

Datová úložiště CESNET

David Antoř

CESNET

20. 9. 2016



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



- ▶ proč budujeme národní datová úložiště pro vědu a výzkum
- ▶ hlavní výhody úložišť
- ▶ typické případy užití úložišť
- ▶ plány rozvoje infrastruktury a služeb

Proč máme úložiště v infrastruktuře

- ▶ uživatelé raději drží data na vlastním zařízení. . .
 - ▶ to ale nedává rozumné garance dlouhodobého uchovávání
 - ▶ používají se nevhodná média, chybí systematická péče
 - ▶ data je lépe uchovávat v živém systému
- ▶ konsolidace zdrojů – národní úložiště vědeckých dat
 - ▶ umožňuje pořídit zařízení dostatečné velikosti
 - ▶ spravované specializovaným týmem
 - ▶ s možností sdílení dat mezi skupinami uživatelů

- ▶ potřeba dlouhodobě uchovávat vědecká data vzrůstá
 - ▶ uchování primárních dat z experimentů a výpočtů
 - ▶ s ohledem na potenciál jejich dalšího využití
 - ▶ k návratu a revizi publikovaných výsledků
- ▶ uchování zajišťujeme na úrovni zachování binárních dat (bitstream)
 - ▶ správce úložiště nerozumí odborné povaze dat
 - ▶ ani nemůže rozhodovat o smysluplnosti jejich archivace
 - ▶ proto sami nezajišťujeme plné LTP (Long-Term Preservation)
 - ▶ bitstream je nicméně předpokladem pro uchování dat
 - ▶ další kopie, geografické repliky, redundance v systému, kontrolní součty
 - ▶ kurátorství dat je vždy role uživatelské skupiny

- ▶ služba je provozována v rámci komunity
 - ▶ CESNET je vlastněn univerzitami a AV
- ▶ data uživatelů zásadně patří uživatelům
- ▶ úložiště jsou na území ČR
- ▶ srovnajte s provozními podmínkami komerčních poskytovatelů

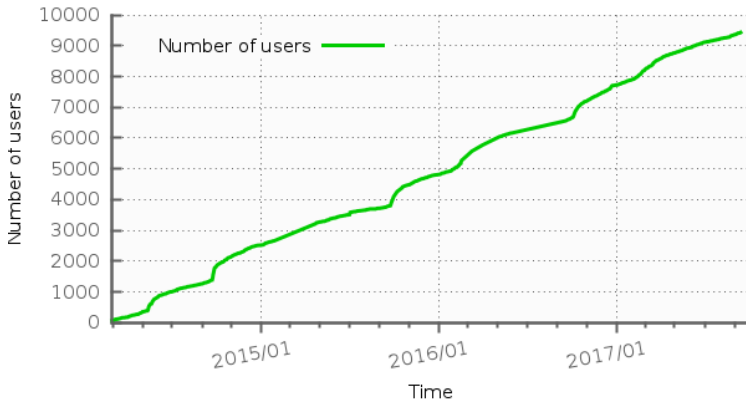
- ▶ FileSender: webová služba pro jednorázový přenos (velkých) souborů
 - ▶ „velkých“: aktuálně 500 GB
- ▶ přispíváme i k jeho vývoji
- ▶ <http://filesender.cesnet.cz>
- ▶ alespoň jedna strana komunikace musí být oprávněný uživatel infrastruktury
 - ▶ autentizace federací eduID.cz
- ▶ oprávněný uživatel může nahrát soubor a poslat mu oznámení
- ▶ lze poslat komukoli pozvánku
- ▶ testovací verze na <http://filesender2.cesnet.cz>

- ▶ pokud hledáte způsob, jak
 - ▶ synchronizovat data mezi svými počítači
 - ▶ i mobilními zařízeními
 - ▶ mít je zároveň dostupná přes web
 - ▶ moci data sdílet
- ▶ a přitom
 - ▶ data nejsou příliš velká
- ▶ ownCloud

- ▶ cloudové úložiště se synchronizací souborů
 - ▶ pro osobní počítače (Windows, Linux, Mac OS X)
 - ▶ mobilní Android, Apple (klient pod 1 Euro)
 - ▶ a webové rozhraní
- ▶ data se synchronizují přes úložiště
 - ▶ na počítači jsou i lokálně, na mobilní při otevření
- ▶ data lze sdílet
 - ▶ konkrétní osobě nebo „kdo zná odkaz“
- ▶ registrace federací na <http://owncloud.cesnet.cz>
- ▶ standardní limit 100 GB na uživatele

CESNET ownCloud users

2017-09-14 07:12:01



Klasické možnosti využití úložišť

- ▶ zálohy
 - ▶ uživatelé mají primární data u sebe
 - ▶ na úložiště odkládají zálohu pro případ havárie
 - ▶ buď pro zálohování jednotlivých strojů
 - ▶ nebo i agregovaně – IT oddělení zálohuje celou katedru
- ▶ archivace
 - ▶ uživatelé na úložiště odkládají cenná primární data
 - ▶ data nejsou často využívána
 - ▶ uživatelé nemají prostředky pro jejich uchování
 - ▶ individuální přístup koncových uživatelů vs. „laboratorní archivář“

- ▶ sdílení dat
 - ▶ distribuovaný tým potřebuje společně pracovat nad většími objemy dat, případně je zveřejňovat
 - ▶ typicky koncoví uživatelé
- ▶ „něco jiného“
 - ▶ distribuce obsahu, jiné speciální aplikace
- ▶ realizováno přístupem k souborovému systému nebo objektovému úložišti
 - ▶ sadou obvyklých přístupových protokolů
 - ▶ máme sadu hierarchických úložišť, celkem cca 22 tis. TB
 - ▶ správa uživatelů je plně integrována do národní e-infrastruktury

- ▶ spolupracovat se skupinami, které mají velká data
 - ▶ ukládání, zpracování
- ▶ komunitní modely budování úložišť
 - ▶ s HSM není snadné podpořit model „a co kdybych vám koupil disky/pásky/...“
 - ▶ jsme monolitický poskytovatel *kapacity*
- ▶ chceme umět podpořit scénář typu „chtěl bych na tři místa uložit 500 TB dat“
 - ▶ „tak si kupte 1500 TB a zapojte se k nám“
 - ▶ „500 budete mít přímo u sebe, další dvě repliky zařídí infrastruktura, zbylou kapacitu úložiště použijeme pro repliky ostatních“
 - ▶ a provozovali bychom dostatečnou infrastrukturu pro pokrytí nárazových a dočasných potřeb

- ▶ `http://du.cesnet.cz`
- ▶ uživatelská podpora: `du-support@cesnet.cz`