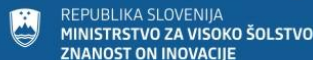


Definicija in življenjski krog raziskovalnih podatkov

Doc. dr. Ana SLAVEC, Fakulteta za matematiko, naravoslovje in informacijske tehnologije Univerze na Primorskem in InnoRenew CoE

Ljubljana in online, Centralna tehniška knjižnica Univerze v Ljubljani,
Usposabljanje podatkovnih strokovnjakov, 14. – 17. 10. 2024





Moja pot v OZ

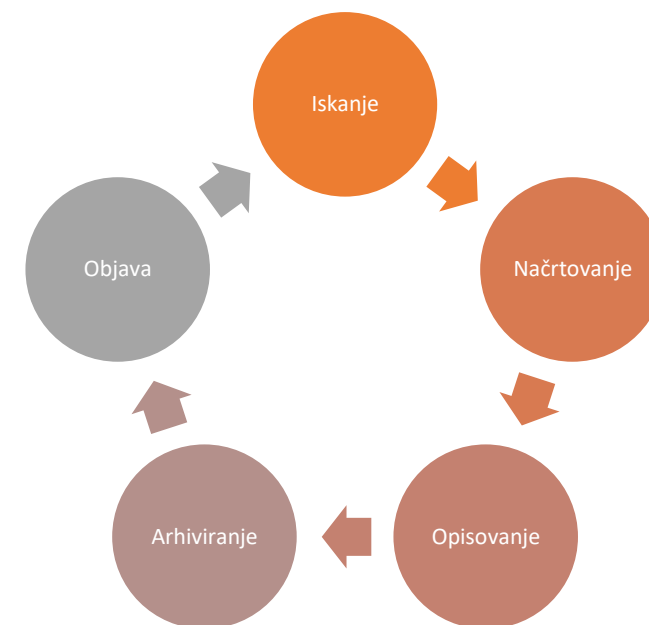


- Univ. diploma s področja sociologije – družboslovne informatike (2010) in doktorat s področja statistike (2016) na **Univerzi v Ljubljani**
- Raziskovalna asistentka v **Arhivu družboslovnih podatkov** (2016 – 2017)
- Svetovalka za statistiko na **InnoRenew CoE Centru odličnosti za raziskave in inovacije na področju obnovljivih materialov in zdravega bivanjskega okolja** (2017 –)
- Koordinatorica za odprto znanost pri **društvu Mlada akademija** (2018 –)
- So-koordinatorica in ambasadorica za odprto znanost pri **Evropskem svetu doktorskih kandidatov in raziskovalcev na začetku kariere (Eurodoc)** (2018 – 2021)
- Ambasadorica za področje tehnike/obnovljivih materialov pri **Združenju za raziskovalne podatke Research Data Alliance** (2019 –)
- Članica DS za načela FAIR pri **Evropskem oblaku odprte znanosti** (2019 – 2020)
- Alumna **pilotne šole CODATA-RDA za podatkovne skrbnike** (2019)
- Implementacija nacionalnega akcijskega načrta odprte znanosti na **Univerzi na Primorskem** (2023 –)



Vsebina

- Definicija raziskovalnih podatkov
- Življenjski krog raziskovalnih podatkov
- Podrobna predstavitev načel FAIR





Kako opredeliti raziskovalne podatke (oz. OZ na sploh)?



Prilika o slepcih in slonu

Vir: [openclipart](https://openclipart.org/)



Kaj so raziskovalni podatki?

Slovenski prevod OECD definicije (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

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



Primarni in sekundarni podatki

- **Primarni podatki** so tisti, ki jih raziskovalci za namen svoje raziskave neposredno zbirajo prek opazovanja ali eksperimentiranja.
- **Sekundarni podatki** so tisti, ki so bili zbrani že prej za nek drug namen. Pridobljeni so iz že obstoječih virov, bodisi s simulacijami, z analizo ali reorganizacijo obstoječih podatkovnih točk (vključno z že objavljenimi opazovanimi in eksperimentalnimi podatki).

1

2



Analogni in digitalni podatki

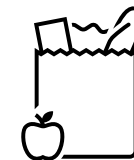
- **Analogni podatki** so zajeti v svoji fizični obliki, kot so papirni zapisi, zvočni posnetki, slike na fotografskih filmih, filmski trakovi, itd. Njihova oblika temelji na zveznem signalu ali fizičnem mediju.
- **Digitalni podatki** so tisti, ki so shranjeni, obdelani in preneseni v digitalni obliki, kot so npr. dokumenti v elektronski obliki, podatkovne baze, itd. Predstavljeni so v diskretni obliki, kar omogoča njihovo enostavno obdelavo z računalniškimi sistemi.





Surovi in obdelani podatki

- **Surovi podatki** so podatki v njihovi izvorni, neprilagojeni obliki, takšni kot so bili zbrani, brez kakršne koli obdelave, organizacije, ali analize.
- **Obdelani podatki** so tisti, ki so shranjeni, obdelani in preneseni v digitalni obliki, kot so npr. dokumenti v elektronski obliki, podatkovne baze, itd. Predstavljeni so v diskretni obliki, kar omogoča njihovo enostavno obdelavo z računalniškimi sistemi.





Mikro in makro podatki

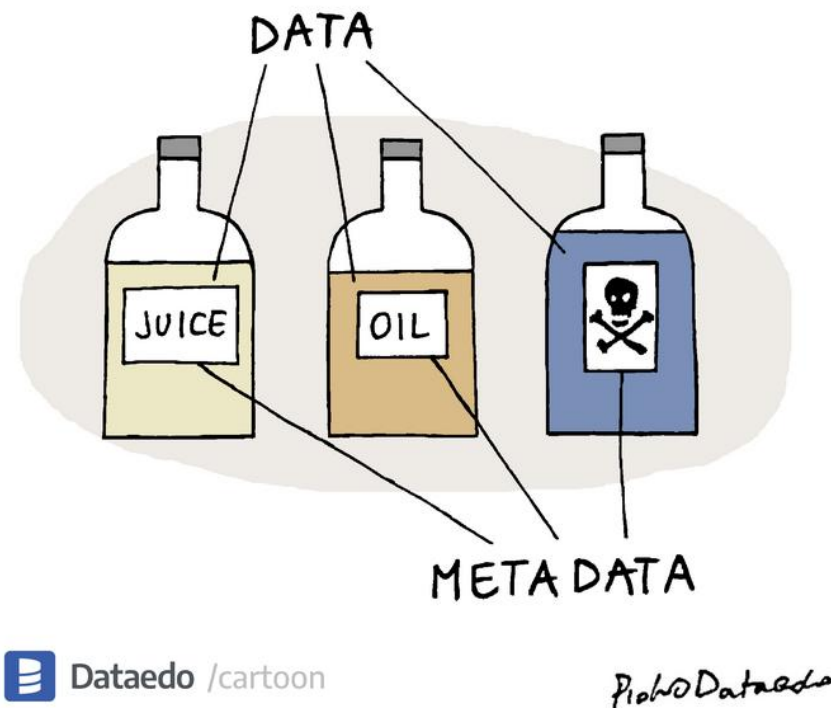
- **Mikropodatki** so neagregirana opazovanja ali meritve značilnosti posameznih enot, ki niso bili združeni ali povprečeni, torej ohranjajo natančnejše informacije o vsaki enoti.
- **Makropodatki (agregirani podatki)** so podatki, ki so zbrani in združeni iz posameznih enot v širše skupine ali kategorije.





Metapodatki (podatki o podatkih)

- **Metapodatki** so podatki, ki opišejo vsebino podatkov, njihovo kakovost, obliko, vir, lokacijo in pogoje dostopa (npr. ID raziskave, avtorji, naslov, ključne besede, datum, licenca, financiranje...).
- Strukturirani so na način, da so človeško in strojno berljivi (uporabljajo se različni formati, kot so XML, JSON, RDF, CSV, idr.)
- Različne metapodatkovne sheme in standardi.

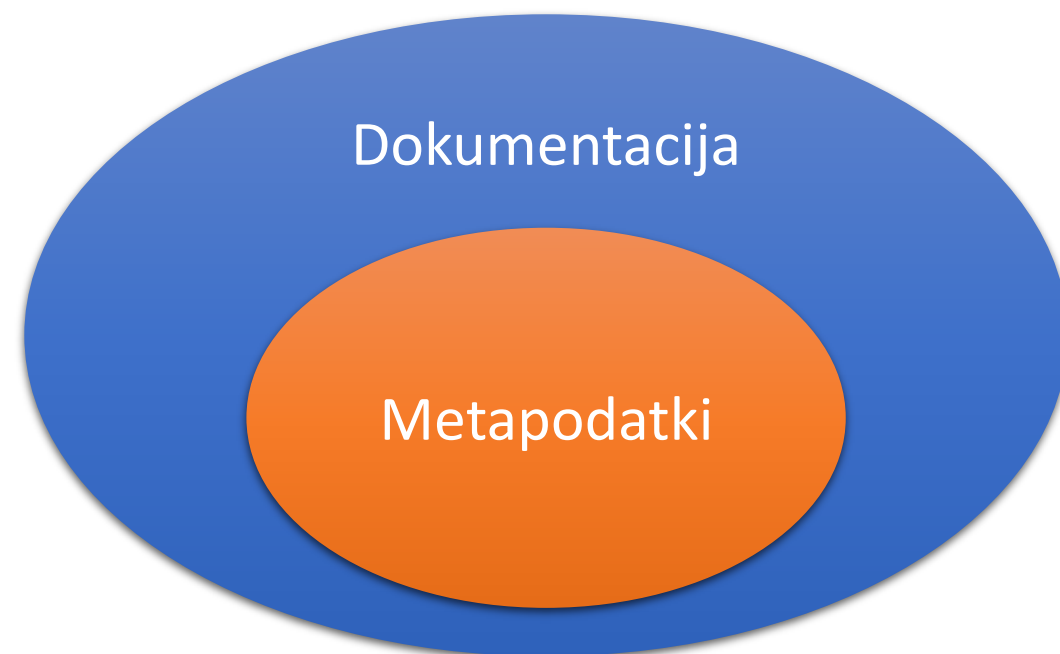


Vir: Piotr Kononow (Dataedo.com)



Dokumentacija in metapodatki

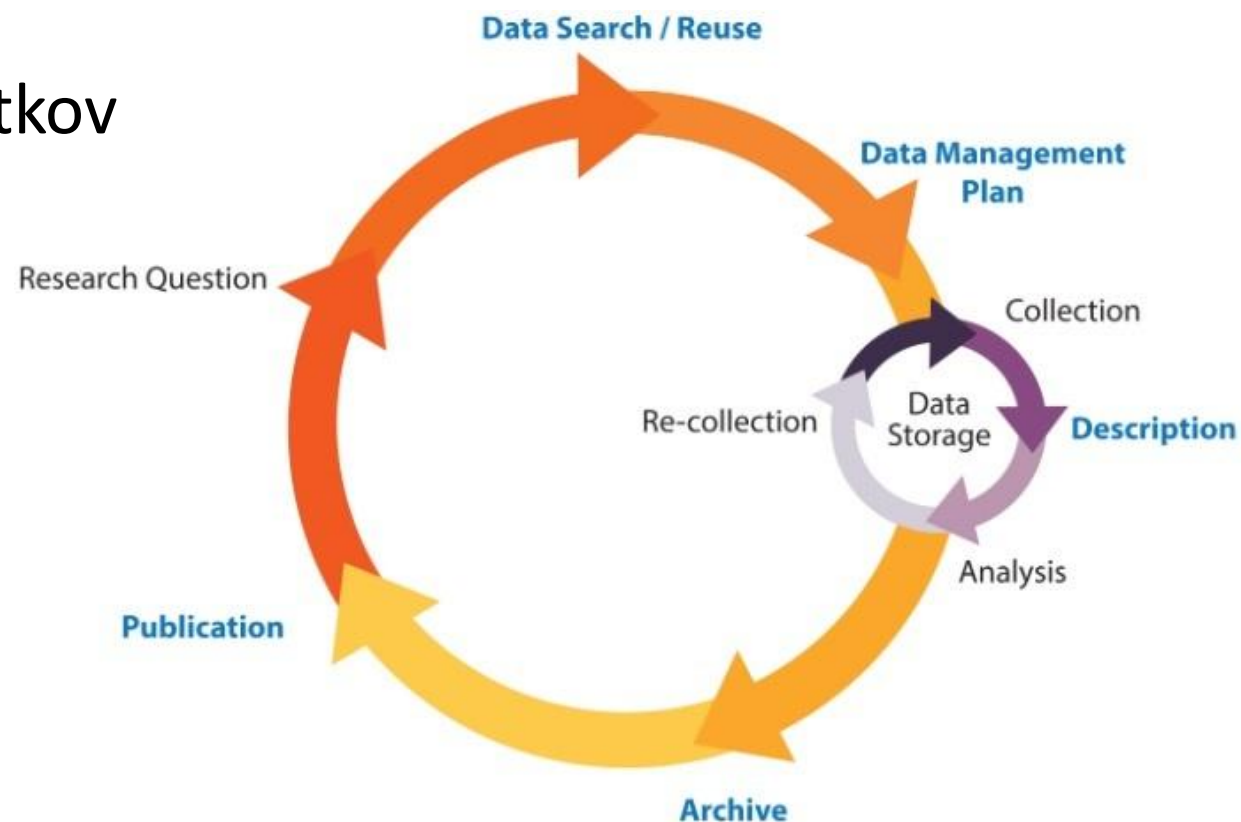
- Metapodatki so del **dokumentacije** raziskave, ki vključuje še vse druge dokumente, ki opisujejo potek raziskave, njeno metodologijo, rezultate, analize in zaključke.
- Dokumentacija nam pove širši kontekst, zakaj so nastali podatki, kaj je bilo narejeno in kako, kakšna koda je bila uporabljena , itd.





Življenjski krog raziskovalnih podatkov

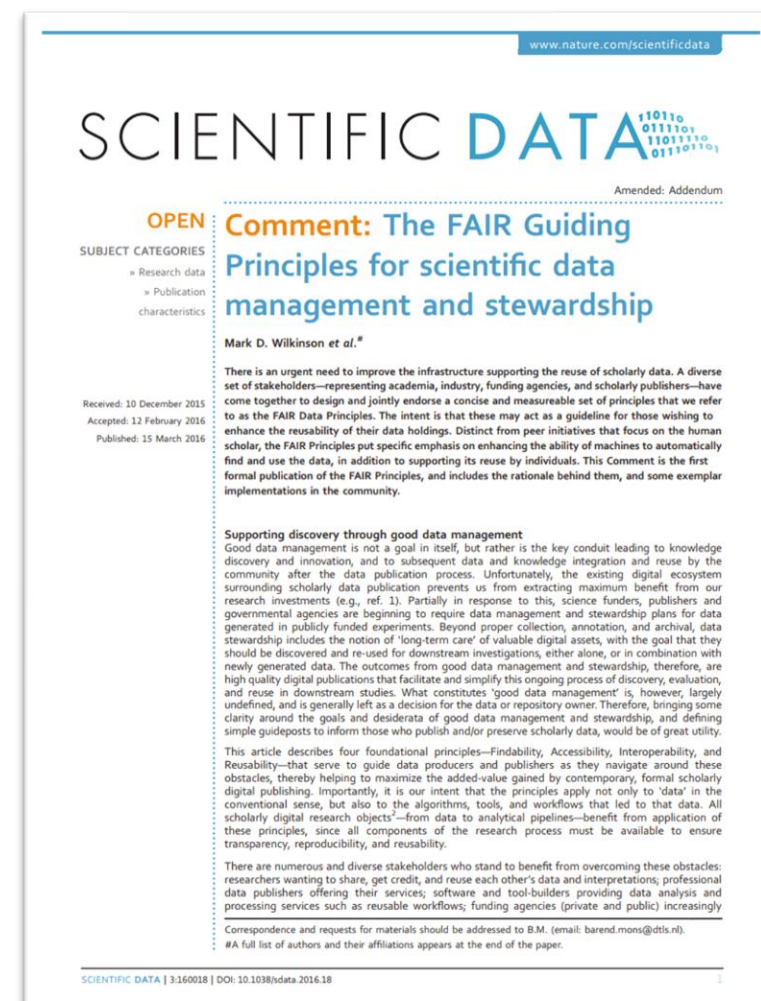
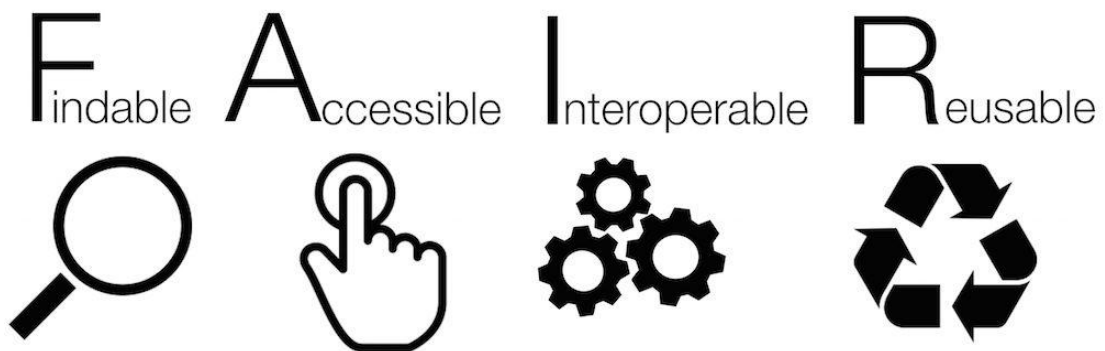
- Iskanje in ponovna raba podatkov
- Načrt ravnanja s podatki
- Opis podatkov
- Arhiviranje podatkov
- Objava podatkov





Načela FAIR

Članek “The FAIR Guiding Principles for scientific data management and stewardship” ([Wilkinson idr. 2016](#))





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.

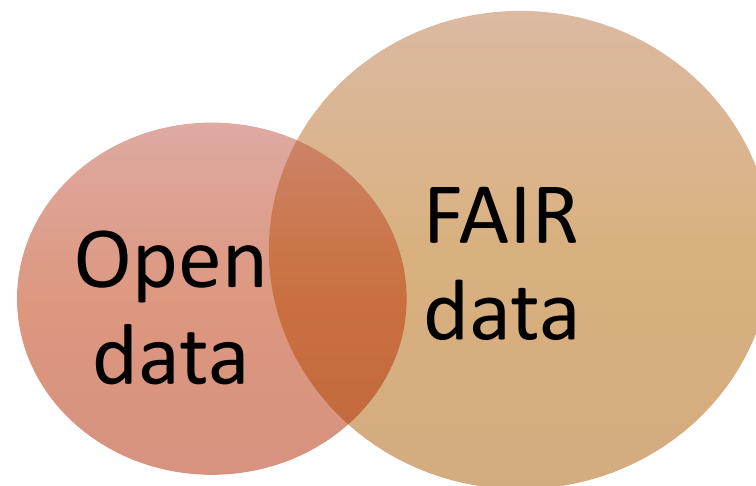




Načela FAIR ≠ Odprti podatki

Kolikor odprto je mogoče, kolikor zaprto je potrebno.

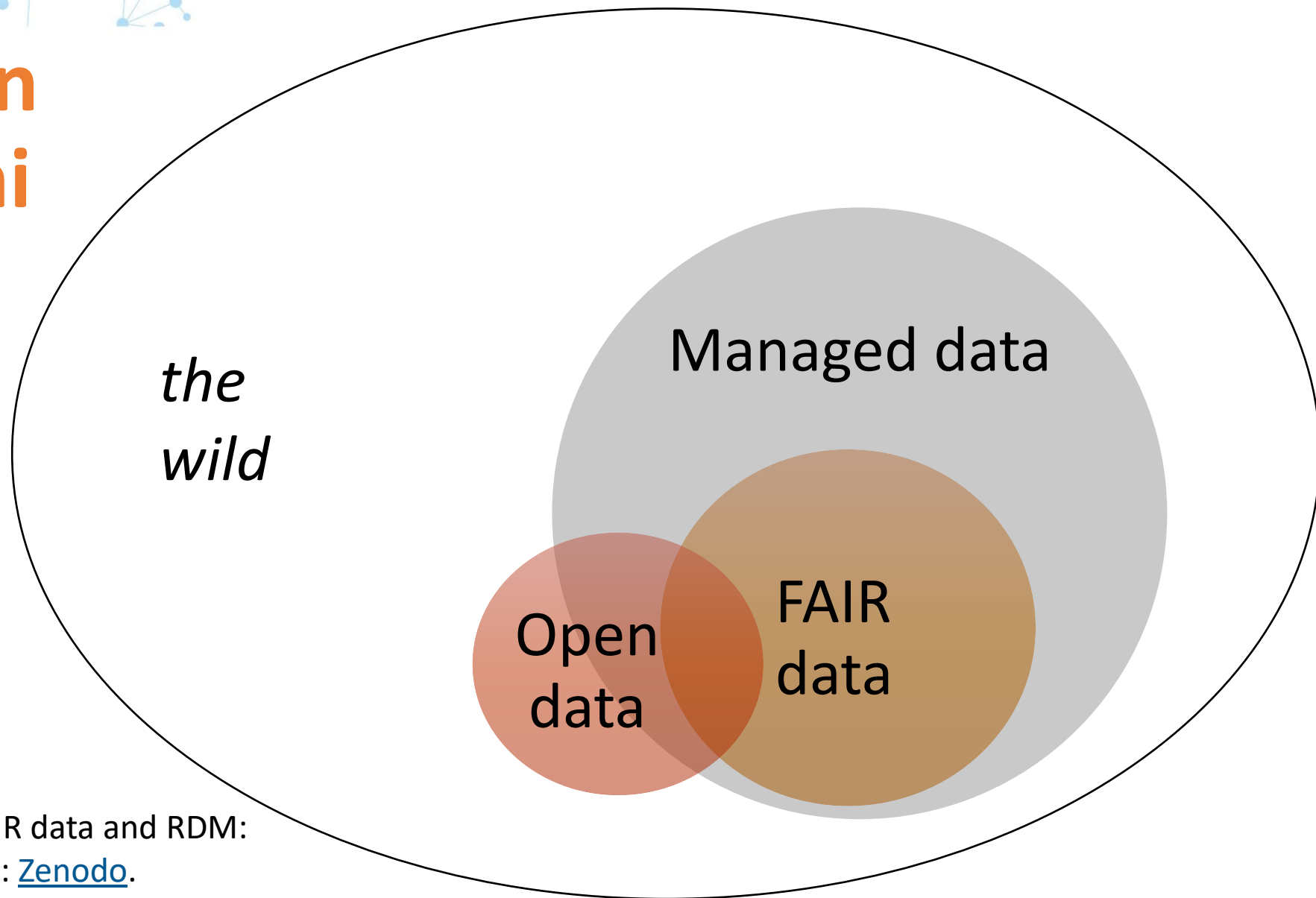
(angl. „as open as possible, as closed as necessary“)



Jones, S. 2018. Open data, FAIR data and RDM: the ugly duckling. Available at: [Zenodo](#).



Upravljeni in neupravljeni podatki



Jones, S. 2018. Open data, FAIR data and RDM: the ugly duckling. Available at: [Zenodo](#).



Kako definirati metrike za načela FAIR?



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





Specifikacija in smernice modela zrelosti

- Smerni Model

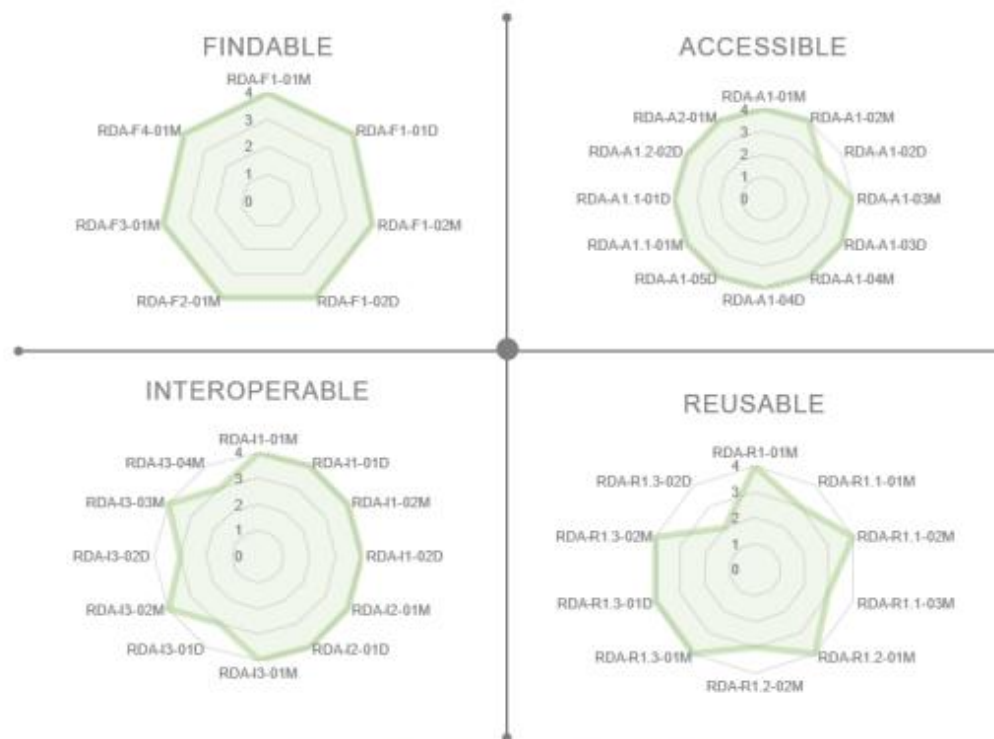
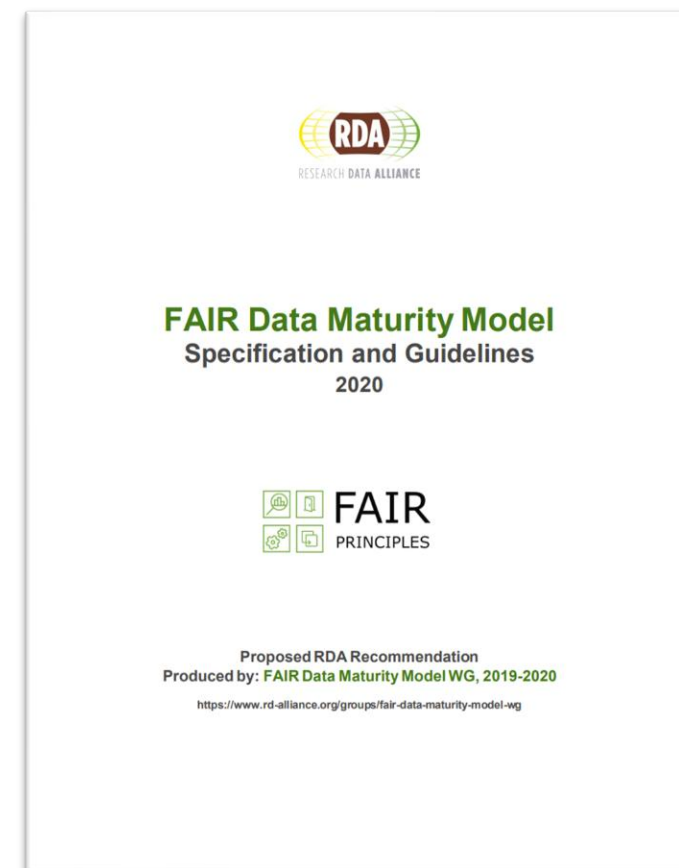


Figure 3 'Measuring progress' visualisation





Priporočila za FAIR metrike za EOSC

- Poročilo, ki smo ga pripravili v delovni skupini EOSC Executive Board FAIR WG ([Genova et al. 2021](#))

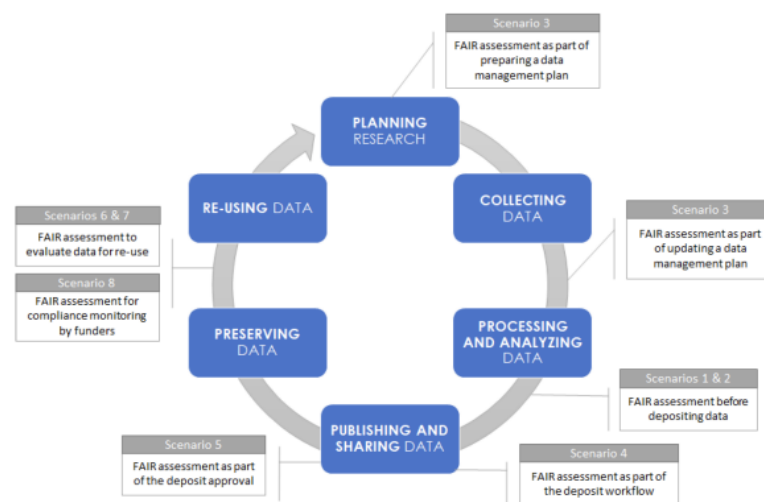
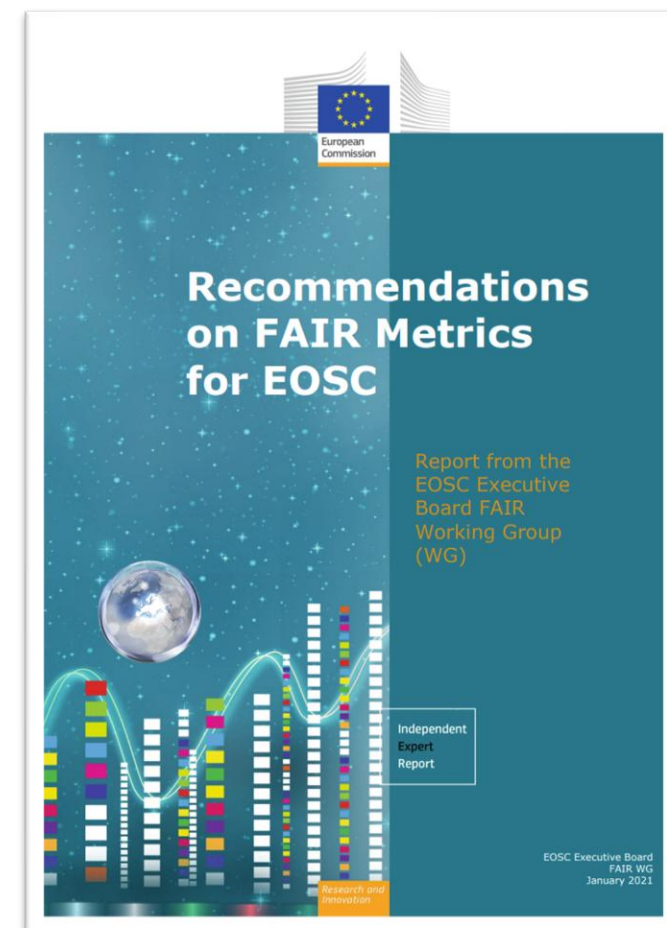


Figure 9: Research data lifecycle (figure adapted from Mosconi et al., 2019) and scenarios of FAIR assessment of datasets therein. Figure 3 from FAIRsFAIR D4.1.



Hvala za pozornost!



Univerza v Ljubljani

