

Data Management Plan

Autor: Mgr. Bc. Aleš Mučka, Ph.D.
Grafické zpracování: Mgr. Martina Suráková

Ústřední knihovna VŠB-TUO
vydáno 10/2025
verze 1.0

Modelová grantová dohoda programu Horizon Europe vyžaduje, aby byl vypracován a pravidelně aktualizován plán správy dat (Data Management Plan, DMP). Příjemcům podpory z programu Horizon Europe se doporučuje používat tento vzor. Při vyplňování sekcí šablony je třeba zohlednit požadavky programu Horizon Europe na správu výzkumných dat, jak jsou popsány v článku 17 a rozebrány v Anotované grantové dohodě.

DMP je živý dokument, který musí jeho autoři pravidelně aktualizovat. Doporučuje se vypracovat aktuální verzi společně s každou průběžnou zprávou o výzkumném projektu. Poslední verze se vždy vytváří krátce po ukončení projektu a vědci v ní popisují, jaké postupy nakonec použili pro správu svých sesbíraných, vypořizovaných nebo vygenerovaných výzkumných dat.

1) DATA SUMMARY | SHRUTÍ DAT

Použijete znovu některá stávající data? K čemu je znovu použijete? Uveďte důvody, pokud bylo opakované použití stávajících dat zvažováno, ale zamítnuto.

Je vhodné uvést, že v projektu budou použita data, která byla předtím vytvořena v příbuzných studiích nebo oborech. Budou používána především pro srovnání s novými výsledky navrhovaného/probíhajícího projektu. Vždy je důležité uvést, na co bude projekt zaměřen.

Jaké typy a formáty dat budou v rámci projektu generovány nebo znovu použity?

Zásadní otázkou je, jaká data se budou shromažďovat a jaké soubory jsou pro ně nejvhodnější. Pro zachování interoperability je nutné se poradit s odborníkem na otevřenou vědu. I když zatím neexistuje žádná oficiální norma, doporučuje se:

- obrázky, nákresy, schémata: gif, jpg, png, tiff, případně pdf
- texty, zprávy, výpočty: doc, docx, pdf, txt
- výsledky měření, statistiky: xls,xlsx

Za jakým účelem se budou data generovat nebo opětovně používat ve vztahu k cílům projektu?

Zde je vhodné uvést, že získaná data budou znamenat vědecký posun v konkrétním vědním oboru, případně uvést konkrétnější dopady.

Jaká je předpokládaná velikost dat, které hodláte generovat nebo znovu použít?

Vědci by měli nejdříve s členy technické podpory projednat, kolik jim v úložištích poskytnou prostoru. Pokud již znají velikost budoucích datasetů, mohou je uvést ve formě textu nebo tabulky. Důležitým limitujícím faktorem je maximální povolená velikost ve zvoleném veřejném úložišti (repozitáři).

Jaký je původ dat, ať už vytvořených, nebo znovu použitých?

Většina dat bude v technických vědách pocházet z různých měření nebo pozorování (např. některých druhů ptáků v přírodě), v sociální oblasti z různých průzkumů a rozhovorů. Je běžné, že výzkumníci používají také starší vědecké projekty a jejich data. Pokud pro svůj výzkum uzavřeli smlouvy s třetími stranami, měli by zde uvést alespoň ty nejvýznamnější.

Pro koho mohou být vaše data užitečná („užitečnost dat“) kromě vašeho projektu?

Získané informace mohou být užitečné všem výzkumníkům v daném a příbuzných oborech a samozřejmě také soukromým společnostem. Praktické využití nashromážděných dat řeší vědci až na konci projektu.

2) FAIR DATA | FAIR DATA

Zajištění dohledatelnosti dat (findability) včetně ustanovení pro metadata

Budou data opatřena perzistentním identifikátorem?

Perzistentní identifikátory většinou přiděluje automaticky zvolený repozitář (nejčastěji DOI nebo Handle). Umožňují přesné dohledání otevřených dat a publikovaných odborných textů, i když dojde ke změně URL adresy, a proto je vhodné používat je také při jejich citování.

Budou poskytnuta bohatá metadata umožňující vyhledávání? Jaká metadata budou vytvořena? Jaké disciplinární/oborové nebo obecné standardy budou dodržovány? V případě, že ve vašem oboru neexistují metadatové standardy, uveďte, jaký typ metadat bude vytvořen a jakým způsobem.

Metadata jsou povinné a nepovinné údaje, které repozitář vyžaduje pro popsání uloženého datasetu. Pro větší přehlednost existuje několik obecných metadatových schémat, přičemž nejvíce používaným je DataCite Metadata Schema. Vědci musí povinně uvést název datasetu; jeho perzistentní identifikátor; typ vloženého souboru; datum publikování a jména, tituly a zaměstnavatele všech autorů. V elektronickém schématu je také mnoho nepovinných položek (např. klíčová slova, stručný popis atd.). Výzkumníci si vždy musí zjistit, jaké schéma používá jimi zvolený repozitář (většinou v sekci About).

Budou v metadatech uvedena klíčová slova pro vyhledávání, aby se zlepšila možnost vyhledávání a následného potenciálního opakovaného použití?

Zadávání klíčových slov či stručného popisu sice zabere čas, nicméně umožňuje potenciálním zájemcům mnohem rychlejší a snadnější nalezení datasetů a zvyšuje citovanost odborné publikace. Proto je vhodné zde uvést, že klíčová slova budou v metadatech zahrnuta.

Budou metadata nabízena takovým způsobem, aby je bylo možné automaticky zpracovávat (harvestovat) a indexovat?

Zde je rovněž vhodné uvést, že si výzkumníci uvědomují význam klíčových slov v přístrojovém vyhledávání výsledků svého projektu, a proto se pokusí uvést, co nejobecnější slova z odborné terminologie, aby případní zájemci měli možnost si je snadno vyhledat prostřednictvím webových služeb, vyhledávačů a dalších online nástrojů.

Making data accessible | Zpřístupnění dat (accessibility)

Repository | Repozitáře (úložiště)

Budou data uložena v důvěryhodném repozitáři?

Nejdříve by se autoři měli obrátit na své kolegy, kteří již mají s „otevíráním“ dat zkušenosti, a zjistit jejich praxi ověřené zvyklosti. Pokud ještě v jejich instituci nemají mnoho zkušeností s politikou otevřené vědy, existují různé „rozcestníky“ (nejznámější je re3data.org), díky kterým si mohou vědci vyhledat vyhovující obecné či oborové repozitáře. Také je možné využívat služeb institucionálních repozitářů jejich vlastní organizace, ačkoli ty se zatím nedoporučují.

Prověřili jste vhodná řešení s daným repozitářem, kde budou vaše data uložena?

Výzkumníci si mohou, většinou opět v sekci About každého repozitáře, zjistit základní údaje (finanční náklady, používané schéma metadat, velikost datasetů v počtech souborů a bitech atd.). Některá veřejná úložiště získala také ověřený certifikát, nicméně ne všechny důvěryhodné repozitáře jimi disponují.

Zajišťuje repozitář, přidělení identifikátoru k datům? Umožní repozitář přiřadit identifikátor k digitálnímu objektu?

Každý důvěryhodný repozitář by měl přiřazovat k datům perzistentní identifikátor (nejběžněji DOI, Handle aj.). Výzkumníci jej musí zjistit v sekci About nebo v New upload.

Data

Budou všechna data volně dostupná? Pokud určité soubory dat nelze sdílet (nebo je třeba je sdílet za podmínek omezeného přístupu), vysvětlete proč, a jasně oddělte právní a smluvní důvody pro záměrné omezení. Mějte na paměti, že v projektech s více příjemci je také možné, aby si konkrétní příjemci ponechali data uzavřená, pokud je jejich zpřístupnění dat v rozporu s jejich oprávněnými zájmy nebo jinými omezeními podle grantové dohody.

V rámci politiky otevřené vědy se musí zveřejňovat všechna výzkumná data, která byla použita k napsání odborných textů. Pokud výzkumníci usoudí, že informace by mohly být užitečné pro další výzkum, mohou zveřejnit všechna FAIR data, samozřejmě s ohledem na jejich velikost a možnosti úložiště.

Pokud je uplatňováno embargo, které má poskytnout čas na publikování nebo na ochranu duševního vlastnictví (např. patentů), uveďte, proč a jak dlouho bude časové embargo platit, s ohledem na to, že významná data by měla být zpřístupněna co nejdříve.

Pokud budou muset vědci využívat embargo, bude nutné ho posuzovat u každého případu individuálně. Také je možné, že nebudou během projektu embargo využívat vůbec. Nejčastěji se používá v případech, kdy výzkumníci sestavují dataset přímo v repozitáři, a proto nastaví časové omezení, dokud neshromáždí všechny potřebné soubory, a potom teprve ukončí embargo a umožní jejich zveřejnění.

Budou data přístupná prostřednictvím bezplatného a standardizovaného přístupového protokolu?

Data se musí zpřístupňovat prostřednictvím bezplatného a standardizovaného kódu, aby byla dodržena jejich interoperabilita, což je jeden ze základních FAIR principů. Možná omezení se musí detailně zdůvodnit a uvést plánovaná opatření.

Pokud existuje omezení k užití dat, jak bude zajištěn přístup k datům? Jak bude nastaven přístup v průběhu projektu, a jak po jeho skončení?

I pro otevřená výzkumná data existují výjimky, nelze totiž zveřejnit data, která by mohla porušit:

- | Ochranu národní bezpečnosti
- | Ochranu osobních údajů (GDPR)
- | Ochranu duševního vlastnictví.

Jak bude ověřena totožnost osoby, která dostane k datům přístup?

O ověření osoby se u dat uložených v repozitáři postarají jeho správci prostřednictvím registrace. Pokud bude třetí osoba žádat o přístup do interních úložišť konkrétní vědecké instituce, budou za její prověření odpovídat opět správci úložiště. Tvůrci výzkumných dat by jim měli pouze povolit zpřístupnění výsledků svého vědeckého bádání.

Je potřeba zřídit komisi pro přístup k datům (např. pro hodnocení/schvalování žádostí o přístup k osobním/citlivým údajům)?

Zde by vědci měli zvážit, kolik a jaká data pravděpodobně shromáždí během svého výzkumu a kolik výzkumných skupin bude do projektu zapojeno. Ačkoli se obojí může během projektu měnit, již před jeho zahájením by si mohli vytvořit alespoň přibližnou představu a poté zvážit, zda bude třeba sestavovat zvláštní komisi nebo vše potřebné zajistí administrátoři projektu a odborníci na otevřenou vědu.

Metadata

Budou metadata otevřeně dostupná a licencovaná pod veřejnou licenci CC-0, jak je uvedeno v grantové dohodě? Pokud ne, objasněte, prosím, proč. Budou metadata obsahovat informace, které uživateli umožní přístup k datům?

Jelikož přístupnost (accessible) je jedním ze základních FAIR principů a metadata by měla být zachována i po smazání původního datasetu, aby se v repozitáři uvolnilo místo, musí být opatřena vhodnou licencí, aby byla k dispozici všem potenciálním zájemcům i po delší době. V opačném případě by nemohly být výsledky výzkumu využity v navazujících projektech a snížila by se jejich citovanost.

Jak dlouho budou data dostupná a dohledatelná? Bude zaručeno, že metadata zůstanou dostupná i poté, co data už nebudou dostupná?

Data by měla zůstat dostupná po neomezenou dobu. Pokud správci veřejného, oborového nebo vnitřního (institucionálního) úložiště budou požadovat smazání souborů, aby uvolnili místo pro nové datasety, měli by na dobu neurčitou ponechat všechna metadata, která umožní potenciálním zájemcům kontaktovat tvůrce výzkumných dat.

Bude zahrnuta dokumentace nebo odkaz na jakýkoliv software, který bude potřebný pro přístup k datům nebo jejich prohlížení?

Dokumentace k projektu může být zveřejněna, ale optimálně jen v případech, kdy je nezbytná pro pochopení průběhu a závěrů „otevřeného“ výzkumu. Podle principů interoperability by se vědci měli vyhýbat používání obecně nedostupných softwarů, nicméně, pokud to bude nutné, mohou je používat. Speciální software by měl být snadno vyhledatelný a zdarma dostupný.

Bude možné zahrnout příslušný software (např. v otevřeném zdrojovém kódu)?

Pokud bude třeba použít speciální software, měl by dataset obsahovat přímý odkaz pro jeho online použití nebo stažení. Pro popsany účel je nejvhodnější zvolit některý oborový repozitář, který se zaměřuje na danou problematiku a většinou umožňuje sdílení otevřených zdrojových kódů pro konkrétní obor.

Making data interoperable | Interoperabilita dat

Jakými slovníky, standardy, formáty nebo metodikami dat a metadat se budete řídit, aby vaše data byla interoperabilní a umožňovala šíření a opakované použití dat v rámci oborů i mezi nimi? Budete dodržovat osvědčené postupy pro interoperabilitu, které jsou podporovány (vědeckou) komunitou? Které to budou?

Zde by se výzkumníci měli řídit osvědčenými oborovými zvyklostmi. Pokud zatím ještě nemají dostatek zkušeností, měli by využívat co možná nejobecnější postupy pro svůj vědní obor, čímž by byla zajištěna interoperabilita zveřejněných dat. Současně by měli sestavit seznam dat, která chtějí zveřejnit, a poradit se s odborníkem na otevřenou vědu, jaké formáty souborů a metadat by pro ně byly nejvhodnější.

V případě, že je nevyhnutelné používat neobvyklé ontologie nebo generovat ontologie a slovníky specifické pro projekt, poskytnete mapování běžněji používaným ontologiím? Budete otevřeně publikovat generované ontologie nebo slovníky, aby se dali znovu použít, upřesnit nebo rozšířit?

Aby byly dodrženy principy interoperability, musí být všechny neobvyklé nebo vygenerované ontologie otevřeně publikovány, nejlépe společně s daty, ve kterých byly využity, aby zůstaly stejně přístupné všem potenciálním zájemcům. Pokud se speciální ontologie používá opakovaně, mohou výzkumníci vytvořit speciální dataset, ve kterém budou unikátní slovníky, a u ostatních dat jen uložit jeho perzistentní identifikátor pro rychlé dohledání. Uvedený postup ještě není v praxi dostatečně vyzkoušený.

Budou vaše data obsahovat odkazy na jiná data (např. jiná data z vašeho projektu nebo soubory dat z předchozího výzkumu)?

Zde musí vědci zvážit, zda budou navazovat na předchozí výzkumné projekty, a pokud ano, potom musí všechna převzatá data odcitovat podle normy.

Increase data re-use | Zvýšení opakovaného použití dat

Jak poskytnete dokumentaci potřebnou k ověření analýzy dat a usnadnění opětovného použití dat (např. soubory readme s informacemi o metodice, workflow, čištění dat, analýze dat, definicích proměnných, měrných jednotkách atd.)?

Informace potřebné k ověření analýzy by měl být společně s aktualizacemi Data Management Planu umísťovány do tzv. administrativních datasetů ve zvoleném repozitáři podle FAIR principů.

Budou vaše data volně dostupná, aby bylo umožněno jejich co nejširší opakované použití? Budou vaše data licencována pomocí standardních licencí pro opětovné použití v souladu s povinnostmi stanovenými v grantové dohodě?

Jelikož opětovné využití (reusable) patří mezi základní pilíře FAIR principů, musí je výzkumníci splnit výběrem vhodné licence pro své datasety (nejčastěji se používají licence Creative Commons).

Bude umožněno použití dat třetí straně během a hlavně po skončení projektu?

Data budou běžně dostupná třetím osobám prostřednictvím repozitářů a institucionálních úložišť během a také v určitém časovém intervalu (např. Univerzita Karlova v Praze stanovila dobu uložení výzkumných dat na 10 let) po skončení projektu.

Bude původ dat důkladně zdokumentován pomocí příslušných standardů?

Opětovně použitá data starších výzkumných projektů budou odcitována podle příslušné normy. U každého datasetu bude uveden kompletní seznam jeho autorů včetně jejich zaměstnavatelů a kontaktů na ně, díky čemuž budou mít potenciální zájemci možnost kontaktovat je a předat jim své dotazy.

Popište všechny relevantní procesy zajištění kvality dat.

Výzkumníci by měli při zpracování FAIR dat spolupracovat s odborníkem na otevřenou vědu. Také mohou používat různé online nástroje pro jejich kontrolu. Rovněž by měli postupovat přesně podle požadavků, které jim pošle zadavatel jako zpětnou reakci na průběžnou zprávu o projektu.

Kromě zásad FAIR by se DMP měly rovněž zabývat jinými výstupy výzkumu než daty a měly by pečlivě zvážit aspekty související s přidělováním zdrojů, bezpečností dat a etickými aspekty.

Administrátoři projektu a vedoucí vědečtí pracovníci by měli již při podání návrhu předpokládat, že mohou vzniknout také jiné výstupy výzkumu než data, a měli by pro ně vyčlenit dostatečnou část finančních prostředků, přičemž by se měli primárně zaměřit na problémy ochrany dat, etické otázky a zamezení zneužití jejich výstupu třetími osobami.

3) OTHER RESEARCH OUTPUTS | DALŠÍ VÝSTUPY VÝZKUMU

Kromě správy dat by příjemci měli také zvážit a naplánovat správu dalších výstupů výzkumu, které mohou být generovány nebo znovu použity v rámci jejich projektů. Tyto výstupy mohou být digitální (např. software, pracovní toky, protokoly, modely atd.) nebo fyzické (např. nové materiály, protilátky, činidla, vzorky atd.).

Pokud vědci předpokládají již na začátku, že jejich výzkum přinese také jiné výstupy než výzkumná data, měli by je zde alespoň obecně popsat. Jestli žádné další výstupy zatím neplánují, stačí uvést, že jejich správa bude podrobně popsána v budoucích aktualizovaných verzích Data Management Planu, pokud budou během projektu vytvořeny další výstupy výzkumu.

Příjemci by měli zvážit, které z otázek týkajících se výše uvedených FAIR standardů se mohou vztahovat na řízení dalších výstupů výzkumu a měli by se snažit poskytnout dostatečné podrobnosti o tom, jak budou jejich výstupy z výzkumu řízeny a sdíleny nebo zpřístupněny k opětovnému použití v souladu se zásadami FAIR.

Také u dalších výstupů výzkumu by vědci měli dodržovat FAIR principy, ačkoli to znamená úzkou spolupráci s administrátory projektu a odborníky na otevřenou vědu.

4) ALLOCATION OF RESOURCES | ALOKACE ZDROJŮ

Jaké budou náklady na vytvoření FAIR dat nebo jiných výstupů výzkumu ve vašem projektu (např. přímé a nepřímé náklady související s úložištěm, archivací, opětovným použitím, zabezpečením atd.)?

Zde by výzkumníci měli uvést, kolik odborníků na otevřenou vědu jsou ochotni platit s využitím svých finančních prostředků. Dále by zde měly být uvedeny náklady na publikování odborných textů (hlavně poplatky APC) a na zřízení a správu institucionálních úložišť pro výzkumná data.

Jak budou tyto náklady pokryty? Upozorňujeme, že náklady související se správou výzkumných dat a publikačních výstupů jsou způsobilé v rámci programu Horizon Europe (pokud jsou v souladu s podmínkami grantové dohody).

Výzkumníci by se vždy měli snažit financovat svůj výzkum z finančních prostředků poskytnutých zadavateli, přičemž při nakládání se svěřenými penězi by měli dodržovat všechny podmínky stanovené poskytovatelem, aby výdaje mohly být vyhodnoceny jako potřebné a způsobilé.

Kdo bude zodpovědný za správu dat ve vašem projektu?

Pokud budou výzkumníci zaměstnávat odborníka na otevřenou vědu, měli by mu svěřit správu výzkumných dat a stručně popsat jeho pracovní činnost: kontrola správy dat, aktualizace plánu správy dat, kontrola zpřístupnění datasetů v repozitáři, podpora členů řešitelského týmu, kontrola dalších výstupů projektu atd. Jestli se rozhodnou nezapojit žádného experta na Open Science, měl by tuto činnost vykonávat alespoň jeden administrátor výzkumného projektu.

Jak bude zajištěno dlouhodobé uchování? Diskutujte o nezbytných zdrojích, jak toho dosáhnout (náklady a potenciální hodnota, kdo a jak rozhoduje, jaká data budou uchovávána a jak dlouho)?

Data se budou uchovávat primárně v interních institucionálních úložištích a také ve veřejných úložištích.

5) DATA SECURITY | BEZPEČNOST DAT

Jaká opatření jsou nebo budou zavedena pro zabezpečení dat (včetně obnovy dat, zabezpečeného úložiště, archivace a přenosu citlivých dat)?

Zde je důležité uvést, zda součástí výzkumu bude rovněž práce s citlivými daty (státní nebo obchodní tajemství, data chráněná GDPR nebo právy k duševnímu vlastnictví). Na každé instituci by měli postupovat podle vlastních předpisů pro kybernetickou bezpečnost a u citlivých dat se dohodnout na zvláštních opatřeních (kde se budou shromažďovat; na jedné či více institucích; jak se budou obnovovat v případě ztráty či smazání; kdo k nim bude mít přístup; musí být někde registrován; kde budou uložena po skončení projektu atd.).

Budou data bezpečně uložena v důvěryhodných úložištích pro dlouhodobé uchování a udržování?

Výzkumníci mohou data ukládat na lokální disky, síťová či cloudová úložiště, přenosná zařízení (např. USB disky), mobilní zařízení (mobily, notebooky, tablety aj.) a různé elektronické pracovní nástroje (např. elektronické laboratorní deníky, měřicí přístroje atd.). Zde by ale neměla být dlouhodobě uložena. Ve veřejných repozitářích jsou sice „otevřena“ dlouhodobě, nicméně zde se v budoucnosti předpokládá, že budou starší datasety mazány, aby se uvolnilo místo novým souborům, a zůstanou pouze jejich metadata. Nejdéle by výzkumná data měla být uložena v interních úložištích jednotlivých vědeckých institucí zapojených do projektu, a to po určitý časový interval (např. Univerzita Karlova v Praze je ukládá na 10 let).

6) ETHICS | ETIKA

Existují nebo mohly by existovat nějaké etické nebo právní překážky, které mohou mít dopad na sdílení dat? Lze o nich také diskutovat v kontextu etického přezkumu. Je-li to relevantní, zahrňte do popisu akce Descriptio of Action (DoA) odkazy na etické výstupy a kapitoly etiky.

Zde se rovněž řeší otázka sdílení citlivých dat, která byla podrobně rozebrána v předchozím textu. Výzkumníci musí vždy používat etické principy, které používá jejich instituce. Pokud by narazili na neočekávané právní překážky, musí se obrátit na právníky svého zaměstnavatele či zaměstnavatelů.

Bude informovaný souhlas se sdílením dat a dlouhodobým uchováváním zahrnut do dotazníků zabývajících se osobními údaji?

Pokud nebude výzkum vyžadovat osobní údaje, měli by to autoři jen stručně uvést do Data Management Planu. Pokud budou muset osobní údaje shromažďovat, je vhodné po poradě s právníky stručně uvést, jak budou při jejich zpracovávání postupovat. Souhlas se zpracováním citlivých dat by měl být součástí každého dotazníku, který vyžaduje citlivé údaje.

7) OTHER ISSUES | OSTATNÍ

Využíváte, nebo budete využívat jiné národní/grantové/sektorové/institucionální postupy pro správu dat? Pokud ano, jaké (uved'te prosím jejich seznam a stručně je popište).

Pokud má již některá instituce zkušenosti s uvedenými postupy, může je využít v daném projektu. Jestliže je zatím nemá, může se pokusit je aplikovat do správy dat. Výzkumníci by měli uvést jejich seznam a stručně a konkrétně popsat, jaký přínos budou mít pro jejich výzkum. Také mohou uvést, že očekávaného přínosu nebylo nakonec dosaženo a postup byl opuštěn.