

DATOVÁ ÚLOŽIŠTĚ

Milan Daneček

15. 5. 2024

Datová úložiště I

- infrastruktura
 - všechna úložiště jsou umístěna v ČR
 - hierarchická úložiště (pásková knihovna & disky)
 - disková pole
 - objektová úložiště (budoucí rozvoj)
 - aktuálně 120 PB instalované kapacity
- organizační pohled
 - provozováno uvnitř komunity
 - data jsou uložena na území ČR s příslušnou legislativní regulací
 - data vždy patří uživatelům
 - srovnajte s provozními podmínkami komerčních poskytovatelů

Datová úložiště II

- v provozu do konce 2024
 - hierarchické úložiště s páskovou knihovnou
 - diskové pole
 - momentálně vše migrujeme na nová objektová úložiště
- od teď a na vždy
 - objektová úložiště CL1...CL5
 - 120 PB instalované kapacity
 - objektová úložiště se budou kontinuálně obnovovat – bez omezení uživatelů

Sdílení objemných dat – Filesender

- FileSender: webová služba pro sdílení „velkých souborů“
 - „velké“: aktuálně 500GB
- <http://filesender.cesnet.cz>
- alespoň jeden uživatel musí být způsobilý pro využití služby
 - t.j. přihlášení pomocí eduID.cz federace
- po nahrání souboru se cílovému uživateli odešle link ke stažení
- alternativně je možné vygenerovat pozvánku pro upload a poslat mailem

Synchronizace dat mezi více zařízeními – ownCloud

- ownCloud: sync'n'share („cloudové“) úložiště
 - klient pro oblíbené rodiny OS (Windows, Linux, Mac OS X)
 - mobilní aplikace pro Android a Apple (placená aplikace cca 25 Kč)
 - webové rozhraní
- data jsou synchronizovaná přes server
 - lokálně je vždy držena kopie
 - na mobilních zařízeních se data synchronizují jen v případě požadavku
- data je možné sdílet
 - mezi uživateli anebo pomocí veřejného linku
- stačí se registrovat na <http://owncloud.cesnet.cz>
- 100GB limit na jednoho uživatele

Objektová úložiště

- postaveno na technologii Ceph
- základní služby S3, RBD, CephFS
 - S3 – základní případy užití: automatizované uploady dat, aplikace vydávající data, LMS, ...
 - RBD – větší objemy s potřebou filesystému, nutnost připojení na Linuxový stroj s veřejnou IPv4 adresou
 - CephFS – specializované aplikace
- všechny objektové clustery jsou umístěné v ČR – možnost geografické optimalizace

- systémy pro ukládání dat s extenzivními metadaty
- podporují FAIR principy
- umožňují přístup přes webové rozhraní a API – strojové operace
- jsou zodpovědné za uložená data – zálohy a redundance
- repozitáře jsou technické a procesní řešení pro dlouhodobé uložení dat a jejich veřejné publikování
- nejedná se pouze o dlouhodobý archiv odložených/historických dat

- Národní Repozitářová Platforma
 - distribuovaný systém, podpora skupin/oborů, institucionalizace
 - v rámci NRP projektu se partneři podílejí na vývoji jednotlivých řešení pro repozitáře
- implementace repozitářů v rámci NRP
 - CESNET Invenio (CESNET), CLARIN DSpace (UK), ASEP/ARL (AV ČR)
- v rámci NRP budou vznikat další institucionální, oborové repozitáře
 - převážně s použitím těchto implementací
- v rámci vyvíjených řešení se předpokládá i podpora specializovaných případů užití – laboratorní deníky, výpočetní workflow, citlivá data, ...
- očekávaná kapacita do 2028: 250 PB instalovaných / 50 PB uživatelských

Pilotní catch-all repozitář, metadatový adresář

- Catch-all repozitář
 - podpora uživatelských rolí (uživatel, kurátor, publisher)
 - přístup přes webové a API rozhraní
 - základní metadatové schéma, DOI, ...
 - upload/download přímo do S3
 - catch-all repozitář:
 - <https://data.narodni-repozitar.cz/>
- Národní metadatový adresář
 - agregátor metadat z dostupných repozitářů v ČR
 - jednoduché vyhledávání dostupných dat z jednoho místa
 - národní metadatový adresář:
 - <https://nma.eosc.cz/>

- základní služby
 - Filesender, ownCloud
 - Objektová úložiště – S3, RBD, CephFS
- Repozitáře pro vědecká data
 - data opatřená metadaty, FAIR
 - NRP – repozitářová platforma, komunity, oborovost
- pilotní provoz catch-all datového repozitáře (instance NRP)
 - <https://data.narodni-repozitar.cz/>
- národní metadatový adresář:
 - <https://nma.eosc.cz/>

Kontakty



- <https://du.cesnet.cz>
- uživatelská podpora: support@cesnet.cz



Co-funded by
the European Union



MINISTRY OF EDUCATION,
YOUTH AND SPORTS

A large graphic consisting of two concentric, dark blue circles. The text 'e-infra.cz' is centered within the inner circle.

e-infra.cz

