

Discover the Data Archiving Guide (DAG) - a training event for new(ish) staff members

13 October 2022 - Online

Morning Exercise: How to

1. Groups of 3-4 participants
2. Define one speaker, who is willing to present the group results
3. Max. 2 groups can treat the same topic / issue ⇒ put your group number behind the issue as soon as you start treating it
4. Treat as many topics / issues as you would like
5. You are asked to share HOW and WHY you are handling the issue a certain way. There is no right or wrong answer as everything will depend on the context of your archive
6. Please do not delete your boards after the exercise (will be used in event reporting)

Find your group on Jamboard:

1. [Group 1](#)
2. [Group 2](#)

(Groups used Jamboards for their discussions. Results are not publically available)

Issue

Issue / Question description	Group A	Group B
Pre-Ingest: 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?	Taken by group 1	

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

INGEST

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?

ISSUES

A. You receive a request to archive data that was collected about 30 years ago. This is a mixed-method project with qualitative and quantitative data, longitudinal, sensitive topic (addictive behavior). Initially, it was not intended to release the data for secondary research. What kind of questions are you asking and why? What kind of problems might arise?	group 1	
B. You receive a request to archive ethnological data (life histories). The data is qualitative, the people involved are no longer accessible, and there are language and cultural barriers for communication. What kind of questions do you ask and why? What kind of problems might arise?		
C. You receive a request to archive sound data. What kind of questions do you ask and why? What kind of problems might arise?		group 2
D. You are handed data for archiving that is already available in another archive. How do you deal with the situation, what solutions do you offer and why?		
F. You receive a request from a researcher to archive data that was collected 20 years ago. The data is very interesting, but it was not completely cleaned at that time; e.g. some variables are not at all documented. Do you accept this data in your archive, why / under what conditions / what additional information, if any, do you have to provide?	group 1	
G. You are dealing with an international project; researchers from different countries, no central office, longitudinal. No clear agreement was made at the start of the project as to whether, who and where the data would be archived. What do you recommend, how do you proceed and why?		
H. You receive a request from a researcher who has done a content analysis - based on images and article excerpts from magazines. The material is protected by copyright. How do you proceed and why?	group 1	

I. You receive a request to archive data consisting of a compilation of various European statistics. No additional data was collected, but existing data was compiled. Partly it concerns old statistics; it is not possible to find out who/how is responsible for a possible release. How do you proceed and why?		
Y. You receive a request to archive data that contains a variable from another project/data set. This is an Open Response in the context of a political survey. How do you proceed and why?		