CESSDA Data Archiving Guide

Soft Launch

Libby Bishop / GESIS - Leibniz Institute for the Social Sciences
Ellen Leenarts / DANS-KNAW
Dimitra Kondyli / EKKE & SoDaNet
Ilze Lace / SND, University of Gothenburg
Iris Butzlaff / AUSSDA, University of Vienna

16 November 2021
DOI: 10.5281/zenodo.5642605

cessda.eu

@CESSDA_Data

CESSDA Data Archiving Guide

A resource for people who love data

Libby Bishop / GESIS - Leibniz Institute for the Social Sciences

16 November 2021, CESSDA Data Archiving Guide Soft Launch
DOI: 10.5281/zenodo.5642605

cessda.eu

@CESSDA_Data

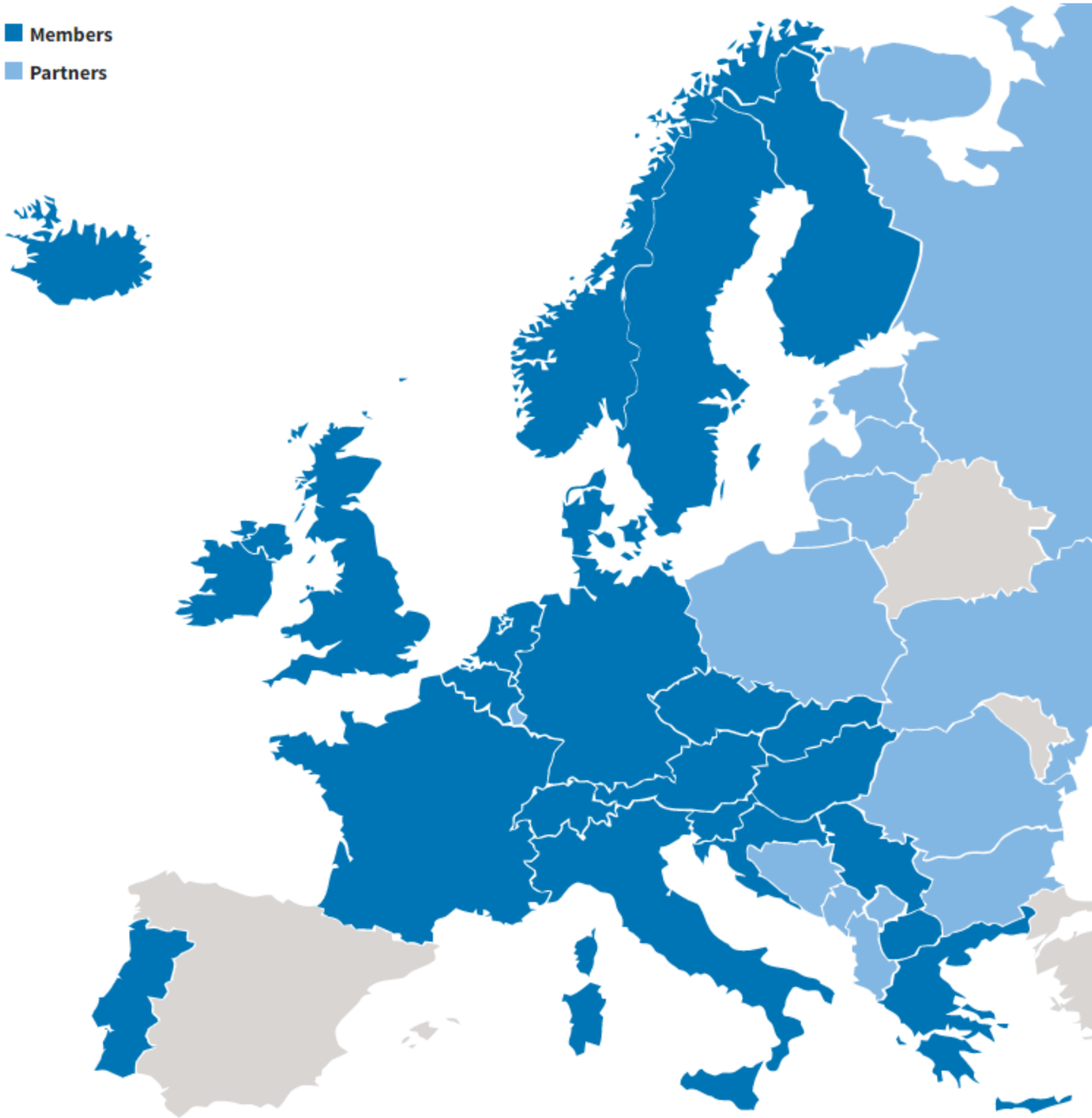
CESSDA®

Our vision is that the provision of access to social science data and metadata is vital – for both science and society.

For this we must offer services to data producers to easily describe and store their data – if needed in a secured environment.

We will adhere to the FAIR (Findable, Accessible, Interoperable, Reusable) data principles to make data findable and provide information about the data, where they are, how they can be accessed.

We will also focus on providing training and enabling the transfer of expertise and sharing of knowledge on data, as well as relevant rules and regulations.



Tools & services

cessda
DC Data Catalogue

cessda
DMEG Data Management Expert Guide

cessda
VS Vocabulary Service

cessda
ELSST Thesaurus

cessda
TRAINING

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.

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

Why do data archives exist?

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

DAG – Past, Present, and Future

- Goal: support the professional development needs of archive and repository staff – diverse backgrounds, domains, facilities

- CESSDA – well positioned to develop this resource
  - European coverage with growing international connections
  - Sustainable infrastructure (an ERIC, NOT a project...)
  - Synergy with other CESSDA projects & services – Resource Directory

- The work to date – idea of DMEG, CESSDA project, 2021-2022 & beyond
  - Selection of chapter topics – need, expertise available
  - Content written and external reviews done, CESSDA web developer

- Future – after the chapter presentations

Process for presentations and feedback

- Presentation of each chapter by an author (10-15 mins)
- Divide in break-out groups – random but you can change
- In groups, discussion and focus on feedback
- *Choose a notetaker for every break-out group*
- Notes captured in Google doc-1 for each group

Time	Topic	Presenter
09:00-09:15	Introduction	Libby Bishop, GESIS
09:15-10:00	DAG Chapter 1: FAQs	Ellen Leenart, DANS-KNAW
10:00-10:45	DAG Chapter 2: Policies	Dimitra Kondyli, EKKE
10:45-11:00	Break	
11:00-11:45	DAG Chapter 3: Pre-ingest	Ilze Lace, SND, University of Gothenburg
11:45-12:30	DAG Chapter 4: Ingest	Iris Butzlaff, AUSSDA, University of Vienna
12:30-13:00	Future content plans	Libby Bishop, GESIS

Looking ahead

- Rather complex integration of DAG with CESSDA general website, and CESSDA Training website
- Work continues through Dec 2021
- Plan two more chapters for next year
    - FAIR
    - Communicating to Funders and Policy Makers

- Other possible topics for chapters?
    - Tools for archiving data
    - More ideas please!

- 2022-2023+
    - Many more archive-specific examples and cases
    - Downloadable PDFs

Thank you!

Libby Bishop / GESIS - Leibniz Institute for the Social Sciences
ElizabethLea.Bishop@gesis.org

cessda.eu

@CESSDA_Data

CESSDA Data Archiving Guide

Chapter 1: Why do data archives exist? FAQs

Ellen Leenarts / DANS-KNAW

16 November 2021, CESSDA Data Archiving Guide Soft Launch
DOI: 10.5281/zenodo.5642605

cessda.eu @CESSDA_Data

Why do data archives exist?

As a new employee at a Social Sciences Data Archive I have all these questions…

FAQ 1. What is an archive?
FAQ 2. What does an archive look like and what does it do?
FAQ 3. Why is archiving important?
FAQ 4. What is data acquisition?
FAQ 5. What is a certified archive?
FAQ 6. How does an archive make data available
FAQ 7. What is archived?
FAQ 8. What is the overall process of archiving from beginning to end
FAQ 9. What are relevant legislations and legal questions in relation to data archiving
FAQ 10. Where can I find a list of relevant terms?
FAQ 11. What are the main tools used by archives?

FAQ 1. What is an archive?

More than just storage of data!

A data archive acquires, processes, documents and stores data files from research projects.

It curates data and ensures that data remains available and accessible for the long term

It serves as a resource of information about methodology and research instruments.

Short explanation of:

- Difference online repositories
- Mission
- CESSDA
- References to sources and policies

Questions new employees could ask their colleagues...

FAQ 2. What does an archive look like and what does it do?

What does it look like?

- It depends.... large, small, in employees or in number of datasets
- Part of the university, university library or other

What does it do? (e.g. curation of data, create and maintain PIDs, making data discoverable)

What are core job titles?

<u>Questions new employees could ask their colleagues...</u>

To what institution is your data archive affiliated? Is your archive rather small or large? What kind of training activities does your archive offer? What are the core positions in your archive? How does your archive make data available to its users?

FAQ 3. Why is archiving important?

Archiving data is beneficial both for producers of the data and for the research community.

What are these benefits for data producers and the research community...

➔ Quality of (meta)data for reuse
➔ (Secure) Access to data

Includes links to FAQ 6 on how data is made available and FAQ 7 on what is archived.

FAQ 4. What is data acquisition?

Data acquisition is the process of acquiring data from data producers who are in this context referred to as data depositors.

➔ obligation to make data available by funders - archives are ‘passive’ ingestors
➔ actively search for data (link to FAQ 3 on benefits)

➔ ingestion process (includes link to FAQ 8 for the overall process)
- self-archiving services of some archives
- link to FAQ 11 on archiving tools

➔ expert tip for acquisition: look for other than core social sciences data

FAQ 5. What is a certified archive?

A certified archive is an archive that has received a formal type of certification under which the archive has been assessed according to a set of specified guidelines.

- ➔ Trustworthy Digital Repositories (TDR)
- ➔ Certification instruments: CoreTrustSeal, Nestor Seal, ISO 16363
- ➔ Advantages of a certified archive
  - curation of the data, ensuring data quality and providing long-term storage

- ➔ Two expert tips
  - Trust principles and
  - Re3data

Comments & feedback

Do you like the content?

- Is it too long, too short
- Style of writing?
- Other comments?

Is there any content missing that you would have expected or would like to see?

Please also include your suggestions for further reading!

What do you think about the questions for the new employee to explore their data archive?

➢ Go online to have a look at the FAQs.

FAQ 6. How does an archive make data available?

How an archive can make data available starts with the characteristics of the deposited data. Archives often support the data producer in preparing the data so it can be made available at a later stage.

- ➔ file formats
- ➔ consent
- ➔ licences & access categories
- ➔ metadata standards

Archive search functionalities and promotion of the reuse of data

FAQ 7. What is archived?

Typically, social science data archives store both qualitative and quantitative research data with affiliated documentation e.g.

- ➔ digitalised survey data
- ➔ administrative data

- ➔ Documentation
- ➔ Long-term preservation formats
- ➔ (Sensitive) personal data

FAQ 8. What is the overall process of archiving from beginning to end?

CESSDA archives are different in size and capacity but they have a common goal making research data more accessible and findable now and in the future.

To fulfill this role, many data archives follow the Reference Model for an Open Archival Information System (short OAIS).

- ➔ Submission Information Package (SIP), Archival Information Package (AIP) and Dissemination Information Package (DIP)?
- ➔ 6 functions in archiving:
  - ingest
  - archival storage
  - data management
  - administration
  - preservation planning
  - access
- ➔ Persistent identifiers in the archiving process (PID)

FAQ 9. What are relevant legislations and legal questions in relation to data archiving?

In European data archives, the archiving process is regulated by both national laws and international regulations. The main legal issues are intellectual property rights (e.g., copyright law) and personal data.

- ➔ Copyright
- ➔ Legal questions when handling personal data
- ➔ Personal data in research data
- ➔ Personal data in administrative data
- ➔ General Data Protection Regulation

FAQ 10. Where can I find a list of relevant terms?

Links to available glossaries

FAQ 11. What are the main tools used by archives?

In this context, we mean systems, software and technology that the various archives use in archiving research data.

- ➔ Ingest
- ➔ Curation, administration, documentation, upgrades, versioning
- ➔ Dissemination

Tools used by various archives such as: AUSSDA, ADP, CSDA, DANS NSD, UKDA

Comments & feedback

Do you like the content?

- Is it too long, too short
- Style of writing?
- Other comments?

Is there any content missing that you would have expected or would like to see?

Please also include your suggestions for further reading!

What do you think about the questions for the new employee to explore their data archive?

➢ Go online to have a look at the FAQs.

Thank you!

And have fun the rest of the day…

Ellen Leenarts | DANS-KNAW

cessda.eu

@CESSDA_Data

CESSDA Data Archiving Guide

Chapter 2: Policies of Data Archives

Dimitra Kondyli / EKKE & SoDaNet, ORCID: 0000-0003-1037-5415

Contributors: Christina Frentzou / EKKE; Marielle Kappeler / FORS

16 November 2021, CESSDA Data Archiving Guide Soft Launch
DOI: 10.5281/zenodo.5642605

cessda.eu @CESSDA_Data

Policies of Data Archives

Structure and the logic behind

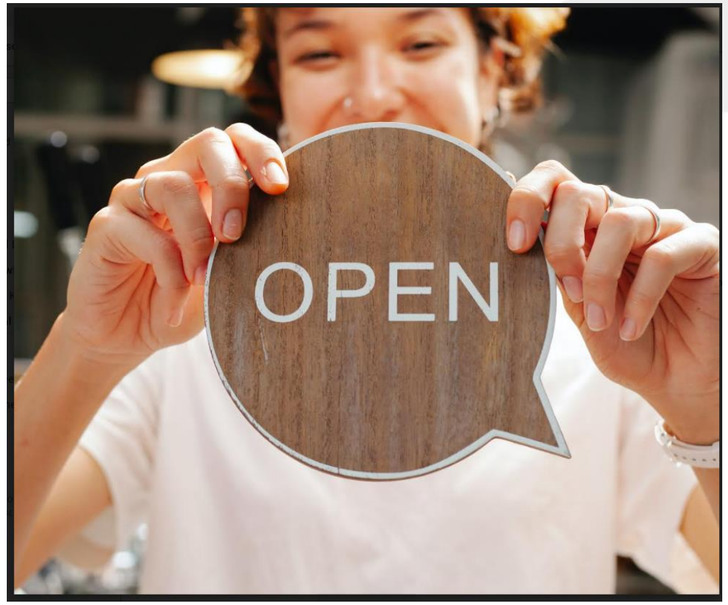
The second chapter of the DAG focuses on “Policies of Data Repositories”. It describes the types of policies (and some related documents) a social science data archive has in place.

- The sections of this chapter are conceptually connected to the sections of the first chapter of this Data Archiving Guide including basic common policies and procedures like Data Collection and Acquisition, Preservation, Curation, Data Access, and Data Dissemination. They cover principles and necessary measures to be implemented in terms of policies to ensure the sustainability of both datasets and repositories.
- Each section is being articulated under the following :
  - A. Introduction
  - B. Relevant FAQs from Chapter 1 (FAQ 9: What are relevant legislations and legal questions in relation to data archiving?)
  - C. Links to examples
  - D. What to consider when creating or adapting a policy or principle
  - E. Sources and further reading

1. Data Collection and Data Acquisition Policies

A. Introduction

- Repositories vary widely in how they receive data. For some, where data producers are encouraged or required to deposit, the problem may be too much data. More typically, repositories face challenges of finding, acquiring, and ingesting the “right” data.
- The definition of the criteria for the right data is done by a Collection Development Policy or an Acquisition Policy.
- Typical criteria might include:
  - Purpose: Research, teaching, replication.
  - Level of curation – short or long term.
  - Alignment with other internal and external policies, e.g., research priorities of funders.
  - Flexibility to deal with exceptions, such as data of high risk of loss.
- A Policy might include some discussion of methods, or shifting strategic objectives



1. Data Collection and Data Acquisition Policies (ii)

B. Relevant FAQs from Chapter 1

FAQ 4: What is data acquisition?

C. Links to examples

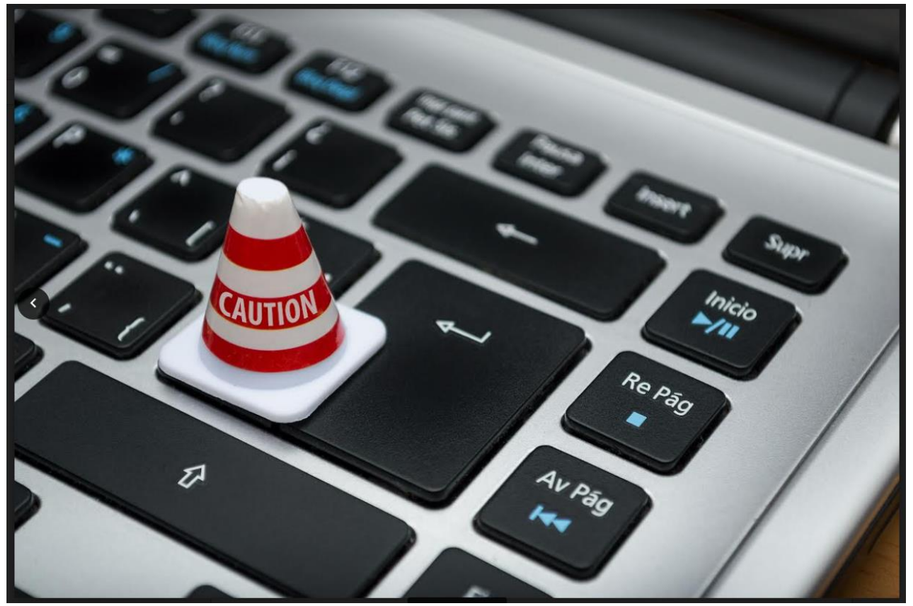
- ie. Archival Acquisition Policy DARIS. FORS. Retrieved from: https://forscenter.ch/wp-content/uploads/2018/10/collections-policy.pdf

1. Data Collection and Data Acquisition Policies (iii)

- D. What to consider when adapting a policy

[E. Sources and further reading]

An inspiring and informative source as an example: A paper that describes FSD's and ADP's data acquisition policies in relation to their general archival development strategy.



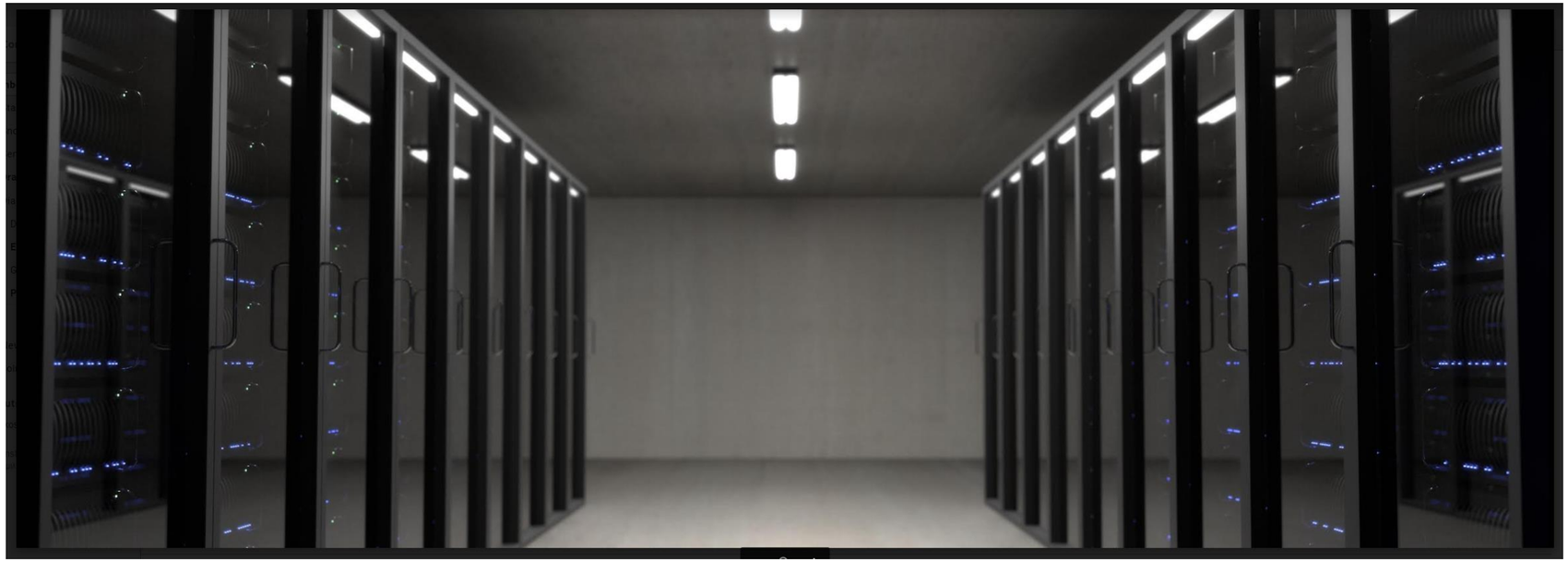
2. Preservation Policies

A. Introduction

- “Digital Preservation is the series of actions and interventions required to ensure continued and reliable access to authentic digital objects for as long as they are deemed to be of value." (JISC (2006). Digital Preservation. Briefing Paper)
- Preservation, is the heart of a repository, and often it is what distinguishes a repository from other facilities that store data. Its content includes the information about:
  - (i) an organizational framework in terms of the mission, selection and acquisition, access and use of data;
  - (ii) potential challenges
  - (iii) principles for the guidelines of the Data Preservation
  - (iv) roles and responsibilities for the different types of work that Data Preservation requires

2. Preservation Policies (ii)

- Preservation Policies should be in line with the core mission of the organization.
- Repositories need to determine the scope of preservation actions: how, for how long, and under which specific access conditions/restrictions.
- Resilient preservation makes an archive trustworthy.



2. Preservation Policies (iii)

B. Relevant FAQs from Chapter 1

C. Links to examples

- FAQ 2: What does an archive look like and what does it do?
- FAQ 5: What is a certified archive?

2. Preservation Policies (iv)

D. What to consider when creating or adapting a Preservation Policy

- A Preservation Policy should clearly define the scope of materials covered, connections with other internal areas, external guidelines and standards are all relevant in view of long-term sustainability.

E. Sources and further reading

3. Data Curation Policy (i)

Data Curation Policy

- A. Introduction
- It can be defined as “a managed process, throughout the data lifecycle, by which data & data collections are cleansed, documented, standardized, formatted and inter-related.
- Repositories should ensure that data are curated and maintained by all appropriate means and tools. In order to uniquely identify dataset/data sources submitted, should include in its data curation policies among others:
    - Authorization to create DOIs.
    - Access conditions for the whole data collection.
    - Specific data sets’ access conditions.

3. Data Curation Policy (ii)

- B. Relevant FAQs from Chapter 1

- FAQ 2: What does an archive look like and what does it do?
- FAQ 5: What is a certified archive?

- C. Links to examples

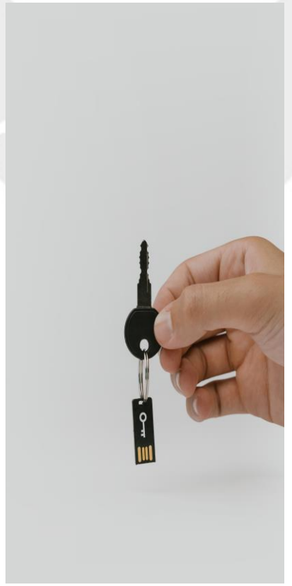
3. Data Curation Policy (iii)

- D. What to consider when creating or adapting a Curation Policy
  - How and for how long deposited data will be distributed, whether unchanged, how edits will be made or added value to the datasets. Assessment of technical means, resources and skilled staff are required when data repositories implement curation policy.

- E. Sources and further reading

4. Data Access Policies (i)

A. Introduction



- Policies that an Archive needs to apply when providing users with access to its digital resources. The main policies to be presented include data access and collection policies for designated user communities. Data Repositories usually maintain different data access conditions and subsequent policies. As of today, there are three levels of access, depending on the nature of the data and agreements with data depositors:
  - 1) Open access to data
  - 2) Restrictive data access (access under conditions): a. Standard access, b. Special conditions access c. Access under special license
  - 3) Controlled data access
  - Unique and persistent identifier (PID) is needed to enable access and related policy for the Repository to meet citation and visibility purposes (see also Ch. 3 Data Collection and Acquisition Policies)

4. Data Access Policies (ii)

- B. Relevant FAQs from Chapter 1
  - FAQ 2: What does an archive look like and what does it do?
  - FAQ 6: How does an archive make data available?

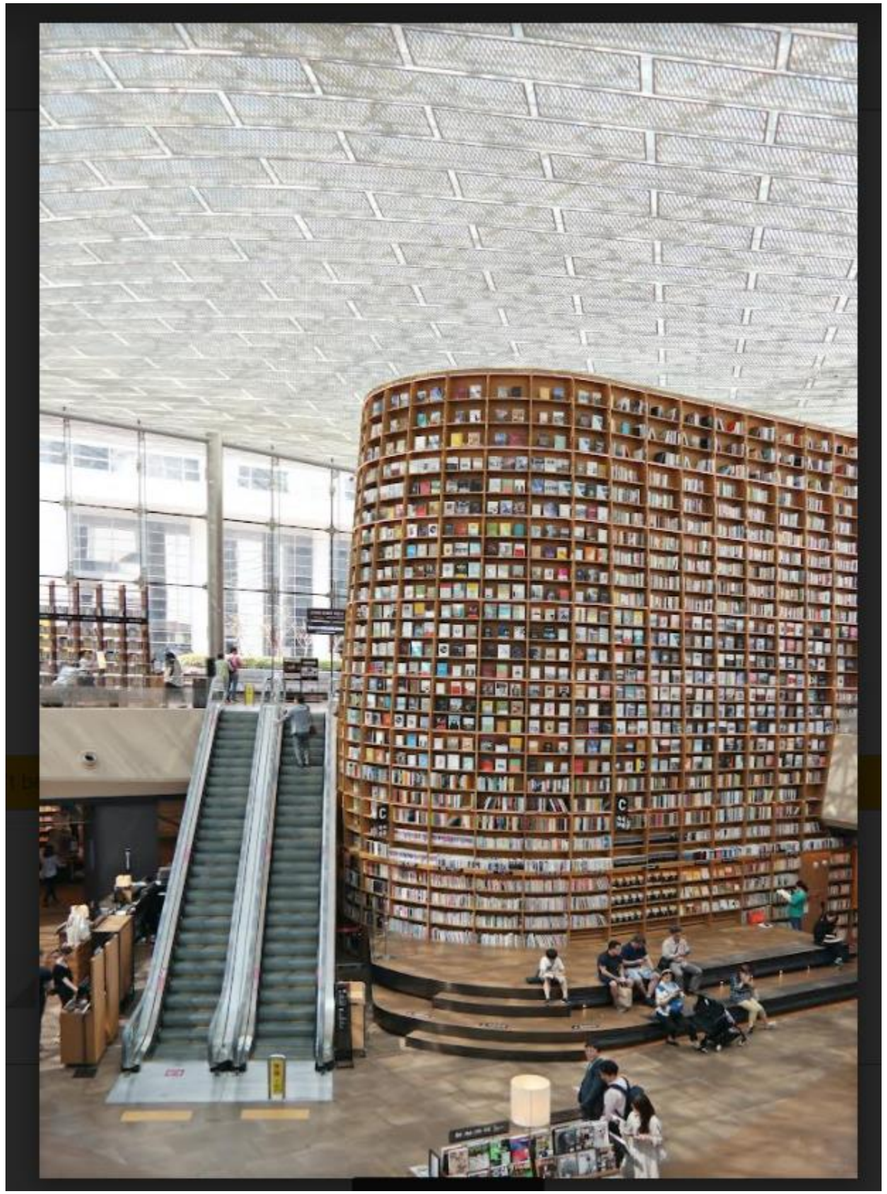
- C. Links to examples

4. Data Access Policies (ii)

- D. What to consider when creating or adapting a policy
  - Sine qua non: License agreement for redistribution in place for each data collection clearly stating the rights owner, the data collector, and access conditions.
- E. Sources and further reading

5. Data Dissemination Policies and Principles (i)

- A. Introduction
  - Data Dissemination Principles are sometimes formal policies but also more general statements about an archive's position on the safeguarding and dissemination of data. They might reference explicit policies, such as access policies, or they might describe the legal framework.
  - These statements declare a general commitment to Open Data, or legal use of data, e.g., compliance with General Data Protection Regulation (GDPR).



- B. Relevant FAQs from Chapter 1
  - FAQ 9: What are relevant legislations and legal questions in relation to data archiving?
- C. Links to examples

5. Data Dissemination Policies and Principles (iii)

- D. What to consider when creating or adapting a policy or principle
  - Data that have received public funding, data under embargo or safeguarded data under restricted access, license agreements for datasets dissemination are also components to be taken into consideration when repositories implement dissemination policies.
- E. Sources and further reading

Thank you!

To be continued in 2021-22

Dimitra Kondyli / EKKE & SoDaNet

cessda.eu

@CESSDA_Data

CESSDA Data Archiving Guide

Chapter 3: Pre-ingest

Ilze Lace | SND, University of Gothenburg

16 November 2021, CESSDA Data Archiving Guide Soft Launch
DOI: 10.5281/zenodo.5642605

cessda.eu

@CESSDA_Data

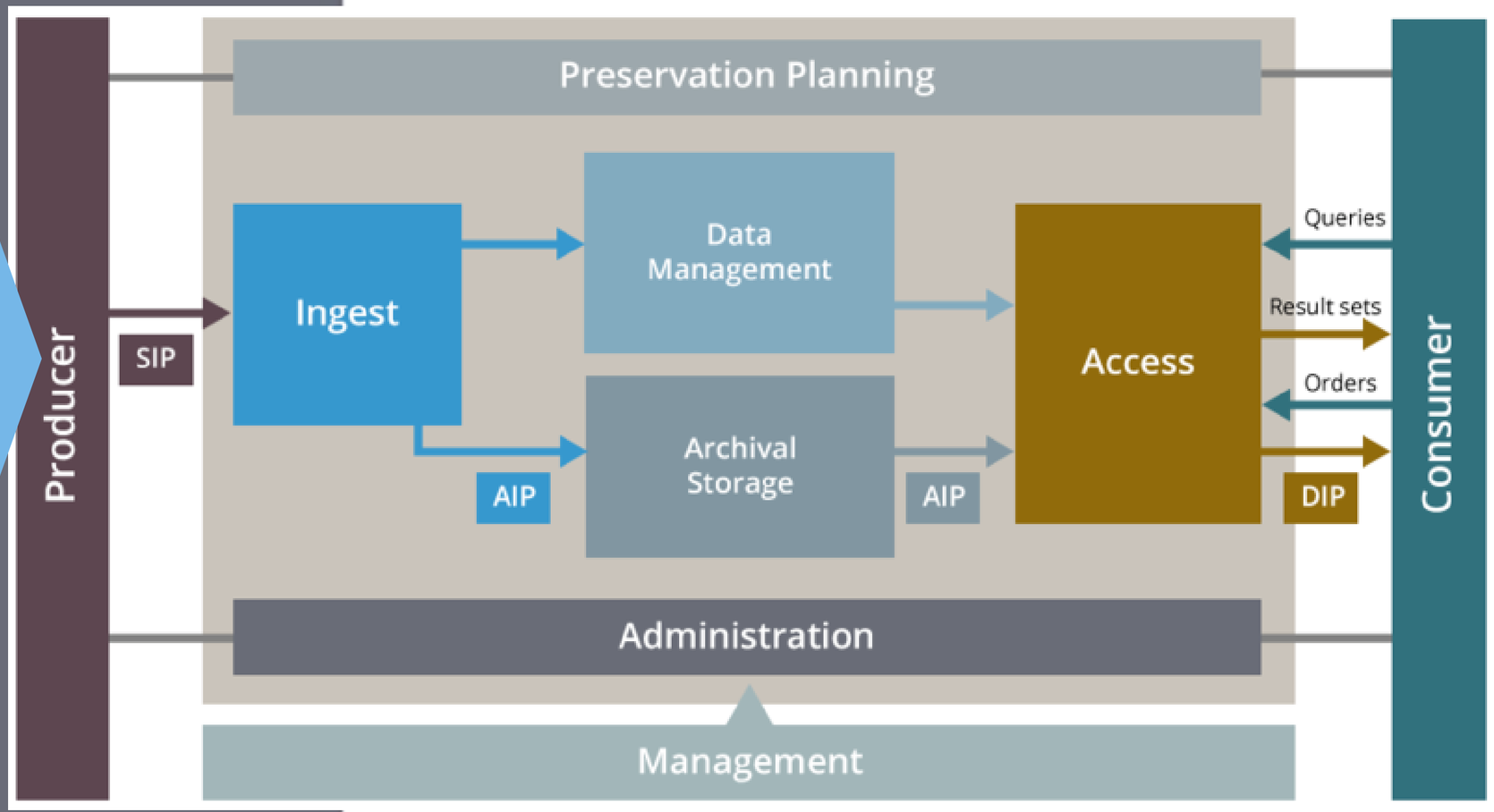
Planning and writing of DAG Chapter 3

- 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
5. Further reading

OAIS Model



CESSDA Tutorial:OAIS (2017)

Online: https://dag.cessda.eu/Chapter%203

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
- Formats and documentation

Support and consult:

- RDM training
- Consultations and support 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

Informing depositors:

- Informing on submission requirements
- Describing of the main steps in the deposit process
- Instructing how to proceed

Ensuring data transfer:

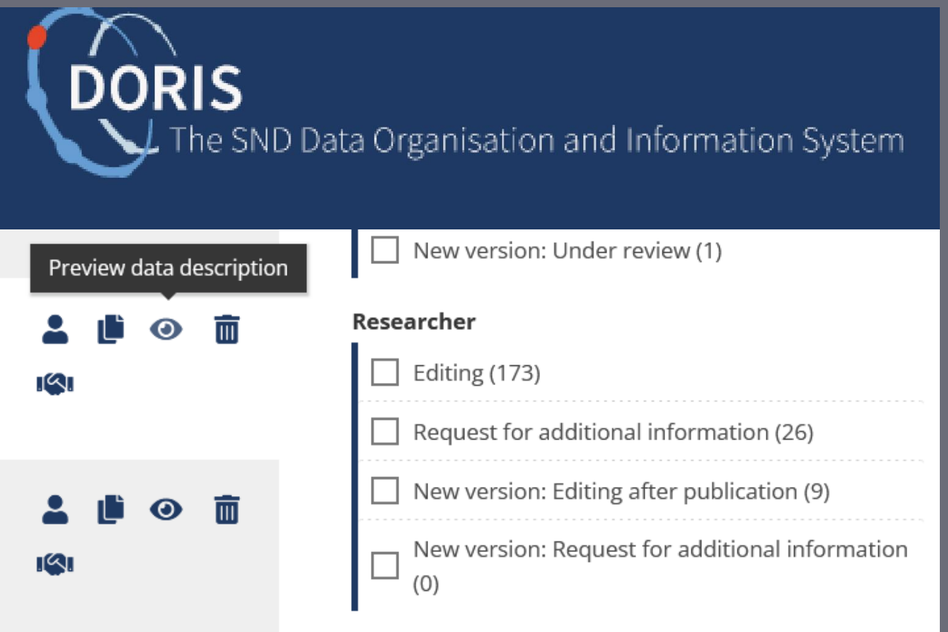
- Finding an acceptable solution for incoming data
- Special cases: sensitive data, big data..

Administering 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 of compliance with acquisition criteria:

- Criteria should be in the Acquisition policy

Control of the submitted material:

- Depositor ID
- Completeness of material
- Virus and readability
- Data description
- Ethical and legal aspects

Review of documentation:

- Main focus to ensure there is enough context information

Communication 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

Accepting data:

start

- Notify researchers and explain further steps
- Ensure there is a legal ground for the data archive to curation and preservation process
- Prepare information (checklists etc.), alert Ingest-team

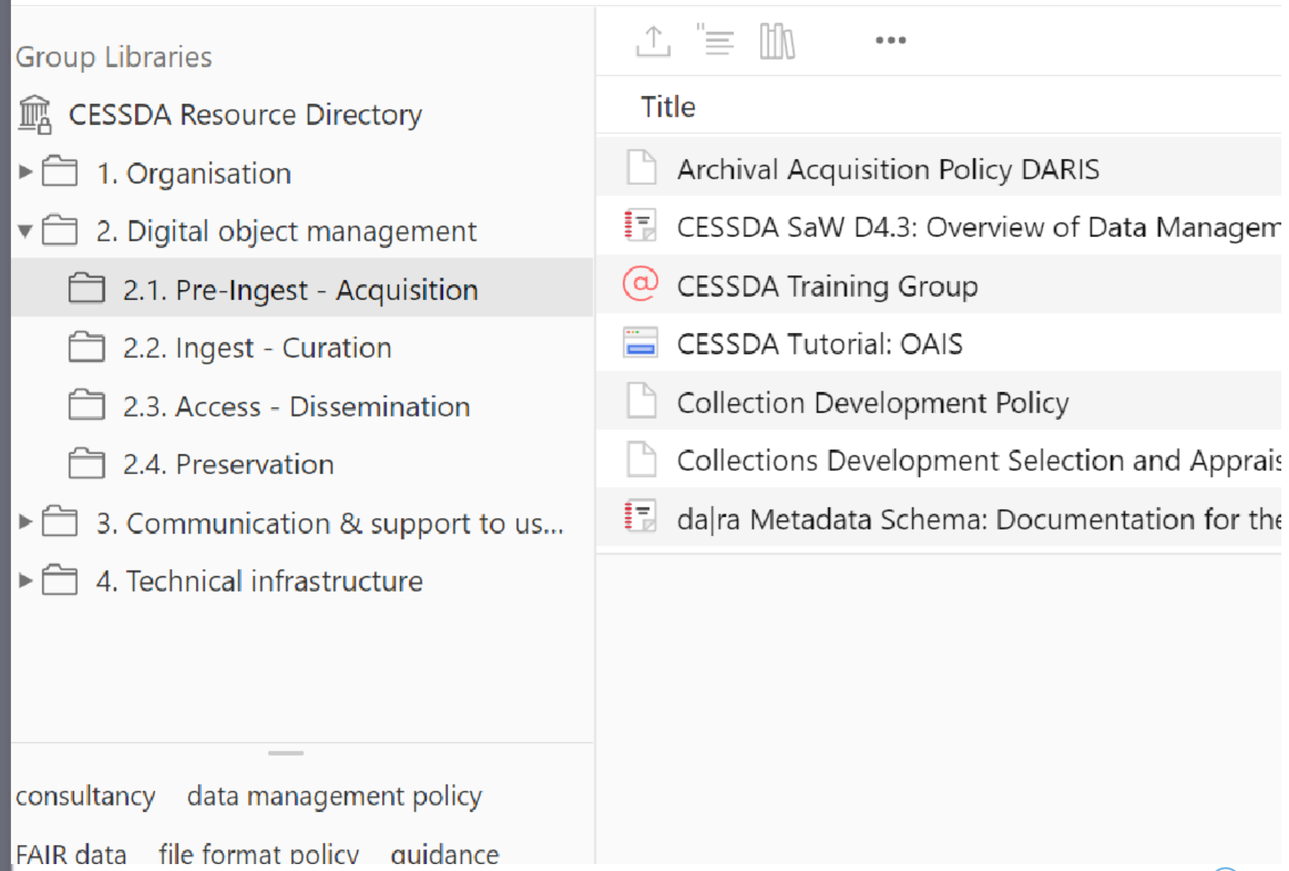
Rejecting data:

- 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]



What next?

- Living document
- Feedback and input from users
- Better integration with other CESSDA Training resources

Thank you!

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

cessda.eu

@CESSDA_Data

CESSDA Data Archiving Guide

Chapter 4: Ingest / Curation

Iris Butzlaff / AUSSDA, University of Vienna

16 November 2021, CESSDA Data Archiving Guide Soft Launch
DOI: 10.5281/zenodo.5642605

cessda.eu

@CESSDA_Data

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

Find the Chapter 4 Ingest/Curation here:
https://dag.cessda.eu/Chapter 4

1. Handover from acquisition/pre-ingest to ingest

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

→ Data Collection Policy

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

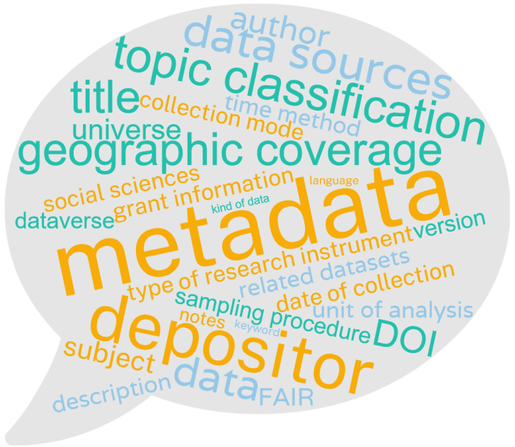
- Data types and formats used in community
  - Quantitative data (Stata, SPSS, SAS, R, Excel)
  - Qualitative data (PDF, text programmes)

- Preferred formats
  - provide a list of your preferred formats

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

- Open Access, Scientific Use, Replication Data
  - Which audience accesses the data
  - Different levels of pseudonymisation
- Special conditions; e.g., embargo or restricted access (internal process)
- Re-use or preservation only

3. Metadata



- Make metadata FAIR, license CC 0
- 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)
- Checks on metadata

4. Quality assurance of data and documentation material

- 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

continued ...

- 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)

→ advisable to have a checklist for all necessary steps in ingest (see sub-section 7. Checklist of general interest procedure steps)

5. 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

6. Adaptation of workflows and processes

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

> Against this background, AUSSDA, with the support of the Austrian Federal Ministry of Education, Science and Research (BMBWF), has launched the project COSSDA - COVID-19 Social Science Data Hub Austria. It provides the research community with a basis for empirical analyses on COVID-19 topics in Austria and beyond. COSSDA builds on the previous project COVID-19 Data Fast Track Publishing and ensures that the reusable research data of COVID social science studies in Austria are collected in one place, archived sustainably and made available as promptly as possible according to the FAIR data principles (findable, accessible, interoperable, reusable).

Thank you!

iris.butzlaff@univie.ac.at
Iris Butzlaff, AUSSDA, University of Vienna

cessda.eu

@CESSDA_Data

Questions to repository staff: “How do you do it?”

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?

See Chapter 4 Ingest/Curation:
https://dag.cessda.eu/Chapter 4