

FAIR data v žádosti o Horizon Europe

V Knihovně ČZU jsme pro vás připravili tento návod, jak do žádosti o podporu z Horizon Europe popsat FAIR data. Návod obsahuje informace o tom, co vše by mělo být v žádosti popsáno. Způsob, jakým zajistíte, aby vaše data byla FAIR, si musíte zvolit individuálně.

Co tento návod obsahuje:

Co jsou to FAIR data?.....	1
Jak vytvořit popis FAIR dat do žádosti?	1
Types of data/research outputs:	2
Findability of data/research outputs:.....	2
Accessibility of data/research outputs:	3
Interoperability of data/research outputs:	3
Reusability of data/research outputs:.....	3
Curation and storage/preservation costs:.....	4

Co jsou to FAIR data?

FAIR data byla poprvé představena v roce 2016 v článku popisujícím tzv. [FAIR zásady pro správu vědeckých dat \(Wilkinson et al. 2016\)](#). Slovo FAIR je akronymem pro čtyři principy: Findability, Accessibility, Interoperability, Reusability. Dodržováním těchto principů lze maximalizovat hodnotu dat po jejich zveřejnění. V současné době se mnoho iniciativ snaží o to, aby se FAIR data stala standardem. **V rámci žádosti o Horizon Europe je nutné stručně popsat (max. 1 A4), jakým způsobem bude zajištěno, aby data vytvořená či použitá v projektu byla FAIR (viz průvodce programem Horizon str. 43-53).** Jednotlivé části FAIR dat se navzájem prolínají a doplňují, proto se nebojte, pokud budete mít některé informace uvedeny ve více částech.

Jak vytvořit popis FAIR dat do žádosti?

V prvním kroku doporučujeme si pročíst tento návod, který vám poskytne přehled toho, na co byste při popisu dat neměli zapomenout. Pro inspiraci doporučujeme podívat se také na dostupné Data Management Plans z jiných výzkumných projektů – využijte např. [Public DMPs na DMPOnline](#) nebo popis [FAIR principů Zenoda](#) (jen podotýkáme, že Data Management Plan a popis FAIR dat do žádosti jsou obsahově podobné dokumenty, nicméně mají jisté odlišnosti).

V druhém kroku začne samotná tvorba popisu FAIR dat. Zkuste popsat data a metadata nejlépe, jak jen to umíte. Zamyslete se nad výběrem repozitáře, licence atd., podle všech otázek, které jsou níže uvedené. V případě zájmu vám můžeme v Knihovně ČZU poskytnout konzultaci v oblasti FAIR dat. Bohužel nejme znalci ve všech oborech, ale můžeme poskytnout knihovnický náhled a možná vycytat případné neduhy. **Pokud máte zájem o konzultaci, kontaktujte nás na openscience@lib.czu.cz.**

Pokud máte zájem dozvědět se o FAIR datech něco více, můžete také navštívit web [Knihovny ČZU](#), [web OpenAIRE](#), případně si můžete prohlédnout [šablonu Data Management Plan pro Horizon Europe](#).

Types of data/research outputs:

Uvedte základní popis výzkumných dat, která v projektu vzniknou nebo je využijete. **Co**

je nutné do žádosti popsat?

- **Jaká data použijete?** ○ Popište, zda využijete již vytvořená data nebo vytvoříte v rámci vašeho projektu data nová. Pokud použijete již dříve vytvořená data, popište zde, jakým způsobem a proč je chcete znova využít.
- **Jaký je formát dat?**
 - Popište formát dat, který budete v projektu vytvářet či používat (budete pracovat např. s textem, obrázky, čísly – k tomu přiřadte odpovídající formát jako je např. txt, csv, jpg, atd.) - **Z jakého důvodu potřebujete data v projektu?** ○ Popište, proč jsou pro vás daná data v projektu důležitá, popište jejich očekávanou velikost, jejich původ a také potenciál pro další použití (tzn. data utility).

Findability of data/research outputs:

Nosnou myšlenkou FAIR dat je jejich vyhledatelnost. Vyhledatelnost dat je možné zajistit tím, že jsou datům přiděleny [persistentní identifikátory \(PID\)](#), a jsou bohatě popsána metadaty. (Meta)data jsou zpřístupněná v online prohledatelném zdroji (např. v katalogu či datovém repozitáři). Je důležité, aby metadata byla strojově čitelná.

Co je nutné do žádosti popsat?

- **Jaké PID budete přidělovat datům?**
 - Persistentní identifikátory zajišťují trvalé a jedinečné identifikování objektů. Příkladem persistentního identifikátoru může být ISBN, ISSN, URL anebo třeba DOI či Handle. Pro data je nejčastěji používaným PID formát DOI či Handle, ale existují i specializované PID v závislosti na oboru. ○ Většina repozitářů přiděluje objektům nahraným do repozitáře nějaký PID. Doporučujeme si proto zjistit u provozovatele repozitáře, jaký PID přiděluje (viz část Accessibility of data). Např. repozitář [Zenodo nahraným datům přiděluje DOI](#).
- **Jednoznačně identifikujte autory** ○ PID jsou vhodné i pro jednoznačnou identifikaci autorů dat. V žádosti doporučujeme zmínit, že data budou provázána s [ORCID \(či obdobným identifikátorem autora\)](#).
- **Jaké metadatové schéma použijete při popisu dat?**
 - [Metadata](#) popisují data a napomáhají tak k vyhledatelnosti, jednodušší citovatelnosti a znovu využitelnosti dat. V rámci žádosti byste měli popsat, které metadatové schéma použijete. Existují obecná metadatová schémata (např. [Dublin Core, DDI](#)), ale také oborově specifická. Pokud si nejste jistí, jaké metadatové schéma je vhodné pro váš obor, podívejte se na [DCC Metadata Directory](#), [RDA Metadata Directory](#) či [FAIRsharing](#).
 - Některé repozitáře na svých webech uvádějí, jaké metadatové schéma používají. Např. [Zenodo](#) používá [DataCite's Metadata Schema](#), které je v souladu s komunikačním protokolem [OAI-PMH](#). Metadata jsou také dostupná skrz [veřejnou REST API](#).
 - Je důležité, aby data byla bohatě popsána a optimalizovaná k tomu, aby bylo možné zajistit jejich potenciální znovu využití (a to jak lidmi, tak i stroji – tzn. musí být strojově čitelná). V rámci projektové žádosti tuto skutečnost popíšete tím, že uvedete, které metadatové schéma použijete.
 - Nebojte se popisu metadat, může to vypadat složitě, ale ve skutečnosti stačí krok po kroku následovat návod při nahrávání dat do repozitáře (většinou stačí poctivě vyplnit všechny informace o datech do předpřipraveného formuláře).

Accessibility of data/research outputs:

Data musí být přístupná za jasně definovaných podmínek a s využitím standardizovaných komunikačních protokolů, a to pro lidi i stroje. Je důležité zmínit, že FAIR data neznamená, že data musí být otevřená.

Co je nutné do žádosti popsat?

- **Kde a jakým způsobem budete data ukládat?**
 - o Důležité je zajistit bezpečnost, přístupnost a popis dat. Rozhodně nedoporučujeme zde popisovat, že svá data schraňuje např. na Google Drive či OneDrive.
 - o Doporučujeme využít repozitáře – buďto obecné, jako je např. [Zenodo](#), [FAIRsharing](#), [Harvard Datavers](#), anebo oborově zaměřené. Nevíte, jaký repozitář je pro váš obor vhodný? Zkuste prohledat databáze repozitářů [Re3data](#) nebo OpenDOAR.
- **Jakou licenci datům přidělíte?**
 - o Doplněte krátkou informaci o licencích, které budete pro data používat (viz část Reusability of data).
- **Budou data otevřená?**
 - o Podle Evropské komise se FAIR data řídí heslem „As Open as Possible, As Closed as Necessary“. Toto heslo znamená, že není nutné sdílet všechna data, ale je nutné zpřístupnit alespoň popis těchto dat (metadata). Pokud data nebudou otevřena, je potřeba zdůvodnit proč (např. že se jedná o citlivá data apod.).
 - o Otevřenost dat je dána licencí, kterou použijete – viz část Reusability. Pokud by se jednalo o citlivá data, je dobré zmínit schválení [Etickou komisí ČZU](#), případně vytvoření speciální komise, která by vyhodnocovala případné individuální žádosti o přístup k datům.
 - o Pokud je pro přístup k datům potřebný speciální software (např. nově vytvořený software či neobvyklý software), popište, o jaký software se jedná, případně popište, že bude vytvořena dokumentace k danému software. Není potřeba popisovat, že budete používat pro csv soubory excel apod.

Interoperability of data/research outputs:

Interoperabilita dat zajistí, že je možné data integrovat s dalšími daty, aplikacemi, pracovními postupy atd. Interoperability je docíleno tak, že data a metadata odpovídají uznávaným standardům a formátům, které jsou typické pro daný obor.

Co je nutné do žádosti popsat?

- **V jakém formátu/standardu budete data vytvářet?**
 - o Data by měla být popsána ve srozumitelných a otevřených formátech/standardech. Ideálním případem je, že pro popis dat bude použit řízený slovník či tezaurus, jednoduše správná terminologie daného oboru (doporučujeme se odkázat na ontologii nebo nějaký konkrétní tezaurus, podle kterého budete postupovat).
- **Jaký formát/standard použijete pro popis metadat?**
 - o I zde platí to samé jako je popsáno o odstavec výše. Metadata by měla dodržovat určitý, všeobecně uznávaný standard (viz část Findability)

Reusability of data/research outputs:

FAIR data jsou taková data, která jsou připravena být znovu využita v budoucích výzkumech (například pro revizi výzkumu, volné navázání na výzkum apod.).

Co je nutné do žádosti popsat?

- **Dokumentace dat** ○ Popište, jakým způsobem zajistíte systematickou dokumentaci dat (popis kódů a zkratk, které budete používat. Ideální je, pokud vycházíte z nějakého světově uznávaného standardu/formátu – viz část Interoperability).
 - Je dobré vytvořit dokumentaci dat (např. dokument README, kde popíšete vše podstatné pro práci s daty).
- **Kontrola integrity dat** ○ Popište, jakým způsobem budete zajišťovat integritu a verzování dat – např. jestli použijete [Github](#) nebo třeba [Open Science Framework](#) nebo nějaké jiné nástroje a funkce kontroly dat.
- **Přidělení licence** ○ Aby bylo možné data znovu využít, je nutné jim přidělit správnou licenci. Každá data mají odlišné způsoby a možnosti sdílení, proto je nutné zjistit, pod jakou licenci můžete vaše data sdílet. Nejčastěji se používají licence [Creative Commons](#). Dejte si pozor na požadavky poskytovatele financí – někteří uvádějí, které licence je možné použít.
 - Většina repozitářů přiděluje určitou licenci automaticky. Proto si při výběru repozitáře zjistěte i informaci o přidělované licenci.
 - Nevíte, jakou licenci přidělit? Zkuste [License Selector](#) nebo [Choose a CC licence](#).
- **Budou data znovu využita?**
 - Pokud už nyní víte, že někdo data další data využije, můžete zde tuto skutečnost uvést.

Curation and storage/preservation costs:

Tvorba FAIR dat je činnost náročná na čas i finanční prostředky. V této části žádosti uveďte, jaké náklady na tvorbu, uložení a celkově správu dat budete potřebovat.

Co je nutné do žádosti popsat?

- **Náklady na tvorbu a správu FAIR dat** ○ Při psaní žádosti o grant je dobré pamatovat i na finance určené na správu FAIR dat. Tyto finance jsou v rámci programu Horizon Europe uznány jako přímé náklady.
 - Náklady spojené s tvorbou a správou FAIR dat jsou např. náklady úložiště dat (repozitář), na pracovníky (Data Steward – specialista na data, data librarians, právník – specialista na licencování, oborový specialita) apod.
 - U určení nákladů je vhodné uvést nejen částku, ale také dobu, po kterou tyto náklady zajistí přístupnost dat (např. pokud bude nutné zaplatit poplatek za uložení dat v určitém repozitáři, tak je vhodné uvést na jakou dobu budou data dostupná, případně jak bude přístupnost dat řešena po skončení projektu).
- **Zodpovědnost za data management** ○ Uveďte osobu/osoby, které budou odpovědné za data management v průběhu celého projektu. Krátce můžete popsat i jejich funkci a povinnosti.
 - Pokud budete v projektu vytvářet specializované pozice pro tvorbu a správu FAIR dat, zde je prostor uvést bližší popis těchto pozic (např. pozice Data Steward).
- **Bezpečnost a etika dat** ○ Popište, jak bude zajištěna bezpečnost dat (viz část Accessibility). Pokud se jedná o eticky citlivá data, popište, jakým způsobem budete tuto skutečnost řešit (viz část Accessibility).

Vydala Česká zemědělská univerzita v Praze v rámci projektu Kancelář pro podporu mezinárodních projektů zaměřených na Life Sciences v rámci Evropského výzkumného prostoru III. (program MŠMT INTER-EXCELLENCE, podprogram INTER-INFORM, č. projektu LTI20001)

© 2021 Česká zemědělská univerzita v Praze