

DATA STORAGE

David Antoř

23. 11. 2022

Data Storage

- typical use cases
- infrastructure
- currently available data storage services
 - with a slight outreach to the (EOSC) future
 - more of which in the next talk
- storage for mid- to long-term archival
- 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
 - hierarchical storage (tape libraries&disks)
 - disk arrays
 - object storage (aka “the way forward”)
 - currently about 100 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

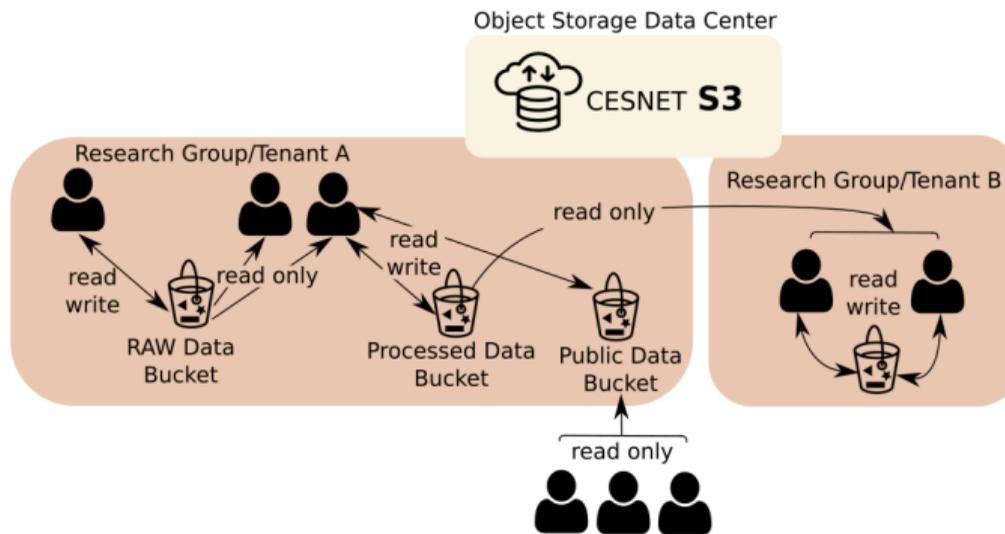
ownCloud

- sync'n'share (“cloud”) storage
 - for personal computers (Windows, Linux, Mac OS X)
 - mobile apps for Android and Apple (client under EUR 1)
 - 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

File and Object Storage I

- backups
 - users keep their primary data locally
 - use the infrastructure for backups
 - for individual users/computers
 - or aggregated (“IT dept. takes care”)
- archiving
 - users store their primary valuable data to the infrastructure
 - data that is not extensively used
 - users do not have means to store the data locally
 - individual access or a “lab archivist”
- data sharing
- “anything else”

File and Object Storage II



Standard File Access

- file system access
 - NFSv4
 - rsync, scp, FTPS
 - usually through Kerberos
 - GUI and CLI clients for all platforms available
 - CIFS (“Windows network disk”)
 - Globus—high-speed transfers

Using File Access/Object Storage

- file system, individual data, TB range
 - available through registration on <http://du.cesnet.cz>
 - code name VO Storage
 - will be transferred to object storage
- requirements exceeding that \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
- 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

Community Storage

- possible resource sharing: significant resources procured by user groups
- integrating the resources to the infrastructure
- user group has a local storage facility
- geographically separated replicas
 - possible to switch to another replica during local failures
- available capacity used for other data

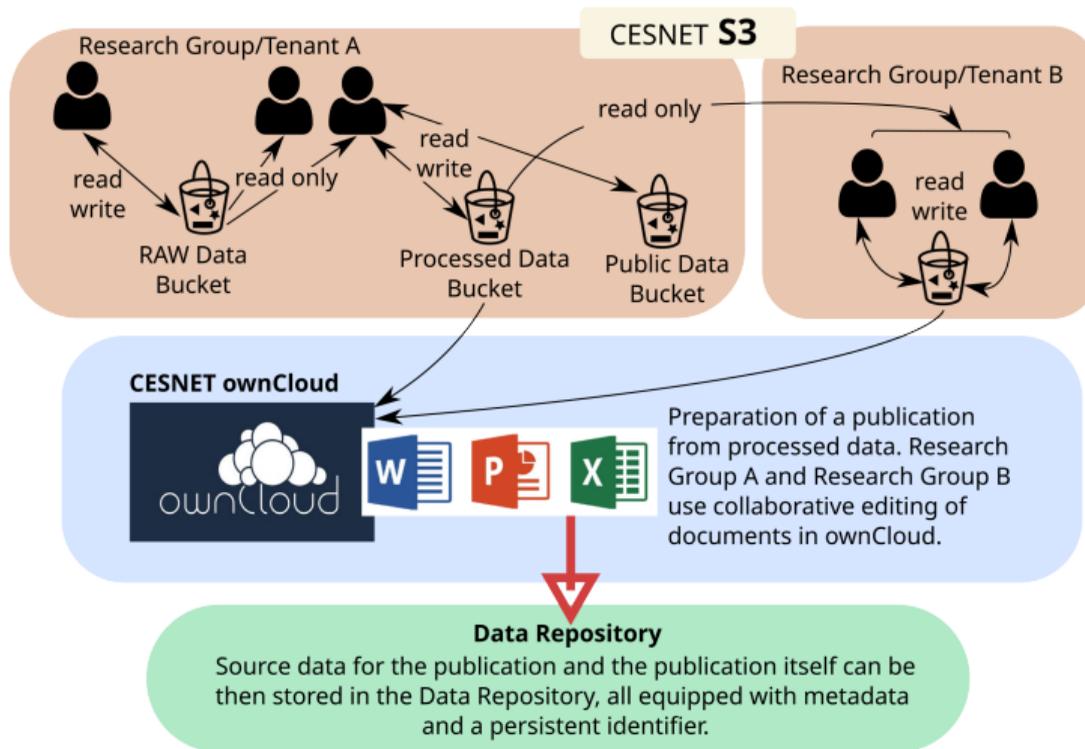
Long-term Storage

- repository for guaranteed long-term storage
 - binary-safe storage
 - accepting Archival Information Packages (OAIS)
 - that is verified in regular intervals
 - audited
 - errors are corrected from other package replicas
- can be used as a bit-safe storage for full OAIS repository

Data Repositories

- repository: data with metadata
- “more in the following presentation about EOSC plans”
- currently
 - prototype of a data repository
 - <https://data.narodni-repozitar.cz>
 - proof-of-concept for the national repository platform
 - curated by the National Library of Technology
 - mostly to gain experience
 - can be used for long-term data storage
 - data stays
 - interfaces/procedures may change

Complex Workflows



Summary

- storage services
 - transferring files—FileSender
 - sync'n'share—ownCloud
 - standard file and object storage
 - long-term binary guaranteed storage
 - repository (pilot)

Contacts

- `http://du.cesnet.cz`
- user support: `support@cesnet.cz`