

# FAIR principi i upravljanje podacima

National Capacity Building NI4OS Training

Vladimir Risojević  
Elektrotehnički fakultet  
Univerzitet u Banjoj Luci



Zašto mi je ovo bitno?



- Preporuka Komisije (EU) 2018/790 o pristupu i čuvanju naučnih informacija
  - (...) primarni podaci nastali tokom istraživanja finansiranih javnim sredstvima treba da postanu i ostanu vidljivi, dostupni, interoperabilni i upotrebljivi i u druge svrhe (načela FAIR – findable, accessible, interoperable, reusable ) unutar sigurnog i pouzdanog okruženja, posredstvom digitalne infrastrukture (uključujući i infrastrukturu Evropskog oblaka za otvorenu nauku – EOSC) (...)

- Reakcije na Platformu za otvorenu nauku MPNTR Republike Srbije:
  - Kako može da se traži da omogućim uvid u svoje neobjavljene podatke, kada postoji opasnost da ih neko drugi iskoristi, objavi itd?
  - Na osnovu čega ću pisati rad i ko će mi takav rad prihvatiti za objavljivanje?
  - Kako smemo da objavimo podatke, kada tu ima i ličnih informacija o ispitanicima?
  - U mojoj oblasti istraživanja ne postoje podaci.
  - ...

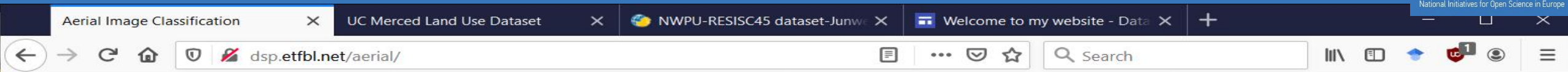
Da li koristite podatke u istraživanju?



Kako pronalazite podatke?







## Aerial Image Classification

The amount of remote sensed imagery that has become available by far surpasses the possibility of manual analysis. One of the most important tasks in the analysis of remote sensed images is land use classification. This task can be recast as semantic classification of remote sensed images.

### Database of RGB+NIR aerial images

[Database](#) of aerial images with both RGB and NIR versions.

### Database of RGB aerial images

The database consists of 606 RGB aerial images of size 128\*128 pixels. They have been obtained by partitioning a larger (4500\*6000 pixels) ortophoto image of the part of Banja Luka, Bosnia and Herzegovina. In this image there is a variety of structures, both man-made, such as buildings, factories, and warehouses, as well as natural, such as fields, trees and rivers.

We manually classified all images into 6 categories. Examples of images from each category are shown in the figure, from left to right: houses, cemetery, industry, field, river, and trees.





[Home](#)[Recent news](#)[Saliency Detection Based  
On Deep Learning](#)[Publications](#)[Group members](#)[Co-saliency database:  
Cosal2015](#)[NWPU VHR-10 dataset](#)[NWPU-RESISC45  
dataset](#)

## NWPU-RESISC45 dataset

NWPU-RESISC45 dataset is a publicly available benchmark for REmote Sensing Image Scene Classification (RESISC), created by Northwestern Polytechnical University (NWPU). This dataset contains 31,500 images, covering 45 scene classes with 700 images in each class. These 45 scene classes include airplane, airport, baseball diamond, basketball court, beach, bridge, chaparral, church, circular farmland, cloud, commercial area, dense residential, desert, forest, freeway, golf course, ground track field, harbor, industrial area, intersection, island, lake, meadow, medium residential, mobile home park, mountain, overpass, palace, parking lot, railway, railway station, rectangular farmland, river, roundabout, runway, seaice, ship, snowberg, sparse residential, stadium, storage tank, tennis court, terrace, thermal power station, and wetland.

Please cite the following paper when publishing results that use this dataset fully or partly:

G. Cheng, J. Han, X. Lu. Remote Sensing Image Scene Classification: Benchmark and State of the Art. Proceedings of the IEEE.

This dataset can be downloaded from [OneDrive \(https://1drv.ms/u/s!AmgKYzARBl5ca3HNaHilzp\\_IXjs\)](https://1drv.ms/u/s!AmgKYzARBl5ca3HNaHilzp_IXjs) or [BaiduWangpan \(http://pan.baidu.com/s/1mifR6tU\)](http://pan.baidu.com/s/1mifR6tU).

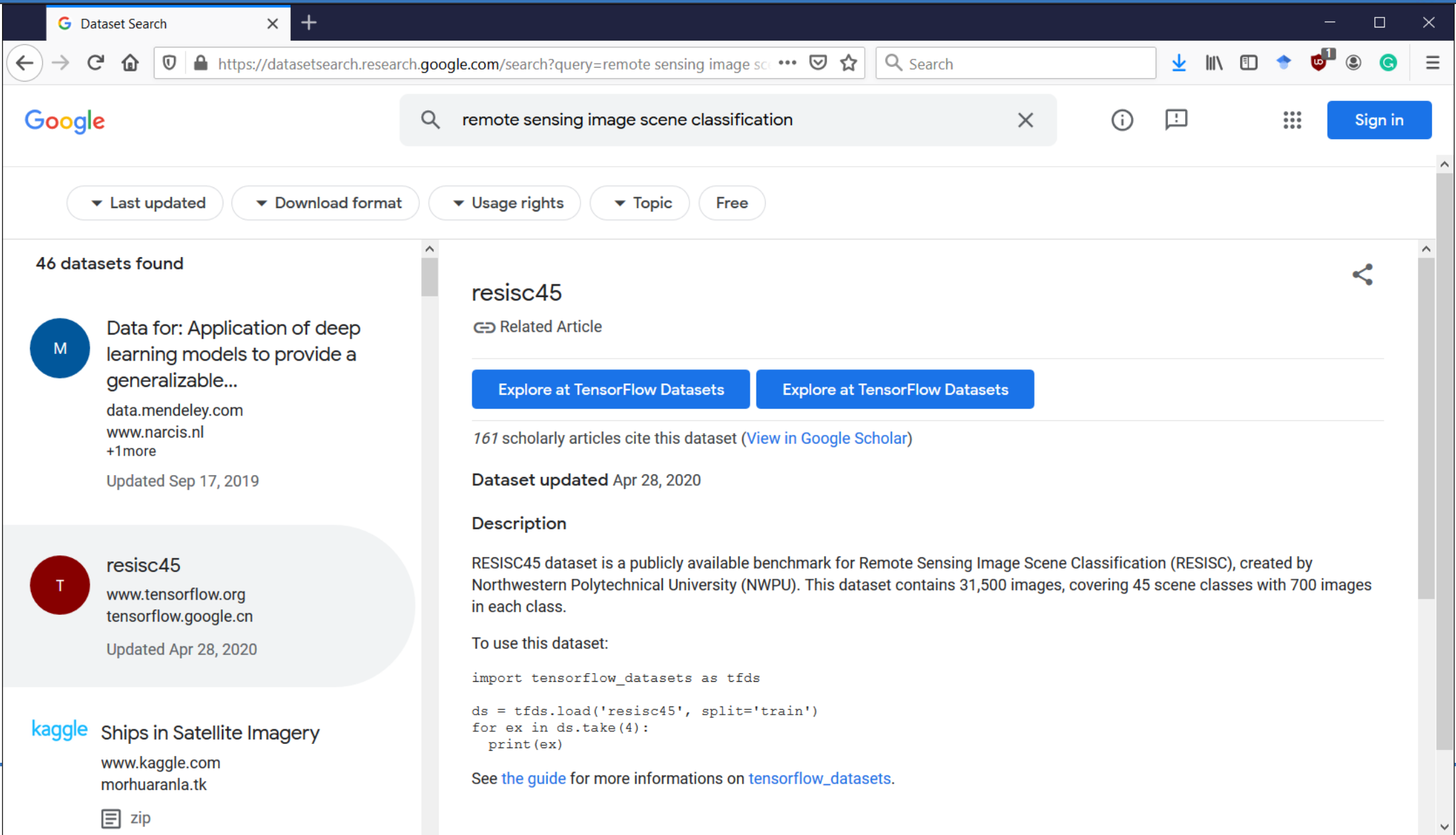
The following figure shows two samples of each class from this dataset.



- ❑ Da li postoji željeni skup podataka?
- ❑ Kako ga pronaći?
- ❑ Kojim alatima?
  - ❑ Postoji li mogućnost filtriranja po različitim kriterijumima da bi se pronašli odgovarajući podaci.
  - ❑ Postoje li metapodaci koji će omogućiti to filtriranje?
  - ❑ U kojem su formatu?
- ❑ Mogu li se pronađeni podaci preuzeti?
- ❑ U kojem formatu?
  - ❑ Mogu li se podaci tom formatu lako integrisati sa postojećim podacima?
  - ❑ Može li se ta integracija uraditi automatski – mašinska čitljivost?
- ❑ Da li se podaci smiju preuzeti?
- ❑ Pod kojim uslovima?
- ❑ Koga treba citirati?

- Da li da finansiramo prikupljanje podataka?
- Možemo li iskoristiti ranije prikupljene podatke za novu namjenu?
- Možemo li kombinovati ranije prikupljene podatke sa novim da bismo analizirali trendove i formirali politike?
- Mogu li građani iskoristiti podatke čije je prikupljanje finansirano javnim sredstvima?
- Kako da iskoristimo podatke za podsticanje razvoja kreativnog biznisa i ekonomije znanja?
- Kako stvoriti uslove za dugoročno čuvanje rezultata naučnih istraživanja?
- Kako da se borimo protiv prevara u nauci?

# Google Dataset Search?



The screenshot shows a web browser window with the Google Dataset Search interface. The search query is "remote sensing image scene classification". The results list 46 datasets found. The first result is "Data for: Application of deep learning models to provide a generalizable..." from data.mendeley.com, updated Sep 17, 2019. The second result is "resisc45" from www.tensorflow.org, updated Apr 28, 2020. The third result is "Ships in Satellite Imagery" from www.kaggle.com, updated Apr 28, 2020. The detailed view for "resisc45" is shown on the right, including a related article link, two "Explore at TensorFlow Datasets" buttons, a link to view 161 scholarly articles in Google Scholar, the update date (Apr 28, 2020), a description of the dataset as a benchmark for Remote Sensing Image Scene Classification (RESISC) with 31,500 images, and a code snippet for using the dataset with TensorFlow Datasets.

Dataset Search

https://datasetsearch.research.google.com/search?query=remote sensing image scene classification

Google remote sensing image scene classification Sign in

Last updated Download format Usage rights Topic Free

46 datasets found

**M** Data for: Application of deep learning models to provide a generalizable...  
data.mendeley.com  
www.narcis.nl  
+1more  
Updated Sep 17, 2019

**T** resisc45  
www.tensorflow.org  
tensorflow.google.cn  
Updated Apr 28, 2020

kaggle Ships in Satellite Imagery  
www.kaggle.com  
morhuanla.tk  
zip

resisc45  
Related Article

Explore at TensorFlow Datasets Explore at TensorFlow Datasets

161 scholarly articles cite this dataset (View in Google Scholar)

Dataset updated Apr 28, 2020

Description

RESISC45 dataset is a publicly available benchmark for Remote Sensing Image Scene Classification (RESISC), created by Northwestern Polytechnical University (NWPUP). This dataset contains 31,500 images, covering 45 scene classes with 700 images in each class.

To use this dataset:

```
import tensorflow_datasets as tfds

ds = tfds.load('resisc45', split='train')
for ex in ds.take(4):
    print(ex)
```

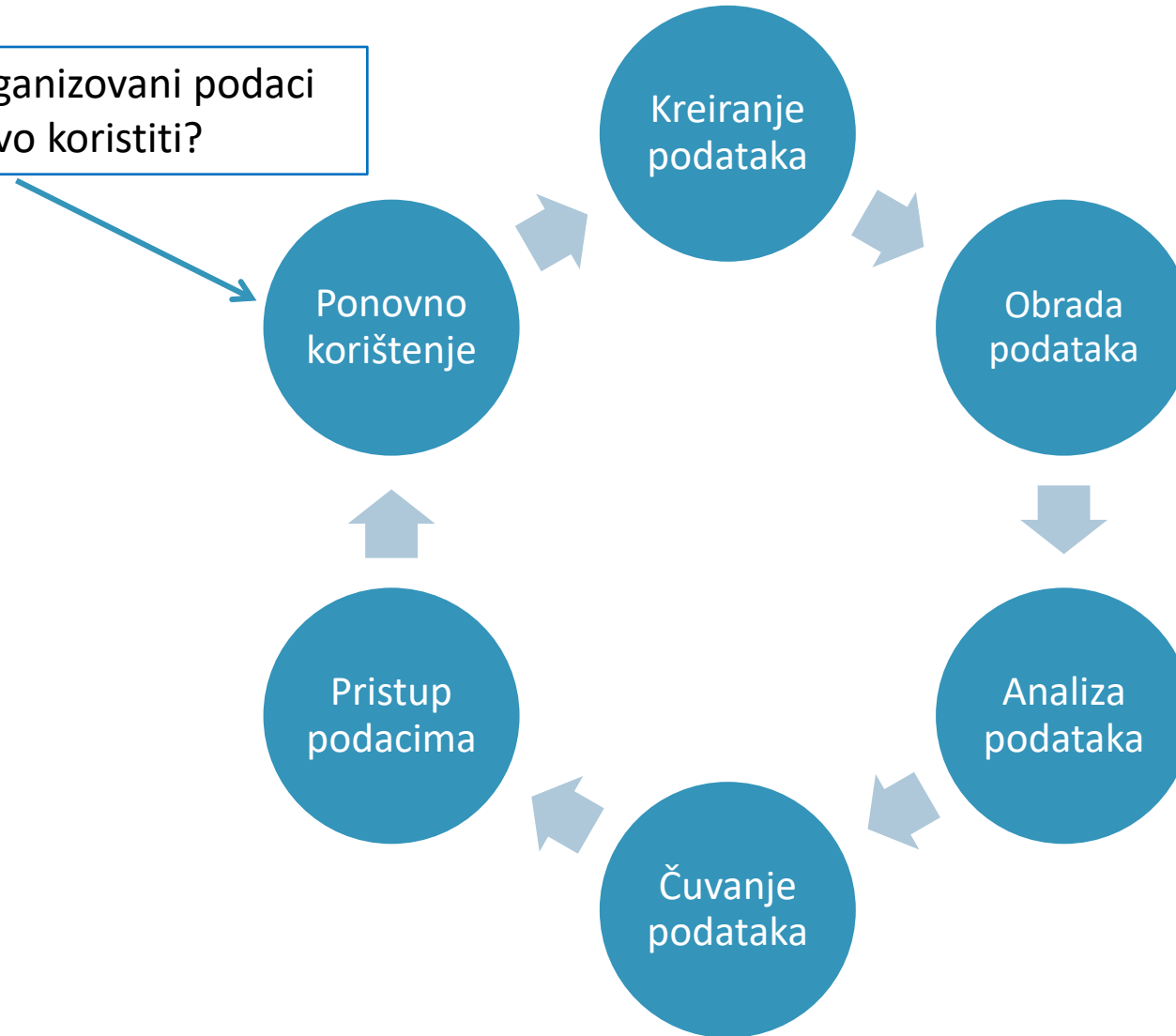
See the guide for more informations on tensorflow\_datasets.

Kako obezbijediti lakši pristup podacima za ljude i mašine?





Kako želimo da budu organizovani podaci da bismo ih mogli ponovo koristiti?



- Preporuka Komisije (EU) 2018/790 o pristupu i čuvanju naučnih informacija
  - (...) izrada plana za upravljanje podacima treba da postane standardna naučna praksa u ranoj fazi istraživačkog procesa, kada se podaci dobijaju i/ili prikupljaju, između ostalog i tako što će se zahtevati izrada plana za upravljanje podacima.



- ❑ Kako se rukuje istraživačkim podacima tokom i nakon kraja projekta?
- ❑ Kakvi podaci se prikupljaju, obrađuju i/ili generišu?
- ❑ Koja metodologija i standardi će se primjenjivati?
- ❑ Da li će podaci biti dostupni/otvoreni?
- ❑ Kako će se brinuti o podacima i kako će se čuvati (i nakon kraja projekta)?

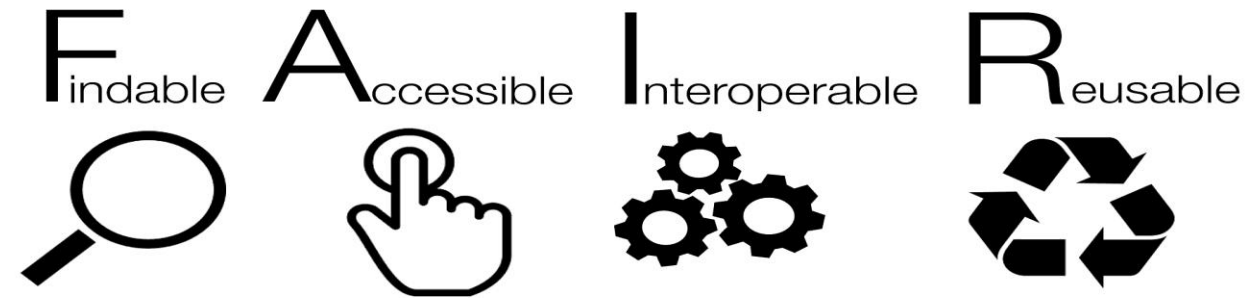
- ❑ Ne mora da sadrži potpuno razrađen plan
- ❑ Trebalo bi da odgovori na neka pitanja:
  - ❑ Koji standardi će biti primjenjeni?
  - ❑ Kako će se podaci koristiti i/ili dijeliti/učiniti dostupnim u cilju verifikacije i ponovnog korištenja?
  - ❑ Ako se podaci ne mogu učiniti dostupnim, zašto?
  - ❑ Kako će se brinuti o podacima i kako će se čuvati?
  - ❑ Kakva je politika konzorcijuma u pogledu eksploatacije rezultata i intelektualnog vlasništva?

- ❑ Inicijalni plan u H2020 projektima treba da bude napravljen u šestom mjesecu realizacije
  - ❑ Obrazac plana: [Horizon 2020 FAIR Data Management Plan \(DMP\) template](#)
  - ❑ Nije neophodno dati detaljne odgovore na sva pitanja
- ❑ Plan je živi dokument
  - ❑ Ažuriranje u toku projekta kada se pojavi potreba
  - ❑ Sadrži broj verzije i tabelu izmjena
- ❑ Plan treba ažurirati kada se:
  - ❑ pojave novi podaci
  - ❑ promijeni politika konzorcijuma (npr. patenti)
  - ❑ promijeni sastav konzorcijuma
- ❑ Ažuriranje u skladu sa revizijama projekta
  - ❑ periodične
  - ❑ finalna

- U H2020 projektima su troškovi upravljanja i obezbjeđivanja otvorenog pristupa podacima prihvatljiva budžetska kategorija
  - [Data Management Cost Guide](#)

1. Data summary
2. FAIR data
  - a. Making data findable, including provisions for metadata
  - b. Making data openly accessible
  - c. Making data interoperable
  - d. Increase data re-use (through clarifying licences)
3. Allocation of resources
4. Data security
5. Ethical aspects
6. Other

- Findable – omogućeno pronalaženje (vidljivi)
- Accessible – omogućen pristup (dostupni)
- Interoperable – omogućena saradnja (interoperabilni)
- Reusable – omogućeno ponovno korištenje (višekratni)
- Mjerljivi principi
- Kontinuum usklađenosti
- Ne radi se o standardu



SangyaPundir / CC BY-SA (<https://creativecommons.org/licenses/by-sa/4.0>)

- ❑ Dodijeliti podacima globalno jedinstven i perzistentan identifikator
  - ❑ DOI, Orcid,...
- ❑ Detaljno opisati podatke metapodacima
  - ❑ DICOM,...
- ❑ Metapodaci jasno i eksplicitno sadrže identifikator podataka koje opisuju
  - ❑ Koristiti standardne šeme za metapodatke (npr. [Dublin Core](#))
  - ❑ Konsultovati [DCC metadata directory](#), [RDA Metadata Directory](#), [FAIRsharing](#),...
- ❑ (Meta)podaci su registrovani i indeksirani u pretraživom repozitorijumu
  - ❑ Koristiti repozitorijum koji dodjeljuje objektima perzistentne identifikatore (npr. [Zenodo](#))

- ❑ Omogućiti pristup (meta)podacima na osnovu njihovog identifikatora korištenjem standardizovanog komunikacionog protokola
  - ❑ HTTP, FTP,...
  - ❑ Ne moraju biti otvoreni (otvoreni koliko je moguće, zatvoreni koliko je potrebno)
  - ❑ Može postojati period embarga
  - ❑ Ali da bi bili FAIR treba da budu vidljivi – omogućiti pristup metapodacima
  - ❑ Definirati način za dobijanje pristupa
  - ❑ Izabrati repozitorijum
    - ❑ [Zenodo](#), [Harvard Dataverse](#), [Dryad](#),...
    - ❑ [re3data.org](#) je registar repozitorijuma
- ❑ Obezbijediti pristup metapodacima i kada podaci više nisu dostupni



# Kako da podaci budu FAIR?

## Interoperable (Interoperabilni)

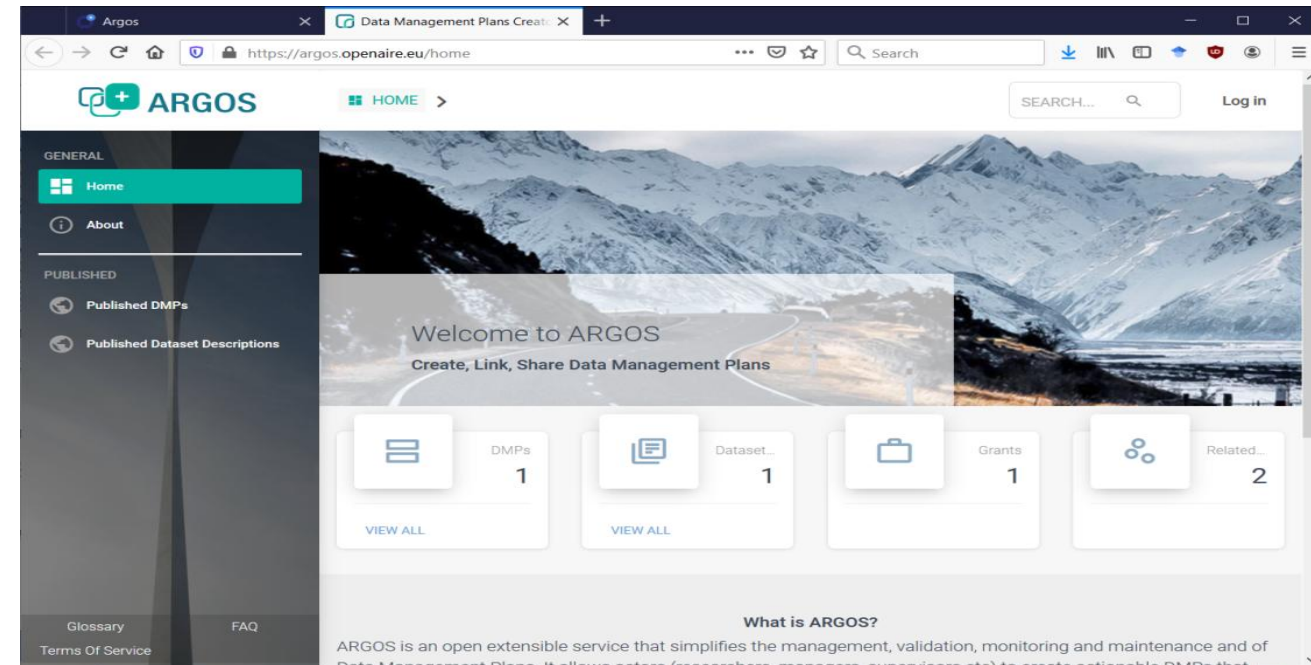
- ❑ Koristiti otvorene formate i standarde za podatke i metapodatke
- ❑ Koristiti kontrolisane rječnike, ontologije, tezaure karakteristične za domen
- ❑ Opisati veze sa drugim skupovima podataka
- ❑ Konsultovati [DCC metadata directory](#), [RDA Metadata Directory](#), [FAIRsharing](#),...

- ❑ Detaljno opisati (meta)podatke mnoštvom tačnih i relevantnih atributa
  - ❑ Dokumentovati podatke, npr. README fajl
    - ❑ Opis šta sadrže pojedini fajlovi
    - ❑ Definicije kolona ili redova u tabeli
    - ❑ Kako su podaci obrađeni?
    - ❑ Sa kojim drugim skupovima su podaci povezani?
    - ❑ Ko je kontakt osoba za podatke?
- ❑ Odrediti licencu pod kojom su podaci dostupni
  - ❑ MIT, Creative Commons,...

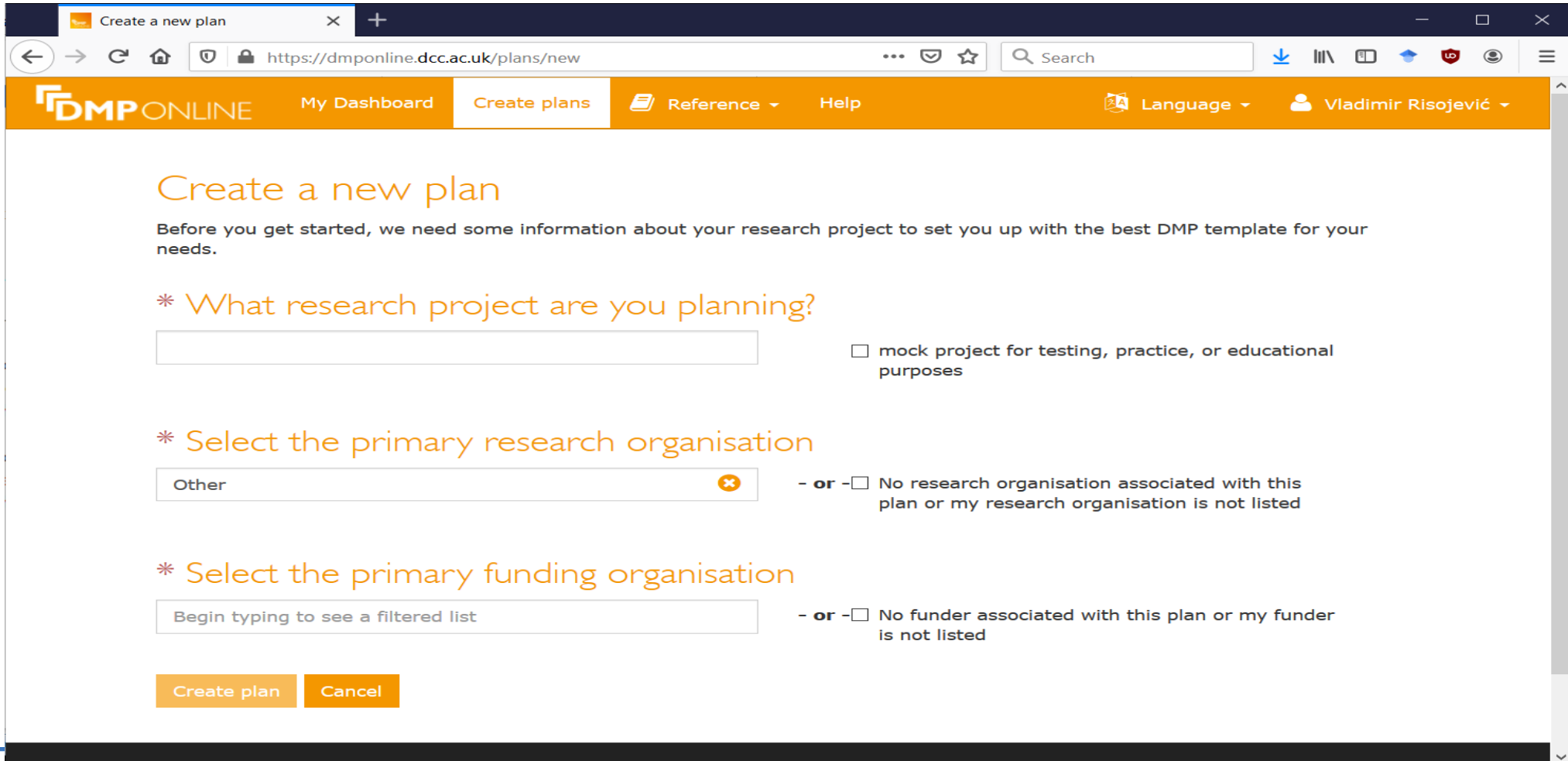
- ❑ Mogu se primijeniti na sve digitalne istraživačke objekte
  - ❑ rječnike,
  - ❑ algoritme,
  - ❑ alate,
  - ❑ radne tokove...
- ❑ Osigurava se transparentnost, ponovljivost i ponovno korištenje

### □ ARGOS (<https://argos.openaire.eu>)

- Onlajn alat za podršku automatizovanim procesima kreiranja, upravljanja, dijeljenja i povezivanja planova upravljanja podacima sa istraživačkim artifaktima kojima odgovaraju



### □ DMPonline, <https://dmponline.dcc.ac.uk>



The screenshot shows a web browser window with the URL <https://dmponline.dcc.ac.uk/plans/new>. The page title is "Create a new plan". The navigation bar includes "My Dashboard", "Create plans", "Reference", and "Help". The user is logged in as "Vladimir Risojević".

**Create a new plan**

Before you get started, we need some information about your research project to set you up with the best DMP template for your needs.

\* What research project are you planning?

mock project for testing, practice, or educational purposes

\* Select the primary research organisation

- or -  No research organisation associated with this plan or my research organisation is not listed

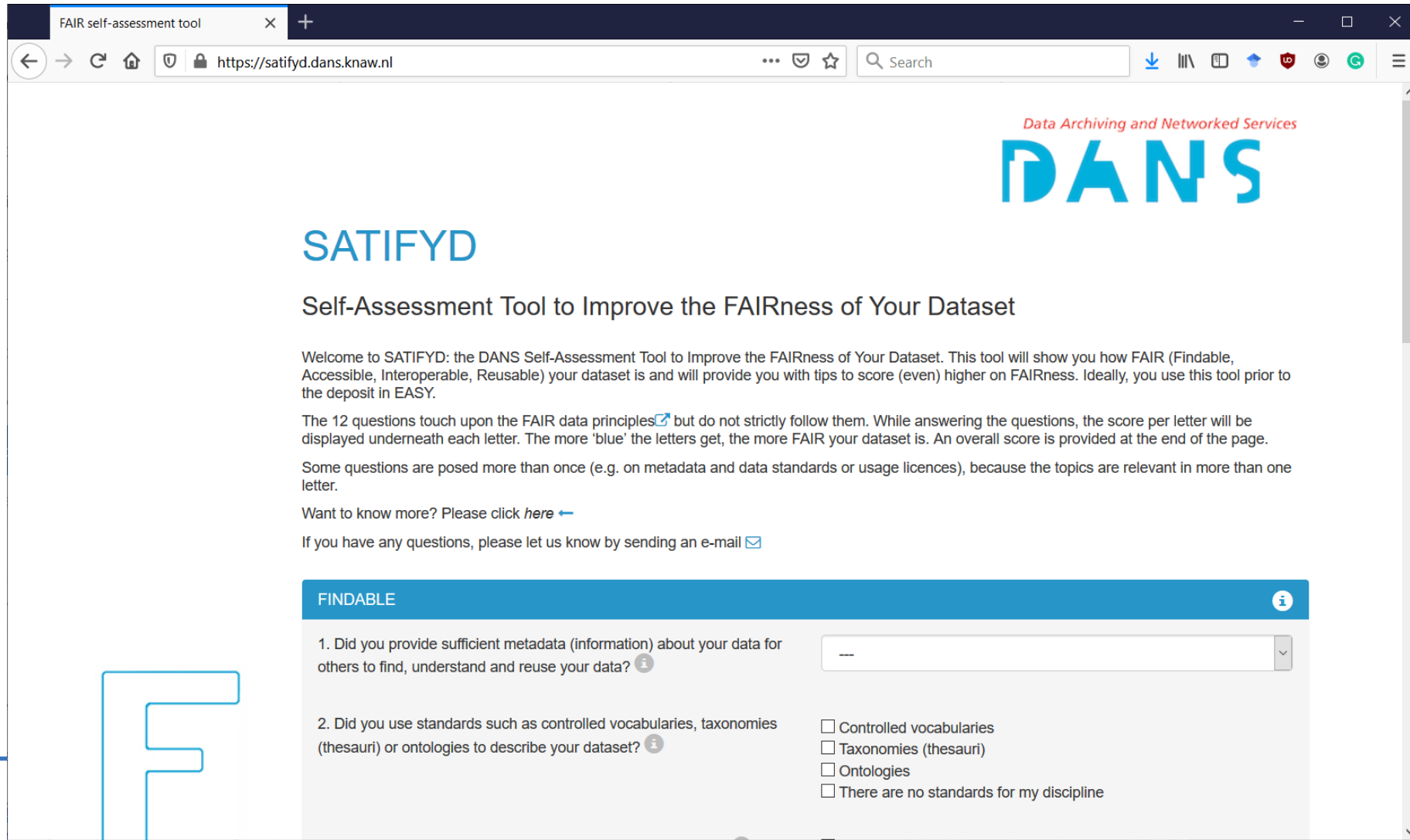
\* Select the primary funding organisation

- or -  No funder associated with this plan or my funder is not listed

# Alati

## Procjena koliko su podaci FAIR

□ SATIFYD, <https://satisfyd.dans.knaw.nl/>



The screenshot shows a web browser window with the URL <https://satisfyd.dans.knaw.nl/>. The page features the DANS logo (Data Archiving and Networked Services) and the title "SATIFYD Self-Assessment Tool to Improve the FAIRness of Your Dataset". The main content includes a welcome message, instructions on how to use the tool, and a section titled "FINDABLE" with two questions and a dropdown menu.

FAIR self-assessment tool

https://satisfyd.dans.knaw.nl

Data Archiving and Networked Services  
**DANS**

## SATIFYD

### Self-Assessment Tool to Improve the FAIRness of Your Dataset

Welcome to SATIFYD: the DANS Self-Assessment Tool to Improve the FAIRness of Your Dataset. This tool will show you how FAIR (Findable, Accessible, Interoperable, Reusable) your dataset is and will provide you with tips to score (even) higher on FAIRness. Ideally, you use this tool prior to the deposit in EASY.

The 12 questions touch upon the FAIR data principles but do not strictly follow them. While answering the questions, the score per letter will be displayed underneath each letter. The more 'blue' the letters get, the more FAIR your dataset is. An overall score is provided at the end of the page.

Some questions are posed more than once (e.g. on metadata and data standards or usage licences), because the topics are relevant in more than one letter.

Want to know more? Please click [here](#)

If you have any questions, please let us know by sending an e-mail

#### FINDABLE

1. Did you provide sufficient metadata (information) about your data for others to find, understand and reuse your data?

2. Did you use standards such as controlled vocabularies, taxonomies (thesauri) or ontologies to describe your dataset?

- Controlled vocabularies
- Taxonomies (thesauri)
- Ontologies
- There are no standards for my discipline

## □ Otvorena nauka

- Preporuka Komisije (EU) 2018/790 o pristupu i čuvanju naučnih informacija, <https://eur-lex.europa.eu/eli/reco/2018/790/oj>  
<http://www.open.ac.rs/docs/Preporuka-EU-2018-790.pdf>

## □ Infrastruktura/Servisi

- OpenAIRE, <https://www.openaire.eu/>
- EOSC, <https://www.eosc-portal.eu/>
- EUDAT, <https://www.eudat.eu/>
- RDA, <https://www.rd-alliance.org/>

## □ FAIR principi

- Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
- Guidelines on FAIR Data Management in Horizon 2020, [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

- <https://www.go-fair.org/>
  - Inicijativa čiji je cilj implementacija FAIR principa
- <https://www.fairsfair.eu/>
  - Projekat čiji su cilj praktična rješenja za korištenje FAIR principa u životnom ciklusu istraživačkih podataka
- <https://fairsharing.org/>
  - Informativni i edukativni resurs o standardima za podatke i metapodatke, bazama podataka i politikama



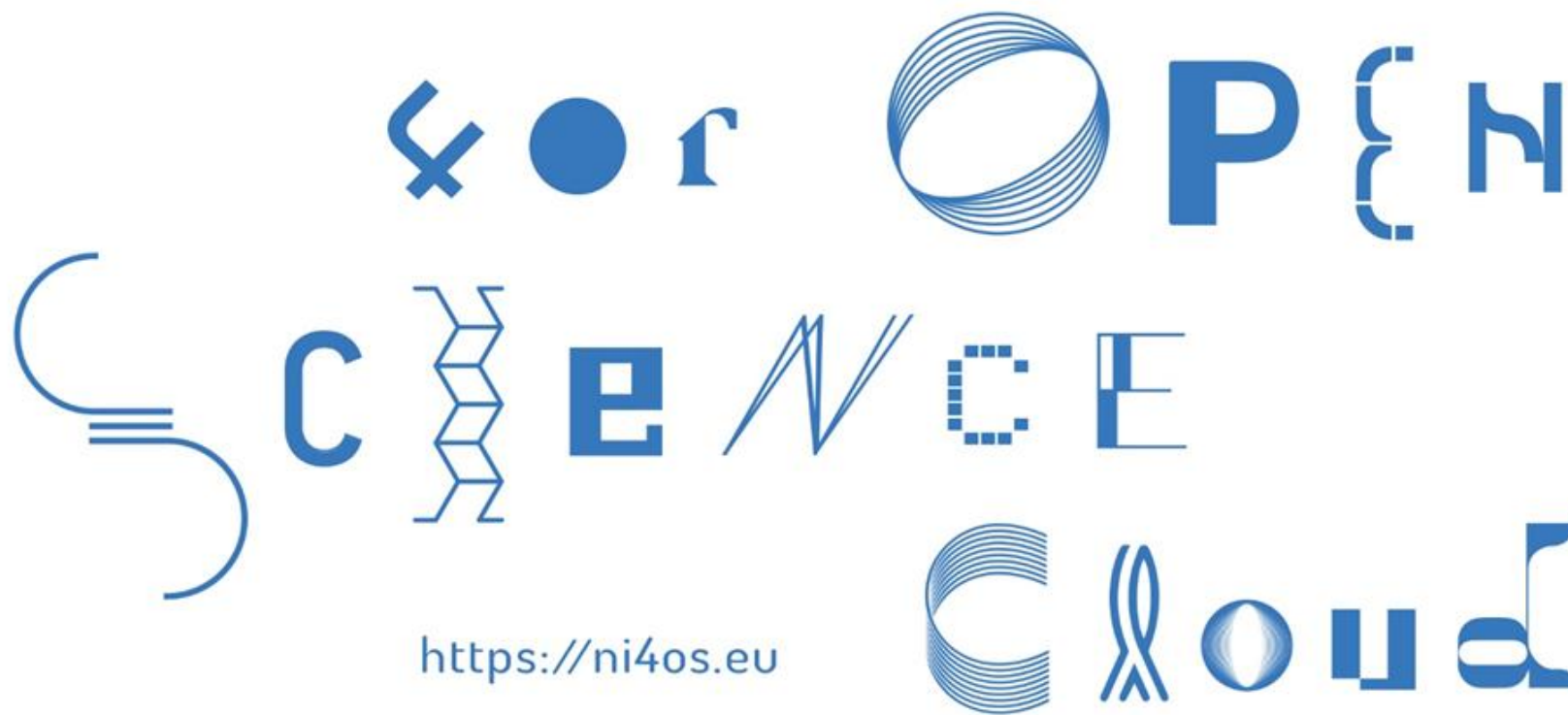
## □ Procjena koliko su podaci FAIR

- Jones, Sarah, & Grootveld, Marjan. (2017, November). How FAIR are your data?. Zenodo. <http://doi.org/10.5281/zenodo.1065991>
- <https://www.surveymonkey.com/r/fairdat>
- <https://satisfyd.dans.knaw.nl/>

## □ Iskustva iz Srbije

- Nacionalni portal otvorene nauke, <http://www.open.ac.rs>
- Platforma za otvorenu nauku, <http://www.mpn.gov.rs/wp-content/uploads/2018/07/Platforma-za-otvorenu-nauku.pdf>

# Hvala!



<https://ni4os.eu>



[@NI4OS\\_eu](https://twitter.com/NI4OS_eu)



[@NI4OS](https://facebook.com/NI4OS)