

IZ NAŠIH KNJIŽNICA

Istraživački podatci hrvatskih autora na platformi Web of Science

J. Bolkovac,^a T. Krajna^{a*} i J. Petrak^b

^a Sveučilište u Zagrebu, Fakultet strojarstva i brodogradnje, Ivana Lučića 5, 10 000 Zagreb

^b Sveučilište u Zagrebu, Medicinski fakultet, Šalata 3, 10 000 Zagreb

Ovo djelo je dano na korištenje pod
Creative Commons Attribution 4.0
International License



Sažetak

Otvoreni pristup istraživačkim podacima važan je dio koncepta otvorene znanosti. Cilj ovoga članka je identificirati hrvatske autore/ustanove koji su prepoznali važnost dijeljenja istraživačkih podataka i koji su im omogućili pristup u nekom od podatkovnih repozitorija zastupljenom u citatnom indeksu *Data Citation Index*. U bazi podataka je ukupno pronađeno 265 takvih dokumenata. Analizirani su i podatkovni članci hrvatskih autora dostupni na platformi *Web of Science Core Collection*. Pronađeno ih je 29, najvećim dijelom multidisciplinarnog usmjerenja. Autori najvećeg broja podatkovnih dokumenata pripadaju Sveučilištu u Zagrebu. Hrvatski znanstvenici arhiviraju podatke u međunarodne repozitorije, jer je infrastruktura za njihovu pohranu na nacionalnoj razini tek razvijena i tek se počinje koristiti.

Ključne riječi

Otvorena znanost, istraživački podatci, podatkovni članci, podatkovni repozitoriji

1. Uvod

Iz velikog broja znanstvenih istraživanja danas proizlazi ogromna količina različitih podataka do kojih se dolazi s pomoću novih metoda i opreme, koji se pohranjuju u različitim oblicima i s kojima se upravlja na različite načine.¹ Podatci iz istraživanja mogu sadržavati kvantitativne informacije ili kvalitativne tvrdnje koje istraživači prikupljaju i bilježe tijekom eksperimenta, promatranja, modeliranja, intervjua i sl., ili to može biti informacija proizišla iz postojećih dokaza.² U područjima prirodnih znanosti najveći broj podataka prikuplja se promatranjem i eksperimentima, u društvenim znanostima vlastitim opažanjima ili preuzimanjem iz drugih javnih izvora (primjerice, podatci o ekonomskoj aktivnosti), a u humanističkim znanostima podatci se najčešće izvlače iz kulturne baštine, arhivskoga gradiva, artefakata i sl.³ Istraživački podatci (engl. *research data*) mogu biti objedinjeni i pohranjeni u tzv. sirovom ili obrađenom obliku u datotekama, mogu tvoriti modele, algoritme, protokole ili biti u obliku programirane računalne potpore analizi podataka.

Prihvaćanjem otvorenog pristupa znanstvenim informacijama kao važne sastavne koncepta otvorene znanosti, od istraživača se sve češće zahtijeva dijeljenje istraživačkih podataka. To prije svega propisuje financijeri znanstvenih projekata⁴ te uredništva znanstvenih časopisa.⁵ Najveća korist dijeljenja podataka ogleda se u a) ponovljivosti znanstvenih rezultata, b) javnoj dostupnosti rezultata istraživanja financiranih javnim sredstvima, c) ubrzanju istraživanja i inovacija.¹ Iako neki autori ističu složenost postupka ponovljivosti znanstvenih rezultata koju omogućava otvoreni pristup istraživačkim podacima,⁶ a neki napominju da je razdvajanje podataka od konteksta iz kojeg potiču rizik

o kojemu treba voditi računa pri pokušajima ponavljanja,⁷ otvoreni pristup tim podacima omogućuje njihovo propitivanje, ocjenu i potvrdu vjerodostojnosti od strane drugih neovisnih znanstvenika. Dijeljenje istraživačkih podataka omogućuje stvaranje većih i velikih uzoraka čime se postiže veća pouzdanost u njihovoj analizi. Štoviše, "zahvaljujući golemim količinama podataka kakvima danas raspolazemo, pozitivni nalazi povezanosti među pojavama mogu se statistički dokazati izvan svake sumnje".⁸ Otvoreni pristup velikoj količini istraživačkih podataka otvara mogućnost postavljanju novih pitanja, potiče meta-analitički pristup i otkriće novih povezanosti.

Javna dostupnost istraživačkih podataka smanjuje troškove istraživanja jer se ne stvaraju redundantni podatci i ne traže odgovori na već riješena znanstvena pitanja a potiče se i njihova translacija u praksu.⁹ Ponajbolji primjer za to su biomedicinska istraživanja, koja svoju primjenu imaju u kliničkoj i farmaceutskoj domeni.

Istraživački podatke dijele radije s onima koji su dio njihove najbliže interesne zajednice, odnosno specijalnosti, nego s općom javnošću¹⁰ jer može doći do zloupotrebe u načinu njihove interpretacije i prikaza.¹¹ Najčešći je oblik dijeljenja istraživačkih podataka njihov prikaz u člancima objavljenim u znanstvenim časopisima. To, međutim, nije dostatno. U znanstvenim člancima prikazuju se samo najvažniji podatci iz istraživanja, najčešće u sažetom obliku. S tim se člancima stoga sve više povezuju organizirani podatci koji su pohranjeni i dostupni na mrežnim mjestima poput repozitorija i raznorodnih arhiva. Mrežno dostupni repozitoriji istraživačkih podataka složene su infrastrukture koje osiguravaju djelotvornu pohranu, pristup i upravljanje skupovima podataka.¹² Podatkovni repozitoriji mogu biti specijalizirani i usmjereni samo na jedno znanstveno polje (npr. *GenBank*, *Crystallography Open Database*) ili mogu okupljati podatke iz više znanstvenih područja (npr. *Figsha-*

* Autor za dopisivanje: dr. sc. Tamara Krajna
E-pošta: tkrajna@fsb.hr

re, Zenodo, Science Data Bank i dr.). Digitalni repozitoriji omogućuju dugotrajno čuvanje podataka u različitim formatima, od Excel-tablica do specifičnih formata karakterističnih za određeno područje i upotrijebljenu znanstvenu opremu (primjerice, GIS shapefiles i sl.). U pohrani podataka u repozitorije najčešće se primjenjuju tzv. FAIR načela.¹³ Podatci se moraju moći pronaći (engl. *findable*), moraju biti dostupni (engl. *accessible*), moraju se primjenjivati standardi za razmjenu podataka (engl. *interoperable*) te se moraju moći ponovno koristiti (engl. *reusable*).

Brojni su izazovi s kojima se suočavaju oni koji organiziraju i pohranjuju podatke kao i oni koji im žele pristupiti. Ti izazovi nisu samo tehničke naravi, kao što su, primjerice, zastarijevanje opreme i softverske potpore. Pohranjeni podatci moraju se detaljno i što potpunije opisati metapodatcima. Metapodatci su strukturirane informacije koje opisuju, objašnjavaju ili na neki drugi način olakšavaju pretraživost, iskoristivost i upravljanje nekom informacijom. Drugim riječima metapodatak je strojno razumljiva informacija, odnosno zapis koji opisuje elektronički resurs.¹⁴ Metapodatci se mogu razlikovati od područja do područja, a izrada metapodatkovnog opisa te njegova standardizacija zahtjevan su zadatak.¹⁵ Izazovna su i važna pitanja izrade plana upravljanja pohranjenim istraživačkim podatcima, pitanje odgovornosti i kontrole, duljine čuvanja podataka, anonimizacije osobnih podataka, odnosno zaštite privatnosti i sl.¹⁶

Dugačku povijest arhiviranja istraživačkih podataka imaju društvene znanosti, a prvi europski arhiv za empirijska društvena istraživanja osnovan je 1960. godine.^{17,18}

Podatci iz biomedicinskih istraživanja također se pohranjuju organizirano i kooperativno dugi niz godina. *Protein Data Bank*¹⁹ uspostavljen je 1971. godine, a *GenBank* 1979. godine.²⁰ Ubrzo poslije osnivanja ti su repozitoriji postali i javno dostupni. Istraživanja u geoznanosti također podupire velik broj podatkovnih repozitorija, poput repozitorija *National Oceanic and Atmospheric Administration* (NOAA),²¹ *GIS Data Repository*²² itd. I u mnogim drugim znanstvenim područjima postoji duga tradicija arhiviranja podataka u svrhu njihove razmjene i ubrzanja daljnjih istraživanja. U Hrvatskoj je arhiviranje podatkovnih skupova započelo u institucijskom repozitoriju Instituta Ruđer Bošković 2013. godine. Danas su istraživački podatci izdvojeni u poseban repozitorij *FULIR Data*,²³ koji je nedavno uključen u registar podatkovnih repozitorija *Re3data* (od engl. *Registry of Research Data Repositories*),²⁴ koji promovira kulturu dijeljenja i otvorene vidljivosti istraživačkih podataka. Nacionalni sustav repozitorija *Dabar* (Digitalni akademski arhivi i repozitoriji)²⁵ ključna je sastavnica podatkovnog sloja nacionalne e-infrastrukture koja ustanovama i drugim dionicima iz sustava znanosti i visokog obrazovanja pruža i uslugu pohrane i čuvanja istraživačkih podataka. Za sad je pohranjeno nešto više od 40 podatkovnih zapisa.

Clarivate Analytics među svoje je citatne indekse dostupne na platformi *Web of Science* 2012. godine uvrstio i *Data Citation Index* (DCI), koji indeksira i čini dostupnima metapodatke iz više od 430 odabranih repozitorija (prosinac 2020.) koji pohranjuju setove istraživačkih podataka iz različitih znanstvenih područja. Postupak odabira repozitorija uključuje kvalitativne i kvantitativne kriterije, među

kojima su znanstveno područje, visoki uređivački standardi, kvaliteta zapisa itd.²⁶ DCI omogućuje pronalaženje istraživačkih podataka i njihovih poveznica sa znanstvenim radovima indeksiranim u ostalim citatnim indeksima u sustavu *Web of Science Core Collection* (WoSCC) te preko citata osigurava uvid u njihov mogući istraživački učinak.²⁷ Citiranje istraživačkih podataka u većini je znanstvenih područja prilično rijetko i nije standardizirano, pa DCI može potaknuti praksu citiranja koja bi osiguravala praćenje istraživačkog procesa od prikupljanja podataka do njihove konačne objave.²⁸ Podatkovni zapisi u DCI-ju raspoređeni su u tri skupine (engl. *document types*): podatkovne studije (engl. *data study*), podatkovne skupove (engl. *data set*) i softvere. *Podatkovne studije* sadržavaju opis studije ili eksperimenta kojem su pridruženi podatci iz istraživanja. U nekim slučajevima postoji poveznica na indeksirani zapis pripadajućeg podatkovnog skupa unutar DCI-ja. Neki od pridruženih podataka su oni o znanstvenicima koji su sudjelovali u istraživanju, o znanstvenicima koji su pohranili istraživačke podatke, podatci o izvoru financiranja, ključne riječi i drugo. *Podatkovni skup* je osnovna informacijska jedinica koja je najčešće dijelom neke podatkovne studije. DCI definira podatkovni skup kao koherentni skup podataka ili datoteka s podatcima koji su u neki repozitorij pohranjeni kao dio jedne zbirke podataka, podatkovne studije ili eksperimenta (u nekim slučajevima postoji poveznica na indeksirani zapis nadređene podatkovne studije unutar DCI-ja), dok je *softver* definiran kao računalni program s izvornim kodom ili u kompiliranom obliku koji se može instalirati na drugo računalo i primjenjivati u analizi istraživačkih podataka.²⁹ DCI razvrstava pohranjene istraživačke podatke i po njihovom obliku (engl. *data type*), primjerice digitalna karta, anketni podatci, tablični podatci, slike itd. Svaki zapis uključuje izravnu poveznicu na repozitorij u kojemu su podatci izvorno pohranjeni, a zapisi najčešće sadržavaju preuzete metapodatke, poput naslova, autora, kratkog sažetka, datuma objave, ključnih riječi i sl. S preuzetom se zapisima pobiru, međutim, i sve specifičnosti i nedostaci tih repozitorija.³⁰

Podatkovni članak (engl. *data paper*) relativno je nov publicistički oblik, koji sadržava ključne informacije o podatcima pohranjenim u podatkovni repozitorij i poveznice za brzo pronalaženje i moguću ponovnu uporabu kao i standardne identifikatore podatkovnih skupova i autora. Drugim riječima, podatkovni članak je pretraživ metapodatkovni dokument koji opisuje određeni skup podataka i koji se nakon obvezatnog recenzijskog postupka objavljuje u specijaliziranom znanstvenom časopisu koji objavljuje samo takve priloge (npr. časopis *Data in Brief*) ili u časopisima koji objavljuju i druge vrste članaka. Ta vrsta članka nema, dakle, klasičnu tekstualnu strukturu i organiziranost, to "nisu članci o istraživanju, nego o podatcima", oni "ne donose novo znanje, nego služe generiranju znanja".³¹ Autori u tom slučaju ne moraju, ne žele ili ne mogu napisati i objaviti klasični znanstveni članak s analizom rezultata ili žele prikazati opširnije podatke čiji opseg ne dopušta da se ugrade u klasični znanstveni članak.³²

Većina podatkovnih časopisa pripada području prirodnih i biomedicinskih znanosti i objavljuju se u otvorenom pristupu. U bazama podataka WoSCC podatkovni članci mogu se filtrirati kao zasebna kategorija indeksiranih radova.

Cilj ovoga rada je identificirati hrvatske autore/ustanove koji su prepoznali važnost dijeljenja istraživačkih podataka i koji su im omogućili pronalaznje i pristup u nekom od repozitorija zastupljenom u bazi podataka DCI. Analizirat će se i njihova pripadnost određenim znanstvenim poljima.

Drugi dio analize obuhvatit će podatkovne članke objavljene u časopisima koje obrađuju bibliografske baze na platformi WoSCC, autore koji su ih potpisali, časopise koji su ih objavili i moguću povezanost s istraživačkim podatcima koji su dostupni preko citatne baze DCI.

2. Metode

Podatci su prikupljeni na dvije razine. Iz baze podataka *Data Citation Index* (DCI) na platformi *Web of Science* (WoS) preuzeti su samo oni zapisi koji su u polju adrese sadržavali pojam Croatia. Temeljno ograničenje tog načina pretraživanja DCI-ja očituje se u malom broju zapisa kojima pretraživanje rezultira. Naime, više od 80 % zapisa u DCI-ju ne sadržava podatke o adresi autora, ni onih koji su rezultate prikupili, ni onoga koji je podatke pohranio. Napomenu o tom ograničenju DCI navodi u prikazu rezultata pretraživanja po polju adrese (osnovno pretraživanje) ili zemlje (napredno pretraživanje/*country search tag*). Probna pretraživanja po imenu autora pružila su isti dokaz. Od ukupno 50 pronađenih zapisa s imenom jednog od hrvatskih autora samo ih je 15 sadržavalo hrvatsku adresu, 12 je sadržavalo adresu koautora, a 23 zapisa nisu uopće sadržavali taj podatak. U obrascu koji treba popuniti pri unosu podataka u repozitorij *Zenodo*, primjerice, popunjavanje polja adrese nije obvezatno. Uvid u taj obrazac bio je dodatni dokaz ograničenja pretraživanja DCI-ja po polju adrese. U literaturi postoje, međutim, primjeri primjene tog načina pretraživanja u bibliometrijskim analizama,³³ pa smo ga, svjesni svih nedostataka, i mi primijenili.

Zapisi za razdoblje od devet godina (2012. – 2020.) eksportirani su 1. veljače 2021. godine u txt formatu te su učitani u Excel-format. Eksportirani zapisi uključuju: naslov dokumenta, naziv repozitorija, vrstu dokumenta, imena autora ili institucije koja je pohranila zapis, godinu pohrane zapisa te kategoriju istraživanja prema klasifikaciji koju primjenjuju WoS baze podataka.

Preko oznake DOI (od engl. *digital object identifier*) svakoga zapisa pristupalo se repozitoriju u kojem su podatci pohranjeni, a iz kojeg je WoS preuzeo odgovarajuće metapodatke. Nadalje, provjeravalo se jesu li nađeni zapisi povezani s člancima u kojima su se koristili istraživački podatci navedeni u zapisu. Za pronađene povezane članke prikupljali su se sljedeći podatci: naslov članka, imena autora, naslov publikacije u kojoj je članak objavljen, indeksiranost na platformi WoS te kategorija i područje istraživanja prema klasifikaciji WoS. Za neke zapise jasno je navedeno da su vezani uz određeni članak, dok su se za većinu zapisa, prema prezimenima autora i/ili naslovima zapisa, pretraživali drugi relevantni izvori (primjerice bibliografska baza *Scopus*, *Google Scholar* itd.).

Druga grupa podataka prikupljena je pretraživanjem baza podataka dostupnih na istoj platformi, ali unutar sustava WoSCC. Zapisi s pojmom Croatia u polju adrese filtrirani su temeljem vrste dokumenta i to isključivo "data paper" sa stanjem 1. veljače 2021. Pronađeni podatkovni članci analizirani su prema časopisima u kojima su objavljeni, ustanovama autora te kategoriji i području istraživanja prema klasifikaciji koju primjenjuju baze podataka WoS. Dodatno je ispitana njihova povezanost sa zapisima u bazi podataka DCI.

3. Rezultati i rasprava

3.1. Data Citation Index

Pretraživanjem baze podataka DCI ukupno je pronađeno je 265 zapisa kojima je barem jedan od autora s hrvatskom adresom.

Tablica 1 – Podatci prema vrsti arhiviranih dokumenata i godini objavljivanja

Table 1 – Data by type of archived documents and year of publication

Godina Year	Broj zapisa No. of records	Vrsta dokumenta Document Type		
		Podatkovne studije Data study	Podatkovni skup Data set	Softver Software
2012.	2	2	0	0
2013.	4	3	1	0
2014.	17	4	13	0
2015.	27	2	24	1
2016.	53	6	45	2
2017.	25	3	22	0
2018.	3	2	0	1
2019.	22	1	13	8
2020.	112	1	103	8
Ukupno	265	24	221	20

Najviše zapisa bilo je u kategoriji podatkovnih skupova (84 %). Broj se zapisa povećava s godinama, ali taj rast nije ravnomjeran (tablica 1). Ono što se može neprijeporno utvrditi jest višestruki porast broja zapisa u posljednjoj analiziranoj godini. Razlog tome može biti i povećan broj uključenih repozitorija u DCI, ali i obveza otvorenog pristupa podacima iz istraživanja koju nalažu financijske ustanove i nakladničke kuće.

U bazi podataka DCI 183 su jedinstvena zapisa s barem jednim autorom/ustanovom potpisanim/om hrvatskom adresom. Preostala 82 zapisa su ponovno uneseni zapisi, s novim oznakama DOI, nove, promijenjene ili dopunjene

Tablica 2 – Broj i vrste zapisa prema podatkovnom repozitoriju
Table 2 – Number and type of documents by data repository

Repozitorij Data Repository	Izvorni zapis First version	Ostale verzije zapisa Other versions	Podatkovna studija Data study	Podatkovni skup Data set	Softver Software	Ukupno Total
Zenodo	78	76	0	138	16	154
Gene Expression Omnibus	78	0	5	73	0	78
European Nucleotide Archive	11	0	11	0	0	11
Harvard Dataverse	5	2	7	0	0	7
Code Ocean	4	0	0	0	4	4
Gfz German Research Centre For Geosciences	2	0	0	2	0	2
Institut Laue Langevin	2	0	0	2	0	2
IEEE Dataport	1	4	0	5	0	5
Metabolights	1	0	0	1	0	1
Data Archiving and Networked Services Dans Knaw	1	0	1	0	0	1
Ukupno	183	82	24	221	20	265

verzije tih jedinstvenih zapisa. Gledajući pojedine vrste zapisa u DCI-ju, najviše je ponovljenih zapisa u podatkovnim skupovima, više od 33 %. Dodavanje novih podataka iz istraživanja koje nije završeno poželjan je postupak, jer se na taj način stječe uvid u tijek znanstvenoga istraživanja, odnosno novi uvid u već objavljene podatke ili njihova moguća korekcija. Nova se verzija od prethodne može znatno razlikovati kako opsegom tako i vrstom uključenih podataka. Po tome se zapisi o istraživačkim podacima razlikuju od onih za članke ili monografije koji su po naravi fiksni.³⁴ To tzv. verzioniranje istraživačkih podataka, odnosno proces stvaranja ili upravljanja promjenama datoteke tijekom istraživanja ili projekta¹³ može se ilustrirati primjerom repozitorija *Zenodo* kojim se, prema našim rezultatima, koristi najveći broj hrvatskih znanstvenika (tablica 2).³⁵ *Zenodo* je podatkovni repozitorij u otvorenome pristupu razvijen u sklopu programa *OpenAIRE* i kojim upravlja CERN (Europska organizacija za nuklearna istraživanja).³⁶ U *Zenodo* se pohranjuju različite vrste građe (publikacije, poster, prezentacije, skupovi podataka, slike, softver, video/audio materijali i ostali interaktivni materijali) proizašle iz istraživanja koja financira Europska komisija, ali je otvoren svim znanstvenicima neovisno o izvoru financiranja ili nacionalnosti i to sukladno FAIR načelima. Kad se zapis unosi prvi put, dodjeljuju mu se dvije oznake DOI, jedna za taj specifični, jedinstveni zapis (engl. *version DOI*), a druga za sve dodatne verzije zapisa koje bi se mogle pojaviti u repozitoriju (engl. *concept DOI*).³⁷

Rezultati ovog istraživanja pokazuju da hrvatski autori zapise o podacima pohranjuju u ukupno 10 repozitorija, najviše u repozitorije *Zenodo* (58 %) i *Gene Expression Omnibus* (29 %). Najviše zapisa o podatkovnim skupovima (63 %) pohranjeno je u *Zenodu*, podatkovne studije pretežito su pohranjene u *European Nucleotide Archive* (46 %), a 80 % zapisa o softverima pohranjeno je u *Zenodu* (tablica 2).

Povezanost podatkovnih studija sa skupovima podataka vidljiva je na primjeru repozitorija *Gene Expression Omnibus*. Pet studija vezuje ukupno 73 podatkovna skupa (tablica 2).

Tablica 3 – Kategorizacija zapisa prema znanstvenim područjima/poljima
Table 3 – Data records by fields of science and technology

WoS kategorije WoS Categories	Ukupno Total	% %
Multidisciplinary Sciences	157	59,2
Genetics & Heredity	89	33,6
Social Sciences, Interdisciplinary	7	2,6
Engineering, Electrical & Electronic	5	1,9
Geosciences	2	0,8
Physics, Particles & Fields	2	0,8
Biochemistry & Molecular Biology	1	0,4
Computer Science, Interdisciplinary Applications	1	0,4
Humanities, Multidisciplinary	1	0,4
Total	265	100

Kategoriziranje zapisa provedeno je na vršnoj razini, pa je najveći broj zapisa svrstan u multidisciplinarna područja (60 %). Taj podatak, međutim, može očitovati i transdisci-

plinarni pristup znanstvenim problemima, koji obilježava današnja znanstvena istraživanja. Genetika i nasljeđivanje (engl. *Genetics & Heredity*) jedina je izrazito distinktivna kategorija koja okuplja oko 34 % svih zapisa (tablica 3).

3.2. Autorstvo

Potrebno je napomenuti da je vrlo teško, ponekad i nemoguće, identificirati ustanovu, odnosno autore podatkovnih zapisa. Kao što je neujednačeno navođenje naziva ustanova u člancima koje potpisuju hrvatski autori u znanstvenim časopisima, tako se i u pridruživanju imena ustanove pohranjenim istraživačkim podacima može primijetiti ista nedosljednost i nepotpunost.

Analizirali smo autorstvo 24 podatkovne studije (tablica 1). Hrvatsko autorstvo na šest podatkovnih studija nismo uspjeli identificirati iako su se zapisi nalazili u rezultatima pretraživanja. Najveći broj (11) potpisuje Sveučilište u Zagrebu. Pri tome su četiri studije ostvarene suradnjom tog sveučilišta s međunarodnim akademskim ustanovama. Dvije studije potpisuju sastavnice (Medicinski fakultet/Fakultet organizacije i informatike te jedna klinička bolnica), a dvije su potpisane samo nazivom sveučilišta. Autori sedam studija dolaze s pojedinih sastavnica Sveučilišta (Filozofski fakultet, Prehrambeno-biotehnološki fakultet, Medicinski fakultet, Agronomski fakultet). Od preostalih 13 podatkovnih studija znanstvenici s Instituta Ruđer Bošković potpisuju četiri, dok po jednu podatkovnu studiju potpisuju Mediteranski institut za istraživanje života, Hrvatski veterinarski institut i udruga građana Documenta.

Najviše studija pohranjeno je u repozitorij *European Nucleotide Archive* (11), te *Harvard Dataverse* (osam iz područja društvenih znanosti).

Na istome uzorku provjerili smo povezanost podatkovnih zapisa u DCI-ju s radovima koji su nastali na tim podatkovnim podlogama, a indeksirani su u WoSCC-u. Od 24 podatkovne studije 13 ih ima vezane članke (po jedan članak) koji su indeksirani u WoSCC-u. Pet članaka potpisuju isključivo hrvatski autori, među kojima su četiri kategorizirana u područje društvenih znanosti. Preostalih osam članaka nastalo je u međunarodnom koautorstvu, a kategorizirani su u društvene i različite discipline prirodnih i/ili biomedicinskih znanosti.

Identificirati istraživače koji su pohranili neki podatkovni skup, nije uvijek moguće. Od ukupno 265 nađenih zapisa u DCI-ju pouzdano možemo tvrditi da su hrvatski autori pohranili istraživačke podatke povezane s 89 zapisa (34 %). Naime, iz metapodataka u nekim repozitorijima, kao što je primjerice *Gene Expression*, jasno se može odrediti osoba koja je unijela zapis. Svih 78 zapisa u tom repozitoriju pohranili su hrvatski autori. Iz metapodataka zapisa u repozitoriju *Harvard Dataverse* također se može vidjeti tko je pohranio podatke, pa su od ukupno sedam zapisa u tom repozitoriju hrvatski autori pohranili dva. To se odnosi i na *IEEE Dataport* (hrvatski su autori pohranili podatke vezane uz svih pet zapisa) te na *Code Ocean*, u koji su hrvatski autori pohranili podatke o softverima (četiri zapisa). U ostalim repozitorijima ne postoje vidljivi metapodatci s

pomoću kojih bi se mogli identificirati oni koji su istraživačke podatke pohranili.

3.3. Podatkovni članci

Drugi dio analize obuhvatio je radove hrvatskih autora objavljene u časopisima indeksiranim u WoSCC-u te klasificirane među podatkovne članke (engl. *data papers*). Pronašli smo 29 takvih radova, od kojih je najveći dio objavljen u časopisima *Data in Brief* (17) i *Scientific Data Nature* (5). Autorstvo osam radova pripada isključivo domaćim autorima, a najznačajnija pojedinačna institucijska pripadnost odnosi se na Sveučilište u Zagrebu (17 radova). Velika većina radova (19) kategorizirana je u multidisciplinarne znanosti.

U podatkovnim člancima naglasak se stavlja na detaljan opis podataka, metode njihova prikupljanja i obrade te na njihovu važnost za specifičnu znanstvenu zajednicu. Podatkovni članci također omogućuju dijeljenje istraživačkih podataka. To se najčešće ostvaruje dopunskim datotekama (engl. *supplementary files*) koje su dio članka ili dodavanjem poveznica na podatke pohranjene u nekom repozitoriju.⁵ Za samo dva rada pronađena je veza s bazom podataka DCI. U svakome slučaju, opisani podatci moraju biti javno dostupni, pa svaki opis mora sadržavati i oznaku DOI koja osigurava brzu dostupnost podacima. U našem skupu identificirali smo tri tipa podatkovnih članaka: a) tipične znanstvene članke (16) koji sadržavaju i dodatne datoteke s podacima, b) članke koji opisuju podatke i upućuju na prethodno objavljene znanstvene članke koji su plod tih podataka (7) i c) članke koji sumarno prikazuju podatke a onda upućuju na repozitorij u kome su pohranjeni u cijelosti (6).

4. Zaključak

Iako je u ovome radu prikazan samo isječak vjerojatne zastupljenosti podatkovnih zapisa hrvatskih autora u bazi podataka DCI, svojstva analiziranih zapisa mogući su pokazatelj rezultata njihove ekstrapolacije na prošireni, veći broj. Na veličinu analiziranoga uzorka utjecala je, prije svega, struktura zapisa u većini repozitorija koje DCI pobire, pri čemu navođenje adrese autora istraživačkih podataka nije obvezatno. Rezultati analize stoga su samo indikativni, ali sukladni zapažanjima po kojima istraživačke podatke češće pohranjuju autori koji sudjeluju u istraživanjima iz kojih proizlazi velika količina podataka, poput, primjerice, genetike³⁸ te onima po kojima se broj otvoreno dostupnih istraživačkih podataka posljednjih godina povećava.³⁹ Hrvatski autori svoje su podatke uglavnom pohranjivali u međunarodne podatkovne repozitorije, jer je potrebna infrastruktura na nacionalnoj razini tek razvijena i počinje se koristiti.

Najava Hrvatske zaklade za znanost da će svaki projektni prijedlog ubuduće morati sadržavati i plan upravljanja istraživačkim podacima⁴⁰ nagovještaj je nove politike njihove sustavne pohrane i otvorenog dijeljenja.

Literatura References

1. C. L. Borgman, The Conundrum of Sharing Research Data, *J. Am. Soc. Inf. Sci. Tec.* **63** (6) (2012) 1059–1078, doi: <https://doi.org/10.1002/asi.22634>.
2. Concordat on Open Research Data. UK Research and Innovation (UKRI), URL: <https://www.ukri.org/wp-content/uploads/2020/10/UKRI-020920-ConcordatOnOpenResearchData.pdf> (8. 4. 2021.).
3. C. L. Borgman, Data, disciplines, and scholarly publishing, *Learn. Publ.* **21** (1) (2008) 29–38, doi: <https://doi.org/10.1087/095315108x254476>.
4. A. Vodopijevac, I. Kranjec, Otvoreni istraživački podatci, u I. Hebrang Grgić (ur.), *Otvorenost u znanosti i visokom obrazovanju*, Školska knjiga, Zagreb, 2018., str. 93–112.
5. S. Reilly, W. Schallier, S. Schrimpf, E. Smit, M. Wilkinson, Report on integration of data and publications, 2011., URL: https://zenodo.org/record/8307#.YH_NpVUzaM8 (25. 3. 2021.).
6. J. P. A. Ioannidis, M. J. Khoury, Improving Validation Practices in “Omics” Research, *Science* **334** (6060) (2011) 1230–1232, doi: <https://doi.org/10.1126/science.1211811>.
7. T. E. Hardwicke, M. B. Mathur, K. MacDonald, G. Nilsson, G. C. Banks, M. C. Kidwell, A. H. Mohr, El. Clayton, E. J. Yoon, M. H. Tessler, R. L. Lenne, S. Altman, B. Long, M. C. Frank, Data availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data policy at the journal *Cognition*, *Roy. Soc. Open Sci.* **5** (2018) Art no. 180448, doi: <https://doi.org/10.1098/rsos.180448>.
8. I. Rudan, Znanost ulazi u posve novu eru istraživanja, URL: <https://www.vecernji.hr/premium/pise-igor-rudan-znanost-ulazi-u-posve-novu-eru-istrazivanja-1463810>, <https://www.vecernji.hr/> (21. 1. 2021.).
9. R. A. Poldrack, K. J. Gorgolewski, Making big data open: Data sharing in neuroimaging, *Nat Neurosci.* **17** (11) (2014) 1510–1517, doi: <https://doi.org/10.1038/nn.3818>.
10. J. C. Wallis, E. Rolando, C. L. Borgman, If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology, *Plos One* **8** (7) (2013) Art no. e67332, doi: <https://doi.org/10.1371/journal.pone.0067332>.
11. S. Hoffman, A. Podgurski, The use and misuse of biomedical data: is bigger really better? *Am. J. Law. Med.* **39** (4) (2013) 497–538, doi: <https://doi.org/10.1177/009885881303900401>.
12. R. Uzwyshyn, Research data repositories: The what, when, why, and how, *Comp. Lib.* **36** (3) (2016) 18–21, url: <https://digital.library.txstate.edu/handle/10877/7597>.
13. D. Celjak, I. Dorotić Malić, M. Matijević, L. Poljak, K. Posavec, I. Turk, Istraživački podaci – što s njima?: priručnik o upravljanju istraživačkim podacima, Sveučilište u Zagrebu, Sveučilišni računski centar, Zagreb, 2020., URL: <https://repositorij.srce.unizg.hr/islandora/object/srce%3A327/datastream/FILE0/view> (15. 3. 2021.).
14. D. Čanić, Metapodaci u upravljanju zapisima, *Knjižničarstvo: glasnik Društva knjižničara Slavonije i Baranje* **20** (2) (2016) 41–56.
15. RDA Metadata Standards Directory, URL: <http://rd-alliance.github.io/metadata-directory/> (10. 3. 2021.).
16. C. Bergerud, E. Borglund, Open research data, an archival challenge? *Arch. Sci.* **20** (3) (2020) 279–302, doi: <https://doi.org/10.1007/s10502-020-09330-3>.
17. L. Corti, Recent developments in archiving social research. *Int. J. Soc. Res. Methodol.* **15** (4) (2012) 281–290, doi: <https://doi.org/10.1080/13645579.2012.688310>.
18. K. Shankar, K. R. Eschenfelder, G. Downey, Studying the history of social science data archives as knowledge infrastructure, *Sci. Technol. Stud.* **29** (2) (2016) 62–73, doi: <https://doi.org/10.23987/sts.55691>.
19. H. M. Berman, M. J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, The Protein Data Bank, *Nucleic Acids Res.* **28** (1) (2000) 235–242, doi: <https://doi.org/10.1093/nar/28.1.235>.
20. GenBank, URL: <https://www.ncbi.nlm.nih.gov/genbank/> (10. 3. 2021.).
21. National Oceanic and Atmospheric Administration, URL: <https://repository.library.noaa.gov/> (10. 3. 2021.).
22. GIS Data repository, URL: <https://www.usgs.gov/media/files/gis-data-repository> (10. 3. 2021.).
23. FULIR Data, URL: <https://data.fulir.irb.hr/> (15. 2. 2021.).
24. The Registry of Research Data Repositories, URL: <https://www.re3data.org/> (15. 2. 2021.).
25. Dabar, URL: <https://dabar.srce.hr/> (15. 2. 2021.).
26. Clarivate, The Repository Selection Process. URL: <https://clarivate.com/webofsciencgroup/essays/the-repository-selection-process/> (11. 6. 2021.).
27. N. Robinson, The Data Citation Index & DataCite, DataCite Annual Conference 2014, doi: <https://doi.org/10.5446/15291>.
28. N. Robinson-García, E. Jiménez-Contreras, D. Torres-Salinas, Analyzing data citation practices using the data citation index, *J. Assoc. Inf. Sci. Tech. (JASIST)*. **67** (12) (2016) 2964–2975, doi: <https://doi.org/10.1002/asi.23529>.
29. L. L. Pavlech, Data Citation Index, *J. Med. Libr. Assoc.* **104** (1) (2016) 88–90, doi: <https://doi.org/10.3163/1536-5050.104.1.020>.
30. Web of Science, Data Citation Index, URL: <https://clarivate.com/webofsciencgroup/solutions/webofscience-data-citation-index/> (15. 4. 2021.).
31. J. Schöpfel, D. Farace, H. Prost, A. Zane, Data papers as a new form of knowledge organization in the field of research data, *Knowl. Organ.* **46** (8) (2019) 622–638, doi: <https://doi.org/10.5771/0943-7444-2019-8-622>.
32. P. M. Shaklee, Data in brief – Making your data count, *Data in Brief*. **1** (2014) 5–6, doi: <https://doi.org/10.1016/j.dib.2014.09.001>.
33. O. B. Onyancha, Open Research Data in Sub-Saharan Africa: A Bibliometric Study Using the Data Citation Index, *Pub. Res. Q.* **32** (2016) 227–246, doi: <https://doi.org/10.1007/s12109-016-9463-6>.
34. S. Konkiel, Assessing the Impact and Quality of Research Data Using Altmetrics and Other Indicators, *Scholarly Assessment Reports* **2** (1) (2020) Art no.13, doi: <https://doi.org/10.29024/sar>.
35. Zenodo, URL: <https://zenodo.org/> (10. 3. 2021.).
36. OpenAIRE, URL: <https://www.openaire.eu/mission-and-vision> (15. 2. 2021.).
37. Zenodo. Frequently asked question, URL: <https://help.zenodo.org/> (11. 6. 2021.).
38. M. Assante, L. Candela, D. Castelli, A. Tani, Are scientific data repositories coping with research data publishing?, *Data Sci. J.* **15** (2016) 6, doi: <http://doi.org/10.5334/dsj-2016-006>.
39. L. Pool, 2020 the year of open data? *Research Information*, April-May 2020, URL: <https://www.researchinformation.info/feature/2020-year-open-data> (21. 6. 2021.).
40. Hrvatska zaklada za znanost. Plan rada, URL: <https://hrzz.hr/wp-content/uploads/2021/01/Plan-rada-2021.pdf>, stranica 10 (15. 2. 2021.).

SUMMARY

Research Data of Croatian Authors on the Web of Science Platform

Jelena Bolkovac,^a Tamara Krajna,^{a} and Jelka Petrak^b*

Open access to research data is an important part of the concept of open science. The aim of this article was to identify Croatian authors/institutions that have recognized the importance of sharing research data, and that have given them access via one of the data repositories covered by *Data Citation Index*. A total of 265 documents were found in the database. Data papers by Croatian authors available on the *Web of Science Core Collection* platform were also analysed. Twenty-nine of them were found, mostly of multidisciplinary orientation. The authors of the largest number of data documents are affiliated to the University of Zagreb. Croatian researchers are archiving data in international repositories, because the infrastructure for their archiving at the national level has only recently been developed, and is just beginning to be used.

Keywords

Open science, research data, data papers, data repositories

^a *University of Zagreb
Faculty of Mechanical Engineering
and Naval Architecture
Ivana Lučića 5, Zagreb*

^b *University of Zagreb
School of Medicine
Šalata 3, Zagreb*

*Original scientific paper
Received May 11, 2021
Accepted June 24, 2021*