

e-INFRA CZ services in practice: how to work with data safely and efficiently

Milan Daneček

CESNET

30. 3. 2026



Co-funded by
the European Union



MUNI
ICS

IPs EOOSC-CZ registration number
CZ.02.01.01/00/22_004/0007682

Data Storage

- typical use cases
- infrastructure
- currently available data storage services
 - with a slight outreach to the EOSC
- short to mid time storage
- and special services
- leaving aside storage directly connected to computing resources
 - MetaCentrum, IT4I—out of scope of this presentation

Data Storage in the Infrastructure

- infrastructure
 - distributed throughout the Czech Republic
 - object storage (the only future service)
 - currently more than 136 PB physical space in total
- organisational view
 - operated within the community
 - stored under local jurisdiction
 - data always belongs to its originators

Simple File Delivery

- FileSender: web service for simple transfer of (large) files
 - “large:” currently 500GB
- `http://filesender.cesnet.cz`
- at least one side of the transfer must be an eligible infra user
 - i.e. using eduID.cz federation
- uploading a file sends a download link
- invitation to upload a file

Syncing and Sharing Files

- searching for a tool to
 - synchronise data between computers
 - and mobile devices
 - have them accessible through web
 - be able to share the data
- and the data is not very big
 - 100's GB
- ~→ownCloud

ownCloud

- sync'n'share (“cloud”) storage
 - for personal computers (Windows, Linux, Mac OS X)
 - mobile apps for Android and Apple
 - and a web interface
- data is synchronised to server
 - and kept locally on computers
 - downloaded on demand on mobile devices
- data can be shared
 - to another user and/or “by link”
- register at <http://owncloud.cesnet.cz>
- 100GB limit per user by default

Using Object Storage

- file system, individual data, TB range
 - available through registration on `http://du.cesnet.cz`
 - code name VO S3 Storage
 - data stored on the CL4 object storage cluster
- requirements exceeding (e.g. group sharing) \rightsquigarrow special Virtual Organisation
 - we need to understand the requirements and use cases
 - discuss storage parameters
 - we prepare the facility
 - user management delegated to the user group

Object Storage

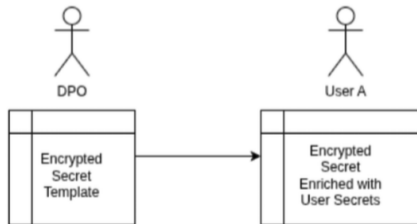
- Ceph technology
- we operate
 - RBD for backups
 - S3 for individual users and backups
 - file system emulation - cephfs ~→ special applications
- access control similar to file storage
- allows for easy integration of user group storage resources
- currently procured resources are of this type
 - storage layer of data repositories as well

S3 service - general purpose service

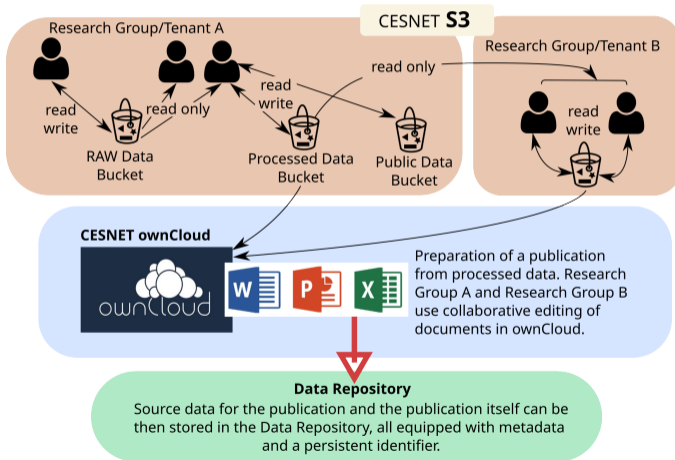
- general service for most users and applications
- https base service put/get
- enable versioning, object locking - WORM
- wide spectrum of clients - enabling client side encryption, mounting as remote device...
 - rclone - command line, encryption utility
 - S3Browser, winSCP, Cloudberry
- perfect for end users and also applications for file distribution or archive
- storage layer for National Repository Platform

Model use case

- laboratory/project wants to store and safely share its data
- dedicated data protection officer (DPO) - extension skill for data steward
 - DPO prepares S3 bucket with encrypted volume, based on rclone tool
 - S3 bucket is shared with cooperation groups
 - DPO distributes encryption template, where the user adds own secrets (access_key, secret_key)



Complex Workflows



Data Repositories

- repository: data with metadata
- “details in the morning sessions”
- currently
 - for users who need to store their research data and do not have any other repository
 - <https://datarepo.eosc.cz>
 - proof-of-concept for the national repository platform
 - can be used for long-term data storage
 - data stays
 - interfaces/procedures may change

What is the NRP I

- National Repository Platform
 - distributed, multi-tenant system for repository instantiation
 - distributed: geographically
 - multi-tenant: not a single big repository, but many tailored repositories
 - repository instantiation: able to build a repository out of pre-fabricated components “as a service”

What is the NRP II

- types of users:
 - repository end-user
 - searches for data, downloads, deposits data
 - is typically interested in a particular repository
 - repository administrator/curator
 - establishes and operates a repository for a particular topic: scientific community or for an institution
 - negotiates properties of the repository with the infrastructure
 - manages user groups and deposited data

Repository in Scientific Workflows: Deposition

- when should data be stored into a repository
- TL;DR: it depends
- aspects to balance
 - as soon as possible
 - when the data doesn't change (any more)
 - when you expect the data to be of future value
 - early deposition makes tracking metadata easier
 - and improves provenance tracking
 - but not sooner
 - e.g. big primary data that is strongly decimated
 - e.g. majority of primary data is wrong anyway

Repository in Scientific Workflows: Accessing the Data

- using data from a repository
- data is typically identified by a persistent identifier
- there will be tools to download datasets resolving PIDs
 - staging to computations, ...
- staging data to computations is similar to standard object storage
- note: repository is responsible for authorisation decisions when accessing data
- repository *is* as a storage system
 - users don't access the underlying storage directly
 - can be technically optimised, but it doesn't change the concept

Summary

- storage services
 - transferring files—FileSender
 - sync'n'share—ownCloud
 - object storage mainly S3 with features
 - NRP, catch-all repository, repository as a service

Contact us

info@eosoc.cz

www.eosoc.cz



Co-funded by
the European Union



MUNI
ICS

cesnet
.....

VSB TECHNICAL
UNIVERSITY
OF OSTRAVA

IT4INNOVATIONS
NATIONAL SUPERCOMPUTING
CENTER

IPs EOSC-CZ registration number
CZ.02.01.01/00/22_004/0007682