European infrastructures for data sharing

Use The Pathogens Portal →

FEGA Sweden

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



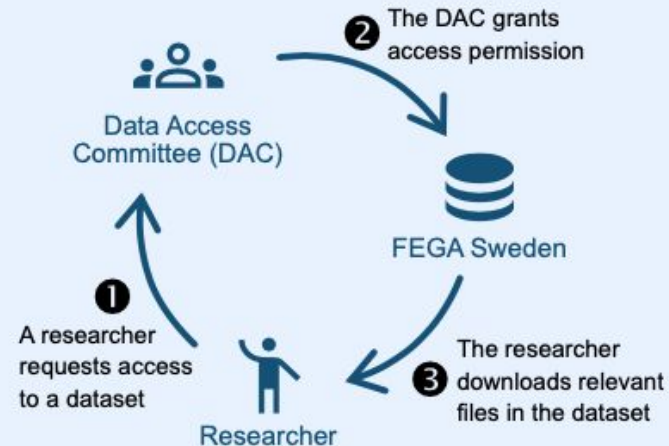
Finding data

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



SciLifeLab Data Repository



- Institutional instance of [FigShare](https://figshare.scilifelab.se/)
- Publish general research data without costs for Swedish researchers
- Available for researchers affiliated to all Swedish universities and institutes working within SciLifeLabs' research areas.
- Possible to apply for a license or embargo and set up a metadata record-only.

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

SciLifeLab FAIR Storage



- Data sharing service, powered with long term object storage (S3)
- OK for sensitive data
- Allocations and projects are coupled with updated data management plans and updated FAIR releases of research outputs

FAIRstorage@scilifelab.se

data.scilifelab.se/services/fairstorage

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

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

- monthly
- 15 minutes Q&A
- recorded

Next occasion: May 22 10:00-11:00



Support throughout the life cycle

