

Příprava datových souborů před vložením do repozitáře ASEP

Příprava dat

- <https://www.lib.cas.cz/asep/pro-autory/priprava-dat/>

ASEP
Evidence výsledků
vědecké práce v AV ČR

Knihovna
Akademie věd ČR

Co máme hledat? Hledat v katalogu Hledat

ASEP Analytika ASEP Pro zpracovatele **Pro autory** Basic Information

▼ Pro autory

- Uživatelské účty
- Zápis publikační činnosti
- Zápis datového záznamu
- ▼ **Příprava dat**

REPOZITÁŘ ASEP >

ANALYTIKA ASEP >

ASEP » Pro autory » Příprava dat

Příprava dat

- [1. Organizace dat, datasety v ASEP](#)
- [2. Příprava datasetů](#)
- [3. Popis datasetu](#)
- [4. Dohoda o uložení](#)

1. Organizace dat, datasety v ASEP

Organizace dat

Výběr souborů k uložení a archivaci je první krok, který musí autor udělat, protože ne všechny soubory je potřeba (a možno) archivovat. Neexistuje univerzální způsob organizace dat, kterým se mohou řídit všichni autoři. Některým autorům pomáhá při organizaci dat (některé projekty ho přímo vyžadují) vytvoření DMP (Data Management Plan). Na Internetu je zveřejněna řada videí a článků, které mohou být inspirací, pokud autoři nemají již vytvořen svůj vlastní systém organizace

Organizace dat

- výběr vhodných souborů k archivaci
- **DMP (Data Management Plan)** podmíněný u projektů EU, nástroj na správu dat (formální dokument, který popisuje způsob jakým se mají zpracovávat údaje během výzkumného projektu i po něm)
- od letošního roku je povinnost ukládat data k výstupům z projektů EU

Organizace dat

1. Organizace dat, datasety v ASEP

Organizace dat

Výběr souborů k uložení a archivaci je první krok, který musí autor udělat, protože ne všechny soubory je potřeba (a možno) archivovat. Neexistuje univerzální způsob organizace dat, kterým se mohou řídit všichni autoři. Některým autorům pomáhá při organizaci dat (některé projekty ho přímo vyžadují) vytvoření DMP (Data Management Plan). Na Internetu je zveřejněna řada videí a článků, které mohou být inspirací, pokud autoři nemají již vytvořen svůj vlastní systém organizace dat.

DMP online: <https://dmponline.dcc.ac.uk/> , Check list of DMP: http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP_Checklist_2013.pdf

Organizace dat: <http://datalib.edina.ac.uk/mantra/#sthash.jZMT0wuz.dpuf> , <https://www2.usgs.gov/datamanagement/>

Datasety v ASEP

- Do repozitáře ASEP lze **uložit datový záznam a dataset**.
- Pokud jsou datasety uloženy v **jiném repozitáři** či **úložišti**, lze vytvořit jen **datový záznam** a uvést **odkaz na dataset**.
- V **bibliografickém záznamu** v ASEP lze vytvořit **odkaz na data** v **jiném repozitáři**.
- **Datové a bibliografické záznamy** v ASEP lze **vzájemně propojit**.

[Seznam repozitářů a příklady jiných úložišť](#)

Příprava datasetů

- zamyslet se nad strukturou (může být více souborů nebo jen jeden, mohou být uloženy vícekrát v rozdílných formátech)
- nepodceňovat vhodné pojmenování
- je vhodné opatřit dataset dokumentací – pokud nestačí popis pomocí metadat, přiložit soubor read_me.txt (Word nebo textový soubor – stručně popisuje co je obsahem datasetu, osnova, struktura)

Příprava datasetů

Velikost

- maximální velikost jednoho datasetu k záznamu je 20GB
- maximální velikost jednoho nahrávaného souboru je 2GB
- maximální počet uložených souborů k jednomu záznamu je 30
- potřebuje-li autor nahrát soubor/dataset větší než je maximální povolená velikost, je nutné kontaktovat správce repozitáře arl@lib.cas.cz
- při ukládání většího počtu souborů doporučujeme zazipování

Příprava datasetů

Formáty

- při volbě formátu je nutné si uvědomit, že je třeba vybrat takový, který bude přístupný i v budoucnosti (vyhnout se formátům v budoucnu nečitelným)
- tytéž datasety je možné uložit vícekrát v různých formátech (eliminujeme tím případnou ztrátu dat)
- seznam doporučených formátů je dostupný na stránkách ASEP



Struktura datasetů a pojmenování souborů

Autor se musí zamyslet, jak budou datasety výzkumných dat **strukturovány**. Datasety mohou být složeny z různých souborů, nebo dataset může obsahovat i pouze jeden soubor, tytéž soubory mohou být uloženy i v několika různých datasetech. K přehlednosti též pomůže **vhodné pojmenování souborů**, které by se nemělo podceňovat.

Dokumentace k datasetu

Obsah datasetů je důležité přehledně popsat, aby bylo jasné i uživateli, který si soubory stáhne pro další využití, jaké soubory dataset obsahuje. Pokud k popisu nestačí pole popisu v datovém formuláři, doporučujeme přiložit soubor **read_me.txt**, ze kterého bude zřejmé, co je v jednotlivých datových souborech uloženo.

Pro ASEP platí:

- **Maximální velikost** uloženého datasetu k jednomu datovému záznamu v ASEP je **20 GB**.
- Přes **webový prohlížeč** lze nahrávat soubor maximální velikosti **2 GB**.
- Maximální počet uložených souborů k jednomu datovému záznamu je **30**.
- Při ukládání **většího počtu souborů**, doporučujeme jejich **kompresi** „zazipování“ (zip, rar).
- Pokud potřebujete nahrát jeden soubor, který je větší než 2 GB, nebo uložit dataset, kde součet velikosti souborů je větší než 20 GB, obraťte se pro konzultaci možností na arl@lib.cas.cz.
- Volba názvu souborů v datasetu a přiložení souboru read_me.txt je na vkladateli.

Volba formátu souborů

Při výběru formátu souboru je dobré se ujistit, že formát, který zvolíme bude v budoucnosti přístupný. Doporučujeme zvolit formát standardní, ale v případě, že se pro data v praxi používají formáty jiné, lze data uložit i v těchto formátech.

U dat, která se jsou vytvořena, upravována či komprimována **speciálním softwarem** je dobré uvést v popisu, jaká verze software byla použita, jaké kódování, komprese apod., pro snadný přístup a další využitelnost uživateli.

Tatáž data lze uložit v několika formátech, soubory v jednom datasetu mohou mít různé formáty.

Pro ASEP platí:

- Seznam doporučených formátů je [zde](#)
- Pokud je potřeba uložit data ve formátu, které nejsou definovány v seznamu formátů ASEP (např. ve specifických oborových formátech), napište svůj požadavek správci repozitáře na arl@lib.cas.cz.

Popis datasetu

Popis polí

- v manuálu je podrobný popis všech polí formuláře
- rozlišujeme **povinná pole** (bez nich nelze záznam zveřejnit), **podmíněně povinná pole** (pokud je to možné doporučujeme vyplnit) a **ostatní pole** (závisí na uvážení autora)

Licence

- pravidla pro uživatele jak mohou s daty nakládat - nutno uvést
- možno použít licence **Creative Commons** (přednastavené) nebo použít vlastní (informovat arl@lib.cas.cz)

Popis datasetu

Přístup k datovým souborům

Stejně jako u bibliografických záznamů, i zde rozlišujeme 4 stupně přístupu:

- veřejně přístupný
- veřejně přístupný s embargem
- přístupný pro ústav
- veřejně nepřístupný (na vyžádání)

Popis datasetu

3. Popis datasetu

V manuálu ASEP je podrobný popis jednotlivých polí formuláře: [Struktura datových záznamů v ASEP](#).

Některá pole jsou **povinná** – bez vyplnění těchto polí nelze záznam zveřejnit. Pole **podmíněně povinná** jsou pole, která doporučujeme vyplnit, pokud je to možné (např. návaznost na projekty). Vyplnění polí **ostatních** závisí na zvážení vkladatele. Správné a úplné vyplnění polí je důležité pro tvorbu **citace**.

Licence a přístup k datasetu

U každého datasetu vkladatel musí uvést licenci, tj. **jaka pravidla platí pro uživatele, kteří si data stáhnou, jak je mohou využívat**. Je možno použít **licenci Creative Commons**, které jsou předdefinované, v případě, že platí **jiná licence**, musí se její text uložit s datasetem. Další informace o licencích [zde](#).

Pokud bude potřeba nějakou speciální licenci přidat, obraťte se na: arl@lib.cas.cz.

Přístup k souboru

Datové soubory mohou být **veřejně přístupné**, to znamená, že uživatel si soubory stáhne a uloží na svém počítači a může s nimi pracovat dle licence, která je zveřejněna u záznamu. Podporujeme také (dočasné) **embargo** na soubory s daty, což znamená, že je veřejně zpřístupňujeme až po určité době, která se může při vkládání nastavit. **Přístupné pro ústav** kde mají k souborům přístup pouze pracovníci ústavu na jehož půdě práce vznikla, ostatní uživatelé mohou o data požádat. **Veřejně nepřístupné** soubory jsou soubory, o které si musí uživatel požádat (vyplní formulář).

Dohoda o uložení

- před předáním dat zpracovateli musí autor odsouhlasit **Dohodu o uložení datasetů v repozitáři ASEP** (součást formuláře)
- obdrží-li KNAV důkaz o porušení autorských práv, bude záznam okamžitě odstraněn
- před zveřejněním dat je potřeba:
 - *mít všechna práva pro zpřístupnění dat*
 - *mít souhlas spoluautorů a ostatních držitelů práv k datům*
 - *mít data dostatečně anonymizována*

Dohoda o uložení

4. Dohoda o uložení

Před předáním dat ke zveřejnění musí autor souhlasit s dohodou o uložení datasetů do datového repozitáře: [Dohoda o uložení datasetů v repozitáři ASEP](#). V případě, že KNAV obdrží důkaz o porušení autorských práv, příslušná položka bude ihned odstraněna.

Co je potřeba si uvědomit před zveřejněním připravených datových souborů:

- Máte všechna potřebná práva pro zpřístupnění dat.
- Máte **souhlas spoluautorů** a ostatních držitelů práv k datům.
- Máte dostatečně **anonymizována svá data**, nebo výslovný souhlas od všech subjektů, jejichž totožnost by mohla být odhalena z dat.

Stránky v angličtině

- připravili jsme zjednodušenou verzi anglických stránek o repozitáři ASEP
- v budoucnu budeme připravovat formuláře v angličtině, aby ukládání dat bylo umožněno i pro vědce – cizince

Stránky v angličtině

ASEP

Evidence výsledků
vědecké práce v AV ČR

Knihovna
Akademie věd ČR

Co máme hledat? Hle

ASEP

Analytika ASEP

Pro zpracovatele

Pro autory

Basic Information

▼ Basic Information

▶ Institutional repository

▶ **Data repository**

▶ ASEP Analytics

[ASEP](#) » [Basic Information](#) » Data repository

Data repository

1. [ASEP data repository](#)
2. [Dataset deposit workflow](#)
3. [Agreement and licensing](#)
4. [Linking](#)



Děkuji za pozornost