

A stylized tree with a light blue trunk and branches, set against a background of soft, blended colors in shades of purple, pink, and blue. The tree's canopy is composed of numerous overlapping circles in various colors, including purple, blue, yellow, and green, creating a vibrant, abstract foliage effect. A large green circle is positioned in the upper right quadrant, containing the title text.

**ZÁSADY
zajištění
FAIRové správy
a využitelnosti
dat**



VYHLEDATELNOST

PŘÍSTUPNOST

INTEROPERABILITA

VYUŽITELNOST

O TÉTO PŘÍRUČCE

Předložená příručka shrnuje zásady vedoucí ke sladění úsilí všech osob a komunit, které se podílejí na sběru, ukládání, správě a užívání výzkumných dat v oblasti humanitních a společenských věd. Smyslem těchto zásad je napomoci co nejlepší využitelnosti těchto dat v daných oborech.

Příručka je dílem více než padesáti účastníků projektu PARTHENOS. Úkolem tohoto týmu bylo prozkoumat existující metody a strategie při správě výzkumných dat. Formou rešerší v literatuře, dotazníkových šetření a rozhovorů s vybranými odborníky shromáždil tým kolem stovky návodů a pokynů pro správu dat (vč. doporučení k preferovaným formátům dat, zásad kontroly dat a příkladů dobré praxe, a to jak formálních, tak neformálních).

Z těchto podkladů sestavil PARTHENOS dvacet hlavních zásad týkajících se kvality (meta)dat a datových úložišť, platných pro různé vědní obory.

Pro snadnější orientaci v jejich obsahu byla každá ze zásad přiřazena k jednomu ze základních principů, které jsou osami FAIR přístupu k datům, totiž jejich vyhledatelnosti (Findability), přístupnosti (Accessibility), interoperabilitě (Interoperability)¹ a využitelnosti (Reusability). Dané principy byly poprvé zveřejněny v manifestu FORCE11 (2016) a od té doby slouží jako vodítko ve společné snaze o zvýšení využitelnosti výzkumných dat. Každá ze zásad je doplněna o konkrétní doporučení pro poskytovatele a uživatele dat na jedné straně a pro správce dat (archiváře) na straně druhé. Ikony u jednotlivých doporučení ukazují, pro kterou cílovou skupinu jsou určena.

Na žádost řady partnerů projektu PARTHENOS pokračoval překlad této příručky do dalších jazyků i po jeho skončení, a to v rámci projektů ARIADNE Plus (Horizont 2020) a SEADDA (COST akce).



Žárovka označuje doporučení pro poskytovatele a uživatele dat, tedy vědce a členy výzkumné komunity v oblasti archeologie, historie, jazykovědy nebo společenských věd.



Ozubená kolečka indikují doporučení pro správce vědeckých infrastruktur a archiváře ve výzkumných a památkových institucích.

¹ Interoperabilita je schopnost různých systémů vzájemně spolupracovat, poskytovat si služby, dosáhnout vzájemné součinnosti. Zdroj: Interoperabilita. In: Wikipedie: otevřená encyklopedie [online]. St. Petersburg (Florida): Wikipedia Foundation, poslední editace 11. 9. 2019 v 11:38 [cit. 2020-07-16]. Dostupné z: <https://cs.wikipedia.org/wiki/Interoperabilita>.

20 ZÁSAD

jak učinit správu dat FAIRovou a jak zajistit jejich trvalou využitelnost



1

Investujte do lidí a vybavení

Důležitým předpokladem, který umožní naplnit dalších devatenáct zásad z této příručky je dostatečná investice do datové infrastruktury a zajištění expertů v oblasti datové správy.



Seznamte se s příklady dobré praxe v oblasti správy výzkumných dat. Projděte si školicí moduly PARTHENOS věnované správě dat nebo příručku CESSDA určenou pro experty v oblasti datové správy (CESSDA Data Management Expert Guide).



Zajistěte dostatečné zdroje pro zapojení a vzdělávání datových expertů a vyhradte odpovídající rozpočet na investice do technické infrastruktury a zaměstnanců.

VYHLEDATELNOST

Výzkumná data by měla být snadno vyhledatelná a čitelná jak uživateli, tak automatizovanými počítačovými systémy. K tomu je nezbytný popis dat a databází metadaty.

2

Používejte persistentní identifikátory

Nalezení dat je základní podmínkou jejich zpřístupnění a využití. K tomu je nutné označit datový objekt unikátním a trvalým označením, tzv. persistentním identifikátorem (PID). PID zůstává funkční, i když se webová adresa datového zdroje změní. Může mít různý formát, obvyklý je např. Handle, DOI, PURL nebo URN.



Ve výzkumných výstupech a publikacích vždy uvádějte PID přidělený vašim datům.



Pro označení dat ve své výzkumné infrastruktuře zvolte vhodné schéma trvalých identifikátorů a přiřaďte PID každému datovému zdroji. Při volbě formátu PID užívejte Příručku pro PID od NCDD.

3

Citujte výzkumná data

Mají-li výzkumná data trvalý identifikátor a jsou citována v souladu s oborovými standardy, je jednodušší použít datové objekty znovu nalézt.



Seznamte se s citačními pravidly vašeho oboru či oblasti práce a v souladu s nimi výzkumná data citujte.



Poskytujte odborné komunitě informace o dobré citační praxi a usnadněte uživatelům dat snadné citování zdrojů, např. pomocí standardizovaného tlačítka „Jak citovat tato data“.

4 Jednoznačně identifikujte autory

Persistentní identifikátor autora (např. VIAF, ISNI nebo ORCID) pomáhá vytvořit vazbu mezi daty, výzkumnými aktivitami, publikacemi a výzkumníky a usnadňuje budoucí vyhledání dat.



Identifikujte se mezi ostatními badateli a výzkumnými týmy. Zažádejte o autorský identifikátor, pokud jej ještě nemáte, a uvádějte jej ve svých datových celcích.



V metadatech odkazujte na identifikátory autorů.

5 Zvolte vhodný metadatový model

Metadata jsou základem pro vyhledání dat, a to zejména ta, která se užívají při citacích a popisu dat. Metadatový model je standardizovaný soubor položek, které zachycují informace o datovém zdroji, např. jeho název, identifikátor, jméno autora či datum vzniku. Použitím již zavedených metadatových modelů zajistíte dodržení mezinárodních standardů pro datovou výměnu.



Pro zajištění vyhledatelnosti datového obsahu popisujte data s maximální konsistencí a úplností. Uvádějte všechny informace potřebné k tomu, aby byla data v budoucnu přístupná a srozumitelná. Je-li to možné, užívejte zavedený metadatový model, který vyhovuje vašim datům.



Jasně popište metadatový model, který užíváte, a doporučujte jej své odborné komunitě. Pro obohacení dat přijímaných do repozitáře zvažte zavedení formulářů ke shromažďování metadat.

PŘÍSTUPNOST

Výzkumná data by měla být bez obtíží přístupná za jasně definovaných podmínek a s využitím standardizovaných komunikačních protokolů.

6

Volte důvěryhodná úložiště

Důvěryhodným místem pro ukládání dat jsou certifikovaná úložiště. Certifikace úložiště je zárukou, že data budou bezpečně uložena, dostupná, vyhledatelná a přístupná v dlouhodobém časovém horizontu. Příklady certifikačních standardů jsou např. CoreTrustSeal, Nestor Seal nebo ISO 16363.



Pro zpřístupnění svých dat zvolte důvěryhodné úložiště. Dodržíte-li v ohledu datových formátů, metadatového modelu atd. standardy takového úložiště, můžete se spolehnout, že vaše data budou zpřístupňována podle principů FAIR.



Na webových stránkách viditelně uvádějte úroveň certifikace svého úložiště. Pokud certifikaci (zatím) nemáte, uveďte, jak hodláte zajistit dlouhodobou dostupnost, vyhledatelnost, přístupnost a využitelnost dat.

7

Jasně určete přístupnost dat

Přístupností dat se rozumí podmínky, za jakých se uživatelé mohou dostat k příslušným datovým celkům. Při ukládání dat do úložiště by mělo být jasné, čím může být přístupnost dat omezoována.



Při stanovení přístupnosti dat berte v potaz stávající zákonné požadavky, pravidla sdílení příslušnou odbornou komunitou a obecné etické zvyklosti. Je-li to možné, dávejte vždy přednost režimu otevřeného přístupu (Open Access). Pokud pracujete s osobními údaji, zamyslete se nad tím, zda data neobsahují informace, které by mohly vést k odhalení identity konkrétních osob, k čemu dali poskytovatelé údajů souhlas a jaká opatření k ochraně dat jste přijali. Pokud data nelze zveřejnit v režimu otevřeného přístupu, měla by být v tomto režimu vedena alespoň metadata, aby informace bylo možné vyhledat.



Podporujte zveřejňování (meta)dat v režimu otevřeného přístupu. Jasně stanovte případná omezení přístupu pro citlivá (meta)data, která v tomto režimu zveřejnit nelze. V takových případech usilujte o jejich zpřístupnění jiným, jasně definovaným a kontrolovatelným způsobem.

8

Užívejte ochranné lhůty, je-li to nutné

Některá data mohou být po dobu určité časové lhůty chráněna před zveřejněním. V dané době může být zveřejňován pouze popis datového celku a plná (meta)data až po vypršení stanovené lhůty.



Jasně uveďte, proč a na jaké období považujete omezení přístupu k datům za nezbytné. (Meta)data zpřístupněte ihned, jakmile je to možné.



Uveďte, zda je dočasné omezení přístupu povoleno a za jakých podmínek.

9

Užívejte standardní výměnné protokoly

Výzkumné infrastruktury mohou k výraznému zlepšení dostupnosti datových zdrojů využít standardizované výměnné protokoly. Pomocí nich lze veřejně zpřístupnit (meta)data a umožnit jejich sklizení, např. webovými vyhledávači.



Používejte standardizované protokoly jako např. SWORD, OAI-PMH, ResourceSync nebo SPARQL. Převádějte metadatové modely do formátů XML nebo RDF. Udržujte veřejný registr přístupových bodů protokolů a přístupových cest k datovým zdrojům.

Pro urychlení vyhledávání dat a získávání nových poznatků by výzkumná data měla být snadno propojitelná s dalšími datovými celky, a to jak prostřednictvím jednotlivých uživatelů, tak počítačových systémů.

INTEROPERABILITA

10

Vytvářejte dobře dokumentovaná strojově čitelná API

Dobře dokumentovaná, strojově čitelná API (application programming interface), tj. soubory definic, protokolů a nástrojů pro vytváření softwarových aplikací, umožňují automatické indexování, získávání a kombinování (meta)dat z různých datových úložišť.



Vedte kvalitní dokumentaci API a zveřejňujte (meta)datové modely. Zvažte uveřejnění příkladů, jak úspěšně získávat data z různých přístupových bodů a kombinovat je do nových datových celků využitelných pro další výzkum.

11

Užívejte kvalitní a dostupné hesláře

Popis (meta)dat by se měl řídit oborovými směrnici, které počítají s široce sdílenými, snadno přístupnými a obecně přijímanými hesláři. Takové hesláře obvykle přesně vystihují významy pojmů a vlastnosti dat.



Používejte relevantní oborové hesláře od počátku výzkumného projektu a popis svých dat podle nich strukturujte.



Nabízejte výzkumné komunitě příklady heslářů vhodných pro využití v jednotlivých oblastech výzkumu.

12

Popisujte metadatové modely

Přehledná dokumentace metadatových modelů pomáhá vývojářům vzájemně srovnávat datové modely a mapovat vztahy mezi daty z různých zdrojů.



Publikujte (meta)datové modely užívané vaší výzkumnou infrastrukturou. Dokumentujte jejich technickou stránku a definujte použité třídy, tedy skupiny záznamů se společnými vlastnostmi, a atributy, vyjadřující vlastnosti určité části metadat či vzájemné vztahy mezi metadaty. Pro účely mapování metadat uvádějte, které údaje jsou povinné a které jen doporučené.

13

Dodržujte interoperabilní datové standardy

Dodržování standardů přijímaných širokou komunitou zvyšuje možnosti sdílení, opakovatelného využití a kombinování dat.



Před předáním dat do repozitáře se informujte, jaké standardy repozitář používá. Podle těchto standardů strukturujte svá data, a to nejlépe již od počátku výzkumného projektu.



Jasně specifikujte, jaké datové standardy vaše instituce používá. Zasazujte se o jejich uplatnění a sdílení odbornou komunitou a udržujte je s ohledem na interoperabilitu. Příkladem dobré praxe jsou standardy CMDI (lingvistika) a SIKB0102 (archeologie).

14

Věnujte se zlepšování kvality dat

Pro zvýšení kvality (meta)dat, a tím i jejich interoperability, zavádějte (automatizované) postupy čištění, odvozování a doplňování (meta)dat.



Zavádějte postupy minimalizující riziko chyby při sběru dat. Např. při zadávání časového údaje vybírejte datum z kalendáře místo ručního zápisu.



Zavádějte nástroje k čištění (meta)dat a jejich převádění do standardizovaných a interoperabilních formátů. Zapojte se do vývoje pracovních postupů a softwarových řešení pro automatizaci těchto operací, např. s využitím nástrojů strojového učení.

15

Užívejte dlouhodobě udržitelné formáty souborů

Veškeré datové soubory v úložišti by měly být dostupné v otevřeném, standardizovaném a mezinárodně uznávaném formátu. To zajistí jejich interoperabilitu z hlediska dlouhodobé využitelnosti, přístupnosti a udržitelnosti.



Od začátku výzkumného projektu přemýšlejte o budoucnosti svých dat a volte dlouhodobě udržitelné formáty souborů. Používejte preferované formáty doporučené datovým úložištěm, nezávislé na specifickém softwaru, vývojáři či dodavateli.



Podporujte používání formátů vhodných pro dlouhodobou archivaci, jako např. PDF-A, CSV a MID/MIF. Snadno dostupnou formou zveřejněte dostatečně podrobný seznam přípustných formátů přijímaných souborů.

VYUŽITELNOST

Výzkumná data by měla být připravována pro budoucí výzkum a budoucí využití v podobě, ve které umožní revizi, zopakování výzkumu a návaznost nových výzkumných aktivit na výsledky těch předchozích.

16

Systematicky dokumentujte data

Aby bylo jasné, co lze a co nelze najít v určitém datovém celku či úložišti, data by měla být systematicky dokumentována. Srozumitelný a úplný popis obsahu zvyšuje důvěryhodnost dat, a tím i pravděpodobnost jejich opětovného využití.



Dávejte k dispozici užité popisné kódy a zkratky, údaje o chybějících datech, způsobu vzniku databáze, použitých metodách atd.

17

Dodržujte pravidla pojmenování souborů

Dodržování přesných, konsistentních a všeobecně sdílených pravidel pojmenování datových souborů podstatně usnadňuje budoucí využití, přístupnost a pochopení/rozpoznání dat.



Až budete volit způsob pojmenování datových souborů, konzultujte metodiku a příklady dobré praxe používané ve vašem oboru.



Jasně uveďte zásady správného pojmenování datových souborů a jejich užívání v praxi.

18

Užívejte běžné formáty souborů

Užívání standardních, široce rozšířených souborových formátů zvyšuje možnosti budoucího využití dat.



Vedle formátů určených k archivaci užívejte pro sdílení dat běžné souborové formáty, jako např. XLSX vedle CSV nebo ESRI SHP vedle MID/MIF.



Publikujte data nejen v archivních formátech, ale i v běžných formátech, pokud nejsou totožné.

**19**

Udržujte integritu dat

Data použitá k výzkumu musí být identická s těmi, která jsou následně zpřístupňována. Pro zajištění autenticity datového obsahu je nutno provádět kontroly integrity.



Zaveďte vhodnou metodu verzování dat. Záruka, že každá změna v revidované verzi datového celku byla řádně podchycena, je zásadní pro autenticitu každého datového celku.



Pro identifikaci případných změn v datových souborech je nezbytné zaznamenávat jejich životní cyklus, tj. původní stav souboru a veškeré další změny v čase, a srovnávat každou kopii s originálem. Kontrolu integrity dat lze provádět pomocí různých nepřímých metod (např. kontrolním součtem) nebo přímým porovnáním verzí. Zaveďte jasné způsoby označování verzí, např. přidáním jejich označení do identifikátoru a mezi vyhledávací parametry.

**20**

Licencujte data pro opětovné využití

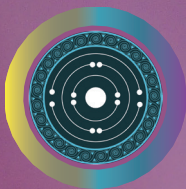
Aby bylo možné (meta)data v co největší míře opakovaně využívat, musí být jasné, kdo je vlastníkem autorských práv a jaké licenční podmínky se k jejich dalšímu použití vztahují.



Před publikováním výzkumných (meta)dat se ujistěte, kdo je vlastníkem autorských práv.



Zřetelně zveřejněte licenční podmínky použití (meta)dat, nejlépe ve strojově čitelném formátu. Pro zlepšení interoperability se pokuste vyjádřit vlastní podmínky v obecně uznávaných pojmech, např. pomocí licencí Creative Commons.



 **ARIADNE** plus

 **SEADDA**
Solving European Archaeology from the Digital Dark Age



Na žádost řady partnerů projektu PARTHENOS pokračoval překlad této příručky do dalších jazyků i po jeho skončení, a to v rámci projektů ARIADNE Plus (Horizont 2020) a SEADDA (COST akce).

Tato příručka (původní verze prosinec 2018; český překlad červenec 2020) je sdílána pod licencí Creative Commons CC BY 4.0. Design: Verbeeldingskr8. Český překlad příručky byl připraven týmem výzkumné infrastruktury AIS CR ve spolupráci s původními autory. <https://doi.org/10.5281/zenodo.3946100>.