

Od začetkov načel FAIR do orodij in pristopov ocenjevanja primernosti podatkov za uporabo

Abstract

From The Emergence of FAIR Principles to the Tools and Approaches for Assessing Data Usability

The FAIR principles were first published in 2016. They provide guidelines for good research data management practices, aiming to make data F – findable, A – accessible, I – interoperable and R – reusable. Each of the initial letters is associated with a set of criteria that elaborate on the main principle, and there's 15 criteria in total. FAIR principles emphasize machine readability - the ability of computer systems to search, access, make interoperable, and reuse data with with minimal or no human intervention. This is important because researchers increasingly rely on computer assistance for data processing due to the increasing volume, complexity, and speed of data production. Although the FAIR principles have their origins in the natural sciences, they are also applicable in other research disciplines, including the humanities and social sciences. The European Union, as well as individual research funders and universities, increasingly support and endorse the FAIR principles. This is reflected in the development of tools and infrastructure for research data management and the development of guidelines for data processing. The purpose of this article is to present the historical development and enforcement of the FAIR principles, as well as the various tools and approaches that have been developed to assess the suitability of data for other purposes.

Keywords: FAIR data principles, historical development, assessment tools

Maja Dolinar is employed as a Senior Administrative Officer at the Slovenian Social Science Data Archives (maja.dolinar@fdv.uni-lj.si).

Povzetek

Načela FAIR so bila prvič objavljena leta 2016. Vsebujejo smernice za dobro prakso ravnanja z raziskovalnimi podatki, katere cilj je podatke narediti najdljive (F – *findable*), dostopne (A – *accessible*), interoperabilne (I – *interoperable*) in ponovno uporabne (R – *reusable*). Vsaka začetnica je povezana z večjim številom meril, ki natančneje določajo vsebino načela, skupaj jih je 15. Načela FAIR poudarjajo strojno delovanje (tj. zmožnost računalniških sistemov za iskanje, dostop, interoperabilnost in ponovno uporabo podatkov brez kakršnega koli (ali z minimalnim) človeškim posredovanjem, saj se raziskovalci pri obravnavi podatkov vse bolj zanašajo na računalniško podporo, kar je posledica povečanja obsega, zapletenosti in hitrosti ustvarjanja podatkov. Čeprav načela FAIR izvirajo iz naravoslovnih ved, so uporabna tudi v drugih raziskovalnih disciplinah, vključno s humanistiko in družboslovjem. Evropska unija, posamezni financerji raziskovalnega dela in univerze čedalje bolj podpirajo in potrjujejo načela FAIR, kar sega od oblikovanja orodij in infrastrukture za ravnanje z raziskovalnimi podatki do oblikovanja politik za delo s podatki. Namen članka je predstaviti zgodovinski razvoj in uveljavljanje načel FAIR ter različna orodja in pristope, ki so se razvili za ocenjevanje primernosti podatkov za drugo rabo.

Ključne besede: načela FAIR, zgodovinski razvoj, orodja ocenjevanja

Maja Dolinar je kot samostojna strokovna delavka, zaposlena v Arhivu družboslovnih podatkov (maja.dolinar@fdv.uni-lj.si).

Uvod

Podatki so bistven vir raziskav. Petnajst načel FAIR na področju podatkov, ki so bila objavljena leta 2016, podpira utemeljeno ponovno uporabo podatkov, tako da omogoča iskanje, dostopnost, interoperabilnost in ponovno uporabo digitalnih virov (Force11, 2021). Kot pravi preambula k načelom, je eden glavnih izzivov podatkovno intenzivne znanosti olajšati odkrivanje znanja tako ljudem kot strojem (Force11, 2021). Objava načel je sprožila teoretične razprave, nekatere skupnosti pa so že začele implementirati FAIR vodene podatke in storitve. Zaradi hitro razvijajočih se področij odprte znanosti in podatkovnega skrbništva¹ je vprašanje, ali so podatki dovolj FAIR, ves čas relevantno.

Nihče ne nasprotuje ideji, da bi morali biti podatki ter z njimi povezani delovni tokovi in storitve najdljivi, dostopni pod natančno določenimi pogoji, interoperabilni brez spreminjanja podatkov in s tem optimalno ponovno

¹ Podatkovno skrbništvo je aktivno in stalno upravljanje podatkov skozi njihov življenjski cikel in se osredotoča na ohranjanje njihove dolgoročne uporabnosti. To vključuje naslednje ponavljajoče se dejavnosti skrbništva: (1) ohranjanje: zbiranje in skrb za raziskovalne podatke, (2) deljenje: razkrivanje potenciala podatkov med domenami in (3) odkrivanje: spodbujanje ponovne uporabe in novih kombinacij podatkov. Podatkovno skrbništvo se ukvarja bolj z vzdrževanjem in upravljanjem metapodatkov kot s samo bazo podatkov, zato se velik del procesa obdelave podatkov vrti okoli zaužitja metapodatkov (na primer sheme, poizvedbe itd.) (Knight, 2017).

uporabni (Barend idr., 2020). Načela FAIR so vodilo za ustvarjanje digitalnih objektov, ki bi bili čim bolj ponovno uporabni. Zasnovana so bila kot referenca pri sprejemanju optimalnih odločitev pri številnih vidikih ustvarjanja podatkov in orodij ter (ponovne) uporabe in dolgoročnega upravljanja podatkov. Namen članka je predstaviti zgodovinski razvoj in uveljavljanje načel FAIR ter različna orodja in pristope, ki so se razvili za ocenjevanje primernosti podatkov za drugo rabo. Članek je razdeljen na štiri dele. V prvem delu predstavljam zgodovinski razvoj načel FAIR in razvoj njihove vsebine, v drugem izpostavljam kontekst razvoja in predstavim nekaj značilnosti samih načel. V tretjem delu predstavim izbrana orodja za ocenjevanje doseganja načel FAIR, v četrtem delu pa podam razmislek o povezavi načel FAIR z delovanjem (zaupanja vrednih) podatkovnih repozitorijev ter predstavim nekaj aktualnih podatkov o tem, kako so podatkovni repozitoriji uspešni pri njihovem doseganju.

Nastanek načel FAIR in njihov razvoj

Elementi posameznih načel FAIR so bili v strokovnih krogih že dolgo predmet razprav, vendar pa je do poimenovanja in poskusa sistematizacije prvič prišlo leta 2014, ko je nizozemska Univerza v Leidnu organizirala delavnico »Designing a Data FAIRport«. Rezultat delavnice je bilo oblikovanje minimalnega nabora skupnostno sprejetih načel FAIR, ki so se osredotočala na upravljanje in skrbništvo nad znanstvenimi podatki in ki vsebujejo smernice za dobro prakso upravljanja z raziskovalnimi podatki, ki izboljšujejo najdljivost (F – *Findable*), dostopnost (A – *Accessible*), interoperabilnost (I – *Interoperable*) in možnost ponovne rabe podatkov (R – *Reusable*). Za razumevanje načel FAIR je pomembno poudariti, da je bila ta delavnica organizirana v okviru naravoslovnih znanosti in v tej luči moramo tudi razumeti njihovo strojno berljivo zasnovo. Predlagana načela so bila prvotno objavljena na spletni strani Force11 in odprta za razpravo v znanstveni skupnosti (Force11, 2021), leta 2016 pa so bila podrobneje opredeljena in operacionalizirana v daljšem znanstvenem članku v reviji *Scientific Data* (Wilkinson idr., 2016). Članek je bil odmeven, saj je bil predmet številnih razprav na različnih blogih ter v strokovnih in znanstvenih krogih, obsežno pa so ga citirali tudi drugod (Barend idr., 2020: 2). Z dodatno operacionalizacijo so vsa štiri načela dobila dodatna merila, ki jih morajo podatki dosegati, da so lahko ocenjeni kot FAIR.

Načela se nanašajo na tri vrste entitet: podatke (ali kateri koli drug digitalni objekt), metapodatke (informacije o tem digitalnem objektu) in infrastrukturo. Prvo načelo se nanaša na *najdljivost* raziskovalnih podatkov,

saj je prvi korak pri ponovni rabi podatkov možnost njihovega najdenja. Metapodatki in podatki morajo biti enostavno najdljivi tako za ljudi kot za stroje. Strojno berljivi metapodatki so bistvenega pomena za samodejno odkrivanje podatkovnih nizov in storitev, zato je to eden bistvenih delov procesa FAIRifikacije. Dodatna merila prvega načela so:

F1. (Meta)podatki imajo dodeljen stalni enoznačni identifikator.

F2. Bogati metapodatki opisujejo podatke (*dodatno definirano v R1. spodaj*).

F3. Metapodatki jasno in izrecno vključujejo identifikator podatkov, ki jih opisujejo.

F4. (Meta)podatki so registrirani in indeksirani v iskalnem viru (*komponenta infrastrukture*).

Drugo načelo se nanaša na *dostopnost*. Ko uporabnik najde zahtevane podatke, mora vedeti, kako bo lahko do njih dostopal. To po možnosti vključuje tudi preverjanje pristnosti in avtorizacijo. Dodatna merila drugega načela so:

A1. (Meta)podatke lahko pridobimo prek njihovega identifikatorja s pomočjo standardiziranega komunikacijskega protokola.

A.1.1. Protokol je odprt, brezplačen in univerzalno sprejet.

A.1.2. Protokol dovoljuje postopek preverjanja pristnosti in avtorizacije.

A2. Metapodatki so dostopni, tudi če podatki niso več na voljo.

Tretje načelo se nanaša na *interoperabilnost*, saj je treba podatke zelo pogosto kombinirati z drugimi podatkovnimi viri, da lahko izvedemo zastavljene raziskovalni načrt. Poleg tega morajo podatki sodelovati z različnimi aplikacijami ali delovnimi tokovi za analizo podatkov, shranjevanje in njihovo obdelavo. Dodatna merila tretjega načela so:

I1. (Meta)podatki uporabljajo formalni, dostopen in splošno sprejet jezik za predstavljanje informacij.

I2. (Meta)podatki uporabljajo kontrolirana besedišča, ki sledijo načelom FAIR.

I3. (Meta)podatki vključujejo kvalificirane reference na druge (meta)podatke.

Četrto načelo se nanaša na *možnost ponovne rabe podatkov*. Končni cilj načel FAIR je optimizacija na tem področju. Da bi to dosegli, je treba metapodatke in podatke dobro opisati, da jih je mogoče replicirati in/ali kombinirati v drugih kontekstih. Dodatna merila četrtega načela so:

R1. (Meta)podatki so bogati v smislu opisa množice natančnih in relevantnih lastnosti.

R1.1. (Meta)podatki so objavljeni z jasno in dostopno licenco, ki določa pogoje uporabe podatkov.

R1.2. (Meta)podatki vsebujejo podrobne podatke o poreklu (*provenienca*).

R1.3. (Meta)podatki ustrezajo skupnostnim standardom.

Če torej pogledamo celoten konglomerat načel in meril, lahko trdimo, da biti FAIR v resnici pomeni, da skrbimo za kakovostno in odgovorno upravljanje z raziskovalnimi podatki, ki jih ustvarjamo.

Načela FAIR so zelo dobro poznana med oblikovalci politik, financierji in ponudniki storitev, manj pa med raziskovalci samimi. Raziskava Digital Science (2020), ki je bila izvedena med raziskovalci z vsega sveta, kaže, da dokaj slabo poznajo načela FAIR v povezavi z odprtimi podatki. Podatki zadnjih treh let kažejo, da poznavanje narašča, a še vedno le eden od štirih raziskovalcev resnično pozna načela, medtem ko jih skoraj 40 odstotkov niti leta 2020 še nikoli ni slišalo zanje. To dejstvo samo po sebi morda niti ni problematično, saj biti FAIR pomeni, da odgovorno ravnamo s svojimi podatki, vendar pa kljub temu nekaj pove o izobraževanju in informiranosti raziskovalcev o pomembnosti načel FAIR v odprti znanosti.

Od idejne zasnove načel FAIR do danes je še vedno veliko zmede in nejasnosti glede tega, kaj pomenijo v praksi. Menim, da je narobe govoriti o raziskovalnih podatkih, ki so FAIR, in tistih, ki niso; pravilneje bi bilo govoriti o kontinuumu, saj bi moral vsak raziskovalec svoje podatke poskušati narediti za čim bolj FAIR. Doseganje načel je ideal, za katerega bi si morali prizadevati vsi, vsekakor pa velja, da ga je v nekaterih situacijah povsem nemogoče doseči. To kažejo tudi razprave o tem, da so potrebe in standardi ter cilji in prioritete različnih disciplin različni. Prav tako je veliko zmede glede razlike med odprtimi podatki in podatki FAIR. Lahko imamo namreč nabor podatkov, ki je stoddostno odprtodostopen, kar pa ne pomeni, da nujno dosega tudi načela FAIR. Raziskovalni podatki so lahko odprti, a če ni dobrega opisa podatkov v obliki metapodatkov, če podatki niso opremljeni s stalnimi enoznačnimi identifikatorji ipd., ne ustrezajo načelom FAIR. Velja tudi obratno. Lahko imamo raziskovalne podatke, ki ustrezajo načelom FAIR, vendar če vsebujejo zelo občutljive (osebne) podatke, ne bodo nikoli na voljo. Termina odprti podatki in FAIR podatki tako zagotovo nista sinonima.

Kontekst in značilnosti načel FAIR

Koristno je podati kontekst nastanka nekaterih konceptov, uporabljenih pri oblikovanju vodilnih načel, ki so v skupnosti zgodnjih posvojiteljev povzročili zmedo. Med njimi sta, povzeto po Jacobsen idr. (2020: 12–13), najpomembnejša dva:

(1) *Strojno delovanje*: Štiri temeljna načela – najdljivost, dostopnost, interoperabilnost in ponovna raba – opisujejo temeljne cilje, ki bi, če bi bili doseženi, strojem omogočili optimalno uporabo podatkovnih virov. Laično rečeno: FAIR zahteva, da »stroj ve, kaj mislimo«. To tehnično dosežemo tako, da vsak digitalni vir strojno preberemo. Tako lahko digitalni vir po tej obdelavi uporabljamo tudi kot sredstvo ali substrat za strojno učenje in pristope umetne inteligence. To vpliva na vsa štiri temeljna načela:

- *Najdljivost*: digitalne vire bi bilo enostavno najti tako za ljudi kot za stroje. Obsežni strojno berljivi metapodatki so bistveni za samodejno odkrivanje ustreznih naborov podatkov in storitev, zato so bistven sestavni del procesa FAIRifikacije.
- *Dostopnost*: Protokoli za pridobivanje digitalnih virov morajo biti izrecno določeni tako za ljudi kot za stroje, vključno z natančno določenimi mehanizmi za avtomatizirano pridobitev dovoljenja za dostop do zaščitenih podatkov.
- *Interoperabilnost*: Če sta dva (ali več) digitalna vira povezana z isto temo ali entiteto, mora biti strojem omogočeno, da informacije združijo v bogatejši, enoten pogled na to entiteto. Podobno velja: kadar je digitalno enoto mogoče obdelati s spletno storitvijo, mora biti naprava sposobna samodejno zaznati to skladnost ter olajšati interakcijo med podatki in tem orodjem. To zahteva, da je pomen (semantika) vsakega sodelujočega vira – ne glede na to, ali gre za podatke in/ali storitve – jasen.
- *Ponovna raba*: digitalni viri so dovolj dobro opisani tako za ljudi kot za stroje, tako da je naprava sposobna odločiti: ali naj bo digitalni vir ponovno uporabljen (tj. ali je to pomembno za določeno nalogo); ali je mogoče digitalni vir ponovno uporabiti in pod kakšnimi pogoji (tj. ali izpolnjuje pogoje ponovne uporabe); in kateri sestavni del digitalnega vira se ponovno uporabi, če ga uporabimo (granularnost identifikatorjev).

(2) *Metapodatki*: pojma »podatki« in »metapodatki« se pojavljata v 15 vodilnih načelih FAIR. V prvotnem prispevku, v katerem so bila načela predstavljena, je navedeno, da se podatki uporabljajo za sklicevanje na vse digitalne vire (ne samo podatke v omejenem smislu, ampak tudi na primer programska orodja). Metapodatki so kateri koli opis vira, ki lahko služi namenu iskanja in/ali ponovni rabi in/ali interpretaciji in/ali oceni tega vira. Načela FAIR vsak podatek in metapodatek obravnavajo izolirano, kar pomeni, da je metapodatek deskriptor in podatek stvar, ki se nedvoumno opisuje v okviru para podatki/metapodatki. To velja tudi, če so v drugem

kontekstu metapodatki tista stvar, ki se opisuje. To že samo po sebi pomeni, da morajo biti metapodatki tudi FAIR digitalni vir.

Načela FAIR imajo nekaj specifičnih značilnosti. Prva je, da se osredotočajo na podatke in na informacije, povezane z njimi; dodatna merila pa se nanašajo na podatke in metapodatke. Druga značilnost je, da v samih načelih FAIR vidimo njihov razvoj od začetne abstrakcije do oblikovanja meritev in izvedb – načela so bila torej oblikovana zelo splošno, nato pa je sledila njihova operacionalizacija prek meril. Danes lahko govorimo o fazi, ko različne skupnosti in pobude oblikujejo načine implementacije. Tovrstne pobude (na primer različni mednarodni projekti) izdelujejo orodja in smernice za raziskovalce, financerji in založniki pa razvijajo instrumente in orodja, ki bi omogočala ocenjevanje upoštevanja načel FAIR pri že obstoječih podatkih. Pojavljajo se tudi orodja, ki pomagajo pri iskanju načinov, kako podatke narediti še bolj FAIR, saj lahko ocena upoštevanja načel FAIR za posamezen nabor podatkov ali digitalni predmet poteka neodvisno od časa, v katerem je nabor oziroma predmet nastal. Lahko rečemo, da smo na poti k ustvarjanju celotnega ekosistema FAIR orodij, podatkov in storitev, kar bo pomenilo izboljšanje tehnične kakovosti podatkov, zaradi česar bodo postali bolj uporabni in s čimer bomo prispevali k odprti znanosti.

Implementacija načel FAIR

Naravoslovna podlaga oziroma izhodišče podatkov FAIR in njihovo hitro sprejemanje na mednarodni ravni sta privedla do številnih interpretacij, saj načela ne določajo natančno, kako jih dosegati (FAIRness), temveč opisujejo kontinuiteto lastnosti, atributov in vedenj, ki digitalni objekt približajo temu cilju. Kot rezultat tega so bile že razvite številne metodologije za ocenjevanje doseganja načel FAIR, različne skupine pa trenutno razvijajo orodja za ocenjevanje. Zaradi pomanjkanja skupnega nabora temeljnih meril za ocenjevanje upoštevanja načel FAIR raziskovalci in organizacije ne morejo oceniti pripravljenosti in ravni izvajanja svojih podatkovnih nizov v skladu z njimi. Poleg tega raziskovalne organizacije in podatkovne infrastrukture ne morejo razviti ali slediti minimalnemu naboru skupnih smernic za vzpon po lestvici FAIR zaradi povečane heterogenosti ponujenih orodij za merjenje FAIR. To je privedlo do tega, da se številne skupnosti sprašujejo »Kaj je FAIR?« in »Kako FAIR smo trenutno?«. Vendar pa so začetniki načel FAIR že naleteli tudi na vprašanje: »Kako lahko postanemo (bolj) FAIR?« Nanj je težje odgovoriti, saj načela ne predpisujejo nobenega posebnega stan-

darda ali izvedbe. Poleg tega še ne obstaja zrel ekosistem orodij, platform in standardov za podporo človeškim in strojnimi agentom pri upravljanju, izdelavi, objavi in porabi podatkov FAIR na uporabniku prijazen in učinkovit (tj. »enostaven«) način (Jacobsen idr., 2020). Velja omeniti, da načela niso namenjena ocenjevanju in kaznovanju, če kdo ni FAIR, ampak jih je treba razumeti kot ideal, h kateremu težimo – namen je ozaveščati deležnike o merilih (na primer raziskovalce, podatkovne storitve itd.), ti pa nato iščejo primerne možnosti za izboljšave glede na vire, ki so jim na voljo.

Za reševanje te problematike je bila pod okriljem Research Data Alliance (RDA) ustanovljena delovna skupina FAIR Data Maturity Model WG (2021), ki je razvila sklop kazalnikov z ravnmi zrelosti; ti so bili v prvi vrsti namenjeni izvajalcem ocenjevalnih orodij za merjenje, kako dobro podatki dosegajo načela FAIR. To delo se osredotoča na to, »kaj« je treba oceniti, in ni namenjeno temu, »kako« bi bilo mogoče kazalnike ovrednotiti v praksi.

Delovna skupina WDS/RDA Assessment of Data Fitness for Use (RDA, 2021) je razvila merila, ki zajemajo načela FAIR ter vidike kakovosti in obdelave podatkov; ti naj bi služili kot dopolnila k zahtevam, ki jih morajo podatkovni repozitoriji izpolniti, da pridobijo certifikat CoreTrustSeal.² Delovna skupina je razvila spletni vprašalnik (Austin idr., 2019), da bi pregledovalci ročno ocenili podatke o merilih.

Podatkovni repozitorij Data Archiving and Networked Services (DANS) je razvil dva prototipa, ki dokazujeta oceno doseganja načel FAIR za podatke različnih zainteresiranih strani. FAIRdat je namenjen pregledovalcem podatkov, medtem ko FAIR Enough? naslavlja manj izkušene raziskovalce s poudarkom na večjem razumevanju pomena podatkov FAIR (Devaraju idr., 2021).

Projekt FAIRsFAIR³ k uvajanju načel FAIR za podatke v Evropskem oblaku odprte znanosti (EOSC)⁴ prispeva z razvijanjem praktičnih rešitev (strokovno

2 CoreTrustSeal je mednarodna nevladna in neprofitna skupnostna organizacija, ki promovira trajnostne in zaupanja vredne podatkovne infrastrukture. Osnovno certificiranje vključuje minimalno intenziven postopek, s katerim podatkovni repozitoriji predložijo dokaze, da so trajnostni in vredni zaupanja. Certifikat CoreTrustSeal je prvi korak v globalnem okviru za certificiranje repozitorija, ki vključuje razširjeno certificiranje (nestor-Seal DIN 31644) in formalno certificiranje (ISO 16363) (CoreTrustSeal, 2021).

3 Namen projekta FAIRsFAIR (Fostering Fair Data Practices in Europe) je zagotoviti praktične rešitve za uporabo načel FAIR v celotnem življenjskem ciklu raziskovalnih podatkov. Poudarek je na spodbujanju FAIR podatkovne kulture in sprejemanju dobrih praks pri oblikovanju FAIR podatkov. Projekt ima ključno vlogo pri razvoju globalnih standardov za FAIR certificiranje podatkovnih repozitorijev in njihovih podatkov, kar bo prispevalo k oblikovanju tistih politik in praks, ki bodo program EOSC spremenile v delujočo infrastrukturo (za več glej FAIRsFAIR, 2021).

4 Zamisel o Evropskem oblaku za odprto znanost (EOSC) se je oblikovala leta 2015 kot vizija Evropske komisije o veliki infrastrukturi za podporo in razvoj odprte znanosti in odprtih inovacij

znanje, priporočila, usposabljanje in orodja), ki olajšajo uporabo načel v celotnem življenjskem ciklu raziskovalnih podatkov. Priporočilo strokovne skupine Evropske komisije je, da je treba razviti in preizkusiti metrike ocenjevanja, ki pripravljajo načela FAIR in orodja za njihovo izvajanje, da bi ljudem in strojem olajšali oceno njihovega doseganja za raziskovalne podatke (Strokovna skupina Evropske komisije za podatke FAIR, 2018). V odgovor na ta priporočila je več skupin predlagalo metrike ocenjevanja za oceno izvajanja načel, izpostaviti velja zlasti delo delovne skupine FAIR Data Maturity Model Working Group (2020).

Trenutno delo na področju ocenjevanja uresničevanja načel FAIR za ocenjevanje podatkov naslavlja vprašanje, »kaj« je mogoče ovrednotiti z meritvami. Še vedno ostajajo nejasnosti glede tega, kako je mogoče te meritve preizkusiti v praksi. Delovna skupina RDA FAIR Data Maturity Model ugotavlja, da je »natančen način ocenjevanja podatkov na podlagi temeljnih meril odvisen od pristopov ocenjevanja ob upoštevanju zahtev posamezne skupnosti« (Herczog idr., 2019). Projekt FAIRsFAIR izvaja in preizkuša metrike ocene podatkov FAIR z več zainteresiranimi stranmi FAIR po iterativnem pristopu, ki temelji na primerih (Devaraju idr., 2021).

Uresničevanje načel FAIR je za posamezen podatkovni predmet mogoče oceniti ročno, polsamodejno ali samodejno v več fazah v celotnem življenjskem ciklu raziskave (glej sliko 1). Da bi bolje razumeli, kaj je treba upoštevati pri izvajanju ocen FAIR, kot je predlagano na sliki 1, so v okviru projekta FAIRsFAIR razvili vrsto scenarijev (Devaraju in Herterich, 2020). Med njimi so izbrali dva primera primarne uporabe projekta. To sta:

- Zainteresirane strani (na primer institucije, ponudniki podatkovnih storitev) ponujajo splošno orodje za samoocenjevanje, s katerim se raziskovalci izobražujejo in ozaveščajo o tem, da so njihovi podatki FAIR še pred objavo – to je orodje FAIR Aware (2021).
- Ponudnik podatkovnih storitev (podatkovno skladišče, podatkovni portal ali register), ki se zavzema za zagotavljanje podatkov FAIR, želi programsko meriti nabore podatkov glede njihove stopnje »FAIRness« skozi čas – to je orodje F-UJI (2021).

Cilj spletnega orodja za samoocenjevanje FAIR-Aware je povečati ozaveščenost raziskovalcev o pomenu podatkov FAIR pred deponiranjem v odlagališče. Priprava podatkov je za številne raziskovalce v različnih disciplinah še vedno nejasen postopek. Da bi jim pomagal premostiti to vrzel v znanju, FAIR-Aware poudarja izobraževanje in ozaveščanje o FAIR podatkih,

v Evropi in širše. Ideja je, da EOSC postane evropsko virtualno okolje za vse raziskovalce, ki bodo podatke hranili, upravljali, analizirali in ponovno uporabljali za raziskovalne, inovacijske in izobraževalne namene (za več glej EOSC, 2021).

ne pa merjenje obsega, v katerem so njihovi nabori podatkov FAIR. Spodbuja praktično razumevanje načel FAIR ter tega, kako lahko povečajo vrednost in vpliv podatkov. Deset ocenjevalnih vprašanj izhaja iz specifikacije meritev objektov FAIRsFAIR (Devaraju idr., 2020) in zajema vsa merila. Informativni nasveti za vsako vprašanje ponujajo dodatna pojasnila ter kontekst s praktičnimi primeri in navodili.



Slika 1: Scenariji izvajanja ocenjevanja FAIR znotraj življenjskega kroga podatkov (Devaraju in Herterich, 2020).

Na drugi strani je orodje F-UJI. Avtomatizirana ocena podatkovnega predmeta »FAIRness« mora zajemati širok spekter ponudbe disciplinarnih ali področnih podatkov, zato se mora osredotočiti na dokaj ozek nabor najboljših praks in standardov, ki so agnostični za področje in so se razvili v zadnjih letih. V ta namen so pri projektu FAIRsFAIR razvili avtomatizirano orodje za ocenjevanje F-UJI (Devaraju in Huber, 2020), ki omogoča ocenjevanje posameznih podatkovnih objektov na podlagi osnovnih meril FAIR. Orodje izvede oceno na podlagi identifikatorja podatkovnega objekta (PID ali URL) ter temelji na obstoječih spletnih standardih in najboljših praksah, ki jih za raziskovalne podatke potrdijo ponudniki trajnih identifikatorjev (PID).

Načela FAIR in pomen zaupanja vrednih arhivov

»Raziskovalni podatki ne bodo postali ali ostali FAIR s čarovnijo. Za izgradnjo, upravljanje in vzdrževanje raziskovalnih podatkovnih

infrastruktur potrebujemo usposobljene ljudi, pregledne procese, interoperabilne tehnologije in sodelovanje» (Kleemola, 2018). Za FAIR podatke niso dovolj le raziskovalci, ki jih ustvarjajo, potrebujemo celoten ekosistem, ki podpira načela FAIR. Morda se to zdi nemogoče, vendar ne smemo pozabiti, da ta ekosistem deloma že obstaja v obliki podatkovnih infrastruktur, ki ravnajo s podatki in jih hranijo – vprašanje je le, kako obstoječe infrastrukture uporabiti za doseganje načel. Če pogledamo načelo najdljivosti, vidimo, da morajo podatki in metapodatki zadostiti štirim merilom: da imajo trajni identifikator, da morajo biti podatki opisani z bogatimi metapodatki, za kar je potrebna obsežna dokumentacija, ter da morajo metapodatki vključevati identifikator za podatke, ki jih opisujejo. Četrto merilo pravi, da so podatki in metapodatki registrirani in indeksirani v iskalnem viru, češar pa ne moremo urediti znotraj podatkovnega niza, temveč potrebujemo repozitorij, ki ponuja ustrezne tehnologije, postopke, strokovno znanje in ljudi. Podatkovni repozitorij, v katerem so podatki, zagotavlja, da je mogoče iskati po podatkovnem nizu. Tudi če pogledamo prve tri metrike pri načelu najdljivosti, lahko ugotovimo, da običajno niso povezane z nečim, kar raziskovalec sam organizira, temveč se pojavijo, ko so podatki deponirani v podatkovnem repozitoriju. Podatkovni repozitorij poskrbi, da podatki dobijo stalni enoznačni identifikator, predpiše ustrezno metapodatkovno shemo ipd. Podatkovni repozitoriji imajo torej resnično pomembno vlogo, ko gre za obdelavo podatkov, ki sledijo načelom FAIR.

Do danes ostaja poročilo Evropske strokovne skupine Turning FAIR into Reality (Strokovna skupina Evropske komisije za podatke FAIR, 2018) skupna referenca v Evropi in širše, prav tako pa je izhodišče za številne delovne svežnje ter projekte FAIR in Evropski portal odprte znanosti. Poročilo opisuje ekosistem FAIR, ki se opira na politike, načrte ravnanja z raziskovalnimi podatki (NRRP), identifikatorje, standarde in podatkovne repozitorije. Podatkovni repozitoriji ponujajo zbirke podatkov in podatkovne storitve ter bi morali biti certificirani v skladu z novimi standardi zanesljivosti delovanja in načeli FAIR. Za doseganje zahtev certifikata CoreTrustSeal je treba spodbujati in podpirati podatkovne repozitorije, ki zagotavljajo dolgoročno skrbništvo nad podatki.

Poleg tega, da digitalni predmeti postanejo FAIR, se morajo sčasoma tudi ohraniti kot FAIR. To je naloga, ki običajno zahteva vire in strokovno znanje. V nadaljevanju razprave je vidik ohranjanja pogosto podcenjen ali pozabljen. Delovna skupina FAIR izvršnega odbora EOSC meni, da je CoreTrustSeal prava raven za podatkovne repozitorije, in priporoča, da je treba temeljito preizkusiti pristop CoreTrustSeal + FAIR, ki ga predlaga projekt FAIRsFAIR. Tabela 1 predstavlja prekrivanje zahtev CTS in 15 meril načel

FAIR. Jasno je, da se veliko zahtev za repozitorije res posredno in neposredno nanaša na upoštevanje načel FAIR (L'Hours idr., 2020).

Tabela 1: Prepletanje zahtev certifikata CoreTrustSeal in načel FAIR (L'Hours idr., 2020: 16). Zahteve CoreTrustSeal pišem v ležečem tisku.

F	R13 R15	<p>F1. (Meta)podatki imajo dodeljen stalni enoznačni identifikator.</p> <p>F2. Bogati metapodatki opisujejo podatke.</p> <p>F3. Metapodatki jasno in izrecno vključujejo identifikator podatkov, ki jih opisujejo.</p> <p>F4. (Meta)podatki so registrirani in indeksirani v iskalnem viru.</p> <p><i>R13. Iskanje in identifikacija podatkov</i></p>
A	R15 R16 R10	<p>A1. (Meta)podatke lahko pridobimo prek njihovega identifikatorja s pomočjo standardiziranega komunikacijskega protokola.</p> <p>A.1.1. Protokol je odprt, brezplačen in univerzalno sprejet.</p> <p><i>R15. Tehnična infrastruktura</i></p> <p>A.1.2. Protokol dovoljuje postopek preverjanja pristnosti in avtorizacije.</p> <p><i>R16. Varnost</i></p> <p>A2. Metapodatki so dostopni, tudi če podatki niso več na voljo.</p> <p><i>R10. Načrt ohranjanja</i></p>
I	R15 R11	<p>I1. (Meta)podatki uporabljajo formalni, dostopen in splošno sprejet jezik za predstavljanje informacij.</p> <p>I2. (Meta)podatki uporabljajo kontrolirana besedišča, ki sledijo načelom FAIR.</p> <p><i>R15. Tehnična infrastruktura</i></p> <p>I3. (Meta)podatki vključujejo kvalificirane reference na druge (meta) podatke.</p> <p><i>R11. Kvaliteta podatkov</i></p>
R	R11 R2 R7 R15	<p>R1. (Meta)podatki so bogati v smislu opisa množice natančnih in relevantnih lastnosti.</p> <p><i>R11. Kvaliteta podatkov</i></p> <p>R1.1. (Meta)podatki so objavljeni z jasno in dostopno licenco, ki določa pogoje uporabe podatkov.</p> <p><i>R2. Licence</i></p> <p>R1.2. (Meta)podatki vsebujejo podrobne podatke o poreklu.</p> <p><i>R7. Integriteta podatkov in avtentičnost</i></p> <p>R1.3. (Meta)podatki ustrezajo skupnostnim standardom.</p> <p><i>R15. Tehnična infrastruktura</i></p>

Vse repozitorske storitve so del določenih organizacij in prav na ta organizacijski vidik je osredotočen osnovni certifikat CoreTrustSeal. Podatke v podatkovnem repozitoriju s certifikatom CoreTrustSeal upravljamo, urejamo in dolgoročno hranimo tako, da so in ostanejo FAIR za določeno ciljno skupnost repozitorija (Mokrane in Recker, 2019).

Poleg CoreTrustSeal se razvijajo tudi (avtomatizirane) ocene FAIR za digitalne predmete, ki sledijo načelom, trenutno pa sta najbolj napredna in sprejeta že omenjena metoda FAIR Maturity Evaluation Service (Wilkinson idr., 2019) in ocenjevalno orodje F-UJI Automated FAIR Data (Devaraju idr., 2020). Projekt EOSC Nordic (EOSC Nordic, 2021) je s temi orodji preizkusil in še naprej preizkuša vzorec zapisov metapodatkov iz nordijskih in baltskih podatkovnih repozitorijev.

Trenutno je več evropskih projektov posvečenih različnim vidikom načel FAIR ter certificiranju podatkovnih repozitorijev, storitev in digitalnih predmetov. Med njimi izpostavljam skupino, ki je bila ustanovljena za povečanje sodelovanja, zmanjšanje podvajanja in spodbujanje upoštevanja poročila Turning FAIR into Reality (Strokovna skupina Evropske komisije za podatke FAIR, 2018), FAIR Synchronisation Force (za pregled njenih rezultatov glej Dillo idr., 2020).

Glavna omejitev podatkovnega repozitorija je, da se osredotoča na objavo svojih podatkov predvsem na človeškem vmesniku in ne posveča dovolj pozornosti strojni berljivosti podatkov. Koncept digitalnega objekta FAIR igra pomembno vlogo pri trajni in inteligentni povezavi podatkov in metapodatkov s povezanimi nabori podatkov, vključno z obratnim sklicevanjem. Drugi ključni element v procesu FAIRifikacije je vse večja odvisnost od podatkovnih znanstvenikov, ki lahko raziskovalcem in analitikom pomagajo pri strukturiranju in učinkovitem ohranjanju raziskovalnih podatkov. Vloga teh podatkovnih skrbnikov postaja vse pomembnejša za organizacije po vsem svetu. Ta razvoj vidimo v mnogih državah, v katerih univerze in raziskovalni inštituti začenjajo zaposlovati skrbnike podatkov.

Spletni seminar »FAIRification STEP 2 on DATA / METADATA Webinar« (FAIRification, 2021), ki je potekal februarja 2021, je predstavil proces evalvacije doseganja načel FAIR pri okoli 100 podatkovnih repozitorijih, pri čemer so v vsakem ročno izbrali deset naključnih podatkov in jih ocenili s pomočjo orodja F-UJI. Ni presenetljivo, da se večina ocenjenih repozitorijev ni izkazala zelo FAIR v zvezi s strojno berljivimi metapodatki. Vzorec je pokazal, da precejšen delež (27 %) preizkušenih repozitorijev ni podpiral metapodatkov, ki jih je mogoče strojno brati. Približno 34 % podatkovnih repozitorijev je imelo malo metapodatkov, ki jih je mogoče strojno brati. Le peščica podatkovnih repozitorijev (15 %) je dosegla več kot 40 %. Povprečna

ocena 75 ocenjenih podatkovnih repozitorijev je bila 22 %. Vendar je bilo opazno večje število točk med repozitoriji, ki se izvajajo na uveljavljenih platformah (na primer Dataverse, Figshare itd.) (38 %), in certificiranimi repozitoriji (29 %).

Sklep

Načela FAIR so zbirka najboljših praks v zvezi z ravnanjem s podatki, ki je dovolj splošna, da je veljavna med različnimi znanstvenimi disciplinami. Pokazalo se je, kako lahko slabo hranjeni podatki ovirajo njihovo ponovno uporabo in upočasnijo znanstveni napredek. Ocena objektov FAIR je sestavni del ekosistema FAIR ter zahteva vzporedno razvijanje in izboljševanje storitev, ki FAIR omogočajo. Zaupanja vredni podatkovni repozitoriji igrajo pomembno vlogo pri zagotavljanju stalnega dostopa do digitalnih predmetov in njihovih metapodatkov ter njihovega dolgoročnega shranjevanja. Vlogo ponudnikov podatkovnih storitev pri omogočanju načel FAIR bi bilo treba priznati in ceniti, saj so »pooblaščenca« med različnimi zainteresiranimi stranmi (na primer raziskovalci in financerji).

Z vidika Evropske komisije so načela FAIR v zvezi s politiko ključna za spodbujanje politike odprte znanosti, ki vključuje široko in zgodnjo izmenjavo znanja (na primer podatkov) in orodij raziskovalcev znotraj disciplin in med njimi ter z družbo na splošno. Načela FAIR so v središču Evropskega oblaka za odprto znanost, pobude za izgradnjo zaupanja vrednega, odprtega in distribuiranega sistema za znanstveno skupnost, ki raziskovalcem zagotavlja nemoten dostop do spleta podatkov in storitev FAIR, zgrajenega na podlagi teh podatkov. Kar zadeva vlogo Komisije kot financerke, so načela FAIR zelo poudarjena v okvirnem programu za raziskave in inovacije Obzorje Evropa (Evropska komisija, 2021). Komisija se bo zavzela za »odgovorno upravljanje podatkov o raziskavah v skladu z načeli FAIR«. Prijavitelji projektov bodo v fazi predloga ocenjeni glede svojih načrtov ravnanja s podatki in zlasti glede tega, kako nameravajo svoje podatke narediti FAIR. To bo še naprej pomembno v času trajanja projekta, saj bodo upravičenci morali o svojih podatkih poročati prek načrta ravnanja s podatki, ki ga bo treba med projektom posodobiti. V tej luči je bistveno izobraževanje in usposabljanje raziskovalcev, da bodo razumeli in znali uveljavljati načela FAIR pri pripravi podatkov.

Zaupanja vredni podatkovni repozitoriji dodajo vrednost, ker omogočajo osnovno raven upoštevanja načel FAIR za podatke, ki jih hranijo (čeprav

so nekateri morda bolj FAIR kot drugi); prav tako prispevajo k ohranjanju in celo povečanju stopnje upoštevanja načel FAIR z ustreznimi storitvami za pripravo podatkov in s skrbništvom. Da bi zagotovili, da bodo podatki postali in ostali FAIR, moramo oceniti tako podatke kot podatkovne repozitorije. Svoje podatke moramo deliti, da lahko odprto znanost spremenimo v resničnost. Načela FAIR nam pomagajo opredeliti visokokakovostne in pregledne prakse upravljanja raziskovalnih podatkov. Mehanizmi certifikiranja, kot je CoreTrustSeal za digitalne repozitorije, nam pomagajo ustvariti zaupanje v raziskovalno podatkovno infrastrukturo, ki jo potrebujemo za zaščito dostopnosti in dostopnosti naših (FAIR) podatkov v prihodnosti.

Literatura

Austin, Claire, Helena Cousijn, Michael Diepenbroek, Jonathan Petters in Marina Soares e Silva (2019): *WDS/RDA Assessment of Data Fitness for Use WG Outputs and Recommendations*. Dostopno na DOI: 10.15497/rda00034.

CoreTrustSeal (2021): *CoreTrustSeal: About*. Dostopno na: <https://www.coretrust-seal.org/about/> (2. marec 2021).

Devaraju Anusuriya, Mustapha Mokrane, Linas Cepinskas, Robert Huber, Patricia Herterich, Jerry de Vries, Vesa Akerman, Hervé L'Hours, Joy Davidson in Michael Diepenbroek (2021): From Conceptualization to Implementation: FAIR Assessment of Research Data Objects. *Data Science Journal* 20(1): 4. Dostopno na DOI: 10.5334/dsj-2021-004.

Devaraju, Anusuriya in Patricia Herterich (2020): *D4.1 Draft Recommendations on Requirements for FAIR Datasets in Certified Repositories*. Dostopno na DOI: 10.5281/zenodo.3678716.

Devaraju, Anusuriya in Robert Huber (2020): *F-UJI – An Automated FAIR Data Assessment Tool*. Dostopno na DOI: 10.5281/zenodo.4063720.

Devaraju, Anusuriya, Robert Huber, Mustapha Mokrane, Linas Cepinskas, Joy Davidson, Patricia Herterich, Hervé L'Hours, Jerry De Vries in Angus White (2020): *FAIRsFAIR Data Object Assessment Metrics (No. v0.3)*. Dostopno na DOI: 10.5281/zenodo.3934401.

Dillo, Ingrid, Marjan Grootveld, Simon Hodson in Sara Pittonet Gaiarin (2020): *FAIR Synchronisation Force. Second Report of the FAIRsFAIR Synchronisation Force (D5.5) (Version 1.0)*. Dostopno na: <https://www.fairsfair.eu/advisory-board/synchronisation-force17> (20. december 2020).

EOSC Nordic (2021). Dostopno na: <https://www.eosc-nordic.eu>.

Evropska komisija (2021): *Horizon Europe*. Dostopno na: https://ec.europa.eu/info/horizon-europe_en (2. marec 2021).

- FAIR-Aware (2021): *FAIR-Aware: Assess Your Knowledge of FAIR*. Dostopno na: <https://www.fairsfair.eu/fair-aware> (20. december 2020).
- Fair Data Maturity Model Working Group (2020): *FAIR Data Maturity Model: Specification and Guidelines*. Dostopno na DOI: 10.15497/RDA00050.
- FAIRification (2021): *FAIRification STEP 2 on DATA / METADATA Webinar*. Dostopno na: <https://www.eosc-nordic.eu/fairification-step-2-on-data-metadata-webinar/> (20. december 2020).
- FAIRsFAIR (2021): *Fosetrng FAIR Data Practices in Europe*. Dostopno na: <https://www.fairsfair.eu/> (2. marec 2021).
- Force11 (2021): *The FAIR Data Principles*. Dostopno na: <https://www.force11.org/group/fairgroup/fairprinciples> (20. december 2020).
- F-UJI (2021): *F-UJI Automated FAIR Data Assessment Tool*. Dostopno na: <https://www.fairsfair.eu/f-uj-automated-fair-data-assessment-tool> (20. december 2020).
- Herczog, Edit, Keith Russell, Vassilios Peristeras in Makx Dekkers (2019): *FAIR Data Maturity Model*. Dostopno na: https://www.rd-alliance.org/system/files/documents/20191023_FAIR_WG_slides_v0.08.pdf (20. december 2020).
- Jacobsen, Annika, Ricarde de Miranda Azevedo, Nick Juty, Dominique Batista idr. (2020): *FAIR Principles: Interpretations and Implementation Considerations*. *Data Intelligence* 2(2020): 10–29. Dostopno na DOI: 10.1162/dint_r_00024.
- Kleemola, Mari (2018): *Being Trustworthy and FAIR Requires People, Processes, Technologies and Collaboration*. Dostopno na: <https://tietoarkistoblogi.blogspot.com/2018/11/being-trustworthy-and-fair.html> (20. december 2020).
- Knight, Michelle (2017): *What is Data Curation?* Dostopno na: <https://www.dataversity.net/what-is-data-curation/> (3. marec 2021).
- L'Hours, Hervé, Anusuriya Devaraju, Ilona Von Stein in Mustapha Mokrane (2020): *FAIRsFAIR Comments Response on RDA FAIR Data Maturity Model Working Group (2020). FAIR Data Maturity Model: Specification and Guidelines*. Dostopno na DOI: 10.5281/zenodo.3827109.
- L'Hours, Hervé, Ilona Von Stein, Frans Huigen, Anusuriya Devaraju, Mustapha Mokrane, Joy Davidson, Robert Huber idr. (2020): *D4.2 Repository Certification Mechanism: a Recommendation on the Extended Requirements and Procedures (Version 01.00)*. Dostopno na: <https://zenodo.org/record/3835698#.YD3X9mhKiUk> (20. december 2020).
- Mokrane, Mustapha in Jonas Recker (2019): *CoreTrustSeal-Certified Repositories. Enabling Findable, Accessible, Interoperable and Reusable (FAIR) Data. 16th International Conference on Digital Preservation iPRES 2019, Amsterdam, The Netherlands*. Dostopno na: <https://osf.io/9da2x/> (20. december 2020).
- Mons, Barend, Erik Schultes, Fenfhong Liu in Annika Jacobsen (2020): *The FAIR Principles: First Generation Implementation Choices and Challenges*. *Data Intelligence* 2(2020): 1–9. Dostopno na DOI: 10.1162/dint_e_00023.

- RDA (2021): *WDS/RDA Assessment of Data Fitness for Use*. Dostopno na: <https://www.rd-alliance.org/groups/assessment-data-fitness-use> (20. december 2020).
- Digital Science, Mark Hahnel, Leslie McIntosh Borrelli, Alan Hyndman, Grace Baynes, Merce Crosas idr. (2020): *The State of Open Data 2020. Digital Science. Report*. Dostopno na DOI: 10.6084/m9.figshare.13227875.v2.
- Strokovna skupina Evropske komisije za podatke FAIR (2018): *Turning FAIR into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR data*. Dostopno na DOI: 10.2777/1524.
- Wilkinson, Mark D., Michel Dumontier, Ijsbrand Jan Aalbersberg idr. (2016): The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci Data* 3: 160018. Dostopno na DOI: 10.1038/sdata.2016.18.
- Wilkinson, Mark D., Michel Dumontier, Susanna-Assunta Sansone, Luiz Olavo Bonino Da Silva Santos idr. (2019): Evaluating FAIR Maturity Through a Scalable, Automated, Community-Governed Framework. *Sci Data* 6: 174. Dostopno na DOI: 10.1038/s41597-019-0184-5.
- Wilkinson, Mark D., Susanna-Assunta Sansone, Erik Schultes, Peter Doorn, Luiz Olavo Bonino Da Silva Santos in Michel Dumontier (2018): A Design Framework and Exemplar Metrics for FAIRness. *Sci Data* (5). Dostopno na DOI: 10.1038/sdata.2018.118.