

Zaključki konference »ODPRTI RAZISKOVALNI PODATKI« V SLOVENIJI

V Mariboru je od 14. in 15. novembra 2019 potekala konferenca o odprti znanosti v Sloveniji v organizaciji **Združenja za raziskovalne podatke** [Research Data Alliance \(RDA\) vozlišča Slovenija](#) v sodelovanju z Ministrstvom za izobraževanje, znanost in šport, projektom NI4OS (National Initiatives for Open Science in Europe), Univerzitetno knjižnico Maribor, zavodom Arnes in društvom Mlada akademija. Konferenca je ena od aktivnosti, ki jih izvaja RDA Slovenija skozi projekt iz razpisa RDA EU 4.0 v koordinaciji [Arhiva družboslovnih podatkov](#). Cilj vozlišča RDA Slovenija je delovati kot dolgoročna osrednja točka srečevanja med RDA in strokovnjaki s področja dela s podatki, raziskovalnimi agencijami in drugimi zainteresiranimi stranmi v Sloveniji. Del programa konference je bil tudi posvečen organizaciji nacionalne iniciative za odprto znanost in kot tak opredeljen kot nacionalni dogodek mednarodnega projekta NI4OS.¹

Poseben poudarek konference je bil na koordinaciji infrastrukturnih storitev za podporo ravnanju z raziskovalnimi podatki, uveljavitvi načel Nacionalne strategije odprtega dostopa do znanstvenih objav in raziskovalnih podatkov in pregledu trenutne situacije glede dostopa do raziskovalnih podatkov po posameznih področjih.

Gostja **Edit Herczog**, so-voditeljica dveh delovnih skupin v RDA, skupine '**Group of European Data Experts in RDA (GEDE-RDA)**'² in '**FAIR Data Maturity Model**'³ je najprej predstavila pomen in način dela v RDA. RDA je svetovna skupnost za sodelovanje vseh, ki imajo opravka z raziskovalnimi podatki. Člani RDA skozi delovne, interesne skupine in nacionalna vozlišča sporazumno oblikujejo priporočila, ki predstavljajo osnovo za izvajanje in prihodnji razvoj podatkovnih storitev.

Ključni elementi v povezavi z raziskovalnimi podatki so, kakor priporoča **GEDE-RDA** skupina, dodeljevanje trajnih identifikatorjev (PID) in dosledno polno citiranje uporabljenih raziskovalnih podatkov pri znanstvenih objavah.

Glavnina njene predstavitve je bila posvečena **FAIR Data Maturity Modelu**. Delovna skupina je izhajala iz originalnih pojmovnih določil FAIR principov in je na začetku pregledala, kakšne indikatorje so do sedaj že uporabljali v različnih poskusih ocenjevanja posameznih vidikov znotraj FAIR. Za vsak podroben vidik so na tej podlagi izluščili izčrpen in neprekrivajoč nabor indikatorjev, ki so sedaj v fazi preizkušanja. Med prvimi interesenti za preizkušanje in uveljavitev modela je Evropska komisija, ki napoveduje upoštevanje modela pri bodočih razpisih za raziskovalne projekte. Poudarek modela je na spodbujanju uporabnosti in dostopnosti podatkov za druge namene, npr. tudi z izčrpno izpolnjenimi področnimi metapodatki.

¹ <https://ni4os.eu/>, 11. 12. 2019

² <https://www.rd-alliance.org/groups/gede-group-european-data-experts-rda>, 15. 11. 2019.

³ <https://www.rd-alliance.org/groups/fair-data-maturity-model-wg>, 15. 11. 2019.

Predlog št. 1: Izkoristiti FAIR Data Maturity Model za izhodišče pri oblikovanju nacionalnih kriterijev za skladnost raziskovalnih podatkov z zahtevami financerja glede odprtosti, dostopnosti in možnosti ponovne uporabe.

Sebastian Dahle, raziskovalec z Biotehniške fakultete Univerze v Ljubljani je pripravil **Predstavitev izkušenj s pripravo načrta ravnanja z raziskovalnimi podatki**. Njegov poudarek pri izboru ustreznega repozitorija za podatke v primerih, ko ne obstaja določen posebej namenjen za tisto področje, je na enostavnosti uporabe. Po njegovih izkušnjah se po enostavnosti in prijaznosti za uporabo izkaže Zenodo. Načrtovanju ravnanja z raziskovalnimi podatki naj bodo podlaga **FAIR** principi pri izboru in pripravi podatkov, pri tem pa je potrebno osveščati raziskovalce in zagotoviti usposabljanje. Na koncu se oboje, tako priprava načrta ravnanja s podatki kot sledenje principu „odpiranje podatkov, kolikor je mogoče“, absolutno izplača. Opaža, da raziskovalci še niso dobro usposobljeni za sodelovanje pri odprti znanosti, pri čemer bi lahko največ pripomoglo podporno osebje na ustanovah in izobraževanje v podiplomskih programih.

Predlog št. 2: Uvesti obveznost Načrta ravnanja z raziskovalnimi podatki (NRRP/DMP) z namenom 'odpiranja podatkov kolikor mogoče' za javno financirane projekte. Obenem zaradi lažjega izpolnjevanja obveznosti zagotoviti informiranje in usposabljanje raziskovalcev za ravnanje s podatki, ter na raziskovalnih ustanovah zagotoviti osebje za podporo pri pripravi podatkov.

V naslednji sekciji so pristop pri podpori raziskovalnim podatkom predstavile nekatere slovenske področne infrastrukturne storitve, ki so povezane v velike evropske infrastrukturne enote. Sodelavci in sodelavke Arhiva družboslovnih podatkov (**ADP**) so prikazali okvir delovanja, ki jim ga omogoča **CESSDA**, Konzorcij evropskih podatkovnih arhivov s področja družboslovja. Znotraj CESSDA se posamezni nacionalni izvajalci podatkovnih storitev povezujejo v delovne skupine, skupaj z drugimi infrastrukturami pa sodelujejo tudi v različnih mednarodnih projektih, vključno s projekti v povezavi RDA in EOSC. V povezavi s temo konference so predstavili enega od stebrov delovanja CESSDA, to je 'Zaupanje' (Trust), ki je osnovni vezni koncept pri ocenjevanju po kriterijih Core Trust Seal (CTS). ADP sodeluje v projektih CESSDA v povezavi s CTS in drugih projektih, namenjenih pregledu in usmerjanju podatkovnih storitev po kriterijih CTS. Rezultat tega sodelovanja je certifikat CTS, pridobljen na podlagi zagotovil o postopkih digitalnega shranjevanja in dostopa do kakovostnih raziskovalnih podatkov za določeno skupnost uporabnikov. Prikazali so zasnovo posodobljenega postopka vrednotenja raziskovalnih podatkov ob prevzemu od ustvarjalca podatkov. Vrednotenje podatkov in z njimi povezanih metapodatkov je vezni člen med CTS in F.A.I.R (Findable, Accessible, Interoperable, Reusable), pri tem pa je poudarek na ocenjevanju opremljenosti in trajni vrednosti podatkov za ponovno uporabo.

Predlog št. 3: Na podlagi FAIR principov naj se pripravijo načini vrednotenja raziskovalnih podatkov, ki upoštevajo področne posebnosti posameznih disciplin, s poudarkom na kakovosti in vrednosti za ponovno uporabo za znanstvene namene in za javno korist.

Tomaž Erjavec z Instituta "Jožef Štefan" je predstavil slovenske infrastrukturne storitve za jezikovne vire in tehnologije CLARIN.SI, ki je del evropske istoimenske raziskovalne infrastrukture **CLARIN** (Common Language Resources and Technology Infrastructure). Znotraj RDA je interesna skupina za jezikovne podatke, ki pripravlja usmeritve za njihovo citiranje. CLARIN.SI si prizadeva za njihovo uveljavljanje v nacionalnem okolju, kjer še vedno obstajajo različne delitve med bolj računalniško usmerjenimi in tradicionalnimi jezikoslovci.

Predstavil je sistem stalnih identifikatorjev (PID – Persistent Identifiers), ki ga uporabljajo, ter notranje (CLARIN B Center) ter zunanje certificiranje po CTS. CLARIN.SI je bil prvi slovenski repozitorij, ki je bil prejemnik predhodnika CTS, tj. Data Seal of Approval (DSA). V pomoč jim je uporaba jezikovnim virom prilagojene splošne repozitorijske platforme DSpace, ki so jo razvili češki partnerji v CLARIN, imajo pa tudi potrebno institucionalno podporo pri tehnoloških vidikih storitev, ki jo zagotavlja Institut “Jožef Štefan”. Predstavil je tudi upoštevanje načela FAIR. Posebej je izziv dosegati interoperabilnost, saj tudi ob uporabi istega metapodatkovnega okvira, kot je TEI (Text Encoding Initiative) še vedno ostajajo razlike v načinu uporabe določenih elementov. Pomembno je tudi usposabljanje za delo s podatki in motivacija ustvarjalcev podatkov za predajo, pri čemer bi kot spodbuda lahko delovalo vrednotenje predanih podatkov in s tem povezano priznavanje prispevka k znanosti v znanstveni bibliografiji.

***Predlog št. 4:** Core Trust Seal (CTS) kriteriji naj bodo usmeritev za delovanje certificiranih področnih podatkovnih središč v Sloveniji pri podpori izvedbe Akcijskega načrta in zagotavljanja skrbništva za raziskovalne podatke v skladu z mednarodnimi standardi.*

Andrej Pančur z Inštituta za novejšo zgodovino je v svoji predstavitvi nanizal izzive in pristope pri vzpostavljanju certificiranega podatkovnega središča za digitalno humanistiko **DARIAH-SI**. Povedal je, da podatkovno skrbništvo ni edina naloga, pač pa se ukvarjajo tudi z zbiranjem in ustvarjanjem digitalnih surogatov kulturne dediščine, ki se jih lahko dodatno obdelata, vizualizira, označi, poveže in interpretira za različne namene. Izziv je raznolikost formatov in tipov gradiv, s čimer je povezana cela množica naborov metapodatkovnih standardov, npr. s teiHeader, MODS, EAD, LIDO, Dublin Core, DataCite, ipd. in je težko najti skupni imenovalec. Pripravljajo kompatibilnost z zunanjimi odjemalci, npr. OpenAIRE, gradiva in vmesnike pa dodajajo tudi preko GitHub, obenem skrbijo za digitalno hrambo s pomočjo orodja Achivematica.

Jure Dimec je predstavil slovensko vključevanje v **ELIXIR**. To je podatkovna infrastruktura, ki pokriva 'omske' podatke iz različnih področij, medicine, bioloških ved, ipd. Pri tem je v slovenskem kontekstu vprašanje, kaj se dogaja z drugimi podatki s teh področij, ki bi jih bilo navsezadnje za doseganje novih spoznanj koristno povezati s podatki genetike. Klinični podatki zahtevajo posebno skrbnost pri ravnanju zaradi svoje občutljivosti. Slovensko partnerstvo v ELIXIR je pretežno pri podpori izobraževanju za delo s podatki, vključno tudi pri pripravah in podpori načrtom ravnanja s podatki.

Zaključni predlog, ki ga lahko izluščimo iz sekcije področnih infrastruktur:

***Predlog št. 5:** Podpora načrtovanju in ravnanju s podatki v smeri izpolnjevanja načel FAIR in 'kolikor mogoče odprtega dostopa' naj v čim večji meri sloni na usmeritvah, svetovanju in izobraževanju, ki jih že sedaj pripravljajo in ponujajo področne podatkovne infrastrukturne storitve, povezane v evropska vozlišča.*

Naslednja sekcija **Projekti in iniciative za vzpostavljanje infrastrukture podatkovnih storitev** je imela niz krajših predstavitev, ki so sledile načrtanim poudarkom konference: iskanja primernih okvirov za vrednotenje podatkov in z njimi povezanih storitev.

Predstavili sta se dve področni iniciativi za izboljšanje podatkovne infrastrukture. **Ana Slavec**, InnoRenew CoE in Ambasadorica RDA za področje tehnike – obnovljivih materialov je predstavila primer usmeritev za **Načrte za podatke o obnovljivih materialih in produktih**.

Načrt sledi smernicam upravljanja raziskovalnih podatkov po načelih FAIR. Ugotavlja, da sta za področje glavna tehnična izziva pomanjkanje standardov za metapodatke in ontologije, potrebno pa bi bilo tudi identificirati primerne repozitorije. Načrtujejo izobraževalne delavnice in seminarje za raziskovalce z namenom seznanjanja raziskovalcev z načeli FAIR in pridobivanja veččin, potrebnih za upravljanje raziskovalnih podatkov. Primere podatkovnih zbirk na **Kmetijskem inštitutu Slovenije** je predstavil **Janez Jeretina**, Kmetijski inštitut Slovenije.

***Predlog št. 6:** Raziskovalnim področjem je glede na potrebe in obstoj posebnih področnih vrst podatkov potrebno zagotoviti finančne in kadrovske pogoje za vzpostavitev področno specifične podatkovne infrastrukture.*

Dunja Legat iz Univerzitetne knjižnice Maribor je predstavila izhodišča **NI4OS** (National Initiatives for Open Science in Europe) projekta, ki je eden od regionalnih projektov v podporo izgradnji Evropskega oblaka odprte znanosti - EOSC. Partnerji projekta so identificirali pregledne sezname, na katere se uvrščajo preizkušene podatkovne repozitorske storitve in z njimi povezani dostopni digitalni objekti, ki so podvrženi preverjanju in nato na voljo drugim. Za zgled omenjajo nekatera orodja projekta OpenAIRE, kot je ARGOS (<https://argos.openaire.eu/>) za zbiranje komponent Načrtov ravnanja z raziskovalnimi podatki, in pa njihov *Validator repozitorske skladnosti* (<https://www.openaire.eu/validator/>). Pri ocenjevanju repozitorijev glede skladnosti njihovih politik ter standardov z uveljavljenimi pristopi po načelih FAIR so omenjali npr. register FAIRsharing (<https://fairsharing.org/>).

Milan Ojsteršek z Univerze v Mariboru je predstavil načrte za Dopolnitev **repozitorijev nacionalne infrastrukture odprtega dostopa** s funkcionalnostmi podatkovnega arhiva za velepodatke. Zavzel se je za uveljavitev z granularnimi stalnimi identifikatorji obogateno metapodatkovno shemo, ki bi zagotavljala sledljivost verzij transformacij podatkov in njihove uporabe. V zvezi z velepodatki je nakazal rešitve, ki povezujejo storitve HPC in storitve trajnega hranjenja. Ustanove, ki bi izkoriščale tovrstne storitve, bi naj bile vključene s ponudbo izobraževanja uporabnikov in podporo procesom v svojem okolju.

Podrobnosti vzpostavljanja robustne komunikacijske in superračunalniške infrastrukture za podporo delu z raziskovalnimi podatki, ki jo pripravlja **Arnes** ter **Sling** in **EuroHPC**, sta vsak posebej predstavila **Avgust Jauk** iz Arnes in **Jan Jona Javoršek** z Institut "Jožef Stefan". Sling ima v svoji organizacijski zasnovi predvideno vzpostavitev skupnosti za sodelovanje tako uporabnikov kot ponudnikov podatkovnih storitev.

***Predlog št. 7:** Na podlagi popisov orodij, storitev in standardov (EUDAT, DataCite, ORCID, SCHOLIX...) ki se preko OpenAire, EOSC in RDA uveljavljajo v mednarodnem okolju, naj izvajalci infrastrukture podatkovnih storitev poročajo o možnostih implementacije v nacionalnem okolju.*

Miro Pušnik s CTK se je zavzel za vzpostavitev okvira za vrednotenje raziskovalnih podatkov, kjer lahko pri področnih posebnostih svojo vlogo odigrajo **OSIC**-i, pri tem pa je izhodišče že obstoječa kategorija v SICRIS tipologiji: 2.20 Zaključena znanstvena zbirka podatkov ali korpus. Dal je pobudo, da se povežejo različni deležniki uporabnikov ali ponudnikov storitev odprte znanosti ter ARRS, da bi dosegli dogovor glede minimalnih kriterijev pri pripravi podatkov, njihovi kakovosti in ocenjeni življenjski dobi. Obenem je menil, da bi morali znanstveni prispevek raziskovalnih podatkov v odprtem dostopu eksplicitno vključiti v

metodologijo vrednotenja raziskovalnih projektov, skladno z deklaracijo DORA (San Francisco Declaration on Research Assessment).

Predlog št. 8: *ARRS naj se poveže z ostalimi deležniki, da bi dosegli dogovor glede minimalnih kriterijev pri pripravi podatkov, njihovi kakovosti in ocenjeni življenjski dobi ter znanstveni prispevek raziskovalnih podatkov upoštevati pri vrednotenju projektov.*

V sklopu programa z naslovom Nacionalna iniciativa za slovensko skupnost odprte znanosti je Dunja Legat iz Univerzitetne knjižnice Maribor najprej s prispevkom »Novo obdobje sodelovanja: predstavitev projekta NI4OS-Europe« predstavila regionalni projekt za JV Evropo »National initiatives for Open Science – Europe«, katerega slovenska partnerja sta Univerzitetna knjižnica Maribor in Arnes. Značilnosti trenutnega stanja v Evropi so razdrobljen dostop (po znanstvenih področjih, razlike v državah in modelih upravljanja, različne politike dostopa), storitve in podatki niso interoperabilni, podatki so večinoma zaprti. Evropski oblak odprte znanosti (ang. European Open Science Cloud – EOSC) obeta, da bo vsem raziskovalcem iz EU zagotovil okolje z brezplačnimi, odprtimi storitvami za analizo podatkov, digitalno shranjevanje in za pripravo podatkov za ponovno interdisciplinarno uporabo. Povezal bo obstoječe in nastajajoče horizontalne in tematske podatkovne infrastrukture. EOSC zaradi svojega obsega, poudarka na podatkovno podprti znanosti, interdisciplinarnosti in možnosti inovacij predstavlja dodano vrednost preteklim naložbam v podatkovno infrastrukturo. Prva faza NI4OS-Europe je vzpostavitev slovenske skupnosti odprte znanosti. Slovenske deležnike na tem področju želi povezati v celovit in pregleden sistem, ki bi deloval strokovno komplementarno, s prevzemanjem nalog kompetentnih partnerjev v skupnosti. Skupnost bo imela vzajemno korist, pri kateri bo učinek celotnega sistema večji. Izhodiščna zamisel je vzpostavitev posameznih segmentov odprte znanosti po načelu t. i. kompetenčnih centrov, kar bi bila tudi osnova za organizacijo medinstitucionalnih s podatki povezanih storitev kot sestavnega dela nacionalne skupnosti odprte znanosti. Nekaj podatkovnih infrastruktur in storitev se je že izoblikovalo, nekatere pa so še v nastajanju ali jih bo potrebno vzpostaviti.

Predlog št. 9: *Vzpostavi naj se slovenska skupnost odprte znanosti po načelu t. i. kompetenčnih centrov, katere pomembni deležniki so RDA vozlišče Slovenije, OpenAIRE nacionalna točka odprtega dostopa, druge že vzpostavljene slovenske področne infrastrukturne storitve, ki so povezane v velike evropske infrastrukturne enote, ter ostale ustanove in deležniki, ki sestavljajo pokrajino odprte znanosti v Sloveniji.*

V drugi predstavitvi z naslovom »Nacionalne iniciative odprte znanosti: primeri dobrih praks« je **Dunja Legat** predstavila nekaj tujih pristopov pri organizaciji in sistemski podpori odprti znanosti.

Brina Klemenčič iz Univerzitetne knjižnice Maribor je v zaključku predstavila prispevek »Odrpta znanost: kje smo?: predstavitev vprašalnika NI4OS-Europe«, v okviru katerega je pojasnila strukturo vprašalnika, s pomočjo katerega v okviru projekta NI4OS rišejo zemljevid pokrajine odprte znanosti v Sloveniji, s tem da bo ponudil vpogled v delovanje večine deležnikov odprte znanosti v Sloveniji. Skupna ugotovitev v razpravi je bila, da se Strategija odprtega dostopa izteka, pa še veliko stvari ni bilo realiziranih, zato je nujen poziv vsem nosilcem aktivnosti [Akcijskega načrta](#) izvedbe Nacionalne strategije odprtega dostopa do znanstvenih objav in raziskovalnih podatkov v Sloveniji 2015-2020⁴, da poročajo o svojih dosedanjih aktivnostih ter kaj načrtujejo v bližnji prihodnosti.

⁴ <http://www.ff.um.si/dotAsset/68527.pdf>, 11. 12. 2019

Predlog št. 10: Nosilci aktivnosti Akcijskega načrta naj na organiziranih sestankih v koordinaciji RDA vozlišča Slovenije poročajo o svojih dosedanjih aktivnostih ter o načrtih za bližnjo prihodnost. Pri tem sta ključna MIZŠ ter ARRS.

V Ljubljani, 6. 12. 2019

Organizatorji konference