

## Slido results, 2nd EOSC EDEN Webinar "User perspectives on long-term data preservation: Reducing pain points for researchers, repositories and archives", 27 January 2026

### 1. To the support staff: How do you engage with your community of users to identify their changing needs for long-term preservation over time ?

meetings, discussions

Request tracker; User meeting; Webinars on community-relevant topics; Hackathon; Conferences

Meetings

Hackaton; Grant

Attend discipline specific conferences; Give trainings on repository use; Join consortia in which users and engineers jointly develop infrastructure

Personal communication; Seminars; Conferences

Partners in our co-op give us their requirements.

Ownership, retention, record types, format; Meetings

Overview of retention periods

Outreach ; Training; Serve on discipline specific boards and advisory groups

On a person to person basis; Personal because otherwise no response

Attending events organized by researchers to present their work

Individual training

surveys, feedback forms, conference engagement, community participation

unable to, they don't care so much...

one to one meetings

Listening

personal communication; ask for feedback

User study

Training

Individual meetings at deposition, engagement with various user groups.; Conferences; Help desk

discuss file formats case by case

exchange with peers ; exchange with community

Feedback; User Training; Community engagements

Surveys, workshops; Meetings
Feedback survey
121 meetings to discuss data deposits to our repository
focus group discussions; find time to engage/talk to the users
Webinar/trainings for researchers (partly by RDM-departments and knowledge exchange with these departments)
Periodic campaigns
Training; Focus groups; Help desk
tech watch; articles, blog posts; Surveys; talk with stakeholders; we have an advisory board with good selection of main stakeholders; good information services; help desk
User case communication; Data curation communication
open hour; ticket system (via contact form)
customer surveys; conferences; peer exchange with other repositories
Helpdesk ticket analysis & gathering feedback in a structured way at the end of a training.; Our users don't tend to think too much about it, it's more the policy makers
conversations with individual researchers or projects when we support them to make their data / research output findable and accessible
Help desk ; Hackathon; Webinars; Guides
small team meetings
<b>2. To all: After deposit, how long should the type of data you typically work with, be kept available? Could they be replaced or reproduced in the future?</b>
10 years or longer
10 years; reproducibility depends on size/computational costs of/ data availability for climate model runs
Indefinitely
Min 15 years
For ever
Permanently; possibly converted to a new format
More than 10 years; Reproduction very hard
Indefinitely
10 to 40 yrs. Digitization of it is a challenge
10-15 yrs
In perpetuity ; Data only added/removed via reaccessioning/deaccessioning
10 years. Some sets can be kept for ever, for which criteria can be defined. Btw: deleting data is not per se bad. Keeping data per se neither.

Indefinitely
10 years; 10 years minimum, systematic appraisal process for materials that should be preserved long term
Indefinitely
10 years but also consider archiving (=appraisal)
forever
10 years from last access. No archival workflow for replacement currently
10 years
decades; at least 10yrs; as long as they are of interest; depends of dataset
at least 10 years
Minimum 10
In perpetuity
permanently, or as close to as possible
minimum 10 years
We are a long time storage but we only take date after appraisal and after projects/records are closed
10 years
10 years; 5 years
appropriately, data could be converted to new formats... but some of the context may be lost - it may not be documented appropriately in line with new standards
permanently; can be re-appraised for either preservation permanently or destroyed/deleted in the future; may need to consider whether to keep forever depending on resourcing and sustainability
indefinitely; long-term experiments (time series etc) cannot be reproduced; data should be available long-term if they are cited/referenced (as part of good scientific practice); most research data should be reproducible by other researchers (to a degree, it won't be the exact same results but the discovery should be reproducible)
10y; No replacement policy
More than 10 years - working on a more Long term solution
as long as relevant (long terms means more than 100 years for us); our data cannot be replaced or reproduced
Same, 10 years policy
Permanently
indefinitely
forever
indefinitely

data are unique and are preserved for the long term (forever 😊); survey data is collected at certain time and can't be reproduced, so needs to be preserved
policy-wise 10 year, practically, I think it should depend on the cost vs usefulness.; should really depend on the data (content, type, accessibility, potential usability & size)
minimum of 5 years for funders but should often be kept available much much longer if it can be used by thirds parties
at least 5 to 10 years depending on size
2 years after end of study, then 15 years in archive, or possibly 20 years (or more) in an "Entrepôt de Données de Santé"
More than 10 years
10 years minimum
Forever; Cannot be replaced or reproduced
Perhaps important to clarify if there is an overlap or difference of meaning between indefinite and forever?
Permanently
<b>3. What is a pain point you may experience in your “ingest and appraisal” journey?</b>
Cost & environmental impact
Awareness/Motivation: Not quite clear whether data are suitable for LTP in our repository.; Preparation: Using the UI of the repository to change licences is difficult.; Data access: Not immediately obvious how to use UI to request access
Too many different ingest channels.
Value for LTP
Determine who actually is the owner
lacking objective criteria for evaluating the value for LTP; handling of sensitive data
Unclear who covers long term costs for LTP ; data preparation should take into account time period for which the data is supposed to be
Standardization of the data
Researchers and archivists should exchange know-how
invalid file formats
Insufficient infrastructure
Time consuming
spend time on data archival (or do not have any time to spare); everything is done last minute; researchers do not see the value of good metadata
Metadata completion and quality
for appraisal, the search for our preservation repository is difficult to navigate

Not enough exchange between Archivists from institutions which archiving without time frame and researchers for them long time storage is something between 10 and 50 years.
unclear about the value; lack of uniform standards
Metadata good practice for output types beyond 'dataset' ; Poor dataset description/abstract; Unclear consent; Have all contributors been acknowledged? (Technicians, data stewards); no suitable repository for sensitive/ on request data and large data
subjectivity in appraisal and records/data to appraised are overwhelming; not everything can be automated, a lot of manual intervention required to fix pre-ingestion issues; incomplete metadata or lack of context - resulting in issues upstream and downstream
heterogenous data in source system (as infrastructure provider, with automated ingest from source system)
Uncompleteness of meta/contextual info
ownership not clear; respondents have not been informed properly, result is we are unable to archive dat
compatibility with repository systems
Selection of data for LTP because curation requires manual work which can be resource-incentive; Data ownership can be unclear, researchers might not agree with each other
being able to contact the depositor to get more context when necessary or to decide on deletion when relevant.; The cost
value for LTP; large data files or many data files
Researches do not wish to deposit data for reuse
Chaotic national rules around GDPR
Unclear policies
Lack of clear strategy what to preserve for how long