

Repozitáře vědeckých dat

Jan Hrabal

Obsah

Úvod a význam problematiky.....	2
Principy přístupu k vědeckým datům	2
Výzva pro akademické knihovny.....	4
Research Data Management.....	4
Seznam vybraných repositářů.....	4
Multioborové zdroje	4
Archeologie	5
Biologie	5
Chemie a chemická biologie.....	6
Health Sciences	6
Vědy o Zemi a životním prostředí	7
Astrofyzika.....	8
Částicová fyzika	9
Obecná fyzika.....	9
Computer Science.....	9
Humanitní vědy	9
Sociální vědy.....	10
Srovnávací statistiky.....	10
Další seznamy.....	10
Zdroje	11

Úvod a význam problematiky

Sdílení a dlouhodobá ochrana vědeckých dat má pro komunitu vědců a LTP komunitu zásadní význam. Data z původních výzkumných a průzkumných prací tvoří významnou součást vědeckého poznání. Umožňují badatelům různých oborů objevovat, nalézat, ověřovat a vyvracet nové i zavedené hypotézy a poznatky. Následující text se v první části věnuje problematice sdílení a uchovávání vědeckých dat na obecné úrovni, jsou uvedeny přínosy a význam této aktivity, principy přístupu k vědeckým datům, návaznost na akademické knihovny a Research Data Management, je zmíněna problematika politiky časopisů a nakladatelů v otázkách sdílení dat autory. O jednotlivých oblastech se více píše v citované literatuře. Ve druhé části textu se nachází seznam repozitářů vědeckých dat, jež je rozdělen dle vědních oborů. Seznam přirozeně není úplný, nachází se v něm známé a populární repozitáře, jež byly vybrány dle již existujících seznamů a dle vlastního průzkumu a vyhledávání. Pro zájemce jsou na konci textu uvedeny odkazy na další seznamy repozitářů.

Archivace a zpřístupnění dat je velmi prospěšné a slouží badatelům různými způsoby (Whitlock, 2011). Veřejně přístupná data umožňují ověřit, zda byla data v proběhlém výzkumu zanalyzována správnými metodami a zda jsou její závěry platné. To má velký význam kupříkladu v oblasti klinického hodnocení při farmakologických studiích, kde lze díky veřejně přístupným datům zkontrolovat správnost klinických závěrů. Díky tomu je věda transparentní. Zřídka se děje úmyslné vědecké pochybení, avšak veřejně archivovaná data mají těmto prohřeškům zabránit. Veřejná data mají také velkou hodnotu ve výuce statistiky, kdy jsou studentům přiblíženy různé konkrétní případy.

Velký význam mají data při provádění meta-analýz. Ne všechny údaje jsou zahrnuty v publikovaných článcích a komplexní data a datasey poskytují materiál pro pečlivé provádění přehledových studií a maximalizují přesnost a preciznost datové analýzy. Články, jejichž data jsou archivovaná, mají v budoucnu větší šanci být součástí meta-analýzy.

Archivovaná data mohou dále vyústit v nové výsledky nových vědeckých otázek, s nimiž původní autor nepracoval. Kupříkladu díky astronomickým datům nasbírané teleskopickým pozorováním objevují astrofyzikové nové planety anebo případná jiná tělesa, a odpadají tak náklady na další pozorování.

Další výhodou archivovaných dat je zabránění jejich ztrátě. Autoři svá data snadno ztratí nebo znehodnotí pouhým zálohováním, přičemž tyto zálohy nejsou přizpůsobeny dlouhodobé ochraně a znemožňují jejich využití za několik desítek let. Pomocí veřejné archivace lze těmto rizikům snadno předejít. Pro autora také znamená archivace dat jakousi dobrou pověst, protože články s veřejně archivovanými daty jsou citovanější. Navíc citovat lze rovněž i datové balíky, takže se zvyšuje i citovanost autora a jeho vědecká prestiž bude posuzována nejen skrze jeho články, ale i přispěnými daty.

Principy přístupu k vědeckým datům

OECD vydala v roce 2007 dokument *OECD Principles and Guidelines for Access to Research Data from Public Funding* (2007), v němž uvádí důvody pro otevřený přístup k výzkumným datům vzniklých z financování z veřejných zdrojů a předkládá principy a doporučení pro vládní politiky vědy. Data z původních výzkumů jsou využívána i v jiných

vědecko-výzkumných či průzkumných projektech a v průmyslu, např. administrativní data OECD o zaměstnanosti jsou využívána v sociálních vědách a při politických rozhodnutích, data vzešlá z organizací zabývajících se péčí o zdraví podporují rozvoj věd o životě (life sciences), [geografická data](#) z výzkumu Země mají význam při environmentálním a dalším výzkumu atd. Přístupnost otevřených dat slouží kupříkladu ke zlepšení mezinárodní spolupráce, posílení otevřeného vědeckého bádání, podpoře k různorodosti provedení datové analýzy a názorů, podpoře nového výzkumu, umožnění testování nových nebo alternativních hypotéz a metod analýzy, podpoře studií o metodách měření a sběru dat, či umožnění vzniku nových datasetů při kombinaci více zdrojů.

Výše zmíněnými principy jsou:

- Otevřenost - jednoduchý a přívětivý otevřený přístup; náklady pouze na diseminaci dat
- Flexibilita - přizpůsobení různým technologiím a systémům či právnímu a socioekonomickému prostředí
- Transparence - informace o organizaci produkující data, podmínky užití dat, informace o politice výzkumných dat na individuální i institucionální úrovni, zřízení standardů pro katalogizaci dat, informace o managementu dat a podmínkách přístupu v rámci sdílení osvědčených postupů
- Právní konformita - přístup a použití dat musí dodržovat práva a zájmy zúčastněných stran a z několika důvodů může dojít k jejich omezení (národní bezpečnost, ochrana osobních údajů, obchodní tajemství a práva duševního vlastnictví, ochrana vzácných nebo ohrožených druhů, soudní proces)
- Ochrana duševního vlastnictví - nastavení přístupnosti k datům musí zohlednit copyright a jiná práva duševního vlastnictví
- Formální odpovědnost - instituce by měla vypracovat pravidla a předpisy pro přístupová opatření, formulovat povinnosti různých stran. Tohoto se týkají otázky autorství a vlastnictví, diseminace dat, omezení použití, finančního ujednání, etických pravidel, licenčních podmínek či dlouhodobé archivace.
- Profesionalismus - určení profesních standardů a hodnot, jež jsou vtělené v kodexech chování zúčastněných vědeckých komunit
- Interoperabilita - technologická a sémantická interoperabilita umožňuje a podporuje mezinárodní a interdisciplinární přístup a použití výzkumných dat
- Kvalita - dodržení standardů kvality dat. Ty by neměly být univerzální, úroveň kvality se liší v závislosti na vědecké disciplíně.
- Bezpečnost
- Účinnost - sběr a sdílení dat zlepšuje celkovou účinnost veřejně financovaného vědeckého výzkumu. Pravidelně by se měly hodnotit náklady a přínosy, zpřístupňovat a trvale uchovávat jen datasety s největším potenciálním využitím (ne všechna data by měla být uchována trvale)
- Odpovědnost - pravidelné hodnocení nastavení přístupu k datům (uživatelskými skupinami, odpovědných institucí a sponzorských výzkumných agentur)
- Udržitelnost - dlouhodobá ochrana dat

Výzva pro akademické knihovny

Dlouhodobou ochranu výzkumných dat řešili na Spolkové vysoké technické škole v Curychu (Neubauer, 2011). Tamní univerzita si uvědomuje potřebu dlouhodobé ochrany u tří typů dat: digitální obsah v knihovně, administrativní záznamy a primární výzkumná data pocházející z vědecké práce. Jejich knihovna se práce ujala a dle počátečního výzkumu zjistila, že komunita vědců skutečně potřebuje dlouhodobou ochranu výzkumných dat. Na základě zjištění vymezili žádoucí pracovní procesy, vybrali a testovali vhodný softwarový nástroj, navrhli service model při zohlednění celého životního cyklu digitálních dat, a určili objem práce a nákladů. V další fázi pak měla proběhnout technická realizace.

Institucionální repozitáře vědeckých dat jsou budovány univerzitami na celém světě. Dalším příkladem je [Stanford Digital Repository](#) spravovaný tamními univerzitními knihovnami. Jak autoři výše zmíněné studie poznamenávají, dlouhodobá ochrana vědeckých dat může spadat nyní nebo v budoucnu pod činnost univerzitních knihoven.

Research Data Management

Problematika sdílení a dlouhodobé ochrany vědeckých dat spadá pod oblast Research Data Management, jež se zabývá životním cyklem výzkumných dat (Surkis, Read, 2015). Do něj patří kromě dlouhodobé ochrany a možnostmi sdílení dat i tvorba, organizace či zpracování dat. Univerzity se principy Research Data Management běžně řídí a nastavují vlastní metody a procesy pro správu vědeckých dat napříč celým jeho životním cyklem.

Některé vědecké časopisy stanovují podmínky pro ukládání dat, kdy je vědec povinen s publikací článku sdílet i data (Whitlock, 2011; Burke, 2014). Příkladem jsou časopisy [Nature](#), [Science](#), [The American Naturalist](#), nebo také nakladatelství [Public Library of Science](#) uděluje pokyny pro sdílení dat a uvádí seznam vhodných repozitářů.

Seznam vybraných repozitářů

Existuje řada projektů, které v podobě webové aplikace veřejně zpřístupňují a archivují vědecká data. Následující seznam je neúplný, obsahuje zejména populární repozitáře.

[Re3data](#) (Registr of Research Data Repositories) - globální registr repozitářů výzkumných dat napříč různorodými disciplínami. Obsahuje přes 1400 repozitářů.

Multioborové zdroje

[Dryad](#) - obsahuje data z přírodních a medicínských věd. Většinou jsou přidružená k recenzovaným článkům, není to však pravidlo, nachází se tam i data z nerecenzovaných zdrojů, např. z monografií či dizertací

[Figshare](#) - obsahuje data napříč všemi disciplínami. Jsou různého typu a formátu a je jim přiřazeno DOI.

[Zenodo](#) - projekt, který iniciovala EU a CERN

Archeologie

[Archaeology Data Service](#) - podporuje výzkum, výuku a vzdělávání v archeologii, poskytuje řadu zdrojů a archeologických dat

Biologie

[BioSharing](#) - katalog databází z oblasti biologických, přírodních a biomedicínských věd
[BioGRID](#) (The Biological General Repository for Interaction Datasets) - data z proteinových a genetických interakcí z organismů a lidí

[European Bioinformatics Institute](#) (EMBL-EBI) - poskytuje data z experimentů z oblasti životních věd, obsahuje mnoho databází pro různé specifické disciplíny

[Global Biodiversity Information Facility](#) - data o nejrozmanitějších životních druzích

[GenBank](#) - mezi biology značně vlivná a používaná databáze nukleotidových sekvencí. Je součástí [International Nucleotide Sequence Database Collaboration](#), jejíž součástí je dále [DNA DataBank of Japan](#), [European Nucleotide Archive](#) a další databáze spravované [NCBI](#).
[TreeBASE](#) - obsahuje všechny druhy fylogenetických dat. Data zde vložená vždy vycházejí z publikovaného příspěvku a dokud příspěvek neprojde recenzním řízením, nebudou data dostupná veřejnosti.

[Data Repository for Kinetic Models of Biological Models \(KiMoSys\)](#) - webová aplikace pro uchovávání a sdílení experimentálních dat a modelů

[Coherent X-ray Imaging Data Bank](#) - databáze zpřístupňující data z experimentů rentgenového záření

[Morphbank](#) - databáze obrázků, jež mají využití při výzkumu druhů ve srovnávací anatomii, morfologické fylogenezi, taxonomii a příbuzných oborech, jež mají zvyšovat poznatky o biodiverzitě

[The Cell: An Image Library](#) - obrázky, videa a animace buněk ze všech organismů

[Biological Magnetic Resonance Data Bank](#) - úložiště dat ze spektroskopie využívající jaderné magnetické rezonance na proteinech, peptidech, nukleových kyselinách a dalších biomolekulách

[Plant Expression Database](#) (PLEXdb) - zdroj genových expresí rostlin a rostlinných patogenů

[YRC Public Data Repository](#) - databáze proteinů

[Global Proteome Machine Database](#) - databáze proteomiky

[PeptideAtlas](#) - kompendium peptidů identifikovaných v experimentech tandemové hmotnostní spektrometrie

[International Molecular Exchange Consortium \(IMEX\)](#) - data z interakcí proteinů

[GenomeRNAi](#) - databáze fenotypů ze snímků RNA interference v octomilkách a lidech

[Database of Interacting Proteins](#) - data reakcí mezi proteiny

[Global Biodiversity Information Facility](#) - mezinárodní infrastruktura poskytující data o 1,6 milionu životních druhů

[Eukaryotic Pathogen Database Resources \(EuPathDB\)](#) - databáze integrující více zdrojů, poskytuje údaje o eukaryotických patogenech v organismech

[Pathosystems Resource Integration Center \(PATRIC\)](#) - podporuje komunitu pro výzkum bakteriálních infekčních chorob, poskytuje data a informace o patogenech

[European Variation Archive](#) - databáze všech typů genetických dat všech biologických druhů
[Worldwide Protein Data Bank](#) - úložiště informací o 3D strukturách proteinů, nukleových kyselin a složitých vazeb. Součástí je [RCSB Protein Data Bank](#), [Research Collaboratory for Structural Bioinformatics Protein Data Bank](#), [Protein Data Bank Japan](#), a [Biological Magnetic Resonance Bank](#) (obsahuje data z experimentů jaderné magnetické rezonance)
[Paleobiology Database](#) - zdroj paleontologických dat a informací o fosilních záznamech
[Movebank](#) - obsahuje údaje o pohybech zvířat v lokálních oblastech a migracích napříč kontinenty
[Human Microbiome Project](#) - obsahuje datasety lidských mikrobiomů
[Human Metabolome Database](#) - databáze obsahující informace o malých molekulách metabolitů v lidském těle

Chemie a chemická biologie

[ChemSpider](#) - velká databáze chemických struktur, její obsah tvoří velká řada datových zdrojů
[ChemSynthesis](#) - databáze chemikálií s informacemi o synthesis references a fyzikálních vlastnostech
[ChemxSeer](#) - digitální knihovna a databáze umožňující vyhledávání dokumentů z domény chemie a dat získaných z chemické kinetiky
[nmrshift2db](#) - databáze organických struktur a jejich spekter nukleární magnetické rezonance
[Crystallography Open Database](#) - kolekce krystalových struktur organických, anorganických a kovových sloučenin a minerálů
[caNanoLab](#) - portál na sdílení dat z oblasti nanotechnologie v biomedicině
[PubChem](#) - databáze molekul menších sloučenin (s méně než 1000 atomy a méně než 1000 vazbami)
[chemotion](#) - repozitář molekul a výzkumných dat

Health Sciences

[Virus Pathogen Resource](#) - databáze virových patogenů
[Influenza Research Database](#) - poskytuje data týkající se viru chřipky, (pozorování na ptácích a savcích, lidská klinická data, fenotypické vlastnosti virů, genomické a proteomické údaje o viru chřipky)
[National Addiction & HIV Data Archive Program](#) (NAHDAP) - archiv dat z výzkumů týkajících se závislostí a HIV
[National Database for Autism Research](#) (NDAR) - repozitář výzkumných dat z oblasti poruch autistického spektra
[The Cancer Imaging Archive](#) (TCIA) - archiv snímků s nádorovými nálezy
[The Virtual Skeleton Database](#) (The SICAS Medical Image Repository) - repozitář obrázků anatomických struktur a klinických dat
[PhysioBank](#) - archiv fyziologických signálů

[National Database for Clinical Trials related to Mental Illness](#) - data z výzkumů sponzorovaných National Institute of Mental Health. Obdobným archivem je [Research Domain Criteria Database](#) (RDoCdb)

[NeuroMorpho](#) - databáze obrázků neuronů

[OpenfMRI](#) - obsahuje datasety a surová data snímků funkční magnetické rezonance

[Melanoma Molecular Map Project](#) - obsahuje databanku (Targeted Therapy Database), v níž se nachází klinická a výzkumná data týkajících se melanomů

[Neuroscience Information Network](#) - je zdrojem dat z oblasti neurověd, obsahuje rovněž ontologie, slovníky, standardy pro anotaci dat, atd.

[Linked Clinical Trials](#) (LinkedCT) - je prvním zdrojem pro klinická datasémantického webu. 5 řídící se principy linked data, takže každá entita je identifikována URI a RDF

[VectorBase](#) - zdroj dat o lidských patogenech

Vědy o Zemi a životním prostředí

[EarthChem](#) - kolekce datových systémů s přístupem ke geochemickým a geochronologickým datům

[Global Index of Vegetation-Plot Databases](#) metadatabáze nabízející přehled 230 databází poskytujících data o vegetaci, zejména mapy rostlinstva (vegetation plots).

[DataONE](#) (Data Observation Network for Earth) - nadace sdružující data z pozorování Země z [26 datových archivů](#). Obsahuje celkem přes 340 tisíc datových souborů (8,6 TB), která se dají vyhledávat ve vyhledávacím nástroji [ONEMercury](#). Sdružuje archivy jako [Knowledge Network for Biocomplexity](#) poskytující data z oblasti ekologie a životního prostředí, [Global Lake Ecological Observatory Network](#) (GLEON) sdílející senzorické údaje o jezerech a interpretuje je v globálním měřítku, a [mnoho dalších](#).

[Tree Ring](#) - největší archiv poskytující data o stromových letokruzích.

[National Centers for Environmental Information](#) - světově nejobsáhlejší archiv dat o počasí a klimatu. Vyvíjí národní datasety USA i datasety globální. Archiv čítá přes 20 petabytů dat. Zpřístupňuje atmosferická, oceanická, geofyzikální data a obsahuje statistické reporty o teplotách či extrémním počasí.

[Pangaea](#) - digitální knihovna umožňující vyhledávání a přístup ke georeferencovaným datům z výzkumu Země a životního prostředí. Je součástí [ICSU World Data System](#) (což je orgán mezinárodní organizace [International Council for Science](#)), jehož hlavními cíli jsou všeobecný přístup ke kvalitním vědeckým datům, datovým produktům, službám a informacím, zajištění jejich dlouhodobé správy a posílení standardů. Mezi jeho členy je 92 vědecko-výzkumných [organizací](#), přičemž 60 z nich jsou [řádní členové](#), jež participují na různých datových službách. Organizace mají široký záběr a ISCU-WDS tak nabízí data zejména z přírodních, ale i sociálních a humanitních věd.

[CISL Research Data Archive](#) - americká organizace zpřístupňující stovky datasetů o životním prostředí

[Data Basin](#) - platforma podporující vzdělávání, výzkum a péči o životní prostředí, poskytuje datasety a mapové ilustrace

[RealClimate](#) (data sources) - katalog datových zdrojů z oblasti vědy o klimatu

[AMEE Discover](#) - jedná se o databázi zpřístupňující globální standardy o emisích a metodologiích

[Global Atmosphere Watch World Data Centres](#) - celkem 6 center, jež sbírají, zpracovávají a archivují atmosferická měření o ozónu a UV záření, sluneční radiaci, skleníkových plynech, aerosolech, aj. Spadají pod World Meteorological Organization.

[National Center for Atmospheric Research](#) - obsahuje rozsáhlou a rozmanitou sbírku meteorologických a oceanografických údajů, určených pro badatelskou činnost

[National Snow and Ice Data Center](#) - poskytuje data a podporuje mrazivý výzkum sněhu, ledu, ledovců, zmrzlé půdy a klimatických interakcí tvořící zemskou kryosféru

[NERC Data Centres](#) - Natural Environment Research Council má šest datových center poskytujících data z disciplín environmentálních věd

[LTER Network Data Portal](#) - data původem z ekologického výzkumu

[Interdisciplinary Earth Data Alliance](#) (IEDA) - projekt poskytující řadu datových služeb pro podporu geověd

[EarthStat](#) - obsahuje geografické datasety sloužící k řešení problémů zemědělské produkce při růstu populace a snížení jejího dopadu na životní prostředí

NASA

[EOSDID](#) (The Earth Observing System Data and Information System) - poskytuje data z pozorování Země naměřená přístroji z různých zdrojů - letadla, satelity, terénní měřidla

[GES DISC](#) (Goddard Earth Sciences Data and Information Services Center) - poskytuje data a informace z oblasti vědeckých dat o Zemi

[GCMD](#) (Global Change Master Directory) - data relevantní ke globálním změnám z oblasti environmentálních věd a věd o Zemi

[Landsat](#) - poskytuje obrázky a datasety z oblasti povrchu Země

Další projekty lze nalézt zde <http://www.nasa.gov/open/data.html>

Astrofyzika

astro.caltech.edu - rozcestník obsahující odkazy na datové archivy, katalogy a atlasy galaxií

[Virtual Observatory](#) - obsahuje odkazy na datové archivy, analytické, vizualizační a další nástroje

[ESO Science Archive Facility](#) - archiv obsahující data z teleskopu European Southern Observatory

[NASA'S DATA PORTAL](#) - katalog rozmanitých datasetů, API, vizualizací, a dalších

[NASA Space Science Data Coordinated Archive](#) - archiv dat z oblasti [astrofyziky](#) a [astronomie](#), [výzkumu Slunce](#), [sluneční soustavy](#), a zdroj čtené řady [obrázkových materiálů](#)

[PDS](#) (The Planetary Data System) - archivuje a šíří data z planetárních misí, astronomických pozorování a laboratorních měření

[Minor Planet Center](#) (MPC) - poskytuje naměřené polohy planetek, komet a přirozených družic planet a ostatní údaje z pozorování

[Solar System Dynamics](#) (SSD) - obsahuje údaje o polohách astronomických objektů (efemeridy), jejich fyzikální vlastnosti, diagramy oběžných drah, a další

Částicová fyzika

[HepData](#) - data z několika tisíců publikací včetně těch z experimentů ve Velkém hadronovém urychlovači (LHC) v CERNu, data z fyziky vysokých energií (high-energy physics)

[ILDG](#) - konfigurace měřidel mřížové kvantové chromodynamiky

[MCDB MonteCarlo Database](#) - centrální databáze pro výstupy experimentů simulací Monte-Carlo

[MCPLOTS](#) - Monte Carlo grafy

[DPHEP](#) - studijní společnost pro dlouhodobou archivaci a dlouhodobou analýzu dat z oblasti fyziky vysokých energií

[National Nuclear Data Center](#) - sbírá, hodnotí a šíří data z oblasti jaderné fyziky

Obecná fyzika

[NIST Physical Measurement Laboratory](#) - vyvíjí a šíří standardy (nejen) pro oblast fyziky, ale i pro různé jiné disciplíny a poskytuje data

[Repozitář NoMaD](#) - schraňuje data o materiálech

Organizace CERN ve Švýcarsku má na svých stránkách seznam [datových repozitářů](#) obsahující data z oblasti částicové, obecné a astrofyziky a nachází se tam i projekty zaměřující se na [dlouhodobou ochranu](#).

Computer Science

[CodePlex](#) - projekt společnosti Microsoft, slouží pro sdílení open-source projektů

[GitHub](#) - rovněž slouží ke sdílení otevřeného softwaru, slouží pro vývoj softwaru pomocí verzovacího nástroje Git

[Google Developers](#) - poskytuje nástroje pro vývoj softwaru, API a technické zdroje

[Launchpad](#) - umožňuje kooperativní vývoj softwaru, slouží ke sdílení projektů, zahrnuje např. Ubuntu projekty

[SourceForge](#) - projekt hostující free a open software

[UMBEL](#) (Upper Mapping and Binding Exchange Layer) - podporuje interoperabilitu obsahu na webu

Humanitní vědy

[Data Archiving and Network Services](#) (DANS) - nizozemský repozitář s téměř 150 tisíci datasey (a milionem publikací) z oblasti sociálních, humanitních, behaviorálních a geoprostorových věd a archeologie

[LINDAT/CLARIN](#) - český repozitář lingvistických dat. (Jako jediný český digitální repozitář má certifikát důvěryhodnosti, konkrétně pečeť Data Seal of Approval.)

[Association of Religion Data Archives](#) (ARDA) - sbírka průzkumů, anket a dalších dat týkajících se náboženství

[Child Care and Early Education Research Connections](#) - podporuje výzkum v oblasti péče o děti a raného vzdělávání a jeho použití v politickém rozhodování, poskytuje výzkumná data z této oblasti

[JSTOR Data for Research](#) - úložiště dat z publikací z JSTOR

Sociální vědy

[ICPSR](#) (Inter-university Consortium for Political and Social Research) - archivuje studie a publikované průzkumy z oblasti sociálních věd a datasety, poskytuje rovněž různé učební pomůcky pro studenty a badatele

[UK Data Archive](#) - kolekce výzkumných dat ze sociálních a humanitních věd

[Dataverse](#) - Dataverse Project vznikl na Harvardu a slouží ke sdílení, uchovávání, citování a analyzování výzkumných dat. Je víceoborový, nejvíce dat však pochází ze sociálních věd.

[Český sociálněvědní datový archiv](#) (ČSDA) - archiv Sociologického ústavu AV ČR slouží jako úložiště výzkumných dat. Řídí se OECD principy a doporučeními pro přístup k výzkumným datům financovaných z veřejných zdrojů.

Srovnávací statistiky

[Český statistický úřad](#)

[OECD Data](#)

[World Bank](#)

[United Nations Data](#)

[Eurostat](#)

Další seznamy

<http://www.nature.com/sdata/data-policies/repositories>

https://www.uwo.edu/bshader/datamanagement/disciplinary_specific_data_repositories.html

<https://www.lib.umn.edu/datamanagement/datacenters>

<http://sciencehackday.pbworks.com/w/page/24500475/Datasets>

<http://library.stanford.edu/research/data-management-services/share-and-preserve-research-data/domain-specific-data-repositories>

<http://pam.sla.org/subjects/data/>

Zdroje

BURKE, Katie L. A Safety Net for Scientific Data. *American Scientist* [online]. 2014, **102**(1), 6- [cit. 2016-02-08]. DOI: 10.1511/2014.106.6. ISSN 0003-0996. Dostupné z: <http://www.americanscientist.org/issues/pub/2014/1/a-safety-net-for-scientific-data>

NEUBAUER, Wolfram. Long-term Preservation of Research Data: Challenge and Opportunity for Academic Libraries in the Future. In: *Proceedings of the International Conference on Library & Information Science* [online]. [cit. 2016-02-08]. Editura Universității Transilvania, 2011, s. 1-6. ISSN 2066-5121. Dostupné z: Library & Information Science Source

OECD Principles and Guidelines for Access to Research Data from Public Funding [online]. Paris: OECD Publishing, 2007 [cit. 2016-02-08]. Dostupné z: <http://www.oecd.org/sti/sci-tech/38500813.pdf>

SURKIS, Alisa a Kevin READ. Research data management. *Journal Of The Medical Library Association* [online]. 2015, **103**(3), 154-156 [cit. 2016-02-08]. Dostupné z: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4511058/>

WHITLOCK, Michael C. Data archiving in ecology and evolution: best practices. *Trends in Ecology & Evolution* [online]. 2011, **26**(2), 61-65 [cit. 2016-02-08]. DOI: 10.1016/j.tree.2010.11.006. ISSN 01695347. Dostupné z: <http://linkinghub.elsevier.com/retrieve/pii/S0169534710002697>