

Osnove ravnanja z raziskovalnimi podatki v skladu z načeli FAIR

doc. dr. Ana Slavec (Fakulteta za matematiko, naravoslovje in informacijske tehnologije Univerze na Primorskem in InnoRenew CoE)

Naravoslovnotehniško fakulteta Univerze v Ljubljani, **3. 6. 2024**



Ozadje

6 ukrepov [akcijskega načrta za odprto znanost](#)

1. Mednarodno skladen nacionalni ekosistem odprte znanosti
2. Vrednotenje znanstvenoraziskovalne dejavnosti v skladu z načeli OZ
3. Skladnost rezultatov znanstvenih raziskav z načeli FAIR in odprtost
4. Nacionalna skupnost za odprto znanost
5. Občanska znanost
6. Odprto dostopno nacionalno znanstveno založništvo



slido



Kakšne vrste digitalnih vsebin ustvarjate v svojem raziskovanju?

ⓘ Start presenting to display the poll results on this slide.

slido



Katere formate uporabljate za shranjevanje svojih podatkov?

ⓘ Start presenting to display the poll results on this slide.

slido



Kje shranjujete svoje raziskovalne podatke?

ⓘ Start presenting to display the poll results on this slide.

Kaj so raziskovalni podatki?

Definicija OECD: „Research data are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated. This term does not cover the following: laboratory notebooks, preliminary analyses, and drafts of scientific papers, plans for future research, peer reviews, or personal communications with colleagues or physical objects (e.g. laboratory samples, strains of bacteria and test animals such as mice).“ ([OECD Principles and Guidelines for Access to Research Data from Public Funding](#))

Slovenski prevod (na dirrosdata.ctk.uni-lj.si): „Raziskovalni podatki so definirani kot stvarni zapisi (numerični rezultati, besedilni zapisi, slikovno in zvočno gradivo), ki se uporabljajo kot primarni viri za namene znanstvenih raziskav in so v znanstveni skupnosti splošno sprejeti kot nujni za potrditev raziskovalnih izsledkov. Nabor raziskovalnih podatkov predstavlja sistematično, delno predstavitev raziskovalne tematike. Ta pojem ne vključuje: laboratorijskih dnevnikov, preliminarne analiz, osnutkov znanstvenih člankov, načrtov bodočih raziskav, strokovnih recenzij, komunikacije s kolegi in predmetov (npr. laboratorijskih vzorcev, bakterijskih sevov in testnih živali, kot so miši).“

Viri raziskovalnih podatkov

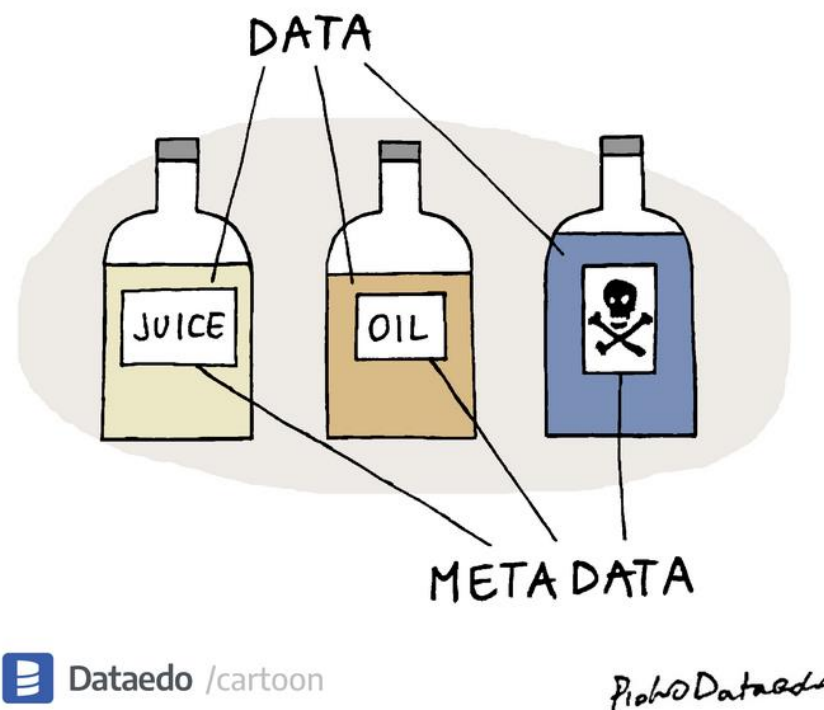
- **Opazovani podatki** so zajeti na podlagi opazovanja, npr. senzorski podatki, anketni podatki, nevrološke slike.
- **Eksperimentalni podatki** nastanejo z aktivnim posegom raziskovalca in običajno vključujejo rabo določene opreme. Npr. zaporedja genov, kromatogrami, podatki o magnetnih poljih.
- **Simulacijski podatki** so ustvarjeni iz testnih modelov. Npr. klimatski modeli, ekonomski modeli.
- **Izpeljani ali sestavljeni podatki** so bi bili preoblikovani iz že obstoječih podatkovnih točk. Npr. sestavljene baze podatkov, podatkovno rudarjenje, 3D modeli.
- **Referenčni ali kanonični podatki** so obsežne zbirke objavljenih zbirk podatkov. Npr. kemijske strukture, portali prostorskih podatkov.



Metapodatki

Metadata so „podatki o podatkih“ in običajno vključujejo naslednje (npr. [DataCite Metadata Schema](#)):

- Trajne identifikatorje (PID)
- Avtorje
- Podatke o objavi
- Naslov
- Opis podatkov
- Licence
- Financiranje
- ...



Dataedo /cartoon

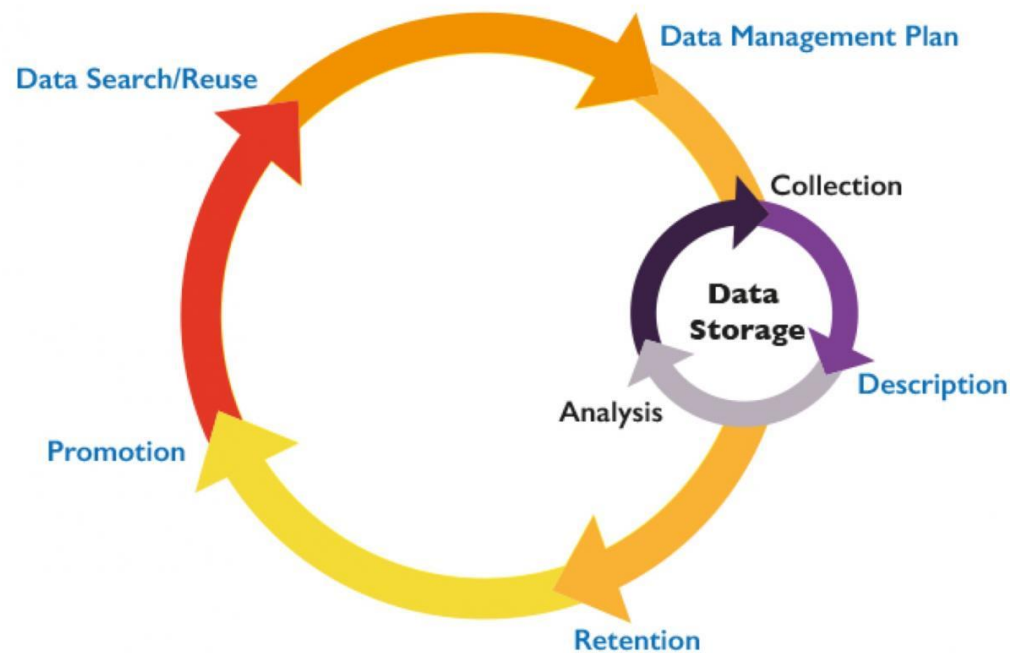
Piotr Dataedo

Vir: Piotr Kononow ([Dataedo.com](#))



Življenjski krog ravnanja z raziskovalnimi podatki

The Research Data Management Lifecycle



- Iskanje in ponovna raba podatkov
- Načrt ravnanja z raziskovalnimi podatki
- Opis podatkov
- Arhiviranje podatkov
- Objava podatkov

Vir: University of California, Santa Cruz, [Data Mangement LibGuide](#)

Načrti ravnanja z raziskovalnimi podatki (NRRP)

- Načrt ravnanja z raziskovalnimi podatki (angl. Data Management Plan, DMP) je formalni dokument, ki opisuje ravnanje s podatki med projektom in po zaključku projekta. Običajno vključuje naslednje elemente:
 - Opis podatkov, ki jih je treba zbrati oziroma ustvariti
 - Standardi/metodologije za zbiranje podatkov in ravnanje z njimi
 - Etika in pravice intelektualne lastnine
 - Načrti za deljenje podatkov z drugimi
 - Strategija dolgoročne hrambe
- Primeri načrtov ravnanja z raziskovalnimi podatki:
 - Sergan, V. 2014. [Orientational order induced by a polymer network in the isotropic phase of liquid crystal.](#)
 - Hong, X. 2016. [Effects of mine waste materials in the north-central Mojave dessert.](#)
 - Dahle, S. et. al. 2019. [Data Management Plan for the MSCA project PlasmaSolution.](#)

Zakaj je ravnanje z raziskovalnimi podatki pomembno?

Učinkovitost

- (Dolgoročen) prihranek časa in sredstev.

Integriteta

- Ravnanje z raziskovalnimi podatki je del odgovornega izvajanja raziskav.

Kakovost

- Dobro ravnanje s podatki je pogoj, da lahko drugi preverijo in ponovijo ugotovitve naših raziskav.

Izzivi pri vpeljevanju ravnanja z raziskovalnimi podatki v skladu z načeli FAIR

Družbene omejitve

- Strah pred zlorabo in napačnim razumevanjem podatkov
 - Strah pred izgubo konkurenčne prednosti
 - Omejitve zaradi sodelovanja z industrijo
 - Pomanjkanje osveščenosti in znanja

Tehnične omejitve

- Na mnogih področjih manjkajo standardi in ontologije
 - Pomanjkanje področno-specifičnih repozitorijev



Smernice okvirnega programa Obzorje

***„Toliko odprto, kot je mogoče,
toliko zaprto, kot je potrebno“***



*(angl. as open as possible,
as closed as necessary)*

Odprti podatki

- **Odprti podatki** se lahko prosto uporabljajo, spreminjajo in delijo s komerkoli za katerikoli namen.
- **Pravno odprti**, tj. na voljo pod licenco za odprte podatke, ki lahko zahteva zagotavljanje pripisa avtorstva (BY) in/ali delitev pod enakimi pogoji (SA).
- **Tehnično odprti**, tj. na voljo za strošek njihove reprodukcije in so strojno berljivi (glej priročnik [Open Data Handbook](#))



Načela FAIR

- Wilkinson idr. 2016: [The FAIR Guiding Principles for scientific data management and stewardship](#).

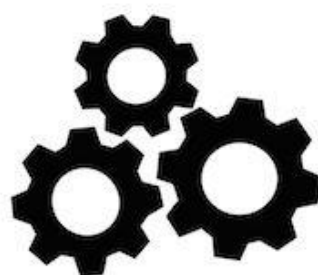
F_{indable}



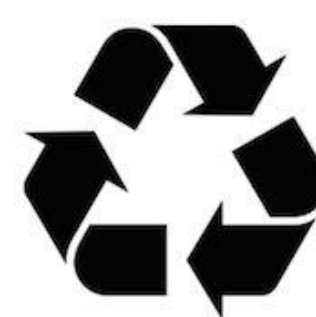
A_{ccessible}



I_{nteroperable}



R_{eusable}



Findability (Najdljivost)



F1. (Meta)podatkom se dodeli globalno edinstven trajni identifikator.

F2. Podatki so opisani z bogatimi metapodatki.

F3. (Meta)podatki jasno in izrecno vključujejo identifikatorje podatkov, ki jih opisujejo.

F4. (Meta)podatki so registrirani ali indeksirani v iskalnem viru.

Accessibility (Dostopnost)



A1. Na podlagi njihovega identifikatorja je do (meta)podatkov možno dostopati z uporabo standardiziranega komunikacijskega protokola.

A1.1. Protokol je odprt, brezplačen in splošno izvedljiv.

A1.2. Protokol po potrebi omogoča postopek preverjanja prisotnosti in avtorizacije.

A2. (Meta)podatki so dostopni tudi, ko podatki niso (več) na voljo.

Interoperability (Interoperabilnost)



- I1. (Meta)podatki za reprezentacijo znanja uporabljajo formalen, dostopen in široko uporabljan jezik.
- I2. (Meta)podatki uporabljajo slovarje, ki sledijo načelom FAIR.
- I3. (Meta)podatki vključujejo kvalificirane reference na druge (meta)podatke.

Reusability (Ponovna uporabnost)



R1. (Meta)podatki so bogato opisani z množico natančnih in ustreznih lastnosti.

R2. (Meta)podatki se izdajo z jasno in dostopno licenco za uporabo podatkov.

R3. (Meta)podatki so povezani s podrobnim poreklom.

R4. (Meta)podatki ustrezajo skupnostnim standardom relevantnega področja.

FAIR ni binaren koncept, ampak spektrum

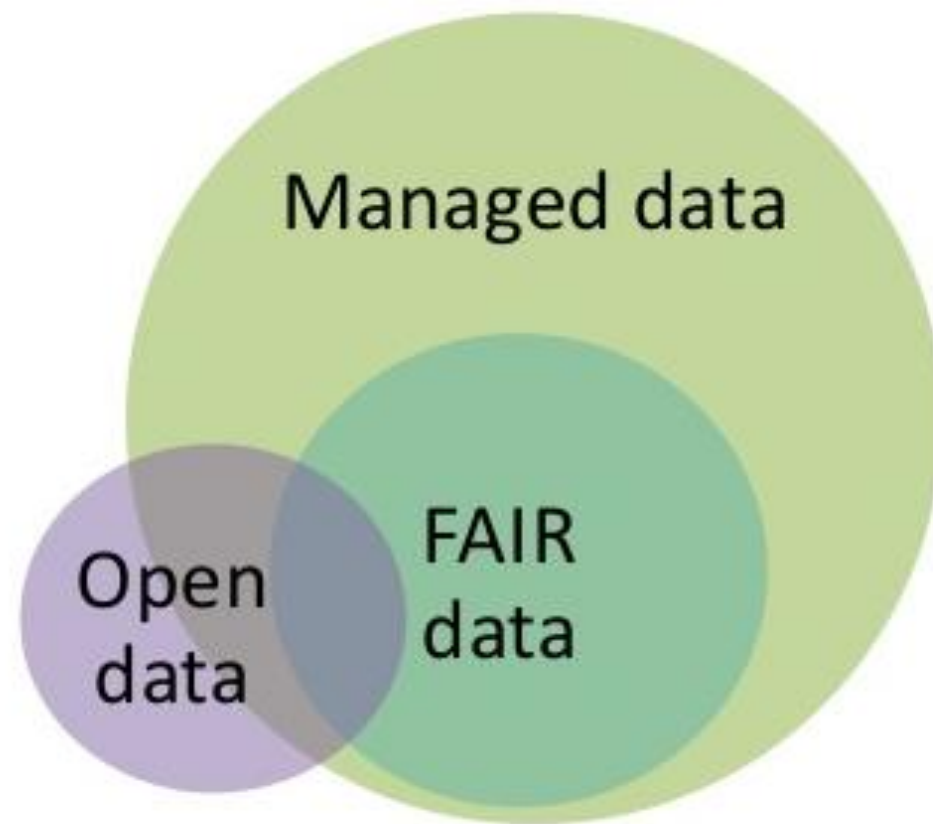


Foto: [Marina Noordegraaf \(Flickr\)](#)



Ravnanje s podatki v kontekstu odprte znanosti

Načela FAIR ≠ Odprti podatki



Jones, S. 2018. Open data, FAIR data and RDM: the ugly duckling. Available at: [Zenodo](https://zenodo.org/record/2572211).

Kje lahko delimo raziskovalne podatke?

- Splošni repozitoriji



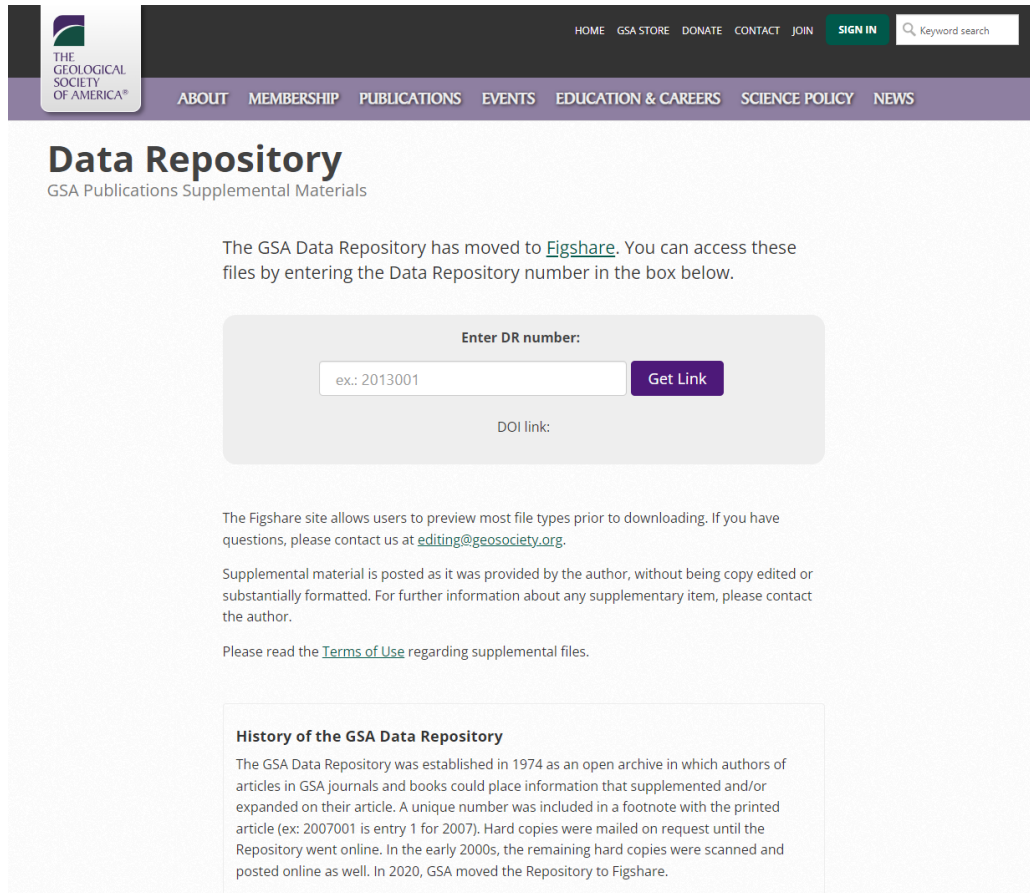
- Institucionalni repozitoriji



- Področno-specifični repozitoriji



Primeri področno-specifičnih repozitorijev



The screenshot shows the homepage of the GSA Data Repository. At the top, there is a navigation bar with links for HOME, GSA STORE, DONATE, CONTACT, JOIN, and SIGN IN, along with a keyword search box. Below this is a secondary navigation bar with links for ABOUT, MEMBERSHIP, PUBLICATIONS, EVENTS, EDUCATION & CAREERS, SCIENCE POLICY, and NEWS. The main heading is "Data Repository" with the subtitle "GSA Publications Supplemental Materials". A text block states: "The GSA Data Repository has moved to [Figshare](#). You can access these files by entering the Data Repository number in the box below." Below this is a form with the label "Enter DR number:" and an input field containing "ex.: 2013001" and a "Get Link" button. Below the form is a "DOI link:" label. A paragraph of text explains that the Figshare site allows users to preview most file types prior to downloading and provides contact information for questions. Another paragraph states that supplemental material is posted as it was provided by the author, without being copy edited or substantially formatted. A final paragraph directs users to read the [Terms of Use](#) regarding supplemental files. At the bottom, there is a section titled "History of the GSA Data Repository" which provides a brief history of the repository's establishment in 1974 and its move to Figshare in 2020.

Data Repository
GSA Publications Supplemental Materials

The GSA Data Repository has moved to [Figshare](#). You can access these files by entering the Data Repository number in the box below.

Enter DR number:
ex.: 2013001 [Get Link](#)

DOI link:

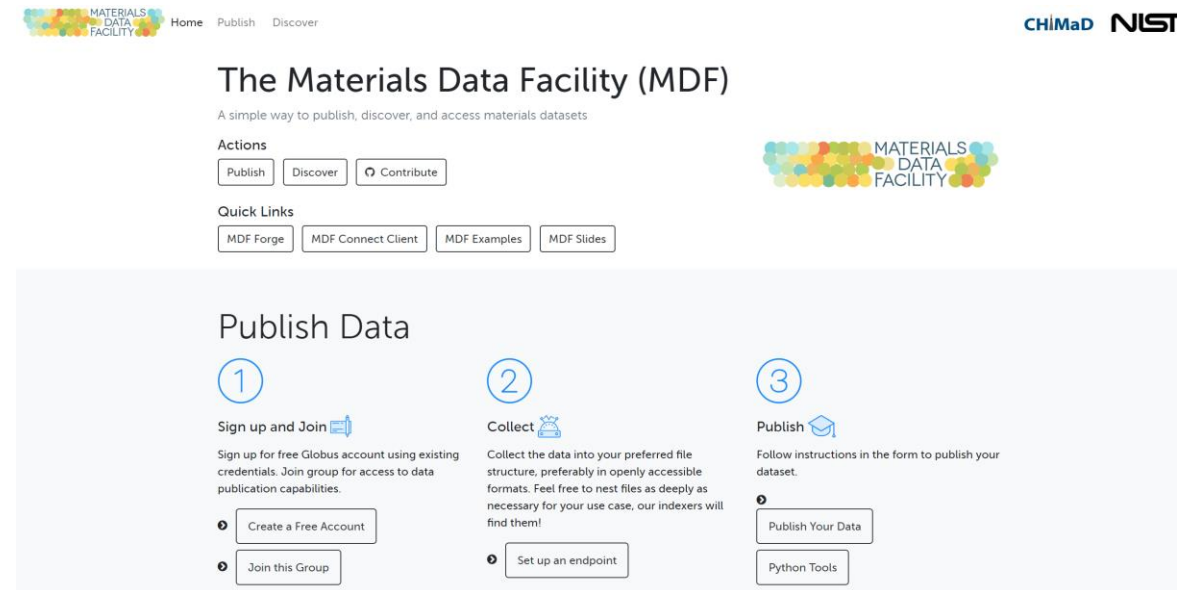
The Figshare site allows users to preview most file types prior to downloading. If you have questions, please contact us at editing@geosociety.org.

Supplemental material is posted as it was provided by the author, without being copy edited or substantially formatted. For further information about any supplementary item, please contact the author.

Please read the [Terms of Use](#) regarding supplemental files.

History of the GSA Data Repository

The GSA Data Repository was established in 1974 as an open archive in which authors of articles in GSA journals and books could place information that supplemented and/or expanded on their article. A unique number was included in a footnote with the printed article (ex: 2007001 is entry 1 for 2007). Hard copies were mailed on request until the Repository went online. In the early 2000s, the remaining hard copies were scanned and posted online as well. In 2020, GSA moved the Repository to Figshare.



The screenshot shows the homepage of the Materials Data Facility (MDF). At the top, there is a navigation bar with links for Home, Publish, and Discover, along with a keyword search box. Below this is a secondary navigation bar with links for ABOUT, MEMBERSHIP, PUBLICATIONS, EVENTS, EDUCATION & CAREERS, SCIENCE POLICY, and NEWS. The main heading is "The Materials Data Facility (MDF)" with the subtitle "A simple way to publish, discover, and access materials datasets". Below this is a section titled "Actions" with buttons for Publish, Discover, and Contribute. Below this is a section titled "Quick Links" with buttons for MDF Forge, MDF Connect Client, MDF Examples, and MDF Slides. Below this is a section titled "Publish Data" with three numbered steps: 1. Sign up and Join (with a button for "Create a Free Account" and "Join this Group"), 2. Collect (with a button for "Set up an endpoint"), and 3. Publish (with a button for "Publish Your Data" and "Python Tools").

The Materials Data Facility (MDF)
A simple way to publish, discover, and access materials datasets

Actions
[Publish](#) [Discover](#) [Contribute](#)

Quick Links
[MDF Forge](#) [MDF Connect Client](#) [MDF Examples](#) [MDF Slides](#)

Publish Data

- 1 Sign up and Join**
Sign up for free Globus account using existing credentials. Join group for access to data publication capabilities.
[Create a Free Account](#)
[Join this Group](#)
- 2 Collect**
Collect the data into your preferred file structure, preferably in openly accessible formats. Feel free to nest files as deeply as necessary for your use case, our indexers will find them!
[Set up an endpoint](#)
- 3 Publish**
Follow instructions in the form to publish your dataset.
[Publish Your Data](#)
[Python Tools](#)

Register repozitorijev:

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Priporočljivi formati

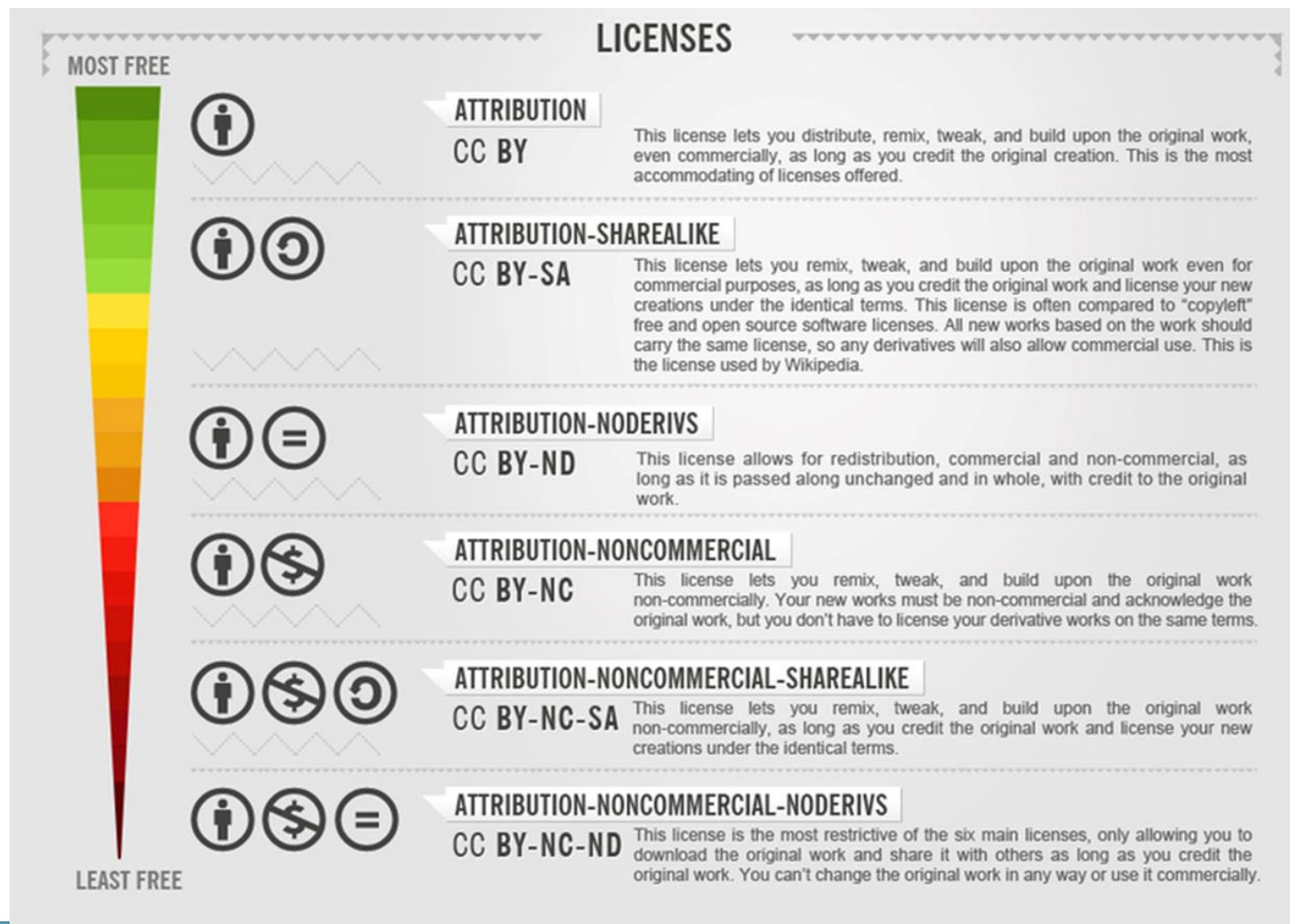
Vrsta podatkov	Priporočljivi formati	Sprejemljivi formati
Tabelarni podatki	csv, .tab, .por, .xml	.txt, xls, .dbf, .ods, .sav, .dta, .mdb
Geoprostorski podatki	.shp, .shx, .dbf, .prj, .sbx, .sbn, .tif, .tfw, .dwg, .gml	.mdb, .mif, .kml, .ai, dxf, .svg
Besedilni podatki	.rtf, .txt, .xml	.html, .doc
Slikovni podatki	.tif	.jpg, .gif, .tif, .tiff, .raw, .psd, .bmp, .png, .pdf
Zvočni podatki	.flac	.mp3, .aif, .wav
Video podatki	.mp4, .ogv, .ogg, .mj2	.avchd
Dokumentacija in skripte	.rtf, .pdf, .xhtml, .htm, .odt	.txt, .doc, .xls, .xml

Vir: [UK Data Service](#)



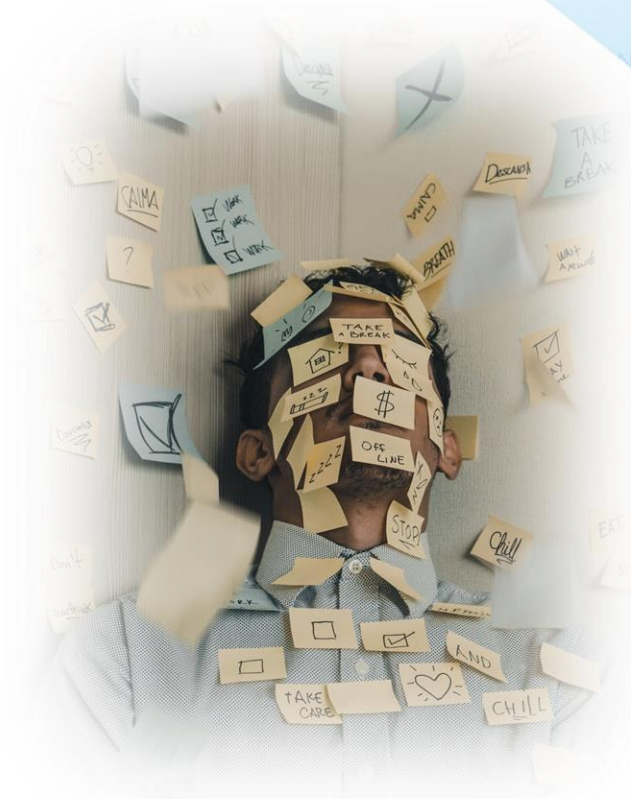
Vir: [Foter.com](https://foter.com)

Odprto licenciranje



Odgovornost za ravnanje z raziskovalnimi podatki

- Pogosto je potrebno strokovno znanje
- **Podatkovni skrbniki (angl. data steward)** raziskovalcem svetujejo, jih podpirajo in jih usposablajo za ravnanje z raziskovalnimi podatki.
- Primeri dobrih praks
 - [Data Stewards and Champions at TU Delft](#)
 - [Data Champions at University of Cambridge](#)
 - [Data Agents and Advisor at Aalto University](#)
 - [Data Stewards at the University of Vienna](#)



Organizacije in projekti s področja ravnanja z raziskovalnimi podatki

- [Research Data Alliance](#)
- [CODATA](#)
- [Digital Curation Centre](#)
- [UK Data Service](#)
- [FAIRsFAIR](#)
- [GO FAIR project](#)
- [FOSTER project](#)
- [Open Science MOOC](#)
- [European Open Science Cloud](#)
- ...



Engineering & Technology

Share    

Composed by: [Marek Cebecauer](#)  (RDA/EOSC Future Ambassador for Materials Sciences and Engineering )

Contributors: List TBA

Comments requested: Please note that this is a new Discipline page, and it is open for comments from the RDA Community. To add your input please use the comments section below.

Downloadable disciplinary info sheet



Materials Sciences and Engineering

Download 

Data sharing in Materials Sciences and Engineering

The Materials Sciences and Engineering (MSE) domain covers a highly heterogeneous community of scientists. These include experimentalists and computational scientists. It comprises also commercial subjects and manufacturing. Open sharing of data is still in its early phase (childhood) in this domain, but several resources and tools have been created to facilitate this transition. A good example are community repositories such as [NOMAD](#) , [DICE](#) , and [Materials Commons](#)  (to name just a few). These sites enable unrestrained data sharing, independent of whether data are open or with a reasonable access restriction.

Material scientists are also very active in the development of electronic laboratory notebooks (ELN) and laboratory information management systems (LIMS) for automatic data and metadata acquisition. [NOMAD Oasis](#) , [KaDi4Mat](#)  and [NexusLIMS](#)  were developed by the MSE research groups. However, their application is not always domain-specific (e.g., KaDi4Mat). Several ontologies are being developed to improve the interoperability of shared data. Among these, [BFO](#) , [EMMO](#) , and [Metadata4Ing](#)  are upper- or medium-level ontologies covering a broader spectrum of disciplines and administrative/interoperability metadata. Others are more specialized, lower-level ontologies for more detailed, disciplined description of data (e.g., [Materials Design Ontology – MDO](#)  or [OM2](#)  for units of measure). Moreover, selected repositories and data resources prefer the use of project vocabularies ([DICE](#)) or knowledge graphs ([openAIRE/Graph](#) ) to facilitate engagement of researchers in data sharing and re-use. All these activities help to implement FAIR principles and Open Science at the increasing number of MSE-oriented universities or other research performing organization. It also leads to digitalization of several inherently non-digital objects, enabling machine-readability. Thus, the outputs can be further studied by the global scientific community or with the help of AI/ML approaches.



From beams to bytes: navigating data manag... 



RESEARCH DATA ALLIANCE

O ravnanju z raziskovalnimi podatki v Sloveniji



Hvala za vašo pozornost.

ana.slavec@famnit.upr.si



Univerza na Primorskem
Università del Litorale
University of Primorska

