

**Open Science
Week**

Jak na data management ve výzkumných projektech

Milan Janíček

Centrum pro podporu open science, Univerzita Karlova

O čem to bude?

- nové požadavky ze strany poskytovatelů financí - správa výzkumných dat
- byla dlouho na obzoru – v Horizon 2020 s možností opt-outu, v Horizon Europe povinně (týká se třeba i ERC)
- v českém prostředí se vyčkávalo (není infrastruktura apod.)
 - co se změnilo a jak to dopadlo/dopadá?
- o RDM mluvíme už pár let – jak se potkávají obecné požadavky s těmi definovanými poskytovateli?

Osnova

- kontext – co je open science a proč se řeší data (změna paradigmatu)
- požadavky českých grantových agentur
 - Horizon - MŠMT / OP JAK
 - TA ČR
 - GA ČR
- jak na ně (na požadavky)
 - DMP
 - FAIR data
 - repozitáře - ukládání & sdílení
 - support – data stewardi

Proč mluvíme o správě dat?

- Rozšiřuje se představa o tom, co jsou výstupy výzkumného procesu
 - Data mají zásadní roli
 - Zvýšila se dostupnost IT prostředků
 - Infrastruktura, přístroje
 - Úložná a výpočetní kapacita
 - Přenosová kapacita
 - standardy výměny dat
 - software (open source)
- => všechno je (snadněji) propojitelné

Proč mluvíme o správě dat?

- Rostou požadavky na kvalitu výstupů
 - Transparentnost
 - Reproducibilita
- ... a pokud už máme všechna ta kvalitně vytvořená data, proč je znovu nepoužít?
 - Znovuvyužitelnost
 - Otevřený přístup (Open Data?)

Proč mluvíme o správě dat?

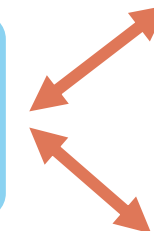
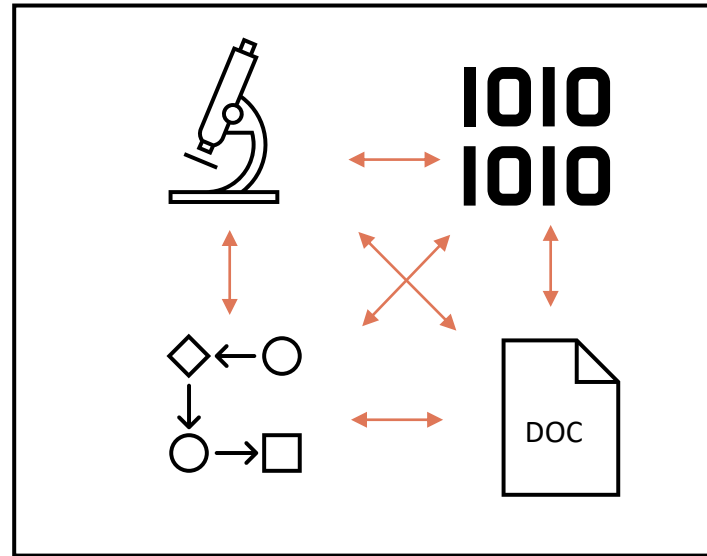
- RDM je součástí Open Science

- podobně lze pracovat i s jinými typy výstupů a procesů ve výzkumu
- workflows, výzkumný software, laboratorní deníky (více odpoledne)
- ... systém hodnocení je ale silně nastavený na publikace jako výstup

=> přechod k „open science“ znamená víc než jen otevírání přístupu k článkům, nebo k datům!

Open science

€



8 pilířů open science

The eight ambitions of Open Science

Open science policy has developed progressively in the EU. It concerns all aspects of the research cycle, from scientific discovery and scientific review to research assessment, publishing and outreach; its cornerstone being open access to publications and research data. Since 2016, the Commission organises its open science policy according to eight 'ambitions':

- Open Data: FAIR (Findable, Accessible, Interoperable and Re-usable data) and open data sharing should become the default for the results of EU-funded scientific research.
- European Open Science Cloud (EOSC): a 'federated ecosystem of research data infrastructures' will allow the scientific community to share and process publicly funded research results and data across borders and scientific domains.
- New Generation Metrics: New indicators must be developed to complement the conventional indicators for research quality and impact, so as to do justice to open science practices.
- Future of scholarly communication: all peer-reviewed scientific publications should be freely accessible, and the early sharing of different kinds of research outputs should be encouraged.
- Rewards: research career evaluation systems should fully acknowledge open science activities.
- Research integrity: all publicly funded research in the EU should adhere to commonly agreed standards of research integrity.
- Education and skills: all scientists in Europe should have the necessary skills and support to apply open science research routines and practices.
- Citizen science: the general public should be able to make significant contributions and be recognised as valid European science knowledge producers.

Open Science (2019)

- **otevřená data, FAIR**
- European Open Science Cloud (EOSC)
- metriky nové generace
- budoucnost vědecké komunikace
- ohodnocení
- integrita výzkumu
- vzdělávání a dovednosti
- Citizen Science – občanská věda

Horizon Europe



June 2021
"Horizon Europe will set a new standard for dissemination of knowledge and new skills across European societies. With clear and immediate open access requirements for beneficiaries, the Open Research Europe publishing platform and a strengthened European Open Science Cloud, we are well underway in making truly open science a reality."

Mariya Gabriel Commissioner for Innovation, Research, Culture, Education and Youth

OPEN SCIENCE

EARLY KNOWLEDGE AND DATA SHARING, AND OPEN COLLABORATION



The challenge is for Europe to embrace open science as the modus operandi for all researchers. Open science consists in the sharing of knowledge, data and tools as early as possible in the Research and Innovation (R&I) process, in open collaboration with all relevant knowledge actors, including academia, industry, public authorities, end users, citizens and society at large. Open science has the potential to increase the quality, efficiency and impact of R&I, lead to greater responsiveness to societal challenges, and increase trust of society in the science system.

- open access, FAIR Data, plán správy dat a další
- + open science součástí evaluace (součást excellence)
- 95,5 mld EUR (2021-2027)

What are open science practices?

- Open access to research outputs such as publications, data, software, models, algorithms, and workflows;
- Early and open sharing of research, for example through preregistration, registered reports, pre-prints, and crowd-sourcing of solutions to a specific problem;
- Use of open research infrastructures for knowledge and data sharing;
- Participation in open peer-review;
- Measures to ensure reproducibility of results; and
- Open collaboration within science and with other knowledge actors, including involving citizens, civil society and end-users, such as in citizen science.



Horizon Europe, open science (2021)

Horizon Europe – požadavky na správu dat



Open science: research data management

The beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles and by taking all of the following actions:

- establish a data management plan ('DMP') (and regularly update it)
- as soon as possible and within the deadlines set out in the DMP, deposit the data in a trusted repository; if required in the call conditions, this repository must be federated in the EOSC in compliance with EOSC requirements
- as soon as possible and within the deadlines set out in the DMP, ensure open access — via the repository — to the deposited data, under the latest available version of the Creative Commons Attribution International Public License (CC BY) or Creative Commons Public Domain Dedication (CC 0) or a licence/dedication with equivalent rights, following the principle 'as open as possible as closed as necessary', unless providing open access would in particular:
 - be against the beneficiary's legitimate interests, including regarding commercial exploitation, or
 - be contrary to any other constraints, in particular the EU competitive interests or the beneficiary's obligations under this Agreement; if open access is not provided (to some or all data), this must be justified in the DMP
- provide information via the repository about any research output or any other tools and instruments needed to re-use or validate the data.

Horizon Europe - DMPs

- Projekty jsou velké, konsorciální
- DMP jsou komplikovanější a specifické
- DMP jsou Deliverables – jsou v systému CORDIS tady!
- katalog vzorových DMP?
 - ... ne 😊



[Open Science
Week

Podmínky českých poskytovatelů financí

Jak to vypadá v ČR?

- nejdříve „disclaimer“
 - ***nemůžeme mluvit za poskytovatele financí, naše interpretace nemusí být nutně zcela správná***
 - na druhou stranu můžeme srovnávat zvenku
- => pokud si nejste jistí interpretací pravidel, zeptejte se pro jistotu přímo daného poskytovatele

MŠMT / OP JAK



Operační program
Jan Amos Komenský

PRAVIDLA PRO ŽADATELE A PŘÍJEMCE
SPECIFICKÁ ČÁST
VÝZVA
ŠPIČKOVÝ VÝZKUM
Programové období 2021–2027

Verze:	1
Vydal:	Řídicí orgán OP JAK
Datum platnosti:	Dnem zveřejnění na webových stránkách OP JAK
Datum účinnosti:	14. 07. 2022



Spolufinancováno
Evropskou unií



OPJAK.cz
MS



2. Žadatel/příjemce je povinen zajistit správu výzkumných dat, shromážděných a vytvořených během projektu, v souladu s FAIR principy, a to zejména těmito prostředky:

- vypracováním Plánu pro správu dat¹⁸ v souladu s FAIR principy a jeho předložením při podání žádosti o podporu, v průběhu trvání projektu a při ukončení projektu¹⁹;
- uložením výzkumných dat do důvěryhodného repozitáře podle Plánu pro správu dat;
- zajištěním otevřeného přístupu k výzkumným datům uloženým v repozitáři za podmínek poslední dostupné verze veřejné licence Creative Commons Attribution International (CC BY) nebo jejího ekvivalentu, a to v souladu se zásadou „otevřená jak jen možno, uzavřená jen jak nutno“ s ohledem na soukromí, ochranu osobních údajů, důvěrnost, oprávněné obchodní zájmy a práva duševního vlastnictví třetích stran, bezpečnost státu nebo jiné oprávněné zájmy a jiná oprávněná omezení (zdůvodnění musí být uvedeno v Plánu pro správu dat a musí být zajištěn pravidelný přezkum tohoto zdůvodnění)²⁰;
- zajištěním otevřeného přístupu k metadatům uložených výzkumných dat, aby byla v souladu s Obecnými doporučeními metadatového popisu²¹;
- poskytnutím informací (prostřednictvím repozitáře) o všech dalších výsledcích výzkumu nebo nástrojích a instrumentech potřebných k opětovnému využití výzkumných dat nebo k jejich validaci (pokud se neuplatní oprávněné důvody pro omezení těchto informací).

MŠMT / OP JAK



Operační program
Jan Amos Komenský

PRAVIDLA PRO ŽADATELE A PŘÍJEMCE
SPECIFICKÁ ČÁST
VÝZVA
ŠPIČKOVÝ VÝZKUM
Programové období 2021–2027

Verze:	1
Vydal:	Rídicí orgán OP JAK
Datum platnosti:	Dnem zveřejnění na webových stránkách OP JAK
Datum účinnosti:	14. 07. 2022



Spolufinancováno
Evropskou unií



OPJAK.cz
MS



2. Žadatel/příjemce je povinen zajistit správu výzkumných dat, shromážděných a vytvořených během projektu, v souladu s FAIR principy, a to zejména těmito prostředky:

- vypracováním Plánu pro správu dat¹⁸ v souladu s FAIR principy a jeho předložením při podání žádosti o podporu, v průběhu trvání projektu a při ukončení projektu¹⁹;
- uložením výzkumných dat do důvěryhodného repozitáře podle Plánu pro správu dat;

¹⁶ Creative Commons licence jsou soubor veřejných licencí, které přinášejí nové možnosti v oblasti publikování autorských děl a posilují pozici autora při rozhodování za jakých podmínek bude dílo veřejně přístupné. Licence [CC BY](https://creativecommons.org/licenses/by/4.0/) - umožňuje dílo volně sdílet (rozmnožovat a distribuovat prostřednictvím jakéhokoli média v jakémkoli formátu) a upravovat (remixovat, změnit a vyjít z původního díla pro jakýkoli účel, a to i komerční) za jediné podmínky uvedení původce. Výklad dalších variant viz souhrnný odkaz <http://www.creativecommons.cz/licence-cc/varianty-licence/>.

¹⁷ Viz <https://doi.org/10.48813/yt6w-6h15>


¹⁸ *Plán pro správu dat* je dokument, který popisuje nakládání s výzkumnými daty během projektu a po jeho skončení. Jedná se o ekvivalent anglického pojmu „data management plan“. Plán pro správu dat musí obsahově odpovídat vzoru pro program HORIZON Europe viz <https://doi.org/10.48813/sstg-4g21>

¹⁹ **Žadatel/příjemce předloží Plán pro správu dat vždy tak, aby uvedené údaje reflektovaly aktuální stav, a to následující formou a v těchto projektových fázích:** a) při podání žádosti o podporu jako přílohu Studie proveditelnosti b) jako přílohu první Zprávy o realizaci následující po uplynutí 24 měsíců od zahájení realizace projektu c) jako přílohu Závěrečné Zprávy o realizaci.


MŠMT / OP JAK

- Požadavky
 - Data = data shromážděná a vznikající v rámci projektu
 - vypracování Plánu pro správu dat v souladu s FAIR principy a jeho pravidelné aktualizování
 - uložení výzkumných dat do důvěryhodného repozitáře podle Plánu pro správu dat
 - zajištění otevřeného přístupu k výzkumným datům v souladu se zásadou “as open as possible, as closed as necessary”
 - zajištění otevřeného přístupu k metadatům uložených výzkumných dat, aby byla v souladu s Obecnými doporučeními metadatového popisu
 - Poskytnutí informací (prostřednictvím repozitáře) o všech dalších výsledcích výzkumu nebo nástrojích a instrumentech potřebných k opětovnému využití výzkumných dat nebo k jejich validaci
 - (shrnutí na <https://openscience.cuni.cz/OSCI-181.html>)

TA ČR, GA ČR

- Nejdříve se musíme podívat na Zákon č. 130/2002 Sb. (aktuální znění - novela 2022)...
 - *Zákon o podpoře výzkumu a vývoje z veřejných prostředků a o změně některých souvisejících zákonů (zákon o podpoře výzkumu a vývoje)*
 - <https://www.zakonyprolidi.cz/cs/2002-130>
- ... jehož novela vychází z implementace Směrnice 2019/1024
 - *Směrnice Evropského parlamentu a Rady (EU) 2019/1024 ze dne 20. června 2019 o otevřených datech a opakovaném použití informací veřejného sektoru*
 - <https://eur-lex.europa.eu/legal-content/CS/TXT/?uri=CELEX:32019L1024>
- „Follow the white rabbit.“ 

Zákon č.130/2002

- „I’d rather not... but“ 
- Zákon č. 130/2002 se v novelizovaných částech věnuje výzkumným datům
 - povinnost zveřejňovat/poskytovat výzkumná data k výsledkům podpořených z veřejných zdrojů a to pro všechny projekty z veřejných soutěží ve výzkumu
 - data = *“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.”*
- Náležitosti smlouvy o poskytnutí účelové podpory na projekt:
 - Způsob správy výzkumných dat (dle § 9 odst. 1 písm. l)
 - Informace o dostupnosti a způsobu šíření výsledků výzkumu a výzkumných dat (dle § 9 odst. 1 písm. m)
- (shrnutí na <https://openscience.cuni.cz/OSCI-230.html>)

TA ČR



**Výzkumná data
a otevřený přístup
k výstupům/výsledkům**

Verze č. 1 (2023)

- implementace podmínek zákona č.130/2022
- Příjemci podpory mají povinnost vytvořit pro svá data **Plán správy dat** (také “Data management plan”) a předložit jej s první průběžnou zprávou. Plán správy dat by měl být pravidelně aktualizován a měl by obsahovat informace o tom:
 - jaká data budou **vytvářena**, **zpracovávána** nebo **sbírána**
 - jaké **metody a zásady** jejich správy budou použity s ohledem na FAIR principy
 - jakým způsobem budou data **sdílěna a publikována**
 - jakým způsobem budou data **ukládána** během řešení projektu a **uchována** po skončení projektu



- shrnutí na: <https://openscience.cuni.cz/OSCI-208.html>

TA ČR – poskytování dat

- „Ze zákona plyne, že příjemce nemusí (ale může) data k projektu automaticky zveřejňovat. Zákon však říká, že pokud jiná osoba požádá o výzkumná data vzniklá z projektu, příjemce musí tato data poskytnout a to vždy, pokud“:
 - data vznikla se 100% veřejnou podporou (i kombinace účelové a institucionální);
 - data nejsou chráněná podle zákonů upravujících ochranu výsledků autorské, vynálezecké nebo obdobné tvůrčí činnosti nebo která jsou chráněna toliko zvláštním právem pořizovatele databáze podle jiného právního předpisu, jehož je příjemce vykonavatelem;
 - a pokud uplynulo 12 měsíců od ukončení poskytování podpory

TA ČR – poskytování dat

- Příjemce a další účastníci projektu **nejsou povinni** zpřístupnit výzkumná data, pokud by jejich zpřístupněním došlo k nepřiměřenému zásahu do:
 - práva na ochranu **duševního vlastnictví**;
 - práva na ochranu **soukromí a osobních údajů**;
 - práva na ochranu **obchodního tajemství, bezpečnosti státu** nebo **jiných oprávněných zájmů** příjemce (např. v případě možnosti obchodního využití).

TA ČR – poskytování dat

- „Ve veřejných soutěžích TA ČR s povinným otevřeným přístupem k vybraným typům výsledkům platí nad rámec zákona povinnost zveřejnit výzkumná data nezbytná k ověření výsledku s otevřeným přístupem, a to nejpozději při ukončení projektu.“
- „...náklady spojené s přípravou a uchováním dat jsou způsobilými náklady po dobu trvání projektu.“

OTEVŘENÁ VĚDA NA GA ČR

Open Science, též označovaná jako „otevřená věda“, je přístup k vědeckému procesu, který klade důraz na spolupráci, efektivní sdílení znalostí, zlepšení dostupnosti a opakovanou využitelnost výsledků výzkumu pomocí digitálních technologií a nových nástrojů. Principy Open Science podporují aktivní zapojení vědecké komunity a usilují o vytvoření otevřeného a transparentního prostředí pro vědecký výzkum.

Mezi základní pilíře otevřené vědy bývá považováno zejména otevřený přístup k recenzovaným vědeckým publikacím (Open Access) a otevřený přístup k výzkumným datům (Open Data).

OTEVŘENÝ PŘÍSTUP K RECENOVANÝM VĚDECKÝM PUBLIKACÍM (OPEN ACCESS)

Otevřený přístup k výsledkům výzkumu financovaného z veřejných zdrojů umožňuje jejich široké zpřístupnění a rychlé sdílení, čímž přispívá k inovacím, posilování konkurenceschopnosti a mezinárodní spolupráci. V oblasti otevřeného přístupu k vědeckým publikacím neexistuje závazný legislativní rámec, který by povinnost publikovat formou otevřeného přístupu upravoval. V této oblasti existuje celá řada iniciativ, deklarací a doporučení (např. Doporučením komise (EU) 2018/790 ze dne 25. dubna 2018 o přístupu k vědeckým informacím a jejich uchování, dle kterého je členským státům doporučeno definovat a zavést jasné strategie otevřeného přístupu k vědeckým publikacím a výzkumným datům, které jsou výsledkem výzkumu podporovaného z veřejných prostředků, včetně detailních akčních plánů). Na úrovni zákona č. 130/2002 Sb., o podpoře výzkumu, experimentálního vývoje a inovací, ve znění pozdějších předpisů není problematika otevřeného přístupu prakticky upravena. Politika Grantové agentury České republiky (GA ČR) ve vztahu k otevřenému přístupu k vědeckým publikacím je postavena na následujících principech:

- **motivace řešitele** publikovat Open Access;
- **zvýšení podílu výsledků**, které budou formou Open Access publikovány (s výjimkou výsledků, kde to z principu není možné, např. monografie);
- zvýšení dostupnosti výsledků a tím **zvýšení efektivity prostředků** vynaložených na řešení projektů;
- zahrnutí nákladů na otevřený přístup k vědeckým publikacím **do způsobilých nákladů projektů GA ČR**;
- podpora transformačních smluv českých institucí s cílem zvýšení podílu vědeckých publikací s otevřeným přístupem;
- podpora **všech formátů Open Access: Green Open Access, Gold Open Access a Diamond Open Access**;
- svoboda řešitelů vybrat, jakou formou a v jakých publikačních mediích budou svoje výsledky publikovat.



- implementace podmínek zákona č.130/2022
- principy podporované GA ČR
 - podpora iniciativy Open Data a FAIR data;
 - implementace směrnice, která přináší povinnost publikovat Open Data u všech projektů podpořených z veřejných rozpočtů;
 - náklady na data management jsou způsobilými náklady projektu;
 - všechny vědecké publikace, které budou odeslány k publikaci řešiteli projektů GA ČR, by měly mít dohledatelná otevřená data se zákonnými výjimkami (ochrana duševního vlastnictví, obchodního tajemství, bezpečnosti státu nebo oprávněných obchodních zájmů příjemce nebo třetí osoby);
 - řešitelé projektů GA ČR jsou povinni předkládat Data Management Plan nejpozději k datu doručení první dílčí zprávy projektu a pravidelně tento plán aktualizovat;
 - data by neměla být jen otevřená, ale měla by být také FAIR, tj. dohledatelná, přístupná, interoperabilní a opětovně využitelná (Findable, Accessible, Interoperable, and Reusable);
 - Data Management Plan by mělo být možné vytvořit co nejnadhěji, optimálně pomocí aplikace, která řešitele provede jednotlivými aspekty jeho tvorby;

GA ČR - DMP

- DMP by měl obsahovat minimálně informace o tom:

- (i) jaká výzkumná data bude příjemce v rámci projektu vytvářet, zpracovávat nebo sbírat;

- (ii) jaké metody a zásady jejich správy příjemce použije;

- (iii) zda a jakým způsobem budou data sdílena, publikována nebo uveřejňována, případně odůvodnění, proč nemohou být výzkumná data šířena a zveřejňována (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);

- (iv) jakým způsobem budou data ukládána během řešení projektu a uchovávána po skončení projektu.

GA ČR – uznatelné náklady

- „... náklady na správu výzkumných dat (např. náklady spojené se zpracováním a ukládáním dat, anonymizací dat) vztahujících se k projektu jsou způsobilým nákladem v rámci projektů financovaných Grantovou agenturou České republiky.“

[Open Science
Week

Základní principy správy výzkumných dat

Data management plan

- dokument, který specifikuje, jaká data a jakým způsobem budou během výzkumu vytvářena, a obsahuje informace o jejich dostupnosti a využití
 - Pomáhá předvídat potenciální problémy
 - Kontrola, že máte všechno potřebné vybavení a podporu
 - Snížení rizika duplicitní práce, ztráty dat a narušení bezpečnosti
 - Zajištění, že data jsou přesná, úplná, přesná a spolehlivá
 - Usnadnění sdílení dat
 - Zajištění kontinuity dlouhodobých projektů, kde se např. střídají doktorandi, a konzistence projektů, do kterých je zapojeno více vědeckých pracovníků

Data management plan



dokumentace
a metadata



bezpečné skladování
a zálohování dat



snadné sdílení dat



kontinuita projektu



etika
a právní otázky



integrita projektu



znovuvyužitelnost
dat



zodpovědnost
a zdroje

Data management plan - šablony

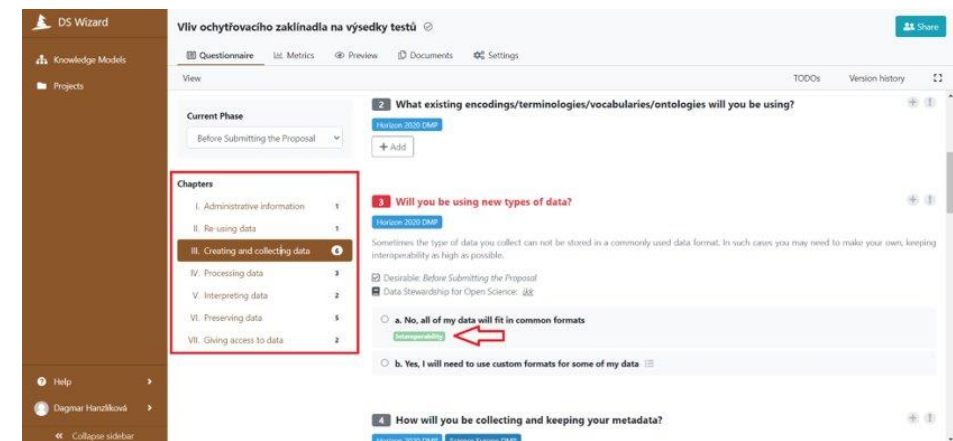
- obvykle se poskytují šablony (template) obsahující základní strukturu dokumentu a specifické otázky
- šablona Horizon Europe
 - + český překlad od NTK
- šablona ERC
- šablona TA ČR
- OP JAK, GA ČR nemají

The image displays two data management plan templates side-by-side. On the left is the 'Data Management Plan template (vzor)' for Horizon Europe, featuring the European Union flag and the text 'Horizon Europe'. On the right is the 'ERC Data Management Plan Template' for the European Research Council, featuring the ERC logo and the text 'European Research Council Executive Agency'. Below the templates, a screenshot of a Microsoft Excel spreadsheet is shown, containing a form for a 'PLÁN SPRÁVY DAT' (Data Management Plan) with fields for project code, name, date, and version, and instructions for completion.

Data management plan - nástroje

- tužka & papír / MS Word
- [DMPonline](#)
- DS Wizard / FAIR Wizard
 - [MU](#)
 - [KNAV](#)
- Argos (OpenAIRE)

• Další – např. [https://rdmkit.elixir-europe.org/data management plan](https://rdmkit.elixir-europe.org/data-management-plan)



Data management plan – český kontext

- často se doporučuje šablona Horizon Europe (případně v překladu od NTK)
- GA ČR má stanovené „minimální požadavky“
- TA ČR má vlastní „zjednodušenou“ šablonu v Excelu
 - v určitých situacích nemusí být omezení na 3000 znaků na pole zjednodušení
 - u veřejných soutěží a mezinárodních výzev lze využít šab. Horizon Europe
- z nástrojů hraje nejvýznamnější roli DS Wizard, vyvíjený na ČVUT ve spolupráci s infrastrukturou ELIXIR (přednáška zítra)

FAIR Data

- 15 principů (<https://www.go-fair.org/fair-principles/>)
- Findable, Accessible, Interoperable, Reusable
- jsou základní principy, jak by měla data vypadat, aby odpovídala současným požadavkům
 - co je potřeba pro budování služeb, která budou propojovat data z různých zdrojů
 - co je potřeba pro sdílení a znovuvyužití dat
 - co je potřeba pro strojové zpracování a data mining
 - co je potřeba k ověření výsledků výzkumu (reproducibilita)
- nestačí vzít svoje soubory a někam je nahrát, kontext je stejně důležitý jako data samotná!
- na druhou stranu – „FAIR“ principy jsou především popisem dobré praxe

Findable

- nalezitelná data
 - cílem je zabezpečit možnost najít data
- významnou roli hrají **perzistentní identifikátory**
 - např. DOI, ORCID, ROR nicméně mohou existovat i specifické oborové identifikátory
 - umožňují jednoznačné propojení různých objektů
- **metadata** datové sady jsou základem pro nalezení v rámci různých vyhledávacích služeb
 - špatná metadata => nenalezitelná data

Findable – v šabloně HE

- Budou data opatřena **perzistentním identifikátorem**?
- 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.
- Budou metadata nabízena takovým způsobem, aby je bylo možné **automaticky zpracovávat** (harverstovat) a indexovat?

Accessible

- přístupná data
 - meta(data) by měla být **získatelná** pomocí svého identifikátoru
- za použití standardního **komunikačního protokolu**, který může použít kdokoli
 - autentizace a autorizace jsou možné
- metadata by měla být dostupná, i kdyby už samotná data nebyla
 - data lze vymazat, ale musí o nich zůstat dostatečná informace

Accessible - v šabloně HE

- 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.
- 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.
- Budou data přístupná prostřednictvím bezplatného a standardizovaného přístupového protokolu?
- Pokud existuje omezení k užití dat, jak bude zajištěn přístup k datům? Jak bude nastaven přístup v průběhu projektu, a jak po jeho skončení?

Accessible - v šabloně HE

- data
 - Jak bude ověřena totožnost osoby, která dostane k datům přístup?
 - Je potřeba zřídit komisi pro přístup k datům (např. pro hodnocení/schvalování žádostí o přístup k osobním/citlivým údajům)?
- 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?
 - Jak dlouho budou data dostupná a dohledatelná? Bude zaručeno, že metadata zůstanou dostupná i poté, co data už nebudou dostupná?
 - Bude zahrnuta dokumentace nebo odkaz na jakýkoliv software, který bude potřebný pro přístup k datům nebo jejich prohlížení?
 - Bude možné zahrnout příslušný software (např. v otevřeném zdrojovém kódu)?

Interoperable

- interoperabilní data
 - mělo by být možné kombinovat data s dalšími daty a využívat další nástroje
- formát by měl být otevřený a interoperabilní pro různé nástroje a služby
- otevřený formát nemusí existovat, ale pokud existuje, měl by být preferován
 - i metadata by měla být interoperabilní
 - je vhodné používat oborové standardy

Interoperable – šabloně HE

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

Reusable

- znovuvyužitelná data
 - optimalizujte data pro znovuvyužití
- důkladný a přesný popis dat umožňuje jak replikaci výsledků tak nové využití dat v novém kontextu
- provenance – (zjednodušeně) informace o původu a úpravách dat
- důležitá je i licence – specifikuje jakým způsobem a za jakých podmínek mohou být data znovu využita
 - volnější licence usnadňuje další využití dat

Reusable - v šabloně HE

- 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.)? Budou vaše data obsahovat odkazy na jiná data (např. jiná data z vašeho projektu nebo soubory dat z předchozího výzkumu)?
- 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?
- Bude umožněno použití dat třetí straně během, a hlavně po skončení projektu?
- Bude původ dat důkladně zdokumentován pomocí příslušných standardů?
- Popište všechny relevantní procesy zajištění kvality dat.

FAIR Data

- velkou část požadavků na FAIR Data lze vyřešit uložením do vhodného repozitáře (F,A)
- naplnění dalších požadavků (I,R) je ale závislé na způsobu práce s daty v průběhu výzkumu a těžko se dohání v momentě publikování dat
- většinou se hovoří o naplnění FAIR principů, pozor není to úplně přenositelné mezi různými obory a druhy dat – ne všechno může být „100% FAIR“

FAIR Data – český kontext

- konkrétní implementace FAIR principů závisí na jednotlivých oborech
 - oborové standardy jsou z principu mezinárodní
- z národního pohledu je důležitá podpora infrastruktury
 - u nás specificky **Národní centrum pro perzistentní identifikátory**
<https://identifikatory.cz/cs/>

 **identifikatory.cz**
Stránky o perzistentních identifikátorech

 English

Identifikátory ▾ Služby ▾ O nás Novinky 🔍

Perzistentní identifikátory (PID) jsou nástroje, které slouží k jednoznačné identifikaci osob, organizací a dalších objektů (např. knih, článků, datových sad) v systému vědecké komunikace. Na tomto webu najdete informace o perzistentních identifikátorech a druzích jejich podpory na národní úrovni.

Oblíbené odkazy

Národní centrum
ORCID

Národní centrum DOI

Česká národní
agentura ISBN
a ISMN

Národní centrum
ISSN

Novinky

Výroční setkání uživatelů služeb
Národního centra DOI 2024

30. září 2024

Národní centrum pro perzistentní identifikátory při Národní technické knihovně srdečně zve zástupce členských institucí i další zájemce na letošní

Výroční setkání uživatelů služeb
národního centra ORCID 2024

30. září 2024

Národní centrum pro perzistentní identifikátory při Národní technické knihovně srdečně zve zástupce členských institucí i další zájemce na letošní,

Trust Markers v ORCID
záznamech: Ověřené e-mailové
domény

27. září 2024

ORCID nyní umožňuje výzkumníkům přidávat ověřené institucionální e-mailové domény do jejich ORCID záznamů jako nový typ tzv. Trust Markeru. Tato...

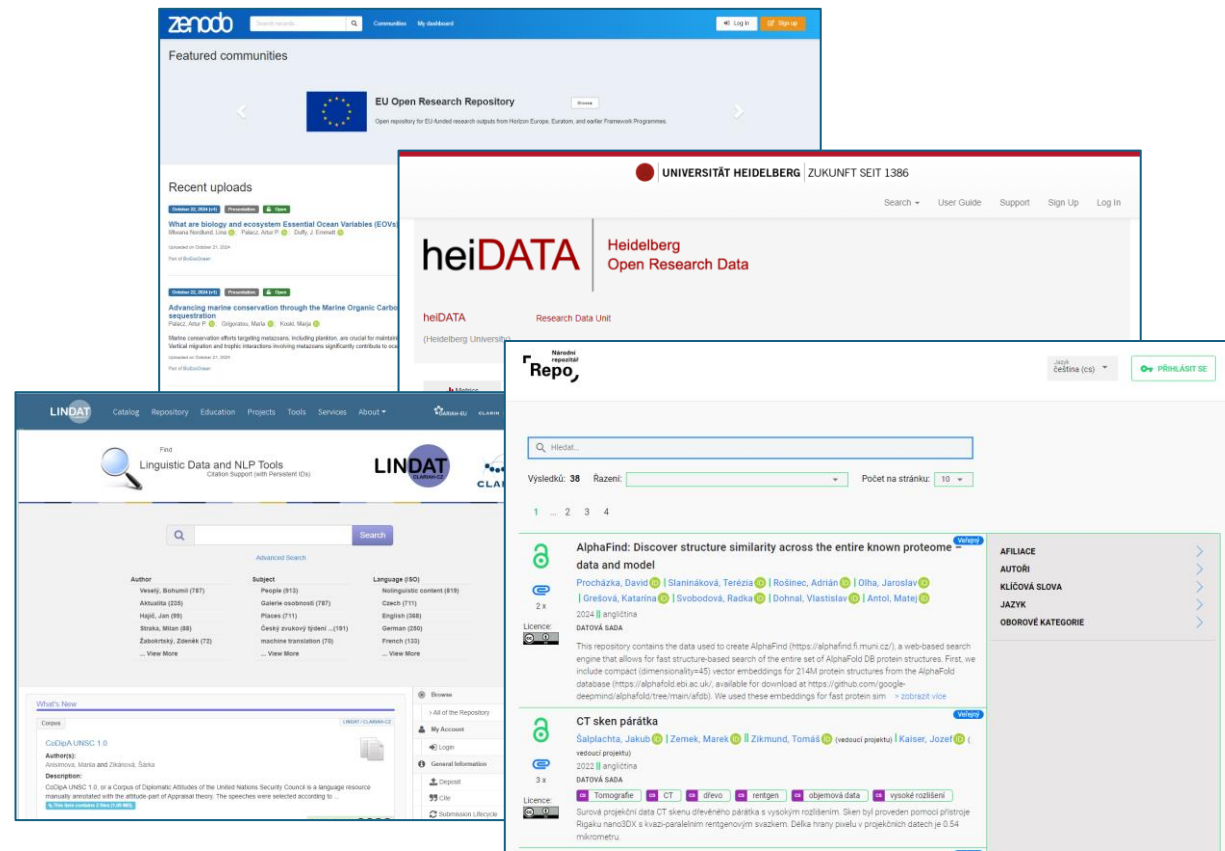
Aktualizace Zásad ochrany
osobních údajů ORCID

27. září 2024

ORCID aktualizoval své Zásady ochrany osobních údajů, aby byly jednodušší a lépe se v nich orientovalo. Základní principy ochrany osobních...

Repozitáře

- „důvěryhodný“ repozitář
 - stabilní, otevřený přístup, identifikátory, metadata, licence, nastavená pravidla správy, udržitelnost,...
- Existují certifikace pro repozitáře, pro některé typy repozitářů jsou ale z povahy certifikace těžko dosažitelné
- Výběr repozitáře
 1. oborový
 2. institucionální
 3. obecný (catch-all)



Repozitáře – český kontext

- lze využívat mezinárodní infrastrukturu
- můžete narazit na kapacitní limity u některých velkých datasetů
 - např. Zenodo má limit 50 GB na dataset
- pro některé oblasti (zatím) neexistují oborové repozitáře
- v ČR se pracuje na výstavbě datové infrastruktury, jejíž součástí je Národní repozitářová platforma (NRP) – (více odpoledne)

Data stewardi

- nová pozice pro podporu práce s výzkumnými daty
- koordinuje sběr, ukládání, zabezpečení a údržbu dat pocházejících z výzkumného projektu
- Odpovědný za nastavení a dodržování správných postupů při práci s daty
- Znalost standardů daného oboru
- Zapojení do výzkumného týmu
- Příprava a aktualizace DMP

Data stewardi – český kontext

- např. GA ČR doporučuje zřídit pozici data stewarda (či obdobnou)
 - způsobilé náklady projektu
- postupně vznikají specializované kurzy
- řada institucí už zaměstnává data stewardy
- Komunita data stewardů – součást projektu NRP
 - Komunitní Discord (dostupné komukoli): <https://discord.gg/WRtuTF9k>
- (více už brzy 😊)

Open Science Week

Na závěr ještě...



Drazí... výzkumníci

- Nevyhýbejte se správě dat, stejně vás dožene. Pokud vás k ní nebudou nutit poskytovatelé, přijdou požadavky ze strany časopisů...
- Zkuste se nedívat na DMP jako na další kus papíru který musíte nějak vyplnit – je to dokumentace a bude tak dobrá, jak dobrou si jí uděláte.
- Je docela dobře možné, že velkou část z toho, co po vás poskytovatelé chtějí už děláte. Open science není o radikální změně toho co se dělá, ale o best practice.
- Ano, je to práce navíc, měla by být oceněna a neměli byste na ní být sami.

Drazí... poskytovatelé

- Nedívejte se na DMP jen jako na další součást projektové administrativy – je to primárně dokumentace a neslouží přímo k vyhodnocení projektu.
- Respektujte oborové zvyklosti a trendy.
- Nesnažte se strojově měřit a následně hodnotit míru naplňování FAIR principů; nebude to fungovat, všechna data si nejsou rovna.
- Do takové míry, jak to jen bude možné respektujte omezení existující infrastrukturou – nejsou k dispozici prostředky na vyřešení všech problémů optimálním způsobem... zatím?

Drazí... data stewardi

- Vítejte v dynamicky se rozvíjejícím prostředí plném „výzev“ a „příležitostí“ 😊
- I cesta může být cíl.
- Zajděte čas od času na Discord Komunitu data stewardů a nebojte se mluvit s ostatními.

[Open Science
Week

Dotazy?

Open Science
Week

Děkuji za pozornost!

Milan Janíček
milan.janicek@ruk.cuni.cz



Odkazy a zdroje

Horizon Europe

- Annotated Grant Agreement
 - podrobnější definice požadavků na open science v HE, včetně požadavků na důvěryhodné repozitáře (str. 373)
 - https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/aga_en.pdf
- Horizon Europe, open science: <https://op.europa.eu/s/zXs5>
- DMP šablona HE v repozitáři NTK: <https://doi.org/10.48813/sstg-4g21>

Odkazy a zdroje

- Zákon č.130/2022 Sb.
 - <https://www.zakonyprolidi.cz/cs/2002-130>
- Směrnice 2019/1024
 - <https://eur-lex.europa.eu/legal-content/CS/TXT/?uri=CELEX:32019L1024>

Odkazy a zdroje

- OP JAK
 - OS v „Specifická pravidla pro žadatele.“
 - nutné zkontrolovat podmínky konkrétní výzvy
 - např.(!) výzva Špičkový výzkum: https://opjak.cz/wp-content/uploads/2022/07/SPpZP_Spickovy_vyzkum-1.pdf
 - Příručka postupů otevřené vědy
 - <https://opjak.cz/dokumenty/otevrena-veda/>

Odkazy a zdroje

- TA ČR
 - Výzkumná data a otevřený přístup k výstupům/výsledkům
 - <https://www.tacr.cz/dokumenty/vyzkumna-data-a-otevreny-pristup-k-vystupum-vysledkum-19>
- GA ČR
 - GA ČR a otevřená věda
 - <https://gacr.cz/ga-cr-a-otevrena-veda/>