

Ukládání dat v e-infrastruktuře a Národní repozitářové platformě

David Antoš

30. 9. 2026

Fundamentální problémy úložišť

- chceme, aby úložiště bylo
 - (nekonečně) velké
 - (zatraceně) rychlé
 - (absolutně) spolehlivé
 - (skoro) zadarmo
 - (snadno) použitelné

Fundamentální problémy úložišť

- chceme, aby úložiště bylo
 - (nekonečně) velké
 - (zatraceně) rychlé
 - (absolutně) spolehlivé
 - (nejlíp úplně) zadarmo
 - (snadno) použitelné
- to má praktická omezení
 - fyziku, stav technologie
 - ekonomiku
- jak z toho ven?
 - vysvětlit uživateli rozdílné použití jednotlivých služeb

Různá úložiště pro různé účely I

- úložiště výpočetních systémů
 - home/scratch
 - většinou souborové systémy přímo dostupné z výpočetních uzlů
 - o těch tohle nebude
- obecná úložiště
 - „objektová“, cca 136 PB fyzické kapacity
 - softwarově-definované úložiště postavené na Ceph
 - pro uživatele S3(, RBD, CephFS)
 - přímý přístup k úložišti

Různá úložiště pro různé účely II

- speciální úložné služby
 - předávání souborů (FileSender)
 - synchronizace a sdílení (ownCloud / CesBox)
 - služby pro větší data s „hezčím uživatelským rozhraním“ než S3 (Onedata)
- repozitáře v Národní repozitářové platformě
 - (velká) data s popisnými metadaty pro dlouhodobé uložení

Objektová úložiště

- úložné clustery CL1–CL5
 - uživatelům dostupná protokolem S3
 - tedy „dostanete přístupové klíče a nastavíte si klienta“
 - případně RBD, CephFS
 - přímý přístup k úložišti
- pro „obecná data“
 - cca 136 PB fyzických médií
 - neplánuje se zásadní navyšování kapacity¹
 - postupně obnovujeme

¹kapacita úložišť za posledních 18 měsíců zdražila cca 5 ×

Použití obecných úložišť

- krátkodobé a střednědobé uložení *vědeckých dat*, např.
 - velká výpočetně zpracovávaná data
 - běžně desítky až stovky TB
 - až PB dat ke zpracování
 - velká dočasná data, která nedávají smysl v repozitáři
 - “to be FAIRified”
- pro pokrytí doby, než si odborné skupiny vybudují repozitáře nad NRP
- nejsou ale přijatelné stovky TB ve formě
zaloha2023/zaloha2020/zaloha-stars/nevim/zbytky
- funkci archivu přebere NRP

Speciální služby úložišť

- FileSender – předávárna souborů
 - <https://filesender.cesnet.cz>
- ownCloud
 - nástroj pro synchronizaci a sdílení dat
 - <https://owncloud.cesnet.cz>
 - dostupný na registraci
 - má webové rozhraní
 - synchronizuje data mezi serverem, počítačem a mobilem
- ... nedává smysl mít v sync-and-share systému například složky z githubu/gitlabu nebo dokonce onedrive

Nástroje pro synchronizaci a sdílení

- ownCloud $\rightsquigarrow$ CesBox
- postaven na Nextcloudu
- budeme postupně oslovovat uživatelské skupiny
 - převádět je na CesBox
 - vybereme skupiny uživatelů, kteří mezi sebou sdílejí data
- novinky v CesBox:
 - „projektové složky“ pro „data patřící týmu/pracovišti“
 - federované sdílení „v evropském EOSCu“
- bude zapojen jako služba národního EOSC uzlu

- platforma pro správu velkých dat nad distribuovanými úložišti
- poskytuje jednotný souborový prostor
 - umí do něj integrovat řadu úložišť různých technologií
- jako úložiště používáme S3
- více možností přístupu ke stejným datům:
 - uživatelsky přívětivé webové rozhraní
 - včetně detailních přístupových práv
 - klienti pro Linux a Windows (beta)
 - REST API a knihovna pro Python (vhodné pro integraci do workflows)
 - S3 protokol
 - ...

Onedata II

01

DATA

+

Marketplace

Sand

Search

Sandbox

3 GB

2

Overview

Files

Shares, Public Data

Transfers

Datasets, Archives

Providers

Members

Harvesters, Discovery

Automation Workflows

e-INFRA CZ

Onezone 25.1

FILES

View provided by  cloud2.onedata.e-infra.cz. [Choose other Oneprovider...](#)

 Sandbox / Data



Files

Size

Modified
(content)



Folder A

363.4 KiB

23 Sep 2026
7:48:33



Folder B

656.2 MiB

23 Sep 2026
9:05:10



file1.tgz

983 B

23 Sep 2026
9:06:43



file2.tgz

5.7 KiB

23 Sep 2026
9:06:42



file3.tgz

1.9 KiB

23 Sep 2026
9:06:42



file4.txt

556 B

23 Sep 2026
9:10:37



- systém je uzpůsoben pro ukládání vědeckých dat
 - vhodné pro sdílenou práci s velkými datovými sadami
 - podpora přidávání metadat ke složkám či souborům
 - snadnější přenos dat k výpočetním prostředí (integrace do Galaxy, Jupyter Notebooks)
- e-INFRA CZ instance systému Onedata:
 - <https://onedata.e-infra.cz>
 - dostupná pro všechny uživatele e-INFRA CZ (sandbox na vyzkoušení)
 - zřízení úložného prostoru na požádání onedata@e-infra.cz

Co je NRP?

- Národní repozitářová platforma
 - distribuovaný, multi-tenant systém pro vytváření repozitářů
 - distribuovaný: geograficky
 - multi-tenant: řada speciálních repozitářů (nikoli jeden velký)
 - vytváření repozitářů: dovoluje vytváření repozitáře z připravených komponent „jako službu“
- vezmeme to ze strany uživatele

Co je repozitář? I

- v první řadě úložiště dat s (obsahovými) metadaty
- podporuje FAIR principy
 - **nalezitelnost/dostupnost/interoperabilita/znovupoužitelnost**
- obvykle má webové rozhraní a API pro strojový přístup
- má odpovědnost za uložená data
- potenciálně certifikovatelný na Core Trust Seal
 - viz např. <https://www.clarin.eu/content/checklist-clarin-b-centres>

Co je repozitář? II

- data se ukládají ve formě datových sad
- datová sada: soubory s daty a popisná metadata
- repozitář musí obsahovat „citovatelné datové sady“
 - zajišťuje jejich neměnnost² a dlouhodobé³ uchování
- repozitář je technický, personální a procesní systém pro dlouhodobé uchování a zpřístupnění citovatelných digitálních objektů

²až na řízené opravy

³na škále „roky až jedno furt“

Co NRP není

- NRP není prostředí pro výpočty
 - na to máte Metacentrum a superpočítačové centrum
- NRP není jen pro „otevřená data“
 - FAIR data: musí být jasné podmínky přístupu (a to nemusí být „veřejně všem“)
 - „data v repozitáři“ $\neq$ „zveřejněná“
 - rozsah přístupu určuje repozitář a uživatel
- NRP nevynucuje podmínky repozitáře
 - politiku repozitáře určuje jeho správce

Repozitáře z pohledu koncového uživatele

- hledání dat
 - hledat datové sady lze v Národním metadatovém adresáři
 - NMA (<https://nma.eosc.cz/>)
 - agregátor metadat
 - dovede uživatele k repozitáři s daty
 - repozitář s daty umí jemnější vyhledávání
- pro ukládání dat je třeba zvolit oborově příslušný repozitář
- ukládání dat
 - přes webové rozhraní nebo API
 - vyplnění metadat je dobré automatizovat
 - detaily závisí na vlastnostech konkrétního repozitáře

Catch-all repozitář I

- <https://datarepo.eosc.cz/>
- „repozitář poslední volby“
 - pro individuální vkladatele, kteří nejsou zapojení ve výzkumných komunitách
 - nebo pro malé komunity
- s obecným metadatovým modelem
- záznamy mohou být zveřejněny/zveřejněna pouze metadata/neveřejné
 - lze nastavit embargo
- lze přiřadit DOI
- pokud reprezentujete větší organizovanou komunitu uživatelů, chcete specifický repozitář

Dataset for Tapinarof as Topical Drug with Potential to Form Pi-stacking and Dimers

Sokolová, Romana¹ 

Show affiliations

Contributors

Researcher (5):

Dostálková, Lucie¹ ; Vacek, Jan² ; Papoušková, Barbora³ 

Zatloukalová, Martina² ; Kabeláč, Martin⁴ 

Show affiliations

Dataset for article Tapinarof as Topical Drug with Potential to Form Pi-stacking and Dimers, which present a mechanistic study on the oxidation of tapinarof (TPN, 3,5-dihydroxy-4-isopropylstilbene) in aqueous and mixed solutions. Oxidation mechanism of tapinarof was proposed. It involves oxidative spectroelectrochemical (SEC) studies, UV-Vis and IR spectroelectrochemistry and theoretical molecular calculations. The aggregation (non-covalent association) behavior and thermodynamic stability of TPN in water and acetonitrile were investigated using empirical molecular dynamics (MD) simulations and density functional theory (DFT) calculations with dispersion corrections.

Methods (English)

Cyclic voltammetry in both aqueous and non-aqueous solutions on glassy carbon and platinum electrodes.

Chronoamperometry for exhaustive oxidative electrolysis, chromatography HPLC-DAD and HPLC-MS/MS for identification of oxidation products.

UV-Vis and infrared spectroelectrochemistry for identification of the first reaction intermediates.

Empirical molecular dynamics (MD) simulations were performed using the AMBER 14 suite of programs. The dimerization stability of these complexes was evaluated using Density Functional Theory (DFT). Two explicit solvent environments were used: Water was modeled using the TIP3P potential, acetonitrile was modeled using parameters

Version v1

Aug 4, 2026

10.48700/bvd89-eg030

Cite all versions? You can cite all versions by using the DOI 10.48700/5c2bb-3e078. This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

Communities



J. Heyrovsky Institute
Community

Keywords and subjects

tapinarof Oxidative transformation
Oxidation mechanism
infrared spectroelectrochemistry
DFT calculations UHPLC-ESI-MS

Details

DOI

DOI 10.48700/bvd89-eg030

Copy to clipboard



Repozitář z pohledu správce repozitáře I

- reprezentujete vědeckou komunitu, která chce provozovat repozitář
 - provozování repozitáře představuje netriviální závazek
- od NRP můžete dostat
 - repozitář jako službu
 - nebo infrastrukturu pro jeho běh jako službu
- v NRP podporujeme tři základní implementace repozitářových systémů
 - CESNET-Invenio, CLARIN-DSpace, ASEP/ARL
 - přesné rozdělení rolí závisí na požadovaných úpravách pro komunitu
 - samotný provoz repozitáře zajistí infrastruktura

Repozitář z pohledu správce repozitáře II

- je možno provozovat i jinou implementaci repozitáře
 - pak správce repozitáře dostane „Kubernetes + S3“
- lze dohodnout individuální rozdělení rolí při použití zdrojů NRP
- organizační aspekty viz
https://docs.nrp.eosc.cz/en/docs/repo_admins/repo_admins

- jaké problémy řeší datová úložiště
- proč máme více typů úložišť
- objektová úložiště
- speciální služby
- repozitáře a Národní repozitářová platforma

Shrnutí a kontakty

- při výběru služby úložišť se podívejte na základní vlastnosti
 - zda se hodí pro zamýšlený účel
- pokud si nejste jisti, napište
- kontakty a dokumentace
 - support@cesnet.cz
 - <https://docs.du.cesnet.cz>
 - <https://docs.nrp.eosc.cz>



**Nejdříve nasadte kyslíkovou masku sobě,
poté pomáhejte ostatním**

The logo for e-infra.cz, consisting of the text 'e-infra.cz' inside a circle, which is flanked by two curved lines.

e-infra.cz