
Výzkumná data

Definice pojmů

Data

- **Jakékoli informace, které byly shromážděny, pozorovány, vygenerovány nebo vytvořeny za účelem ověření či reprodukování výsledků výzkumu**
- dokumenty
- tabulky
- audio/vizuální nahrávky
- obrázky
- fotografie
- dotazníky
- přepisy rozhovorů
- software
- laboratorní deníky
- terénní poznámky
- vzorky
- uplatňují se standardní formáty pro texty (xml, pdf/a, html, json, txt, rtf), obrázky (jpg, gif, tiff, png), video (MPEG, AVI, MKV) nebo audio (WAVE, MP3, FLAC)
- strukturovaná kolekce výzkumných dat se označuje spojením **datová sada**

Metadata

- **Data o datech čili informace popisující atributy datové sady, které umožňují identifikaci, vyhledávání a správu dat v budoucnu**
- data o respondentech
- čas a místo sběru dat
- vysvětlivky
- číselníky
- vzor informovaného souhlasu
- licenční podmínky
- uplatňují se různé formáty, od volného textu po strukturovaný strojově čitelný obsah (některé obory nebo datová uložiště mohou mít specifické požadavky)

Data Steward

- **Osoba, která poskytuje podporu v otázkách správy výzkumných dat, a to buď na úrovni fakulty nebo výzkumného týmu**
- **Fakultní Data Steward**
- zodpovídá za to, aby byla fakultní správa výzkumných dat v souladu s univerzitními i mezinárodními standardy, a to skrze přemosťující kontakt s datovou komunitou na straně jedné, s fakultními výzkumníky a širší podporou na straně druhé
- asistence při tvorbě data management plánů, výběru (meta)datových formátů, vyhledávání vhodných uložišť a repozitářů nebo identifikace překážek pro zveřejnění výzkumných dat
- neplní roli datového analytika, neprovádí oborový rozbor výzkumných dat
- **Projektový Data Steward**
- zodpovídá za vypracování data management plánu, ukládání, zabezpečení, zálohování a sdílení výzkumných dat, vytvoření metadat a finální uložení dat do repozitáře
- pro úvazek projektového data stewarda se doporučuje vyčlenit úvazek 0,1 nebo 0,2 FTE
- pozice je vypsána pouze u některých projektů (GA ČR, OP JAK)

Management výzkumných dat

Datová sada výzkumných projektů má svůj specifický životní cyklus.

- **Vznik dat:** sběr dat a jejich uložení, generování části metadat
- **Zpracování dat:** digitalizace, validace, anonymizace a uložení dat, generování dalších metadat
- **Analýza dat:** interpretace a analýza dat, příprava publikace

- **Ochrana dat:** zálohování, migrace formátů, dokumentace
- **Sdílení dat:** řízení přístupu, copyright a licence
- **Opakované využití dat:** nový výzkum, partnerství, výuka a vzdělávání

Základní tipy při organizaci dat:

- při pojmenování souborů neopomeňte datum (YYYYMMDD), vyhněte se speciálním znakům (! @ # & % * \$), uveďte iniciály účastníků výzkumu, u anonymních respondentů unikátní kód
- u tabulkových dat popisujte sloupce na jednom řádku (počínaje A1), pojmenujte a popište každý sheet, naopak nepoužívejte barvy pro přenos informace a nespojujte buňky

Data se liší podle kategorie a zvolené přístupnosti:

- **veřejná data:** zpřístupněna komukoliv bez omezení
- **interní data:** určena pouze pro vnitřní potřebu volně vymezené skupiny osob (korespondence, zápisy z jednání, vnitřní regulace a předpisy)
- **diskrétní data:** určena výhradně pro vnitřní potřebu přesně vymezené skupiny osob, vyžadují regulaci nebo ochranu, a to ze zákona (GDPR) nebo na základě smlouvy/licence (ekonomické a personální údaje osobní povahy, čísla identifikačních průkazů, rodná čísla)
- **citlivá data:** určena striktně pro vnitřní potřebu přesně vymezené skupiny osob, vyžadují regulaci nebo obzvláštní ochranu, a to ze zákona či na základě smlouvy/licence (zdravotní data, osobní údaje odhalující rasový či etnický původ, politické názory, náboženské nebo filozofické přesvědčení, biometrické údaje zpracovávané za účelem identifikace člověka)
- **otevřená data:** přístup bývá zpravidla realizován prostřednictvím elektronických datových repozitářů, data musí být zpřístupněna ve formě, která umožní další využití z hlediska technického i právního; přístup, využití, reprodukce a šíření dat musí být bezplatné
- **přístup s embargem:** správce dat uvede v repozitáři datum, od kdy se datová sada zpřístupní veřejnosti
- **přístup s omezením:** správce dat určuje podmínky, za kterých udělí přístup k datovým sadám v repozitáři; správce dat si za udělení přístupu nesmí účtovat poplatky
- **uzavřený přístup:** z důvodu obchodního tajemství nebo ochrany duševního vlastnictví; uzavřená data mohou být uložena v repozitáři, aniž by do něj byl přístup

Doporučuje se zejména, aby přístup k datům naplňoval takzvané principy FAIR:

- **naležitelnost** (*to be findable*): zajištění strojově čitelných metadat s identifikátorem (DOI)
- **dostupnost** (*to be accessible*): zajištění otevřeného přístupu, ideálně skrze repozitář
- **interoperabilita** (*to be interoperable*): uplatnit standardizované výrazy k popisu dat
- **opětovná využitelnost** (*to be reusable*): zajistit vhodnou licenci výzkumných dat

Žádný z principů FAIR nevynucuje, aby byla data otevřená nebo zdarma, požaduje stanovení jasných a transparentních podmínek možného přístupu a opětovného využití dat. FAIR data tedy nemusí být otevřená, musí pouze mít přidělenou licenci.

Principy, aplikované na oborové standardy, následují zásadu: ***As open as possible, as closed as necessary***.

Data Management Plan (DMP)

Dokument, který shrnuje jednotlivé fáze správy výzkumných dat během projektu i po jeho skončení.

- **Co obsahuje DMP?**
- Administrativní údaje (název, výzkumný tým, poskytovatel, abstrakt)
- Sběr dat (metody, formáty, objemy, software)
- Organizace dat (kontrola kvality, dokumentace, identifikátory)
- Ukládání dat (zabezpečení, přístupy, zálohování)
- Zveřejnění dat (metadata, licence, embargo)
- Etické a právní otázky výzkumných dat
- Náklady na správu výzkumných dat (APC poplatky, projektový Data Steward, věcné náklady)
- **Jak vytvořit DMP?**
- dokument lze vytvářet v klasickém textovém editoru, dokonce i napsat ručně na papír
- doporučuje se nicméně uplatnit online nástroje, které navigují pomocí předpřipravených otázek, a tím zkracují celé úsilí
- **FAIR Wizard CUNI** : nástroj usnadňující tvorbu DMP, zprostředkovávaný pro zaměstnance a studenty Univerzity Karlovy
- nástroj prochází jednotlivé okruhy správy výzkumných dat pomocí stromových dotazníků
- univerzitní Centrum pro podporu Open Science zpracovalo rozsáhlý a krokový návod pro seznámení s nástrojem, fakultní podpora pak připravila cvičný vzor konkrétního DMP

- **Proč vytvořit DMP?**

- vyžadují jej v určité (nikoli jednotné) formě mnozí interní i externí poskytovatelé finanční podpory: GA ČR , TA ČR , Horizon 2020 , Horizon Europe , PRIMUS , OP JAK , program EXCELES
- DMP může mít i praktický užitek: pomáhá předvídat potenciální problémy, snižuje riziko ztráty dat, usnadňuje jejich sdílení, a tím i kontinuitu dlouhodobého výzkumu
- na obecné úrovni má datová politika tvořit standardy „dobré vědecké praxe“, krom jiného posílit replikovatelnost nebo integritu výzkumu

Datové repozitáře

Online platforma pro ukládání, zveřejňování a uchovávání dat, souvisejících metadat a dokumentace.

- **obecné repozitáře** zveřejňují data z jakéhokoli vědecké oblasti
- **oborové repozitáře** jsou preferované
- **institucionální repozitář** se na úrovni Univerzity Karlovy aktuálně připravuje

Volba vhodného repozitáře záleží na typu dat. Důvěryhodné repozitáře se poznají podle toho, že poskytují otevřený přístup, přiřazují trvalé identifikátory, používají standardizovaná a strojová metadata, datovým sadám umožňují přidělovat licence a mají certifikaci.

- **Zenodo**: obecný repozitář, vytvořený platformou OpenAIRE a společností CERN, umožňuje vkládat soubory do velikosti **50 GB** (maximálně 100 souborů)
- **Národní datový repozitář**: obecný repozitář, provozovaný sdružením CESNET, pilotní režim
- **Harvard Dataverse**: obecný repozitář, provozovaný Harvardovou univerzitou, umožňuje vkládat soubory do velikosti **1 TB**
- **Český sociálněvědní datový archiv**: oborový repozitář, provozovaný Sociologickým ústavem AV ČR, nemá stanovenou maximální velikost souborů
- **LINDAT/CLARIAH-CZ**: oborový repozitář pro lingvistická data a nástroje, zpravovaný Ústavem formální a aplikované lingvistiky MFF UK, nemá stanovenou maximální velikost souborů
- **Re3data.org**: registr nabízí přehled existujících mezinárodních datových repozitářů
- **OpenDOAR**: databáze otevřených repozitářů

Právní aspekty výzkumných dat

Zákon č. 130/2002 Sb., o podpoře výzkumu a vývoje z veřejných prostředků definuje výzkumná data v § 2 odst. 2 písm. o:

„informace, s výjimkou vědeckých publikací, v elektronické podobě, které jsou shromažďovány nebo vytvářeny v průběhu výzkumu nebo vývoje a jsou používány jako důkazy v procesu výzkumu nebo vývoje nebo které jsou obecně akceptovány výzkumnou obcí jako nezbytné k validaci zjištění a výsledků výzkumu nebo vývoje“.

Pro účely poskytování finanční podpory dle předmětného zákona se tedy výzkumnými daty rozumí data v digitální podobě. Univerzitní Centrum pro podporu Open Science v obecné rovině vnímá výzkumná data šířeji, tedy včetně těch nedigitálních.

Zákon č. 130/2002 Sb. byl v roce 2022 novelizován (č. 241/2022 Sb.) v úsilí implementovat do českého práva **Směrnici Evropského parlamentu a Rady (EU) 2019/1024, Otevřená data a opakované použití informací veřejného sektoru** . Novela stanoví povinnost, a to u projektů podpořených z veřejných prostředků, zveřejnit informace o způsobu správy výzkumných dat. Dále ukládá:

- dle **§ 9 odst. 1 písm. l**: *„vedení záznamů o výzkumných datech, známé rovněž jako ‘data management plan’, je základním předpokladem pro efektivní nakládání s výzkumnými daty v průběhu vědeckého projektu“*
- ustanovení ukládá obecnou povinnost zahrnout do smlouvy o poskytnutí podpory část, v níž bude upraven způsob správy výzkumných dat příjemcem
- dle **§ 9 odst. 1 písm. m**: *„informace o dostupnosti a způsobu šíření výsledků výzkumu a výzkumných dat (...) v souladu se zásadou, že výsledky výzkumu a výzkumná data nejsou zveřejňovány pouze v odůvodněných případech“*
- odůvodněné případy mohou zahrnovat například data třetích stran, citlivé údaje o účastnících výzkumu s lidskými subjekty nebo obchodní tajemství
- dle **§ 12 odst. 1**: *„Podporu lze poskytnout pouze za podmínky zveřejnění pravdivých a včasných informací příjemcem i poskytovatelem o prováděném výzkumu, vývoji a inovacích a o jejich výsledcích a výzkumných datech prostřednictvím informačního systému výzkumu, vývoje a inovací“*
- podle stanoviska Univerzitního Centra pro podporu Open Science odstavec zakládá povinnost publikovat v **IS VaVal** informace o výzkumných datech, „*jejich metadata, nikoli data samotná*“
- dle **§ 12 odst. 3**: *„Příjemce nezveřejní informace o výzkumných datech, jejichž zveřejněním by došlo k nepřiměřenému zásahu do práva na ochranu duševního vlastnictví, obchodního tajemství, bezpečnosti státu nebo oprávněných obchodních zájmů příjemce nebo třetí osoby“*
- příjemce má v těchto případech povinnost nejméně jednou za rok po dobu pěti let od ukončení poskytování podpory přezkoumat, zda důvody pro nezveřejnění trvají
- dle nově vloženého **§ 12a odst. 1** má příjemce povinnost bezplatně poskytnout výzkumná data na základě žádosti, která *„nejdou chráněna podle zákonů upravujících ochranu výsledků autorské, vynálezecké nebo obdobné tvůrčí činnosti“*

- **§ 12a odst. 2:** „Příjemce může odmítnout poskytnutí výzkumných dat v případech, kdy by jím došlo k nepřiměřenému zásahu do práva na ochranu soukromí a osobních údajů, práva na ochranu obchodního tajemství, bezpečnosti státu nebo jiných oprávněných zájmů příjemce“
- datové povinnosti dle § 12a se nevztahují na projekty vyhlášené či podpořené před **1. 9. 2022**

Pro výzkumná data neexistuje jednotný systém právní ochrany. Jejich detailnější přehled poskytuje [Centrum pro podporu Open Science](#).

Etické aspekty výzkumných dat

Ochrana dat propojuje několik oblastí, u kterých se předpokládá vzájemný dialog.

Základem je vyhodnocení několika otázek:

- Bude potřeba informovaný souhlas subjektů?
- Existují bariéry pro zpřístupnění dat jiným výzkumníkům?
- Jak bude nakládáno s diskrétními a citlivými data, aby se zajistilo, že jsou bezpečně uloženy?
- Kdo bude data ukládat a kdo k nim bude mít v průběhu projektu přístup?
- Jak dlouho po skončení projektu budou data uchovávána?

Právní konzultace poskytují příslušné kontakty v rámci [Centra pro podporu Open Science](#), případně [Centra pro přenos poznatků a technologií](#). Etické posouzení má v gesci fakultní [Komise pro etiku ve výzkumu](#). [Oddělení informačních technologií](#) zajišťuje řešení vhodných síťových a cloudových uložení.

Komise pro etiku ve výzkumu byla na FHS UK zřízena **Opatřením děkana č. 10/2018**, [Statut Komise pro etiku ve výzkumu Fakulty humanitních studií Univerzity Karlovy](#), a to v návaznosti na celouniverzitní **Opatření rektora č. 74/2017**, [Statut Komise pro etiku ve výzkumu Univerzity Karlovy](#). Postup pro fakultní přijímání žádosti o etické posouzení upravuje Čl. 5 příslušného Opatření děkana.

Fakultní Komise dále poskytuje v součinnosti s [Hlavní spisovnou FHS UK](#) možnost uložit citlivé údaje, originály podepsaných informovaných souhlasů a výzkumná data na jejich zajištěném pracovišti.

Mezi další (méně doporučené) typy uložení patří například:

- **přenosná média** (flash disky, paměťové karty, CD)
- **lokální disky** (počítače, notebooky)
- **síťová uložení** provozovaná na infrastruktuře UK (OneDrive)
- **cloudová uložení** provozovaná externími subjekty mimo infrastrukturu UK (Sharepoint)

Variety fakultních uložení a jejich vhodné užití naznačuje tabulka níže:

	Fyzické uložení	Síťové disky (interní)	Cloudové služby (smluvní)
Kategorie dat	veřejná, interní, diskrétní, citlivá	veřejná, interní, diskrétní, citlivá	veřejná, interní
Kapacita uložení	není stanovena	do 100 GB na uživatele	do 5 TB na uživatele
Zálohování	Hlavní spisovna FHS UK	OIT	ÚVT

Fakultní a univerzitní kontakty

V případě jakýchkoli dotazů lze kontaktovat fakultní i univerzitní podporu.

FHS UK

- **Martin Mišúr:** Data Steward, datasteward@fhs.cuni.cz
- **Miriam Vojtisková:** fakultní koordinátorka Open Science, miriam.vojtiskova@fhs.cuni.cz
- **Tomáš Renner:** tajemník Komise pro etiku ve výzkumu, tomas.renner@fhs.cuni.cz
- **Roman Sukdolák:** správce Hlavní spisovny FHS UK, roman.sukdolak@fhs.cuni.cz
- **Alena Matuszková:** ředitelka fakultní knihovny, alena.matuszkova@fhs.cuni.cz

UK

- konzultace v souvislosti s autorskoprávní problematikou, openlaw@cuni.cz
- pověřenkyně pro ochranu osobních údajů, gdpr@cuni.cz
- komercializace a duševní vlastnictví, research.data@cuip.cz
- technická (ICT) podpora, openict@cuni.cz
- Ústav výpočetní techniky, office@uvt.cuni.cz
- Centrum pro podporu Open Science, researchdata@cuni.cz

Důležité dokumenty

[Zákon č. 130/2002 Sb., o podpoře výzkumu a vývoje z veřejných prostředků](#)

Zákon č. 241/2022 Sb., Zákon, kterým se mění zákon č. 106/1999 Sb., o svobodném přístupu k informacím, ve znění pozdějších předpisů, zákon č. 123/1998 Sb., o právu na informace o životním prostředí, ve znění pozdějších předpisů, a zákon č. 130/2002 Sb., o podpoře výzkumu, experimentálního vývoje a inovací z veřejných prostředků a o změně některých souvisejících zákonů (zákon o podpoře výzkumu, experimentálního vývoje a inovací), ve znění pozdějších předpisů

Směrnice Evropského parlamentu a Rady (EU) 2019/1024, Otevřená data a opakované použití informací veřejného sektoru

Opatření děkana č. 10/2018, Statut Komise pro etiku ve výzkumu Fakulty humanitních studií Univerzity Karlovy

Opatření rektora č. 74/2017, Statut Komise pro etiku ve výzkumu Univerzity Karlovy

Užitečné odkazy

Centrum pro podporu Open Science

FAIR Wizard CUNI

How to make your data FAIR

GA ČR a otevřená věda

Jak bezpečně nakládat s daty