

# Discover the Data Archiving Guide (DAG)

A training event for new(ish) staff members

Speakers:

*Ilze Lace / SND*

*Iris Butzlaff / AUSSDA*

*Kim Ferguson and Maaïke Verburg / DANS-KNAW*

Organisers:

*Gry H. Henriksen / Sikt*

*Marielle Kappeler / FORS*

*13 October 2022*

*DOI: 10.5281/zenodo.7180525*



Licence: CC-BY 4.0



cessda.eu



@CESSDA\_Data

# CESSDA

- Consortium of European Social Science Data Archives (Service Providers).
- European Research Infrastructure (ERIC)
- 22 member and 1 observer countries

Consortia of universities departments

Departments within universities

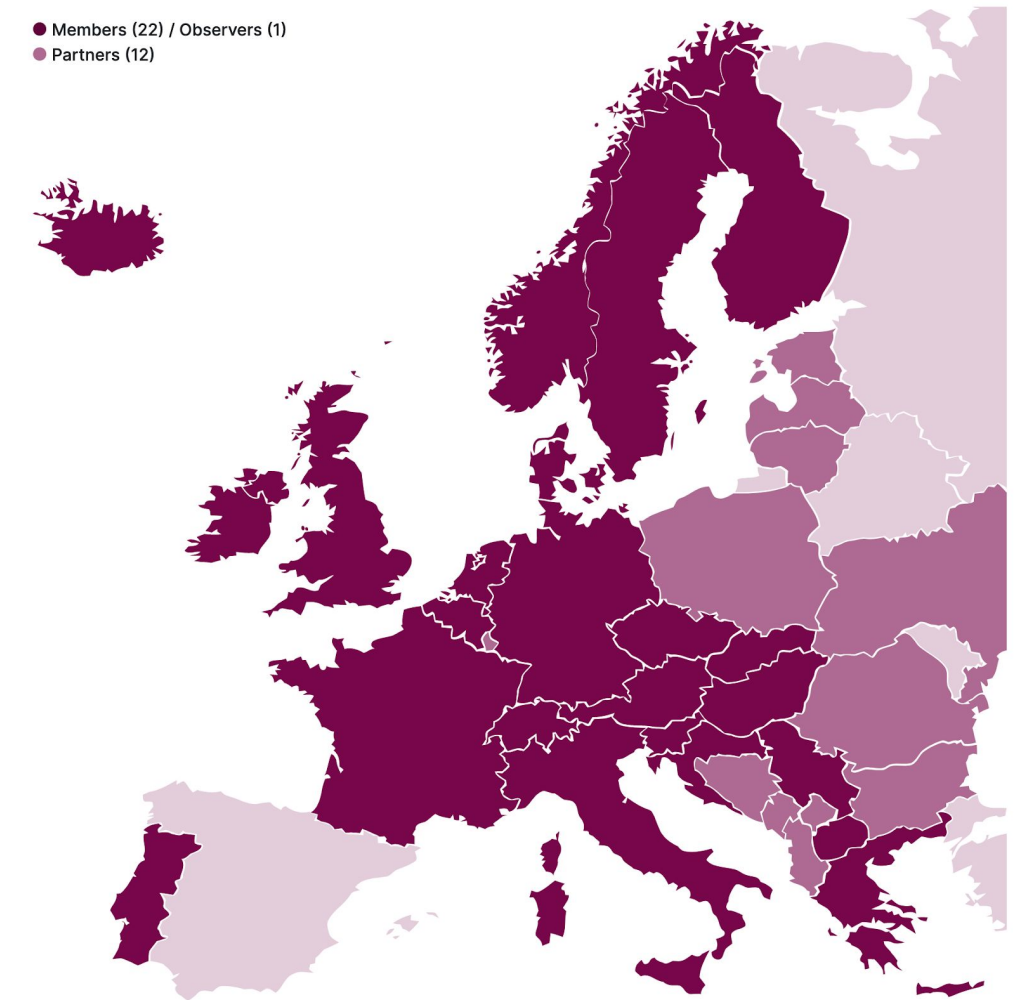
Public Research bodies

Consortia along with SSH ERICs of a country

## MISSION

- to provide a sustainable research infrastructure that enables the research community to conduct high-quality research in the social sciences
- to contribute to effective solutions to the major challenges facing society today.

1. Austria
2. Belgium
3. Croatia
4. Czech Republic
5. Denmark
6. Finland
7. France
8. Germany
9. Greece
10. Hungary
11. Iceland
12. Ireland
13. North Macedonia
14. Netherlands
15. Norway
16. Portugal
17. Serbia
18. Slovak Republic
19. Slovenia
20. Sweden
21. Switzerland (Observer)
22. UK
23. Italy




Countries aiming at membership:

- Bosnia and Herzegovina
- Bulgaria
- Estonia
- Lithuania
- Poland
- Romania
- Spain



# Services Supporting Research

### Training



Learn how to discover, use, manage and preserve research data

### Upcoming Events


**Tue 6 Sep 2022**  
**Dataverse train-the-trainer event in September**  
In September, DANS will be organising a train-the-trainer event on using Dataverse as an archiv...

**Thu 15 Sep 2022**  
**New Data Types in Data Management and Archiving**  
New Data Types in Data Management and Archiving workshop will focus on the management, archivin...


[Event Calendar](#)

### Featured Resources

The [Data Management Expert Guide \(DMEG\)](#) is a comprehensive guide to Research Data Management and FAIR data principles.


 [Data Management Expert Guide](#)

The [Data Archiving Guide \(DAG\)](#) is designed to provide new employees at social science data archives with a general understanding of the work a data archive performs.

 [Data Archiving Guide](#)

## CESSDA Digital Tools

### CESSDA Data Catalogue




Search tens of thousands of social science research studies from our European Service Providers.

### ELSST Thesaurus



The European Language Social Science Thesaurus is a broad-based multilingual thesaurus for the social sciences.

### European Question Bank




The EQB is a cross-national question bank for social science and humanities research.

### Vocabulary Service



Search, browse and download controlled vocabularies in a variety of languages.

### Resource Directory



Access resources for data archives and data professionals from CESSDA, its Service Providers and partners.

### Metadata Validator



Validate metadata for compatibility with the CESSDA Data Catalogue and the European Question Bank.



cessda  
**Roadshow on COVID-19**  
SAVE THE DATE



cessda  
**Roadshow on migration**  
08 October 2021 | SAVE THE DATE




cessda  
**Roadshow on Cancer and Chronic Diseases**  
SAVE THE DATE



cessda  
**Roadshow on Circular Economy**  
SAVE THE DATE  
28 October 2021  
14:00-15:30 CEST  
13:00-14:30 BST

## Trust and Standards

Each national Service Provider works towards achieving the Trustworthy Digital Repository (TDR) standard selected by CESSDA: the [CoreTrustSeal](#).



The CoreTrustSeal defines repository requirements in terms of organisational infrastructure, digital object management, technology and security. Trust is essential between CESSDA Service Providers and with their data depositors and users.

Since CESSDA selected the internationally recognised CoreTrustSeal it has been acknowledged as important for enabling FAIR (Findable, Accessible, Interoperable, Reusable) data and has become the recommended certification approach for data repositories within the European Open Science Cloud (EOSC).



## Start

### Introduction

### Chapter 1: Data Archives - a Quick Introduction

### Chapter 2: Policies of Data Archives

### Chapter 3: Pre-ingest - from early contact to data transfer

### Chapter 4: Ingest and Curation

### About this Guide

# CESSDA Data Archiving Guide

The Data Archiving Guide (DAG) is designed to provide new employees at social science data archives with a general understanding of the work a data archive performs.

**Note: the DAG is under development.**

The information in the DAG was collected by experts from several [CESSDA social science data archives](#) reflecting the procedures and policies at their local archives. While the context of these archives varies and may well be different from your own archive - in size, the underlying technical architecture or in the support you provide to researchers - the DAG focuses on common ground and is a useful tool for professionals new to data archiving.

[In this guide](#), you will learn what a data archive is, what it does, and why data archives play an important role in the scientific infrastructure. After reading the DAG, you will have a basic understanding of the working procedures of a data archive and you will also be able to name the set of policies that a data archive needs to have in place to function optimally.

## Chapter 1

### Data Archives - a Quick Introduction

Gain an overview of why data archives exist and what their mission, function and way of operating is.

[Read this chapter](#)

## Chapter 2

### Policies of Data Archives

Information about the types of policies (and some related documents) a social science data archive has most typically has in place.

[Read this chapter](#)

## Chapter 3

### Pre-ingest

What data archives do to ensure that incoming data meet the criteria of data collection and quality requirements.

[Read this chapter](#)

## Chapter 4

### Ingest and Curation

The steps required to guarantee high quality data, documentation material and metadata for secure archiving and dissemination.

[Read this chapter](#)

# Outline for today

Time	Topic	Presenter
09:30-09:45	Welcome & Logistics	Marielle, Gry
09:45-10:00	Presentation: Pre-Ingest	Ilze
10:00-10:15	Presentation: Ingest	Iris
10:15-11:00	Group Work	Intro: Marielle
11:00-11:15	Coffee / Tea	
11:15-12:00	Presentations & Discussion	All
12:00-13:30	Lunch Break (incl. small homework)	
13:30-14:00	Welcome back, rapid fire homework	All
14:00-15:10	SneakPeak New Fair chapter + Exercise	Maaïke & Kim
15:10-15:25	Coffee / Tea	
15:25-16:00	Glossary Quiz	Intro: Gry
16:00-17:00	Wrap-up & Closing discussion	Intro: Gry & Marielle, All

# Logistical information for this workshop

- The presentations for the Workshop will be **recorded** (Breakout Sessions will not be recorded / Exercises: will be recorded, but not shared)
- Asking **questions** around presentations:
  - Type in the chat *during* presentations
  - Raise hand *after* presentations
- **Materials** from the event:
  - Slides will be published on Zenodo
  - Link to the recording and materials will be sent to registered participants by e-mail
- We kindly ask you to fill out the **evaluation survey** after the workshop (will be sent via e-mail)



# Word Finder game

<https://thewordsearch.com/puzzle/4155175/cessda-dag-workshop-13-october/>

## The WordSearch

### CESSDA DAG workshop 13 October

L	E	N	T	U	D	C	E	A	I	N	S
R	L	D	A	S	I	E	M	E	A	A	P
R	B	O	A	E	S	S	E	A	I	O	O
A	A	C	R	R	S	S	T	D	A	G	N
C	R	U	E	C	I	D	A	E	N	A	O
C	E	M	G	O	M	A	D	S	E	C	I
E	P	E	N	N	I	R	A	F	L	U	T
S	O	N	A	D	N	E	T	A	B	R	I
S	R	T	M	I	A	U	A	I	A	A	S
A	E	A	A	T	T	S	D	R	D	T	I
B	T	T	T	I	I	A	L	O	N	I	U
L	N	I	A	O	O	B	E	I	I	O	Q
E	I	O	D	N	N	L	I	O	F	N	C
L	S	N	L	I	C	E	N	S	E	M	A

DAG  
DISSIMINATION  
ACCESSABLE  
DOCUMENTATION  
ACQUISITION  
DOI  
FAIR  
CURATION  
DATAMANGER  
INTEROPERABLE  
LICENSE  
USERCONDITION  
METADATA  
CESSDA  
REUSABLE  
FINDABLE

 Download / Print Puzzle  Puzzle Settings  Game Theme  Report a bug

# Thank you!

---

[marielle.kappeler@fors.unil.ch](mailto:marielle.kappeler@fors.unil.ch)

[gry.henriksen@sikt.no](mailto:gry.henriksen@sikt.no)

 cessda.eu

 @CESSDA\_Data

  Licence: CC-BY 4.0



# CESSDA Data Archiving Guide

---

## Chapter 3: Pre-ingest

*Ilze Lace, Swedish National Data Service*

# Setting DAG Chapter 3 Pre-Ingest in context

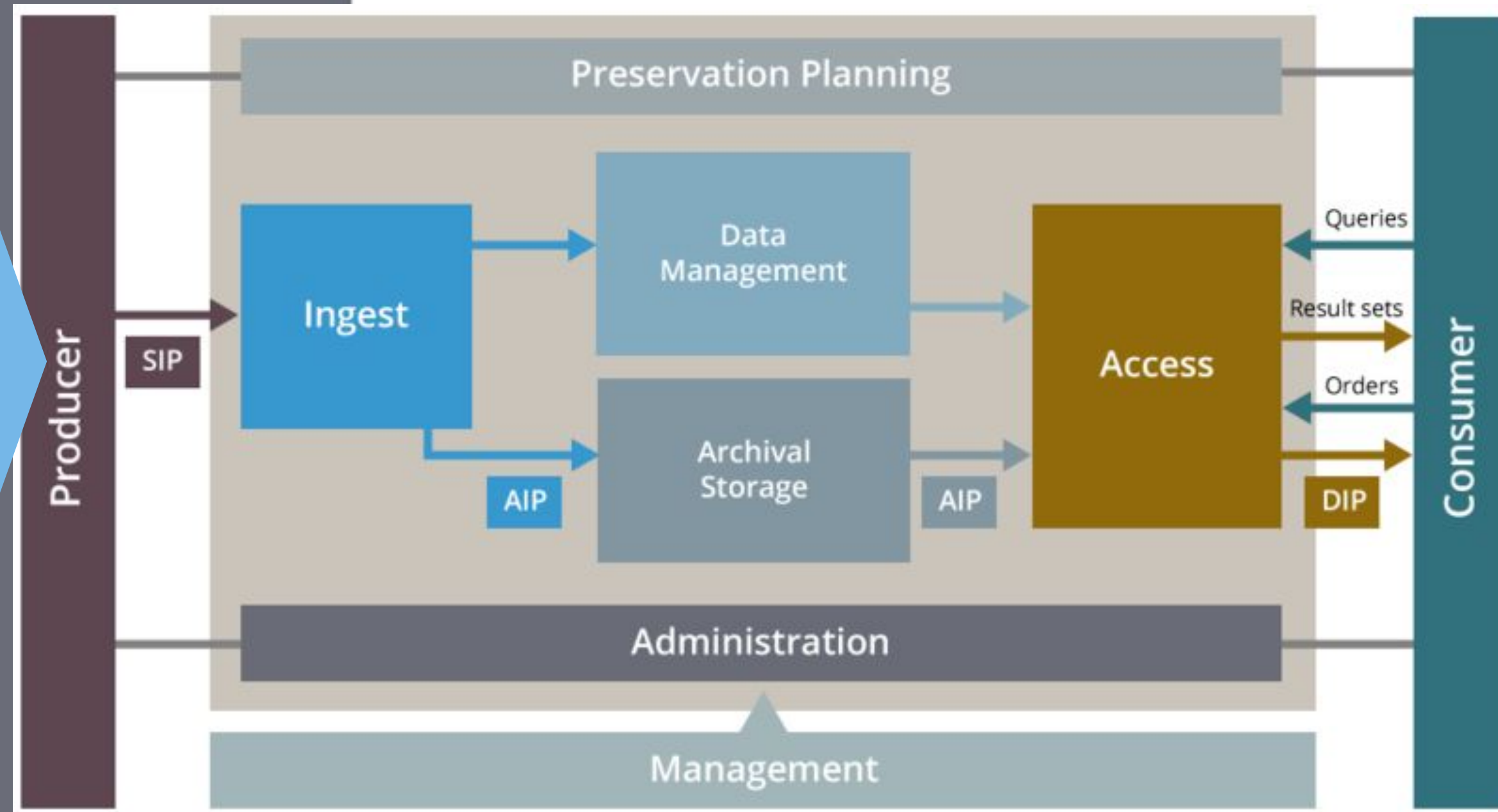
- Scope:
  - from early contact to data transfer
  - “before-OAIS”
  - accumulated “best practice” and acceptable solutions
- Target groups:
  - new staff of social sciences data archives
  - staff of new social sciences data archives
- Structure:
  - timeline-based (though not entirely)
- Links to other CESSDA Training resources:
  - CESSDA DMEG
  - CESSDA Resource Directory

# Pre-ingest: From early contact to data transfer

1. Outreach and support
2. Data submission
3. Data review and appraisal
4. Accepting or rejecting data



OAIS Model



*CESSDA Tutorial:OAIS (2017)*

5. Further reading



# Pre-ingest: From early contact to data transfer

## 1. Outreach and support

## 2. Data submission

## 3. Data review and appraisal

## 4. Accepting or rejecting data

## 5. Further reading

### **Identify** and localise data:

Systematic data inventories, targeting specific data, rescuing data, following requests

### **Advocate** for data sharing:

Focus on benefits and opportunities

Address fears and practical constraints

### **Negotiate** data sharing:

Data access levels and embargos; legal and ethical obligations as well as formats and documentation

### **Support and consult:**

RDM training & consultations on data archive services

# Pre-ingest: From early contact to data transfer

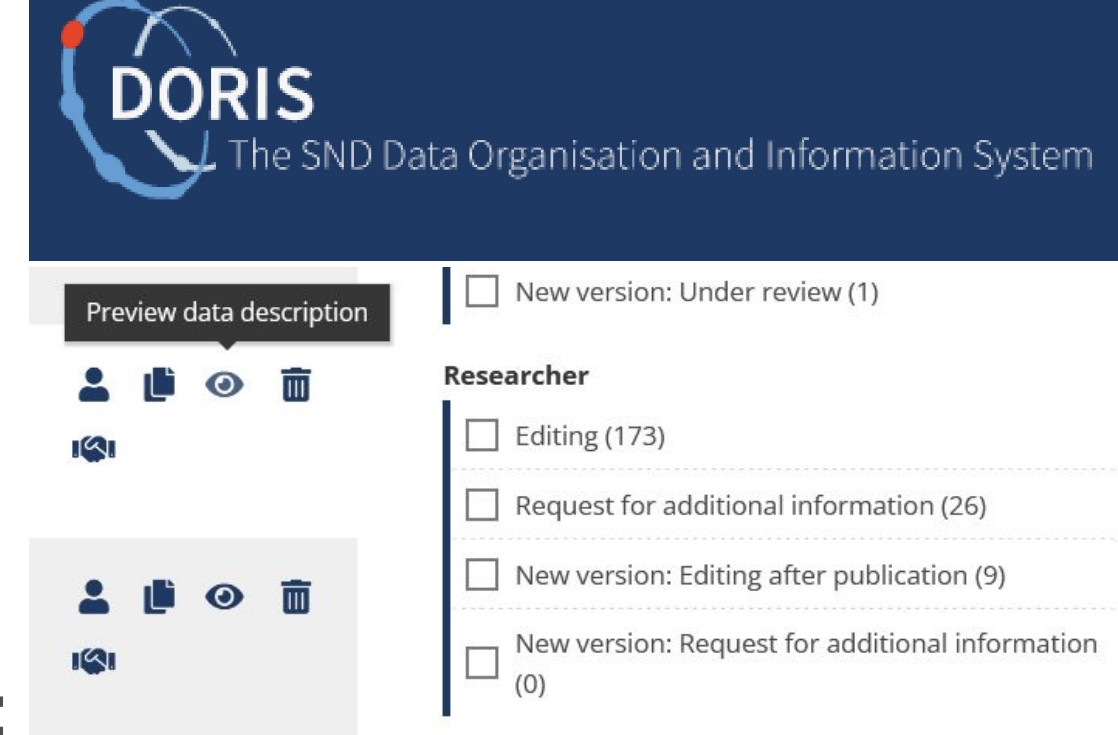
1. Outreach and support

**2. Data submission**

3. Data review and appraisal

4. Accepting or rejecting data

5. Further reading



**Inform** depositors:

- .. on submission **requirements**
- .. describing of the **main steps** in the deposit process
- .. **instructing** how to proceed

Ensure data **transfer**:

- ..an **acceptable solution** for incoming data, including special cases

**Administer** incoming data:

**Alerts** & assigning new cases

**Receipt** to the depositor

**Transfer** data to temporary storage

# Pre-ingest: From early contact to data transfer

1. Outreach and support

2. Data submission

**3. Data review and appraisal**

4. Accepting or rejecting data

5. Further reading



Control compliance to **acquisition** criteria:

Criteria should be in the Acquisition policy

**Control** the submitted material:

Depositor ID; completeness of material; virus and readability; data description; ethical and legal aspects

**Review** the documentation:

..ensure there is enough context information

**Communicate** with researchers:

..on-going process

..goal-oriented

..proactive and supportive



# Pre-ingest: From early contact to data transfer

1. Outreach and support
2. Data submission
3. Data review and appraisal
4. **Accepting or rejecting data**
5. Further reading

## Accept:

- .. **notify** researchers and **explain** further steps
- .. ensure there is a **legal ground**
- .. **alert** Ingest-team

## Reject:

- .. **inform** depositor explaining reasons
- .. **suggest** a more appropriate repository
- .. **remove** data from temporary storage

# Pre-ingest: From early contact to data transfer

- 1. Outreach and support
- 2. Data submission
- 3. Data review and appraisal
- 4. Accepting or rejecting data
- 5. Further reading**

zotero

CESSDA Resource Directory | Zotero

Group Libraries

CESSDA Resource Directory

▶

1. Organisation

▼

2. Digital object management

2.1. Pre-Ingest - Acquisition

2.2. Ingest - Curation

2.3. Access - Dissemination

2.4. Preservation

▶

3. Communication & support to us...

▶

4. Technical infrastructure

consultancy

data management policy

FAIR data

file format policy

guidance

↑

"

≡

📖

...

Title

📄

Archival Acquisition Policy DARIS

📄

CESSDA SaW D4.3: Overview of Data Management

@

CESSDA Training Group

📄

CESSDA Tutorial: OAIS

📄

Collection Development Policy

📄

Collections Development Selection and Appraisal

📄

da|ra Metadata Schema: Documentation for the

# How can Ch3 of DAG be used for solving real life issues with data in your organizations?

A: Working with data depositors and research communities?

B: Developing routines for better quality data?

Focus on data types you are now working with - examples might be survey data; qualitative interview data; audio & video data; archeological data; big data; sensitive data or any other type of data.

C: Organizing and structuring Pre-ingest activities, like data submission, review and appraisal, communication with researchers?

Discuss Ch3 from perspective of case A, B or C:

What seems useful?

What seems less applicable?

What is missing?

If you have had an opportunity to look at DAG before: How well is material presented, regarding structure, granularity, level of detail, etc.?



# How can DAG be used for solving real life issues with data in your organizations? **Choose A, B or C.**

If you have had an opportunity to look at DAG before:

How well is material **presented**, regarding **structure**, **granularity**, **level of detail**, etc.?

**What seems useful?**

**What seems less applicable?**

**What is missing?**

**Case A:** Working with data depositors and research communities?

**Case B:** Developing routines for better quality data?

Focus on data types you are now working with - examples might be survey data; qualitative interview data; audio & video data; archeological data; big data; sensitive data or any other type of data.

**Case C:** Organizing and structuring Pre-ingest activities?

Focus on processes like data submission, review and appraisal, communication with researchers or other.

# Thank you!

---

*Ilze Lace | SND, University of Gothenburg*  
*Ilze.Lace@gu.se*

# CESSDA Data Archiving Guide

---

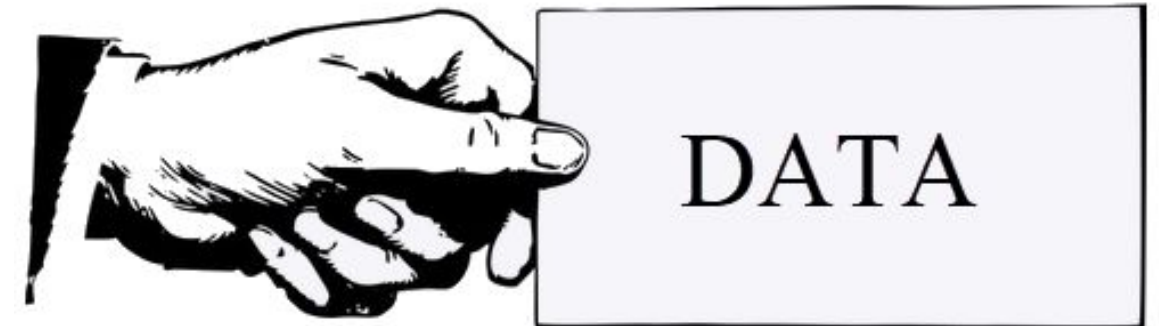
## Chapter 4: Ingest and Curation

*Iris Butzlaff, Austrian Social Science Data Archive*



# Structure

1. Handover from acquisition/pre-ingest to ingest
  2. Different services (workflows) for different data deposit agreements
  3. Metadata
  4. Quality assurance of data and documentation material
  5. Updates and versioning
  6. Adaptation of workflows and processes
  7. Checklist of general ingest procedure steps
- 
1. *Sources and further reading*
  2. *Annex*



<https://dag.cessda.eu/Chapter-4>

# Handover from pre-ingest to ingest (i)

- Checks on completeness of data and documentation material;
- Checks on formats;
- Checks on sensitive content;
- Checks on encrypted or password protected data.

*check out: Data Collection Policy of your repository*



# Handover from pre-ingest to ingest (ii)

## Transfer:

*How we do it: Transition from pre-ingest to ingest*

### Notification that data is ready for ingest

- Option A: Automated notification through software system
- Option B: E-Mail from pre-ingest agent to ingest agent
- Option C: Other form of notification

### Transfer of data

- Option D: Downloading from data transfer system (archive material has been transferred directly from depositor to this system)
- Option E: Email the data (Forward)
- Option F: Download from server (where pre-ingest agent has stored it)

## Keep track with software:

*How we do it: examples*

- Project Management Software (e.g., Teamwork)
- Web Application
- Excel Lists ([see Annex](#))

# Data deposit

- **Mandatory**

- Signed contract (clarified property rights, and signatory powers)
- Metadata information (if not already entered directly in platform)
- Data files
- Codebook
- Methods report
- Instruments of data collection (e.g., questionnaire with interviewer instructions, information material for respondents, ...)
- Informed consent form (if applicable)

- **Optional**

- Project report
- Data Management Plan (DMP) of project proposal
- Project proposal
- Interviewer guidelines
- Interview cards etc
- Documentation about incentives, contacts, ...
- Recoding protocol
- Syntaxes
- Coding instructions

Showcards

- Data types and formats used in community

- Quantitative data (Stata, SPSS, SAS, R, Excel);
- Qualitative data (PDF, text programmes);

- Preferred formats

- Provide a list of your preferred formats.



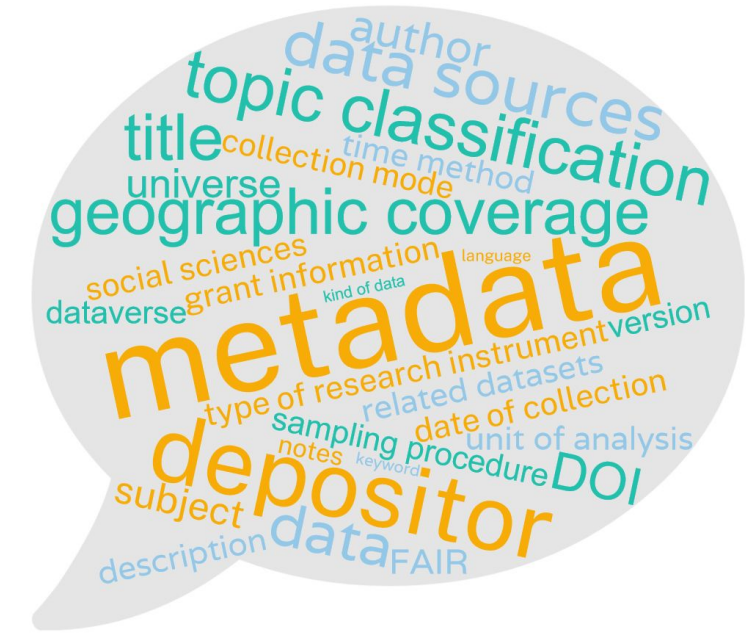


# Different services (workflows) for different data deposit agreements

- **Open Access** (OA, Public Use) – maximal re-use of the data;
- **Scientific Use** (SUF) – restricted account-based access aims at a re-use by researchers with a scientific interest in the data (also restricted controlled access);
- Replication Data;
  - Which audience accesses the data;
  - Different levels of pseudonymisation.
- Special conditions; e.g., embargo or restricted controlled access (internal process).
- Re-use or preservation only?

# Metadata

- Make metadata **FAIR**, license CC;
- Use **Vocabularies**:
  - international standards set by the Data Documentation Initiative (DDI) and the Consortium of European Social Science Data Archives (CESSDA ERIC),
  - mandatory metadata items for harvesting the CESSDA Data Catalogue (CDC),
  - different languages.
- Be clear about free text fields (Capital letters, British vs. American English, Abbreviations, Names of institutions, persistent identifiers in links).



# Check metadata

- Did you receive all [relevant metadata information](#)? Are all mandatory fields filled in? Do you have more information on optional metadata fields?
- Metadata fields should not contain any separators (e.g., like dots or commas as thousands separators), depending on the tool you use for dissemination so that easy harvesting of your entered metadata is assured.
- Are the metadata fields filled in correctly, according to CESSDA CMM and DDI standards?
- For topics and keywords; adhere to the CESSDA Topic Classification and CESSDA ELSST Thesaurus. Is there a maximum of topics the depositors can pick?
- Give the researchers assistance in choosing the right term, neither too broad nor too narrow.

# Quality assurance of data and documentation material (i)

- A. Risks for integrity (checks for compliance with General Data Protection Regulation - GDPR).
  - Removal of all direct identifiers and (strong) indirect identifiers;
  - Check or delete answers to open questions;
  - Check data whilst keeping population group and sample size in mind;
  - Spellcheck of variable names, labels and string variables.
- B. Checks for compatibility with other formats.
  - Data: Technical setting, length of variables, scanning for unlabelled values;
  - Documentation: PDF/A standard (ISO 19005-2); ASCII text format.
- C. Plausibility (checks on logical errors, improbable values/outliers, duplication errors, correct filters).
- D. Comparison of data and documentation material



## Quality assurance of data and documentation material (ii)

- E. Give feedback to depositors (does depositor or ingest staff do adaptations).
- F. Adding DOI and version as variables to the dataset.
- G. File naming and managing files (user can derive various information about file type and content).
- H. Conversion to other formats and provenance (data and documentation material is transformed; who does this? Ingest or preservation staff?, keep track).
- I. Dissemination (software systems used for data dissemination; e.g. Dataverse, Nesstar).

*It is advisable to have a checklist for all necessary steps in ingest (see "Checklist of general interest procedure steps").*



# Updates and Versioning

- Set clear **rules for workflows** in case of **updates, new versions, changes** to the data entry to ensure **reproducibility**:
  - Tracking (internal recording; errata files);
  - Be clear about structure according to OAIS model;
  - Major vs. minor version changes.

The ingest team needs a clear structure on **naming** of different versions. It must be clear and reproducible for the data curators what are the most recent data in the DIP and also in the SIP to not mix things up by accident.

# Adaptation of workflows and processes

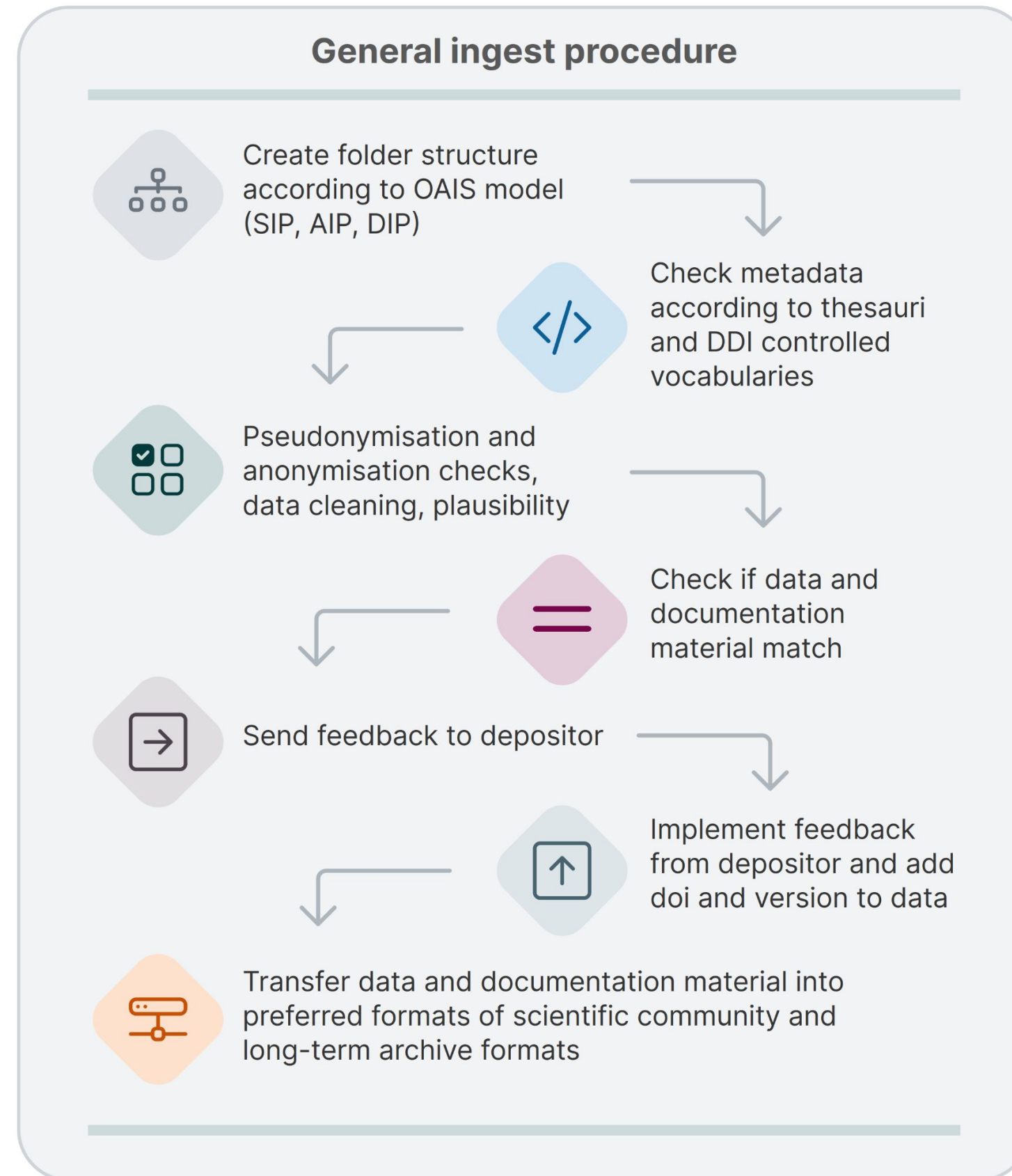
Upcoming incidents or situations could require new workflows (e.g., COVID-19 pandemic, or a handover of a bulk of datasets in form of a [pre-mortem] bequest):

Against this background, several national data archives have released portals or platforms for accelerating COVID-19 research, so that researchers and the public easily can find, share and work with research data related to the corona virus. The [CESSDA data catalogue](#) (CESSDA n.d. [Accessed July 30, 2022c]) has collected the COVID-19 activities from CESSDA service providers (see this [list](#) with links) (CESSDA n.d. [Accessed July 30, 2022b]).

National data archives in Austria, Belgium, Czech Republic, Denmark, Finland, France, Germany, Netherlands, Portugal, Slovenia, Sweden, Switzerland, and the United Kingdom provide datasets related to the Coronavirus and the pandemic situation and therefore have established new (panel) survey data or added related questions to already existing panel surveys (often in English and/or the national language). The archives launched fast-track procedures for depositors to be able to offer quick access to the data.



# Example of ingest process





# Checklist

- ☐ Transfer archive material into working area (correct drive);
- ☐ Create folders according to OAIS model (SIP, AIP, DIP);
- ☐ Are all mandatory metadata fields filled in correctly?
- ☐ Are topics and keywords chosen from the respective thesauri and vocabularies;
- ☐ Pseudonymisation/Anonymisation checks;
- ☐ Data cleaning (variables and values labelled, missing values);
- ☐ Plausibility checks;
- ☐ Check if data and documentation material complete and match;
- ☐ Include suggested citation and information (DOI, version) in data and documentation;
- ☐ Send feedback to depositor;
- ☐ Transfer the data and documentation material into preferred formats of scientific community and long-term archive formats;
- ☐ Make clear all undertaken steps are traceable and documented;
- ☐ Publish archive material in chosen platform (e.g., Dataverse);
- ☐ Clean up all folders and move to repository.



# Thank you!

---

*Iris Butzlaff, AUSSDA, University of Vienna*



# Questions for the discussion

## 1. Hand-over from acquisition/pre-ingest to ingest

- Which programmes/software does your repository use to notify the ingest staff that data has ‘survived’ pre-ingest and is now ready for ingest?
- How does pre-ingest agent transfer the data:
  - from data depositor to repository
  - from pre-ingest agent to ingest agent
- Which software/tools does your repository use to keep track of status the data is in?
- Which data formats do you use in case your archives process qualitative formats?

## 2. Different services (workflows) for different data deposit agreements

- Does your repository use different workflows for different data deposit agreements? If yes, how do these workflows differ from each other?
- Are there any other special conditions that your repository offers to the data depositors, except for embargo or restricted access (or controlled access)?
- Does your repository offer to archive data without making it accessible to the community? If yes, is this data handled any different from ‘normally’ archived data?

## 3. Metadata

- Is there anything to add that needs to be considered except for using correct vocabularies, clear rules for free text fields and checks on metadata?

## 4. Quality assurance of data and documentation material

- Which checks do you do in your archive except for mentioned ones in sections A. to D.?
- Are data conversions done by ingest or by preservation staff in your archive?
- Which software systems do you use for dissemination? (other than Dataverse and Nesstar?)
- Does your software system (which you use for dissemination) allow for online analysis of the data?

## 5. Updates and versioning

- How do you record necessary updates?
- How do you comply with the OAIS model in case you receive new versions?

## 6. Adaptation of workflows and processes

- How often do you update your workflows?
- Is this done by management level or by ingest staff?

# Chapter 5: FAIR-enabling and Trustworthy Qualities of Archives

---

What is FAIR and how is it enabled within organisations and throughout designated communities?

*Maike Verburg / DANS-KNAW (NL)*

*Kim Ferguson / DANS-KNAW (NL)*



# Agenda

- Introductions - who we are and where we're from
- Overview of DAG Chapter 5\*
- Hands-on session with FAIR-Aware
- Questions & wrap-up

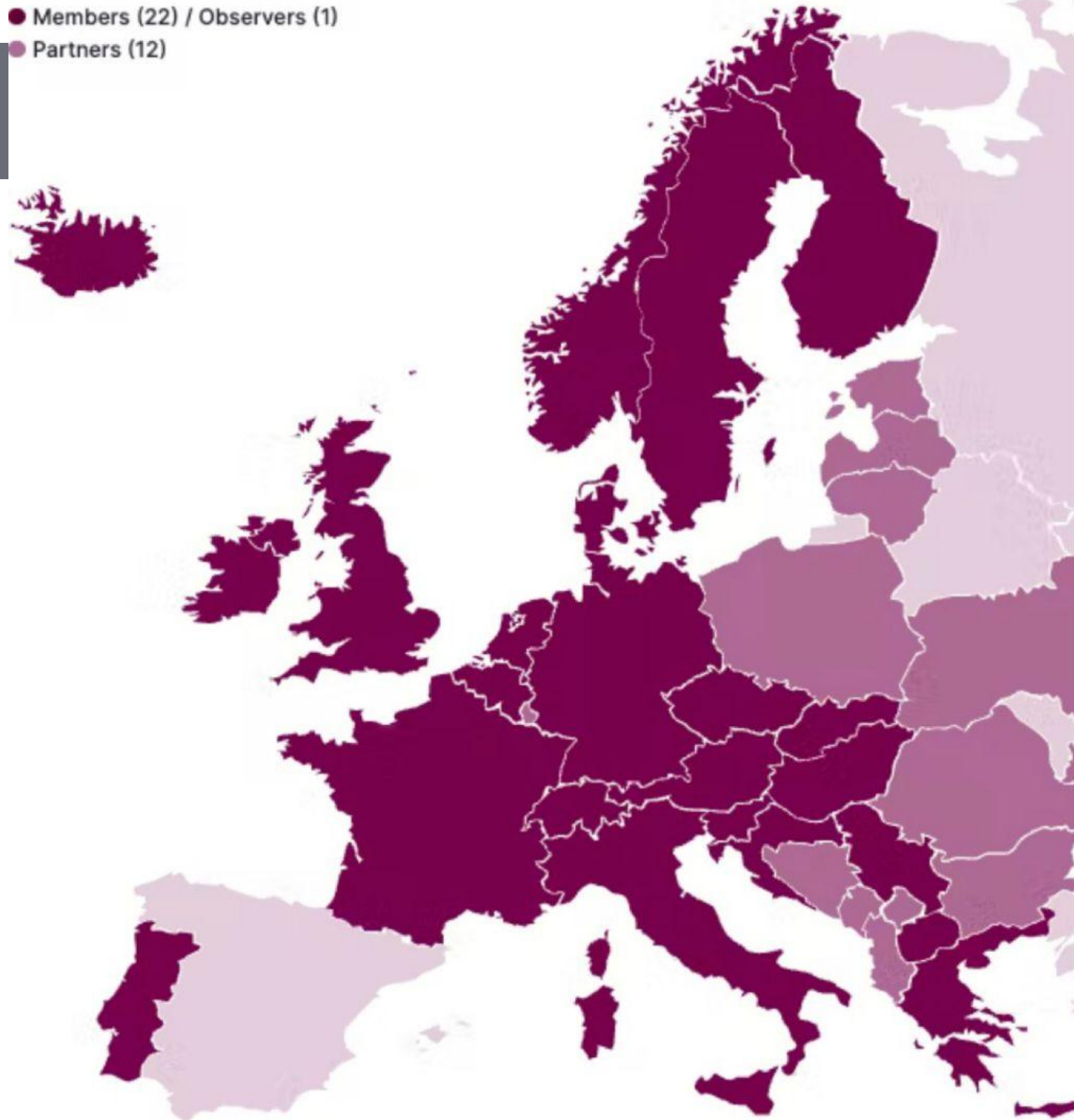
\* CESSDA Training Team (2022). CESSDA Data Archiving Guide version 1.0. Bergen, Norway: CESSDA ERIC.



Let's head over to menti:

Mentimeter information

● Members (22) / Observers (1)  
● Partners (12)



In which country or region is your archive/repository based?

Data Archiving and Networked Services  
**DANS**

switzerland

norway

estonia

slovenia

4







# What is your role within your organisation?

datamanager

project officer

research data specialist

data archivist

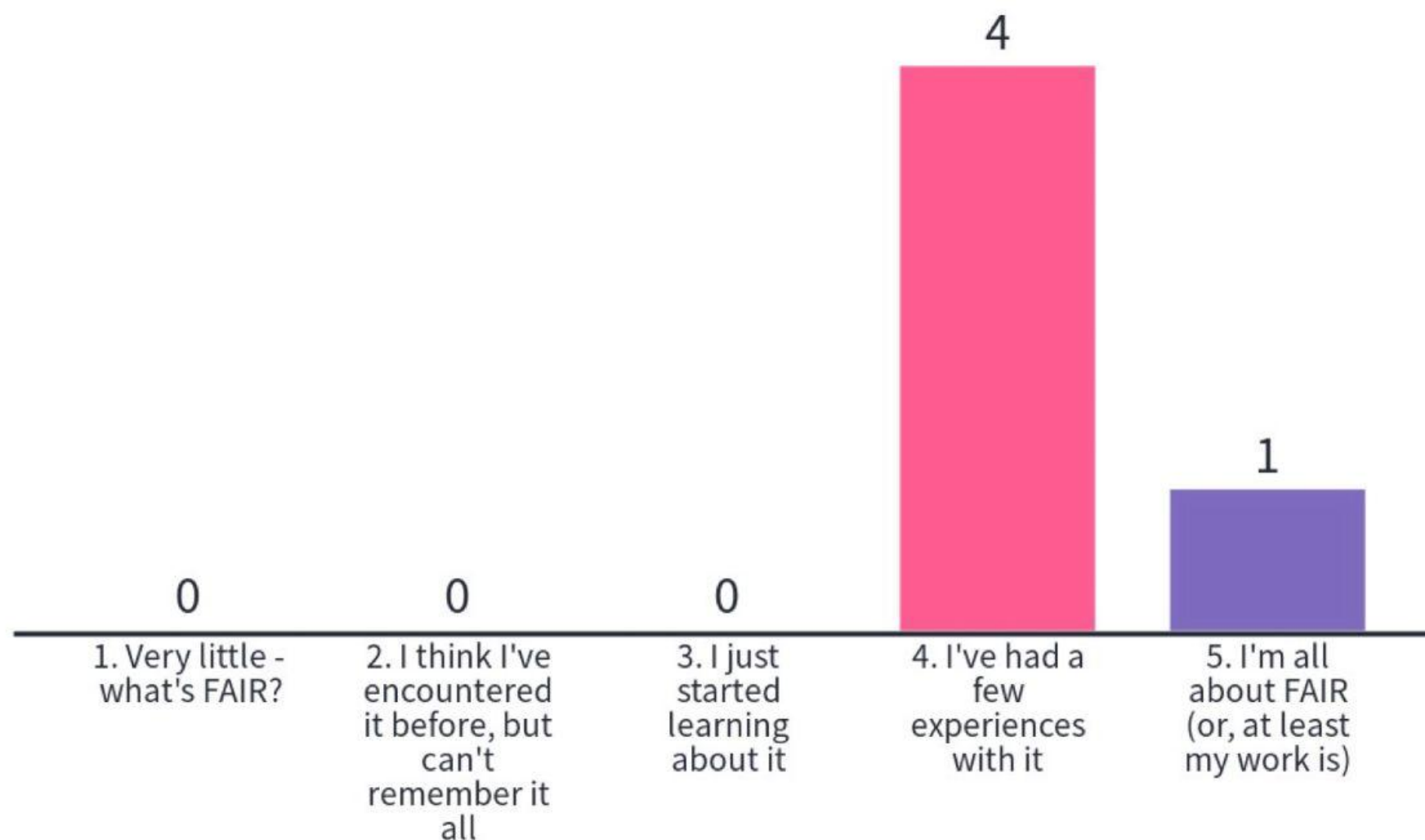
data steward

data protection advicer



# What is your experience with regards to FAIR?

## Scale from 1 to 5





# What is in this chapter?



5.1 Introduction to FAIR

5.2 Enable FAIR in your organisation

5.3 Promote FAIR within your designated community

5.4 Trustworthy data archives

# What is in each section?



- Information about the topic
- Examples from CESSDA Archives
- Questions to increase your understanding
- Expert tips

# 5.1 Introduction to FAIR

- The FAIR Guiding Principles
- Why were they created?
- One term of many
- Becoming FAIR-enabling



## **F**indable

To aid automatic discovery of relevant datasets, (meta)data should be easy to find by both humans and machines and be assigned a persistent identifier.

## **A**ccesible

Limitations on the use of data, and protocols for querying or copying data are made explicit for both humans and machines.

## **I**nteroperable

(Meta)data should use standardised terms (controlled vocabularies), have references to other (meta)data and be machine actionable.

## **R**eusable

(Meta)data are sufficiently well described for both humans and computers to be able to understand them and have a clear and accessible data usage license.

Image: CC-BY-SA CESSDA DMEG <https://dmeg.cessda.eu/>

## 5.2 Enable FAIR in your organisation

From the soon-to-be released chapter, “Archives differ, whether by domain, country, or even concerning their approach to FAIR principles.

An organisation can be FAIR-enabling in a number of ways, including through policy, infrastructure, mission statements, assessments, training protocols, or simply the work culture.

This section will introduce a variety of tools available to assess FAIR qualities of archives, followed by some examples from different CESSDA Service Providers.”

## 5.2 Enable FAIR in your organisation *cont.*

How?

- Assessment frameworks
  - Manual (e.g. FAIRsFAIR framework)
  - Automatic (e.g. F-UJI)
- Examples from other archives
  - CROSSDA (Croatia)
  - FSD (Finland)



## 5.3 Promote FAIR within your designated community

“While Section 5.2 addressed ways that organisations can enable FAIR internally, this section focuses on the role of a data archive in educating their designated communities on FAIR principles.

There is not one approach that fits every archive, but rather you should determine what makes sense for your archive and your designated community.

We will introduce the different perspectives as well as give some examples from CESSDA Service Providers.”

## 5.3 Promote FAIR within your designated community *cont.*

A few ways to approach this:

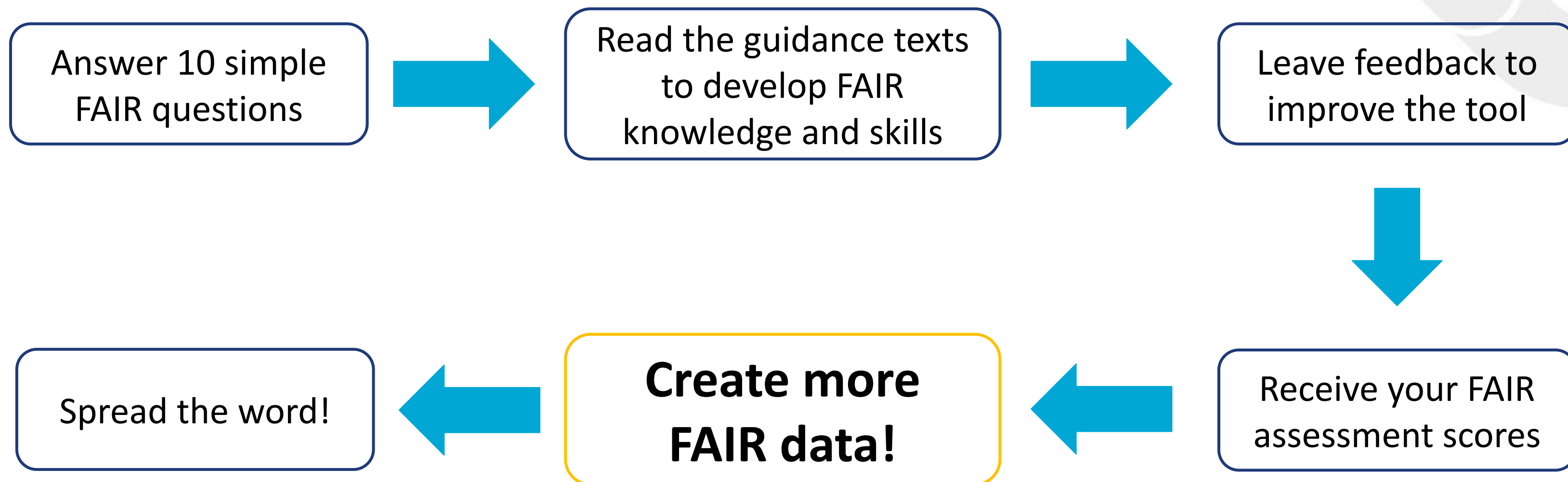
- **Promoting** FAIR, such as through user guides, training, events, tools, or news bulletins
  - Educating users, making them aware of FAIR
- **Embedding** FAIR, such as through infrastructure and depositing requirements
  - Supporting users, without burdening them with extra information

Both approaches are possible, but you can also focus on one instead

## 5.4 Trustworthy data archives

- Why should an archive be trustworthy?
- How can an archive display trustworthiness?
  - Certification
- CESSDA Trust Support
- FAIR + Trustworthiness = FAIR-enabling Trustworthy Digital Repository







# Let's get hands-on!

- Section 5.1 suggests the **FAIR-Aware** tool as a means to assess and increase your understanding of the FAIR Principles
- Aimed at researchers and data stewards, but relevant to all data professionals!
- How is the archive presented in the guidance information? What is expected of you to help create and maintain FAIR data? Let's have a look!



# FAIR-Aware Exercise

What elements of FAIR are in the hands of the archive or repository? Let's find out!





Are you aware that the data repository providing access to your data(set) should make the metadata describing your data(set) available in a format readable by machines as well as humans?

What does this mean?

This question refers to the capability to make the metadata accessible online in a standard and machine-readable format. **Machine-readability** means that the data is presented in a structured **format** that computers can read and process. Facilitating this process is a responsibility of the data repository you deposit your data(set) in.

Why is this important?

Ensuring that the metadata describing your data(set) is machine-readable, it will be accessible to the systems that collect (also known as **harvesting**) and aggregate data from search engines or databases (e.g., Google Search, Web of Science, or a university library collection). This improves your chances of having your data(set) cited and reused, because it will reach a larger audience. Without machine-readable metadata, your data(set) will only be found by people searching the data collection of the specific repository you deposited in or those that have a direct link to your data(set).

How to do this?

Most digital data repositories will have some kind of protocol for making metadata machine-readable. Two protocols that support FAIR (because they are open, free, and universally implementable) are OAI-PMH [↗](#) and REST API [↗](#). Therefore, even though it is the responsibility of the data repository to carry out this task, it is your responsibility to select the right data repository to meet this requirement. You can search for such a data repository on a registry such as Re3data [↗](#) by filtering on 'API'.

### 3. Metadata and (machine/human) readability



0  
No

0  
Uncertain



# How is this communicated to your organisation's users?

not sure if we communicate about it

Via website, agreement, open policy

When users deposit data

User guide. Also you can add machine friendly metadata yourself





## 4. Controlled access via licences

Can access to your data(set) metadata should include licence information? Can your data(set) be reused?

?

be public domain and openly accessible. Legitimate reasons not to share data (e.g. legal constraints). As such, it is your responsibility to whom and when, and to take appropriate measures as closed as necessary.



0  
No

0  
Uncertain





# What kind of licenses are available at your archive?

CC and Closed Contract (with option to restrict teaching/research and add prior agreements)

CC-licences, closed contracts

CC-0, CC-BY

CC licenses

CC, GNU, MIT...

CC, restricted,



Are you aware that the metadata describing your data(set) should use controlled vocabularies?

What does this mean?

There are many different ways you can describe the same information when filling out metadata for your deposit. To prevent ambiguity and facilitate better findability, interoperability, and machine-readability, you should use a **controlled vocabulary** for your metadata.

Controlled vocabularies are lists of terms that are created for specific uses or contexts. They are a type of **semantic artefact** and can take the form of, for example, an ontology, a thesaurus, or taxonomy. Each type of vocabulary comes with a different degree of sophistication (e.g. in their level of expressiveness, structure, and inferential power).

Why is this important?

Using controlled vocabularies, the discovery, linking, understanding, and reuse of research data are improved. Using controlled vocabularies in metadata facilitates advanced data search because people will not have to guess the exact terms you used to describe your data(set) to find it. It also helps facilitate better interoperability of data from different sources, since it will be clear that data(sets) using the same terms cover the same information.

Data repositories should provide support for the use of controlled vocabularies in metadata by offering relevant functionalities. They will often display which controlled vocabularies they support on their website. When controlled vocabularies are included in metadata, your data repository of choice may be able to publish the metadata in a machine-readable format, thus greatly increasing their machine actionability.

## 6. Are you aware that the metadata describing your data(set) should use controlled vocabularies?



0  
No

0  
Uncertain





# Vocabulary Service

SDA Vocabulary Service enables users to create and control vocabularies in a variety of languages in SKOS, HTML and PDF formats, in the language of the service also contains an Editor, where authorised users can create and edit the vocabularies. Access rights to the service are managed by the administrator. The tool can also be used for in-house

lish) vocabularies have been created by the service. (DDI) is an international standard for describing observational methods in the social, behavioural

## Any thoughts on controlled vocabularies?

its useful and helpful

unfortunately not all areas have them

Not always easy to use

yes, biology is the worst offender

Difficult to find and choose

Problem with collaboration project



your data(set) should be in a format that is open and supports long-term preservation?

## 9. File formats and long-term preservation

?

...ds for encoding digital information...  
...e three or four letters at the end of  
...ary in its format. It is highly recommended  
...that others can easily access and



0  
No

0  
Uncertain



What are some file formats that your archive will accept? (multiple submissions allowed)

word pdf stata  
cvr doc docx jpeg  
txt sav  
r rtf CSV png  
for pictures



# Take home messages

- FAIR-Aware tells data owners what to look for in an archive
- Be FAIR and trustworthy to be a depositor's no.1 choice!
- Chapter 5 of Data Archiving Guide will soon appear online
- Contains more content than discussed today
- Interested in becoming an earlier reader of our chapter?

# Thank you!

**Kim Ferguson:** kim.ferguson@dans.knaw.nl | @kfergy

**Maike Verburg:** maike.verburg@dans.knaw.nl | @MaikeVerburg

**DANS:** dans.knaw.nl/en | @DANS\_knaw\_nwo

**FAIR-Aware:** fairaware.dans.knaw.nl | #FAIRAwareTool

# Resources

FAIR-Aware: <https://fairaware.dans.knaw.nl/>

F-UJI: <https://www.f-uji.net/>

FAIRsFAIR Assessment Framework: <https://doi.org/10.5281/zenodo.6656431>

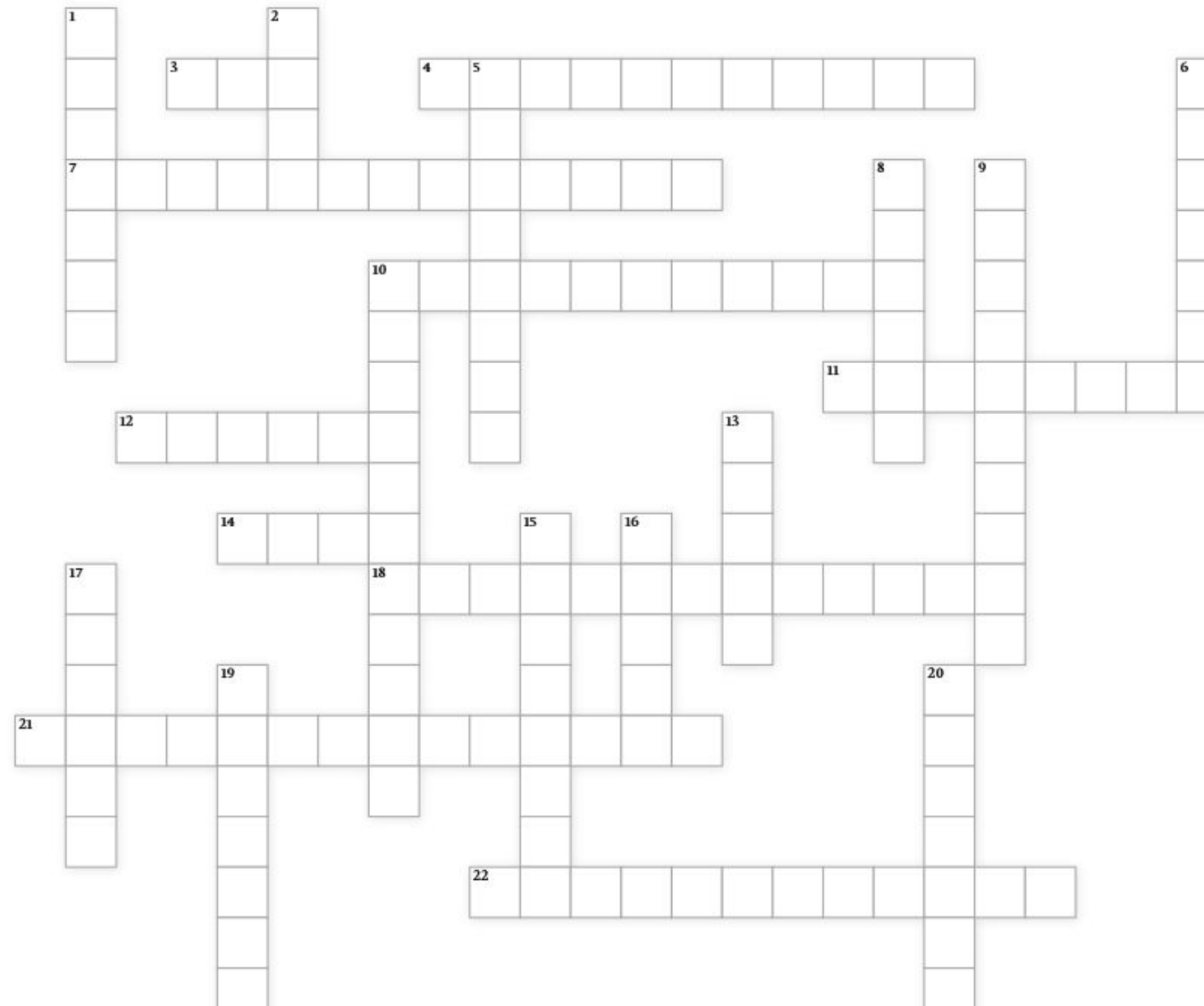
CESSDA Vocabulary Service: <https://www.cessda.eu/Tools/Vocabulary-Service>



# Crossword

<https://share.eclipsescrossword.com/play/c17538a4/cessda-dag-13-oct>

CESSDA DAG 13 Oct



 Share EclipseCrossword

## Across

- 3. Persistent identifiers
- 4. The way data archives gets data to archive
- 7. The I in FAIR
- 10. a person or organisation responsible for data management
- 11. The R in FAIR
- 12. The Consortium of European Social Science Data Archives
- 14. Four principles for research data
- 18. CTS
- 21. Data with potentially harmful effects in the event of disclosure or misuse
- 22. A person or organisation responsible for data curation
- 23. Core purpose for archives
- 24. Unrestricted access to information, documentation or information