

FAIR/OPEN DATA
PRO UŽIVATELE
CF MAFIL

o M U N I

2023-02-22



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY



FAIR/Open data pro uživatele CF MAFIL

FAIR data, praktická implementace a proč to děláme – požadavky projektových výzev

Michal Růžička, ÚVT MU
ruzicka@ics.muni.cz

2023-02-22

Obsah přednášky

1. Kontext
2. Životní cyklus výzkumných dat
3. FAIR principy
4. Repozitáře pro výzkumná data praktičtěji
 - Typy repositářů
 - Jak najít vhodný repositář
5. Data Management Planing
 - DMPlanning vs. DMPlan
 - DSW MUNI
6. PID
 - PID pro osoby, instituce, články, data, ...

MUNI
ICS

Kontext

Open Science a projekty

– Podmínky

- Horizon Europe
- Exceles
- OP JAK – Špičkový výzkum

– Povinné/nepovinné principy

- **Povinné**: Open Access, Data Management (DMP)
- **Nepovinné**: Citizen Science, Open Source, Preprints, ...

Životní cyklus výzkumných dat

Životní cyklus výzkumných dat



- Jaká data (znovu) používáte,
 - včetně licencí, které vám to umožňují,
- jaká data generujete a jakým způsobem,
- kde je ukládáte, zálohujete, dlouhodobě uchováváte,
- jak je trvale a jedinečně identifikujete,
- jak je zpracováváte,
- jak je analyzujete,
- kde je zveřejňujete a sdílíte,
- kdo tohle všechno zaplatí;
- o čem data skutečně jsou,
- k čemu jsou data (ne)vhodná,
- kdo může data znovu použít,
- jaká konkrétní data podporují vaše výsledky,
- jak je použít pro opakování vašich experimentů atd.

Zdroj: ELIXIR RDMkit, <https://rdmkit.elixir-europe.org/>

MUNI
ICS

FAIR principy

FAIR principy – Findable

The first step in (re)using data is to find them. Metadata and data should be **easy to find** for both **humans** and **computers**. **Machine-readable metadata** are essential for automatic discovery of datasets and services, so this is an essential component of the [FAIRification process](#).

- *F1. (Meta)data are assigned a **globally unique and persistent identifier**.*
- *F2. Data are described with **rich metadata** (defined by R1 below).*
- *F3. **Metadata** clearly and **explicitly include the identifier of the data they describe**.*
- *F4. (Meta)data are **registered or indexed in a searchable resource**.*

Zdroj: GO FAIR, <https://www.go-fair.org/fair-principles/>

FAIR principy – Accessible

Once the user finds the required data, she/he/they need to know how can they be accessed, possibly including authentication and authorisation.

- A1. (Meta)data are *retrievable by their identifier using a standardised communications protocol*.
 - A1.1 The *protocol is open, free, and universally implementable*.
 - A1.2 The *protocol allows for an authentication and authorisation procedure, where necessary*.
- A2. *Metadata are accessible, even when the data are no longer available*.

Zdroj: GO FAIR, <https://www.go-fair.org/fair-principles/>

FAIR principy – Interoperable

The data usually need to be **integrated with other data**. In addition, the data need to **interoperate with applications** or **workflows** for analysis, storage, and processing.

- I1. (Meta)data use a **formal, accessible, shared, and broadly applicable language for knowledge representation**.
- I2. (Meta)data use **vocabularies that follow FAIR principles**.
- I3. (Meta)data **include qualified references to other (meta)data**.

Zdroj: GO FAIR, <https://www.go-fair.org/fair-principles/>

FAIR principy – Reusable

The ultimate goal of FAIR is to optimise the **reuse of data**. To achieve this, metadata and data should be **well-described** so that they can be replicated and/or combined in different settings.

- *R1. (Meta)data are richly **described with a plurality of accurate and relevant attributes**.*
 - *R1.1. (Meta)data are released with a clear and accessible **data usage license**.*
 - *R1.2. (Meta)data are associated with **detailed provenance**.*
 - *R1.3. (Meta)data meet **domain-relevant community standards**.*

Zdroj: GO FAIR, <https://www.go-fair.org/fair-principles/>

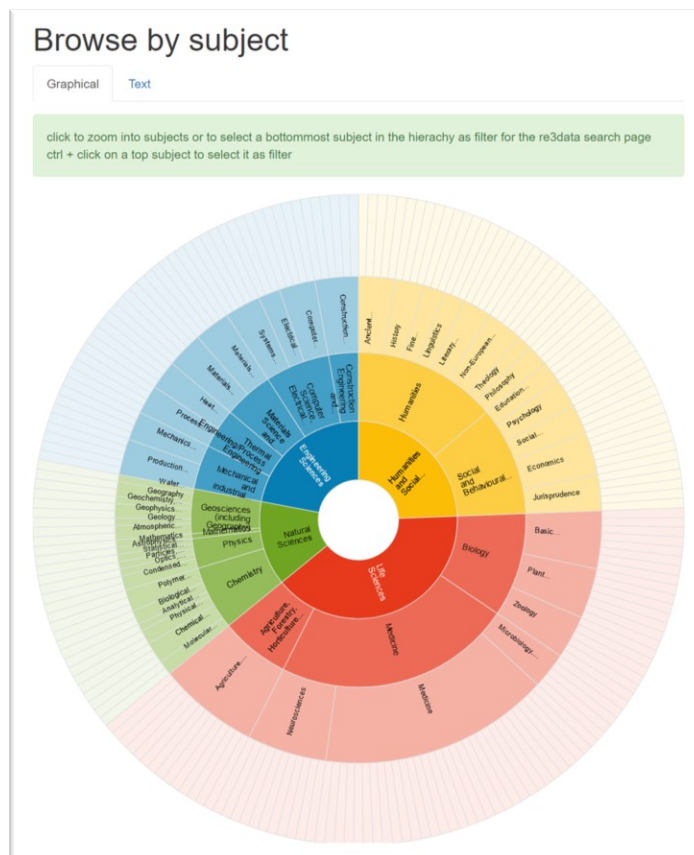
Repozitáře pro výzkumná data

Typy repozitářů

- **Oborové, národní, institucionální, catch-the-all, ...**
- Běžný postup práce s výzkumnými daty je **dělení po oborech, nikoliv po státech či institucích.**
 - Nezohledňujeme národní/institucionální/... specifika, ale oborová specifika.
- Pokud řešíme projektový/institucionální/národní **standard**, musíme být **kompatibilní s mezinárodními oborovými zvyklostmi.**
 - Oborové standardy řeší např. EOSC Task Forces: <https://www.eosc.eu/task-force-faq>
- Pokud chceme shromažďovat **informace napříč obory**, musí to být **formát dostatečně obecný**, který možná přesáhne rámec výzkumných dat.
 - Např. aplikační profil formátu DCAT ([DCAT-AP](#)), který se používá pro popis otevřených dat např. ze státní správy.

Jak najít/vybrat datový repozitář

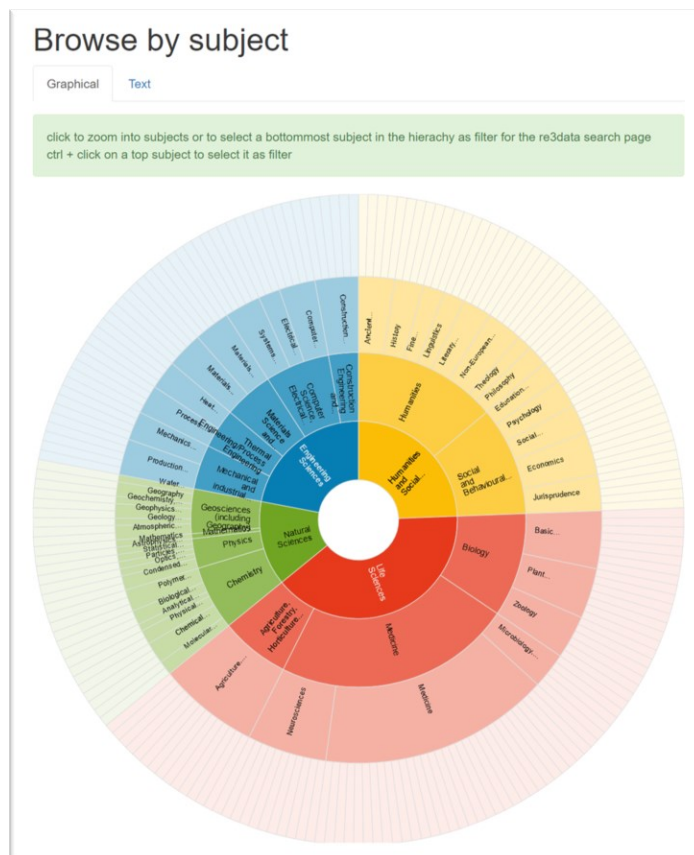
- OpenAIRE: [Jak najít důvěryhodný repozitář pro vaše data](#)
 - Preferovány jsou důvěryhodné certifikované repozitáře.
 - [CoreTrustSeal](#) (seznam certifikovaná repozitářů).
 - [Nestor Seal](#) (verifikace dle DIN 31644).
 - [ISO 16363](#).
 - Např. ale známé [Zenodo](#) žádnou certifikaci nemá...
- Mezi nejpoužívanější obecné repozitáře patří
 - [Zenodo](#),
 - [Figshare](#) nebo
 - [Dryad](#).



Zdroj: <https://www.re3data.org/browse/by-subject/>

Jak najít/vybrat datový repozitář (2)

- Adresáře repozitářů:
 - Open Access repozitáře: [OpenDOAR](#)
 - Datové repozitáře: [re3data.org](#)
- [ELIXIR RDMkit | Data publication](#)
 - [EMBL-EBI's data submission wizard](#).
 - [ELIXIR Deposition Databases](#).
 - [Scientific Data journal's recommended repositories](#).



Zdroj: <https://www.re3data.org/browse/by-subject/>

Na co dát pozor při výběru repozitáře

- Přidělí repozitář vašim datům **perzistentní identifikátor** (např. DOI)?
Díky perzistentním identifikátorům jsou vaše data snáze vyhledatelná a citovatelná.
- Je repozitář **důvěryhodný/certifikovaný**?
U certifikovaných repozitářů máte větší jistotu, že je o vaše data dobře postaráno.
- Poskytuje vybraný repozitář **otevřený přístup** k uloženým datům?
Pokud chcete svá data sdílet otevřeně, pak je toto klíčová informace.
- Opatří repozitář vaše data **licencí**? Uvádí jasné podmínky, za jakých lze data uložená v repozitáři využívat?
Je důležité, aby uživatelé vašich dat věděli, jak s nimi mohou nakládat.
- Poskytne repozitář pro vaše data **úvodní stránku s metadaty** (tzv. landing page)?
Metadata pomohou ostatním data najít, zjistit, co je to za data, a také jak je citovat.
- Umožňuje repozitář **verzování**?
Pokud svůj dataset aktualizujete, můžete aktualizovanou verzi nahrát jako novou verzi k původnímu datasetu. Nový dataset dostane přidělený svůj vlastní perzistentní identifikátor a uživatelé snadno zjistí, jaká je nejnovější verze nebo jaká verze byla použita v původní studii.

Data Management Planing

DMPlanning vs. DMPlan

1. Data Management Planing

„**Proces** plánování, popisu a informování o životním cyklu dat a **činností** spojených s jejich správou v průběhu výzkumu.“

2. Data Management Plan (DMP)

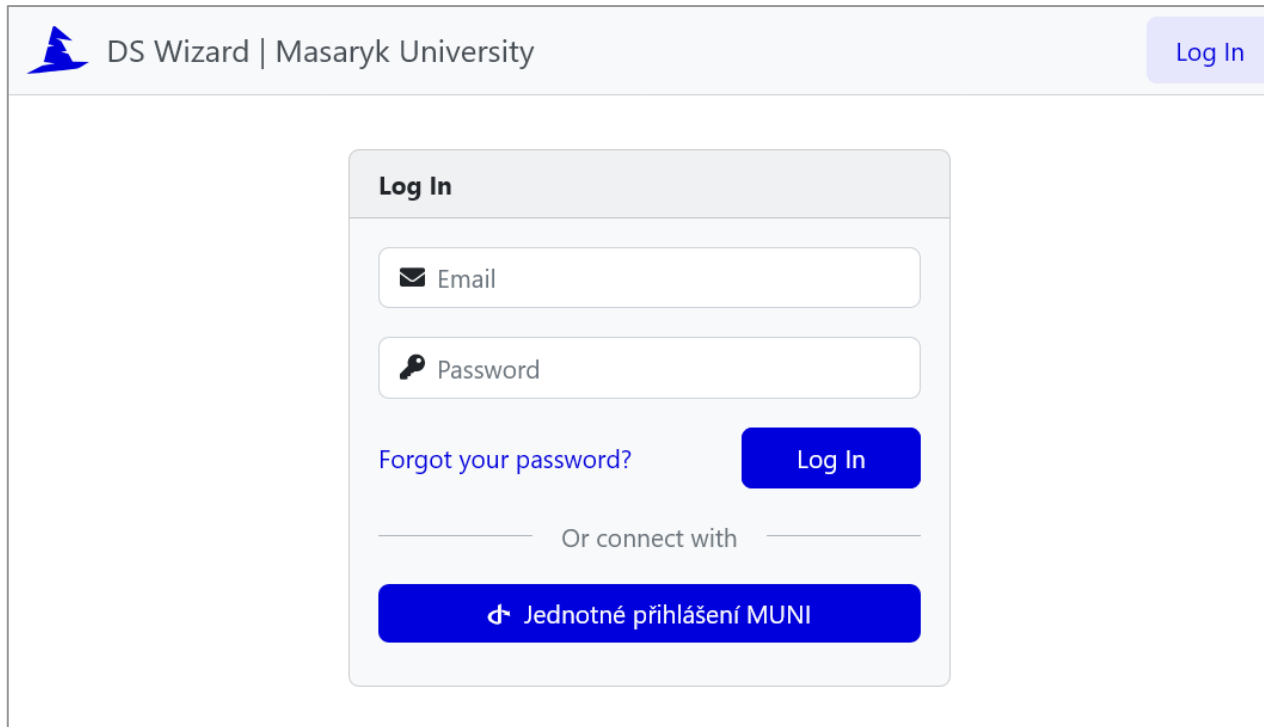
„**Dokument**, který **popisuje** tyto **činnosti** (dokumenty často vyžadují poskytovatelé dotací či grantů).“

Zdroj: Petra Dědičová, Data management a jak psát data management plan
<https://www.slideshare.net/butlibrary/data-management-a-jak-psat-data-management-plan-41441697>

DMP jako projektová podmínka

- DMP je stále častěji **vyžadován** jako standardní **součást** podání/řešení/finalizace **projektů**.
- Např. **Horizon Europe** požaduje
 - iniciální DMP,
 - aktualizace DMP uprostřed projektu,
 - finální revize DMP na konci projektu.
- DMP je **živý dokument** – měl by být průběžně aktualizován po celou dobu projektu.
- Obdobně v **OP JAK**, **Exceles**; nově i v **TAČR** a očekáváme i **GAČR** od roku 2023, nejpozději 2024.

DSW MUNI – naše univerzitní instance DSW



DS Wizard | Masaryk University Log In

Log In

Email

Password

Forgot your password? Log In

Or connect with

↪ Jednotné přihlášení MUNI

<https://dsw.muni.cz/>

[\(návod na používání\)](#)

DSW MUNI – naše univerzitní instance DSW

The screenshot displays the DSW MUNI interface for a project titled "DMP Example: Integration of RNA Biology for Next-Generation Scientists (INTEG-RNA)". The interface is divided into several sections:

- Left Sidebar:** Contains navigation options for "Knowledge Models", "Projects" (with sub-items "List" and "Importers"), and "Document Templates". At the bottom, there is a user profile for "Michal Růžicka, Data Steward" and a "Collapse sidebar" button.
- Top Navigation:** Includes "Questionnaire", "Metrics", "Preview", "Documents", and "Settings". On the right, there are "Warnings", "Comments", "TODOs", and "Version history" indicators.
- Main Content Area:** Shows a list of "Chapters" on the left, with "III. Creating and collecting data" selected. The main area displays three questionnaire questions:
 - 2.a.2:** "Is this a standard data format widely used by researchers in this field?" with options "a. No" and "b. Yes" (selected). A "Desirable: Before Submitting the DMP" checkbox is present. The answer is "b. Yes" with a green "Interoperability" tag. It was answered 5 months ago by Michal Růžicka.
 - 2.a.3:** "Does this data format enable sharing and long term archiving?" with options "a. No" and "b. Yes" (selected). A "Desirable: Before Submitting the Proposal" checkbox is present. The answer is "b. Yes" with a green "Interoperability" tag. It was answered 5 months ago by Michal Růžicka.
 - 2.a.4:** "What volume of data of this type will you be working with?" with options "a. So small that it is not a problem" (selected), "b. I can specify the total amount", and "c. I can specify the number of files/subjects and the size of each". A "Desirable: Before Submitting the Proposal" checkbox is present. It was answered 5 months ago by Michal Růžicka.

PID – perzistentní identifikátory

Perzistentní identifikátory (PID)

- Mají zajistit oddělení identifikace objektu jako takového,
 - osoba,
 - instituce,
 - publikace,
 - dataset,
- od jeho momentálního fyzického umístění.
- Příklad – datová sada *https-set*
 - Identifikátor datové sady: <https://doi.org/10.48791/4mxxp-r725>
 - Současné fyzické umístění: <https://ucnmuni.sharepoint.com/teams/mu-UVT-https-set/Shared%20Documents/Forms/AllItems.aspx?id=%2Fteams%2Fmu%2DUVT%2Dhttps%2Dset%2FShared%20Documents%2Fhttps%2Dset%2Dv1%2E0%2E0&p=true&ga=1>
 - Fyzické umístění se bude pravděpodobně v budoucnu měnit – zvažován přesun do [pilotně provozovaného datového repozitáře CESNET](#).
 - Změny nevadí – uživatelům je jako odkaz na data *vždy* prezentováno [DOI 10.48791/4mxxp-r725](https://doi.org/10.48791/4mxxp-r725), které je vždy zavede na aktuální umístění.

Perzistentní identifikátory (PID) (2)

- Mají zajistit **jednoznačnost**.
- **Příklad** – jména fyzických osob
 - Více forem zápisu jména **jedné fyzické osoby**.
 - Více různých fyzických **osob** se **stejným jménem**.
- Mají zajistit **trvalost** (perzistenci).
 - **Metadata** fyzicky umístěna **u třetí strany** nezávisle na fyzickém umístění odkazované entity.
 - Vlastník identifikátoru se stará o **aktualizaci metadat** a aktualizace **směřování** na aktuální umístění.
 - Třetí strana **pečuje** o **zachování** poslední verze a historie **metadat** a **existenci identifikátoru** jako takového, i pokud se vlastník identifikátoru o něj starat přestane. A dokonce i tehdy, pokud identifikovaná entita jako taková nebude zachována.

<input type="checkbox"/>	NovakD (1)	Novák, David (1)	Join	Delete
<input type="checkbox"/>	NovakE (5)	Nováková, Eva (5) Nováková, E. (0)	Join	Delete
<input type="checkbox"/>	NovakJ (52)	Novák, Josef (38) Novák, Jos. (1) <u>Novák, J. (13)</u>	Join	Delete
<input type="checkbox"/>	NovakJ2 (7)	Novák, Jiří (7) Novak, Jiri (0)	Join	Delete
<input type="checkbox"/>	NovakJ7 (19)	Novák, Josef (16) <u>Novák, J. (3)</u>	Join	Delete
<input type="checkbox"/>	NovakK (2)	Novák, Karel (2)	Join	Delete
<input type="checkbox"/>	NovakM (2)	Novák, Mirko (2) Novak, Miroslav M. (0) Novak, M. M. (0) Novák, M. (0)	Join	Delete
<input type="checkbox"/>	NovakM2 (2)	Nováková, Markéta (2)	Join	Delete
<input type="checkbox"/>	NovakM3 (1)	Novák, Miroslav (1)	Join	Delete
<input type="checkbox"/>	NovakO (2)	Novák, Ondřej (2)	Join	Delete
<input type="checkbox"/>	NovakP (1)	Novák, Petr (1)	Join	Delete
<input type="checkbox"/>	NovakS (1)	Novák, Stanislav (1)	Join	Delete
<input type="checkbox"/>	NovakV (57)	Novák, Vítězslav (55) Novák, V. (2)	Join	Delete
<input type="checkbox"/>	NovakV2 (7)	Novák, Vilém (7) Novák, V. (0)	Join	Delete
<input type="checkbox"/>	NovakV3 (53)	Novák, Vladimír (53)	Join	Delete
<input type="checkbox"/>	NovakZ (1)	Novák, Zdeněk (1)	Join	Delete

Zdroj: Autoritní databáze projektu [DML-CZ](#)

Řada typů PID

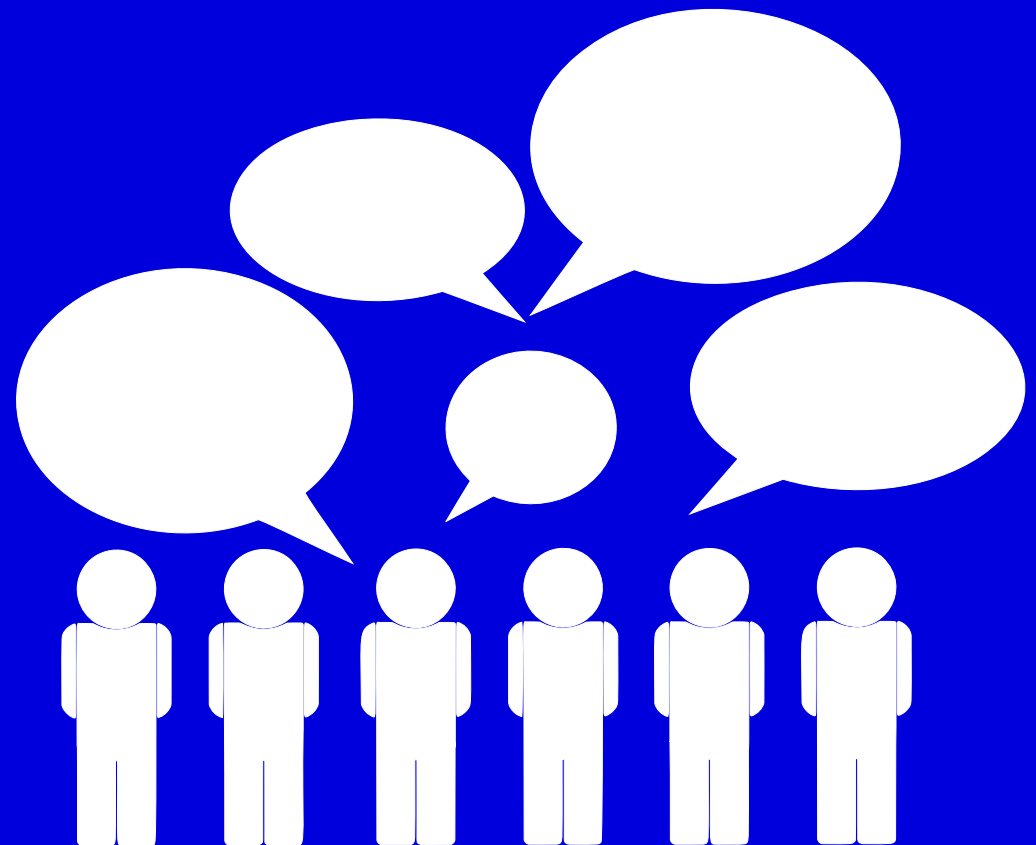
CODEVALUE	PREFLABEL_EN	HREF	Maturity	Globally resolvable	For which object type	Comments
ADSbibcode	Astrophysics Data System - Bibliographic Reference Code	https://ui.adsabs.harvard.edu/	High	Needs token	Publication	
ARK arXiv	Archival Resource Key arXiv identifier scheme	https://arks.org/ https://arxiv.org/	High High	Local Yes	Other (in comments) Publication	Everything
ASIN	Amazon Standard Identification Number	https://sellercentral.amazon.ca/gp/help/external/200317470?language=en-CA&ref=mpbc_200576730_cont_200317470 https://indico.cern.ch/event/780651/attachments/1776614/2888642/Conference_PIDs_and_Crossmark.pdf	High		Other (in comments)	Things sell by Amazon
ConfID Crossref DOI	Conference identifier			?	Event Publication	Not clear if it is a Crossref service
Crossref_funders Crossref_grants	Crossref Funder Registry Registering research grants	https://www.crossref.org/services/content-registration/grants/ https://www.crossref.org/community/grants/	?	Yes	Organisation Other (in comments)	Grants
DataCite DOI						
DOI	Digital Object Identifier	https://www.doi.org/	High High	Yes	Other (in comments) Publication	28 different resources and outputs Services supporting PIDs and metadata for 40+ resource and output types
EAN13	The 13-digit International Article Number	https://www.gs1.org/standards/barcodes/ean-upc	High	?	Other (in comments)	Physical product identifier. A Whole famili of id: UPC-A, UPC-E, EAN13, EAN8
eISBN	electronic International Standard Book Number	https://www.isbn-international.org/			Publication	
eISSN	Electronic International Standard Serial Number	http://portal.issn.org/	High		https://portal.issn.org/resource/ISSN/0376-4583/	Publication
GRID	Global Research Identifier Database	https://www.grid.ac/	Closed	?	Organisation	Transitioned to ROR
Handle ISN	Handle	http://www.handle.net/	High	Yes	Dataset	It is the base of DOI also
ISAN	International Geo Sample Number	https://www.igsn.org/	High	Yes	Other (in comments)	Physical Samples and Sampling Features
ISAN	International Standard Audiovisual Number	https://www.isan.org/		Yes?	Publication	
ISBN	International Standard Book Number	https://www.isbn-international.org/	High		Publication	
ISLI	Identifies the links between different entities	https://www.isbn-international.org/content/isli-introduction	?	Yes	Other (in comments)	Link bw. entities
ISMN	International Standard Music Number	https://www.ismn-international.org/		No	Publication	
ISNI	International Standard Name Identifier	https://isni.org/page/search-database/			Person	Contributors to creative works and their distribution and organizations
ISSN	International Standard Serial Number	http://portal.issn.org/	High		https://portal.issn.org/resource/ISSN/0376-4583/	Publication
ISTC	The International Standard Text Code	http://www.istc-international.org/	Defunct	No	Publication	Identifies various types of serial publications (eg. journals, websites, blogs) Ceased in 2017
LSID	Life Sciences Identifier	http://www.lsid.info/	?	Yes	Other (in comments)	Metadata for life science items
ORCID	Open Researcher and Contributor ID	https://orcid.org/	High	Yes	Person	
PIC	(EC) partner identity code	https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq/1055				
PMID	PubMed ID	https://www.ncbi.nlm.nih.gov/pmc/pmctopmid/			Publication	
PURL	persistent uniform resource locator	https://archive.org/services/purl/	High	Yes	Other (in comments)	Resources on the Web
QID	Wikidata identifier	https://www.wikidata.org/wiki/Wikidata:Identifiers			Other (in comments)	Knowledge item
RAID	Persistent Identifier for research projects Unique numerical identifier applied to organizations in the scholarly supply chain	https://www.raid.org.au/		Needs token	Other (in comments)	Research projects
Ringgold ROR RRID	Research Organization Registry Research Resource Identifier	https://www.ringgold.com/ https://ror.org/ https://scicrunch.org/resources	High	Yes	Organisation Organisation	
ScopusAuthorID SWHID	Scopus Author ID SoftWare Heritage persistent Identifiers	https://service.elsevier.com/app/answers/detail/a_id/11212/supporthub/scopus/ https://docs.softwareheritage.org/devrel/swh-model/persistent-identifiers.html	?	No? Local	Publication Source Code	
UPC	Universal Product Code	https://www.gs1.org/standards/barcodes/ean-upc			Other (in comments)	Synonym of EAN13? Product identifier
URI	Uniform Resource Identifier					
URL	Uniform Resource Locator					
URN	Uniform Resource Name					
VAT-number	VAT number	http://ec.europa.eu/taxation_customs/vies/vatRequest.html			Organisation	

Zdroj: Interní pracovní materiály [EOSC Task Force PID Policy and Implementation](#)

Populární PID

- Osoby
 - ORCID: <https://orcid.org/>
- Instituce
 - ROR: <https://ror.org/>
- Publikace
 - DOI: <https://www.crossref.org/>
- Datasetsy
 - DOI: <https://datacite.org/>
 - Handle: <https://handle.net/>
- EOSC publikace k PID:
 - A Persistent Identifier (PID) policy for the European Open Science Cloud (EOSC):
<https://doi.org/10.2777/926037>
 - PID architecture for the EOSC : report from the EOSC Executive Board Working Group (WG) Architecture PID Task Force (TF):
<https://doi.org/10.2777/525581>

Otázky?



Zdroj: [Communicate_communication_conference_2028004](#) od [OpenClipart-Vectors](#) z [Pixabay](#)

FAIR/OPEN DATA
PRO UŽIVATELE
CF MAFIL

o M U N I

2023-02-22



EVROPSKÁ UNIE
Evropské strukturální a investiční fondy
Operační program Výzkum, vývoj a vzdělávání



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY