

FAIR Data Management

CLAUDIA, 14 September 2023

Department of Chemistry and Bioscience



AALBORG
UNIVERSITY

Agenda

- Introduction to *Knowledge for the world*, goal 2.2

- Introduction to CLAAUDIA
 - What services do CLAAUDIA offer?

- Introduction to the FAIR-principles

- Introduction to Data Management Plans
 - Introduction to Horizon Europe projects

- Tips and tools from CLAAUDIA

- Questions

CLAAUDIA
RESEARCH DATA SERVICES



DIRECTIVE (EU) on open data and the re-use of public sector information (PSI), 2019/1024

Article 10

Research data

1. Member States shall support the availability of research data by adopting national policies and relevant actions aiming at making publicly funded research data openly available ('open access policies'), following the principle of 'open by default' and compatible with the FAIR principles. In that context, concerns relating to intellectual property rights, personal data protection and confidentiality, security and legitimate commercial interests, shall be taken into account in accordance with the principle of 'as open as possible, as closed as necessary'. Those open access policies shall be addressed to research performing organisations and research funding organisations.
2. Without prejudice to point (c) of Article 1(2), research data shall be re-usable for commercial or non-commercial purposes in accordance with Chapters III and IV, insofar as they are publicly funded and researchers, research performing organisations or research funding organisations have already made them publicly available through an institutional or subject-based repository. In that context, legitimate commercial interests, knowledge transfer activities and pre-existing intellectual property rights shall be taken into account.



Goal 2 Research, 2022

2.2 All departments conduct research based on digital data

► Goal description

*Digitalisation is currently transforming both research and the surrounding society and will be a strong lever in achieving our strategic goals. We want **to be among the best at taking advantage of the opportunities** offered by digitalisation to achieve new insights for developing our problem-oriented learning model and our service functions.*

*We thus want **to integrate digitalisation into all university functions**. Understanding data management and using infrastructure to manage data contributes to modifications to research paradigms in most sciences; mission-oriented research will also require the ability to collaborate on data and infrastructure – both nationally and internationally. AAU is already well-anchored in the national and European setup for digital infrastructure; we thus have the prerequisite to take the lead and become a sought-after participant in various missions.*





Objectives for 2022

Goal A1: Store, expose, search and retrieve datasets (in a local AAU repository).

Goal A2: Test the repository-system with 1(-2) pilot tests pr. main area with significant diversity in the datasets.

Goal A3: Unique ID on all researchers and new datasets

Goal A4: Research assistance to EU required DM-FAIR planes.

Goal B: Hiring a CLAUDIA data scientist with health skills.

Goal C: Help to search for competitive compute-cycles nationally and og internationally.



Objectives for 2023

1. New data is guaranteed a minimum level in relation the FAIR, in that it is [...] described with a set of basic metadata. It is the ambition that the basic metadata takes place automatically at system level.
2. The level of fulfillment of FAIR, in addition to the basic requirements for metadata as described above, is defined by the individual research area in relation to international norms and standards for the area. It is the goal for 2023 that the work has begun for **at least 2 research areas at all institutes**.
3. Data that a researcher or AAU wish to have stored under increased security, must be stored in a technically secure data infrastructure in accordance with the FAIR-principles.
4. Four new Data Stewards are hired in CLAAUDIA – as a supplement to the existing resources – **to support the faculties and departments**, and who can assist researchers with the preparation of FAIR data management plans based on objective 2 in the strategy.





How can we help the departments in 2023?

- ▶ Each department should point at two research areas, that can be included in the work with basic requirements for metadata based on the international norms and standards for the specific research area.
- ▶ We help ensure that your researchers become more FAIR, for example by develop and maintain www.fair.aau.dk
- ▶ We prepare a catalogue of subject-specific taxonomies and ontologies and an overview of subject-specific repositories from a FAIR point of view.
- ▶ We can help you create awareness of requirements and opportunities and assist with implementation of work processes
- ▶ Assist with the implementation of own goals in relation to *Knowledge for the World 2022-2026*.

CLAAUDIA is a practice-oriented support unit that among other things can assist researcher in development of Data Management Planes, technical specifications for applications for grant agreements and much more in relation to research infrastructure, data management and data science.





The CLAUDIA Team



Jeanette Dahl-Sørensen
Team Leader



Nina Simone Marstrand
Sander Aagaard
Chief Consultant



Rasmus D. Jensen
Digitalization Partner



Dagmar Knudsen Fallesen
Project Consultant



Rainer Bohm
Enterprise Architect



Pelle Rosenbeck Gøeg
Data Scientist



Robert Smith
Data Scientist



Gergely István Barsi
Data Scientist



Sighvatur Sveinn Davidsson
Data Scientist



Carina Ollerup Christensen
Data Manager



Thomas Andersen
Data Steward
Primary @ TECH



Freya Vamberg Delfs
Data Steward
Primary @ SSH

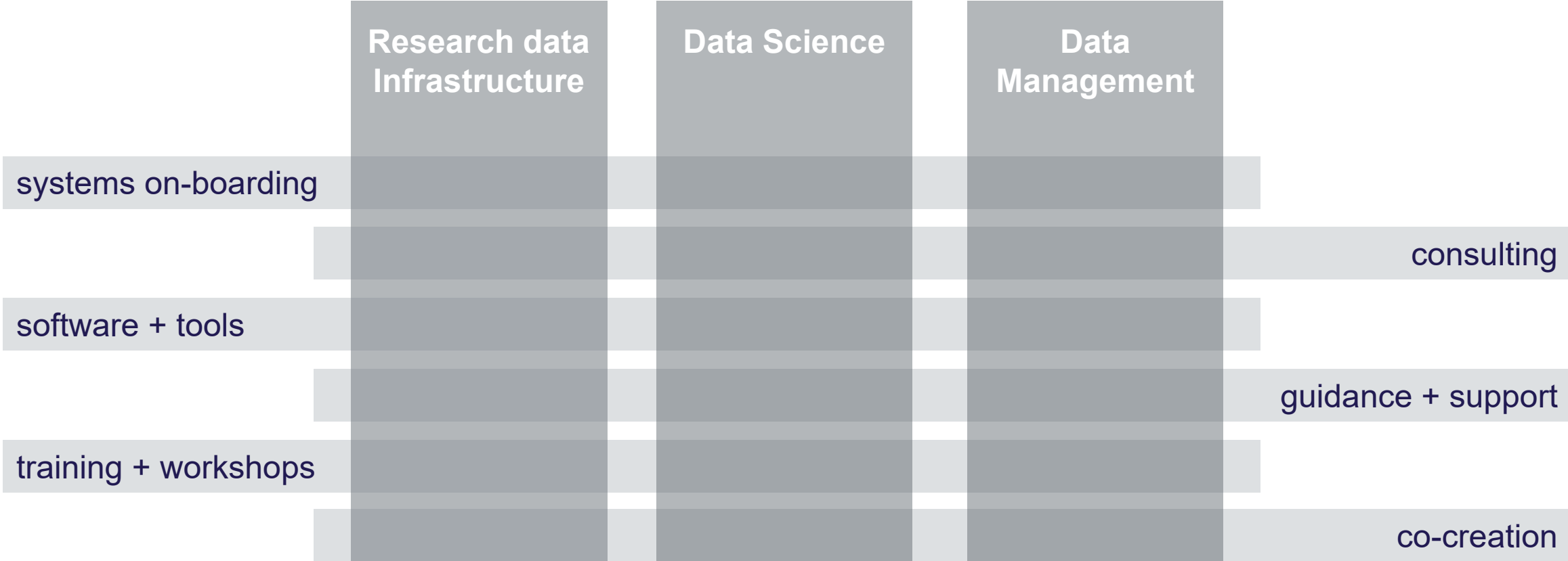


Kamilla Hall Kragelund
Data Steward,
Primary @ SUND



Dennis Aagaard Pedersen
Data Steward
Primary @ ENG

CLAUDIA RESEARCH DATA SERVICES



CLAAUDIA RESEARCH DATA SERVICES

CONSULTING

- Choice of cloud platform
- Access, onboarding and set-up of internal platforms
- Applications to work on external platforms
- Data Science methods
- Configuration of infrastructure
- Technical requirement specifications
- Data storage, collaboration and sharing
- FAIR data management
- Data management plans
- Grant applications (technical requirements and data management)

SUPPORT AND DEVELOPMENT

- Strato
- AI Cloud
- U Cloud (HPC type 1)
- Computerome
- Sohia
- Genome DK
- LUMI
- DataDeposit
- DMP Online
- Nvidia Jetson Nano Kit
- REDCap
- Whisper
- Sundhedsdatanettet

C
O
M
P
U
T
I
N
G

COURSES AND WORKSHOPS

- Introduction to Strato
- Introduction to AI Cloud
- Introduction to U Cloud
- Intro and workshop: FAIR and DM
- Ph.d. school: Intro to DM & CLAAUDIA
- Ph.d. school: FAIR Data Management
- Ph.d. school: Scientific software development
- Ph.d. school: Python

Existing HPC possibilities at AAU

AI Cloud

You can use up to 16 GPUs at a time on one server.

Strato

You can apply for a project to use virtual machines with up to 4 GPUs each.

U Cloud

You can start virtual machines with up to 4 GPUs each.

EuroHPC facilities

EuroHPC JU is an organisation in the European Union developing and managing access to high-performance and supercomputer facilities in Europe.

Facilities (with GPUs):

- LUMI (world's 3rd most powerful supercomputer): 10,240 AMD MI250x GPUs
- Leonardo: 13,824 NVIDIA custom Ampere-arch. GPUs
- Marenostrum 5: ? NVIDIA Hopper-arch. GPUs (coming)
- Vega: 240 NVIDIA A100 GPUs
- Meluxina: 800 NVIDIA A100 GPUs
- Vega: 560 NVIDIA A100 GPUs
- Deucalion: ? NVIDIA Ampere-arch. GPUs

Access to LUMI

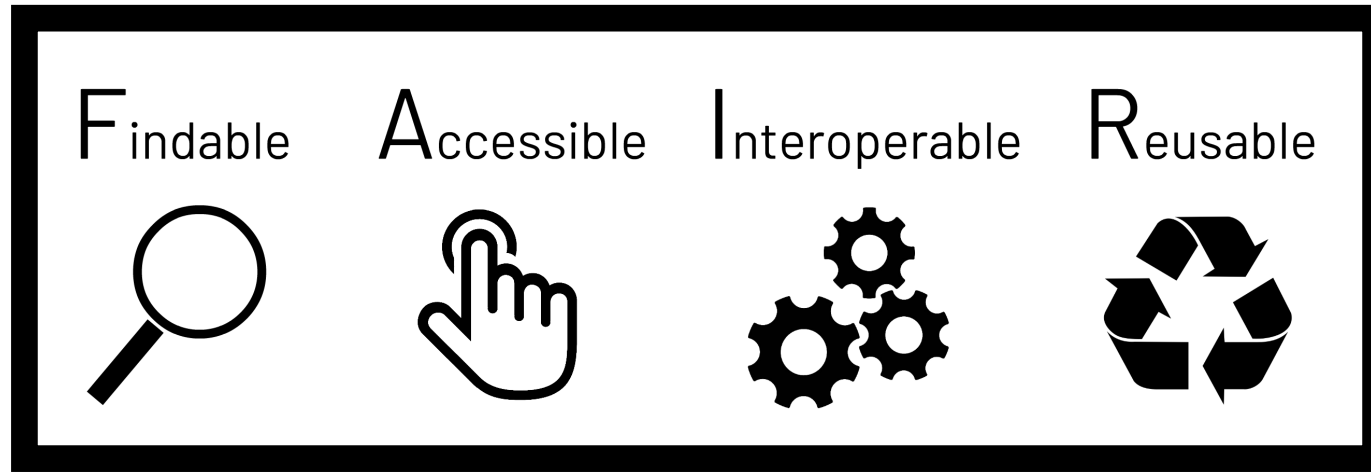
The LUMI supercomputer can be accessed in 2 ways:

1. DeiC semiannual calls for applications (the deadline for applications was on September 5th, 2023)
2. AAU local resources: AAU has a pool of resources for LUMI that CLAUDIA can distribute locally.

How can CLAUDIA assist with applications to HPC facilities?

- **Assessment of needs:** Most suitable facility, extent of computational requirements, necessary software. This could mean, that you figure out you need something else than an application to DeiC (Danish e-infrastructure cooperation).
- **Feedback on application draft:** Focus on correctly addressing the various aspects of the application form.
- **Quick, preliminary access** to the desired HPC facility for test and benchmarking of the planned workload. For example, this has also helped 1 researcher clarify that LUMI was not the suitable facility for their application. Due to software incompatibility; important to determine before applying.

The FAIR principles



- Each letter consists of a range of principles
- The principles are guidelines for good data management practice.
- As open as possible as closed as necessary!
- All open data must be FAIR, but not all FAIR data must be open.
- FAIR is a continuum – not a firm set of rules

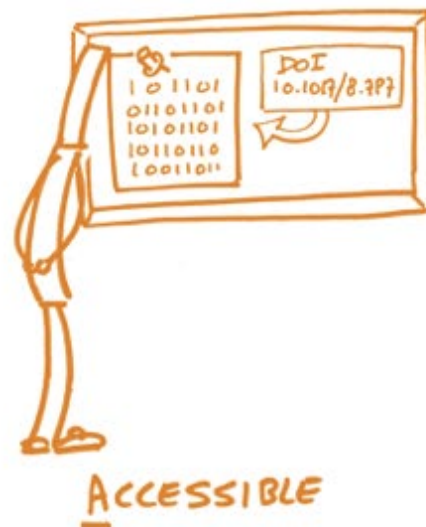
The FAIR principles

FAIR DATA PRINCIPLES



Humans (and machines) can find out the data exists. Because you:

- Publish searchable metadata.
- Assign a unique persistent identifier.



... can find out how to get access to the data. Because you:

- Upload to public data repository.
- Define access conditions for data and metadata (if any).



... can open the data and work with it. Because you:

- Use open file formats.
- Use community standards, keywords and ontologies.



... understand how the data were created and how to reuse it. Because you:

- Attach sufficient documentation.
- Add usage license.

WHY FAIR?

The FAIR principles are at the core of Danish and international strategies for research data because of a demand for:

- Increased research impact
 - Enhanced transparency, reliability and reusability of your research.
 - Potential for more citations and new collaboration possibilities due to higher visibility of research data/metadata.
 - Secure data storage and prevention of data loss.
- Increased data reuse
 - Knowledge discovery and accelerating research.
 - Making data understandable to humans and machines (machine-actionable data).

And because of this, more and more funders require you work with the FAIR-principles.

Start your journey towards FAIR data

1. Store your (meta)data in a **trusted data repository** and attach a **DOI** (digital object identifier).

A trusted data repository is a digital archive where your data can be preserved and maintain usability. A trusted data repository, allows you to decide the visibility of both your data and metadata.

2. Make **metadata** descriptions.

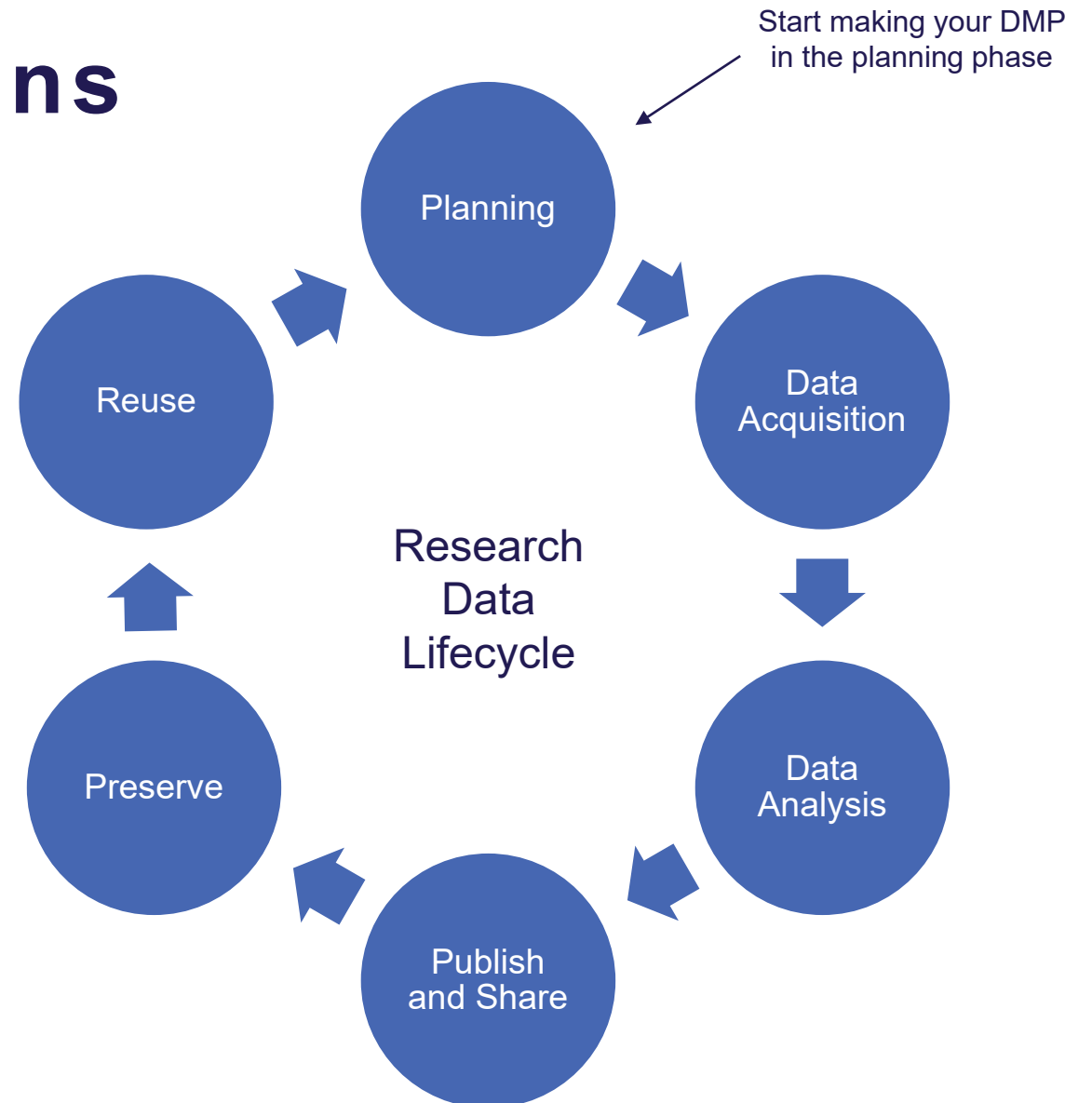
Metadata are descriptions of your data, that follows standards from your research field. Standardised metadata enhances the **F**indability, **I**nteroperability, and **R**eusability of your dataset.

3. Put a **data license** on your data.

This will describe how others may handle data. Access to metadata alone can be of immense value for your research impact.

Data Management Plans (DMP's)

- A DMP is a living document – a reflection tool. You change it according to your research and circumstances.
- It focuses on data and data handling practices, and it helps you make informed choices on how you handle your data.
- A way of creating an overview over all relevant data management topics.
- A help to identify key actions and strategies to ensure high-quality and secure data.
- A help to make your data FAIR.



DMPonline

https://dmponline.deic.dk

The screenshot shows the DMPonline website interface. At the top, a green navigation bar contains the 'DMP ONLINE' logo and menu items: 'Home', 'Public DMPs', 'DMP Templates', and 'Help'. Below the navigation bar is a light blue notification banner that reads 'Notice: Signed out successfully.' The main content area features a 'Welcome to DMPonline' heading, followed by a paragraph explaining that this is the Danish installation of DMPonline, a tool for writing data management plans. It states that the service is provided by the Danish e-infrastructure Cooperation (DeiC) and administered jointly by the Royal Danish Library and the Technical University of Denmark (DTU Library). A second paragraph explains that new users must sign up with a valid email address and create an account, which can then be linked to institutional credentials for 'WAYF' login. A 'Contact us here' link is provided. On the right side, there is a sign-in form with tabs for 'Sign in' and 'Create account'. The form includes fields for 'Email' and 'Password', a 'Forgot password?' link, a 'Remember email' checkbox, and a 'Sign in' button. Below the form is a '- or -' separator and a button for 'Sign in with your institutional credentials'. The footer contains copyright information for 2017-2023 Danish e-infrastructure Cooperation, links for 'About', 'Contact us', 'Terms of use', 'Privacy policy', and 'Github', and the DeiC logo.

DMPonline

Project Details Plan overview Write Plan Share Download

expand all | collapse all 0/14 answered

- 1. Data Collection (0 / 2) +
- 2. Data Processing (0 / 1) +
- 3. Documentation and Metadata (0 / 1) +
- 4. Ethics and Legal Compliance (0 / 2) +
- 5. Storage, Backup and Access (0 / 2) +
- 6. Selection and Preservation (0 / 2) +
- 7. Data Sharing (0 / 2) +
- 8. Responsibilities and Resources (0 / 2) +

DMPonline

5. Storage, Backup and Access (0 / 2) +

6. Selection and Preservation (0 / 2) +

7. Data Sharing (0 / 2) -

How will you share your finished data?

B I [List Icons] [Link Icon] [Table Icon]

Save

Comments & Guidance

Guidance | Comments

AAU

Guidance:
Consider how other researchers will be able to find your data and metadata in collections or search engines. How your data will be available to others might depend on your choice of repository and possible restrictions due to type, size, complexity and sensitivity.?

Questions to consider:

- Where will the data be deposited? You can find different repositories at re3data.org.
- How will it be possible to access the data in the repository or other places data are stored?
- Will you pursue getting a persistent identifier for your data? This might be part of your repository workflow.

Are there any restrictions on data sharing?

B I [List Icons] [Link Icon] [Table Icon]

Save

Comments & Guidance

Guidance | Comments

AAU

Guidance:
Outline any expected restrictions and difficulties in sharing your data, along with causes and possible measures to overcome. Also, consider how people might credit and acknowledge the reuse of your data.

Questions to consider:

- Are any restrictions on data sharing required? E.g., limits on who can use the data, when and for what purpose, if so, indicate who will be

DMPonline- Horizon Europe template

Project Details Contributors Plan overview **Write Plan** Share Download

expand all | collapse all 0/41

- 1. Data Summary (0 / 6) +
- 2. FAIR data (0 / 24) +
- 3. Other research outputs (0 / 2) +
- 4. Allocation of resources (0 / 4) +
- 5. Data security (0 / 2) +
- 6. Ethics (0 / 2) +
- 7. Other issues (0 / 1) +

Example: Data overview

WP partner	Data type, format, and short description	How created/source	Data volume	Classification, security, Ethics	Availability and restrictions
WP2	<p>Simulations:</p> <ul style="list-style-type: none"> Models for predicting climate change effects in urban areas (python scripts) Statistical analysis of several thousand model calculations (mostly txt and csv files) Software package for the design of smart watering solutions 	<p>Models will be developed by the P.I. at AAU. Calculations will be carried out at the university's HPC cluster. Additionally, [commercial software] will be used for benchmark studies.</p> <p>The software package will be developed in collaboration with [external company]</p>	<p>Total size of files unknown but estimated to be in the order of a few 100 GB</p>	<p>Use of [commercial software] requires purchase of a usage license.</p> <p>IPR to the software package will be shared between AAU and [external company]. Details will be outlined in a cooperation agreement set up by AAU Technology Transfer Office.</p>	<p>Open access and commercialized (closed)</p> <p>Own models and results will be made openly available:</p> <ul style="list-style-type: none"> Models and documentation will be deposited in [code repository] and available for free use under public domain. Results from model calculations will be published in [public repository]. <p>The software package will be commercialized.</p>
WP3	<p>Interviews:</p> <ul style="list-style-type: none"> Interview guide (pdf) Audio recordings (mp3) Transcriptions (doc) Metadata file (txt) 	<p>Interviews will be conducted with 20 participants in [country1] and [country2]</p>	<p>Total size around 200 MB</p>	<p>Audio recordings contain personal, sensitive information. The recordings will be deleted at project end. Transcriptions will be pseudonymized.</p> <p>All participants will have to sign an informed consent form prior to the interview</p>	<p>Confidential and Open access</p> <p>Personal and sensitive information will only be available to the P.I. during the project.</p> <p>The interview guide and a completely anonymized summary of the interviews will be published in [general repository] at the end of [WP].</p>
WP5	<p>Existing data:</p> <ul style="list-style-type: none"> Compiled results from presidential elections in the US between 1788 and 2020. Articles from online news portals in the US regarding the presidential elections in 2020. Scholarly publications investigating potential biases of online media during the presidential elections in 2020. 	<p>Election data are retrieved from [public database].</p> <p>Articles are accessed via [online archive].</p> <p>A systematic review is carried out for available literature using [database 1] and [database 2].</p>	<p>NO</p>	<p>All data from [public database] are freely available under a CC-BY-NC license.</p> <p>Articles and publications are subject to the terms and conditions specified by their respective copyright holder</p>	<p>Open access</p> <p>All original data are already publicly available. Therefore, the project will only make relevant documentation (e.g., information on search methodologies) and metadata files (including references and access details) openly available through [general repository] at the end of the project.</p>

Examples: Metadata

Each dataset has metadata associated to it, for descriptive, structural and administrative issues. The embedded metadata in data will be part of the data collection and documentation process, which are elaborated before the publication of the metadata, both in terms of sanitizing and for adding metadata. Some metadata will be created during the publication procedure to a given repository and follow the (metadata) standard(s), for the specific repository. As for the specific format of metadata, Zenodo utilizes DataCite which is a de facto standard for describing datasets.

All partners are requested to link to the grant number 101058643 in the metadata standard in the grant section, which is supported by DataCite, or to include an acknowledgement of funding in their metadata, mentioning the name of the project and grant agreement, in case of metadata standards that does not have a separate grant section.

Partners are encouraged to provide documents such as e.g., readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, whenever possible. Whenever possible, the data and publications will be interlinked using metadata in Zenodo or other repositories.

Examples: Storage and access

All raw data (including transcripts of interviews, which will not be provided openly) will be retained for at least five years on the servers of the project partner responsible for gathering this raw data (this data is often in the local language of that partner, and it is better not to transfer unnecessary personal data between partners) for validation.

Each partner should ensure that there will be people curating the raw data during this period. The servers for storing the raw data will have necessary security mechanisms, such as password protection and restricted access only to the project partner. Access to this raw data will only be provided for validation. Access will only be provided based on a positive review from the partner that is keeping the data. If such does not exist at the specific partner, the review is done by the coordinator.

Why should I write a DMP?

- Reflect on your data handling practices.
- Make active decisions before they become a problem.
- Structure saves time in the long run.
- Structure could increase the quality of your research.
- Awareness of your data's FAIRness.
- Increase your research impact.



A good DMP?

We suggest you:

- Start as early as possible writing your DMP.
- Know that the DMP changes with the research project.
- Use a DMP-template (required by funder or AAU-template).
- Know each DMP is unique, and the content, composition, and structure can vary greatly - keep it simple.
- Contact CLAAUDIA for feedback.



Horizon Europe's requirements for data management

Submission

- Part B, section 1.2; Research data management and management of other research outputs (1 page).

In case of successful applications:

- **Data Management Plan** at 6 and 36 months.
- **Manage data responsibly**, in accordance with legislation, contracts information security requirements, etc. and according to **the FAIR-principles**.
- **Make data accessible by default**, but “as open as possible, as closed as necessary”.



HORIZON EUROPE

How can CLAUDIA assist with your DMP?

- Consult/assist when you make the DMP.
- Feedback on your DMP draft, focusing on correctly addressing the various aspects of the DMP template.
- We have a written guides to both the AAU generic DMP and the Horizon Europe DMP template on DMPonline.
- Workshop: Practice-based introduction to Data Management Plan
- Ph.D.-course: FAIR Data Management



DataDeposit

AAU's new local archiving solution:

- Deposit your finalized research data.
- Deposition of any data type – also sensitive
- Public metadata exposed on AAU's VBN/PURE
- A step towards making FAIR-data
- 4 data stewards in CLAUDIA ready to assist.

<https://datadeposit.claudia.aau.dk/>

The screenshot shows the DataDeposit website interface. At the top, there is a blue header with the "DataDeposit" logo and "DarkArchive: 1.0.1" below it. On the right side of the header, there is a user profile icon and the text "@its.aau.dk". Below the header, there is a "Home" button. The main content area features a "Dataset Search" section with a search bar containing the placeholder text "Type here to search in Datasets..." and a "Search" button. Below the search bar, there are two main sections: "Communities" and "DataDeposit Licensing Terms". The "Communities" section contains a circular icon of a classical building and a button labeled "Communities". The "DataDeposit Licensing Terms" section contains a circular icon of a gavel and a button labeled "DataDeposit Licensing Terms". At the bottom of the page, there is a footer with the text "DataDeposit DarkArchive: 1.0.1" on the left and a "Back to top" link on the right.

Data Management tips and tools from CLAAUDIA

- researcher.aau.dk – New! First version of an attempt to gather all relevant information, guidance and contacts for researchers at AAU.
 - [Research Data - Aalborg University \(aau.dk\)](https://researcher.aau.dk) – Entrance to CLAAUDIA's data science and management services.
 - FAIR training, Data management plans, Metadata, Repositories, High performance computing, Data science assistance and more.
- datadeposit.claudia.aau.dk: AAUs local repository for finalized data.
- [DMP Online](#): Data management plan templates with guides.
- Whisper transcription – NEW! AI transcription that can handle sensitive data.
- Decision tree for publication and licensing of research data (under development).

REPOSITORIES		
<small>DISCLAIMER</small> None of the repositories listed is certified as a trusted repository meaning that they, for example, do not live up to an ISO standard for security. AAU cannot guarantee data security in these repositories: the uploaded data is your responsibility. AAU encourage not publishing data that entails confidential and/or personal data in the listed repositories. You can safely publish all types of data, including confidential and personal, in AAU's repository DataDeposit.		
DOMAIN SPECIFIC	GENERIC	LOCAL
BioRxiv LIFE SCIENCE	APOLLO MULTIDISCIPLINARY	DataDeposit MULTIDISCIPLINARY
BMRB BIOLOGY	ArXiv MULTIDISCIPLINARY	
DRYAD MEDICAL	BRITISH LIBRARY RESEARCH MULTIDISCIPLINARY	
EUROPE PMC LIFE SCIENCES LITERATURE	FIGSHARE MULTIDISCIPLINARY	
GESIS SOCIAL SCIENCES	HARVARD DATAVERSE MULTIDISCIPLINARY	
LawArxiv LAW	Hyper Article en Ligne MULTIDISCIPLINARY	
MedRxiv MEDICINE AND HEALTH SCIENCES	MENDELEY DATA MULTIDISCIPLINARY	
PsyArxiv BEHAVIOURAL SCIENCES	OSF MULTIDISCIPLINARY	
RCSB PDB BIOLOGY	OSF MULTIDISCIPLINARY	
SocArxiv SOCIAL SCIENCES AND HUMANITIES	ZENODO MULTIDISCIPLINARY	
	<intr>Dok MULTIDISCIPLINARY	

It is possible to find more repositories through the following link:
[Repositoryfinder.datacite.org](https://repositoryfinder.datacite.org)



Contact info



claudia@aau.dk / support@its.aau.dk



claudia.aau.dk + researcher.aau.dk



CLAAUDIA community on yammer.com/aau.dk
(now Viva Engage)



[linkedin.com/company/claudia/](https://www.linkedin.com/company/claudia/)

