

NASTAVKI ODPRTIH PODATKOVNIH ZBIRK KOT PODLAGA ZA DRUŽBOSLOVNO IN HUMANISTIČNO RAZISKOVANJE

Pregledni znanstveni članek | 1.02

Izvleček: Odprt dostop do raziskovalnih podatkov na vseh področjih znanosti prinaša nove pristope in na njih temelječa spoznanja. Avtorja nakažeta, kako večja količina in kakovost podatkov, njihova dostopnost, možnosti analiz in interdisciplinarno povezovanje odzvanjajo v humanistiki in družboslovju. Razmere dodatno osvetlita z analizo intervjujev, opravljenih med slovenskimi raziskovalci in nosilci podatkovnih storitev. Podrobneje poudarita prednosti, ki jih prinašajo odprte podatkovne zbirke.

Gljučne besede: odprti podatki, raziskovalna infrastruktura, podatkovne storitve, digitalna humanistika in družboslovje

Abstract: Open access to research data in all fields of science brings new approaches and the resulting insights. The authors of this paper outline how a larger quantity and quality of data, their accessibility, analysis possibilities, and interdisciplinary cooperation can be detected in social sciences and the humanities. The existing situation is further highlighted by their analysis of interviews with Slovene researchers and media data services. The benefits of open databases are discussed in greater detail.

Key Words: open data, research infrastructure, data services, digital humanities and social sciences

Uvod

Pri postavljanju podatkovne infrastrukture v Sloveniji se prav na področju družboslovja in humanistike dogajajo premiki k večji institucionalizaciji in s tem dolgoročno vzdržnemu delovanju. V naravoslovju je pomen raziskovalne infrastrukture že dolgo jasen in ga razumejo predvsem kot veliko raziskovalno opremo. V družboslovju in humanistiki je izražena potreba po raziskovalni infrastrukturi razmeroma nova in se kaže predvsem v obliki reprezentativnih podatkov, pripravljenih in urejenih za raziskovalne namene. Razvoj tovrstne infrastrukture s seboj prinaša vrsto izzivov, obenem pa se odpirajo nove možnosti tako za raziskovanje kot pri organiziranju podatkovnih storitev. Podali bomo pregled ponudbe podatkovnih storitev, ocenili njihovo prepoznavnost in problematizirali dostopnost in možnosti širše izrabe. Zadržki pri odpiranju podatkov v tu obravnavanih disciplinah so nekoliko posebni zaradi same narave gradiva. Povezani so s skrbjo za etično ustrezno pripravo podatkov in pomisleki o njihovi uporabnosti, če so izvzeti iz konteksta nastanka. Zanimalo nas bo, kakšne rešitve se ponujajo za premoščanje teh zadržkov.

Pregled domačih in z njimi povezanih mednarodnih projektov podatkovne infrastrukture

Za graditev infrastrukture podatkovnih storitev so odločilnega pomena domači in mednarodni projekti in pobude. V zadnjem času je glavno gibalno razvoja sprejetje *Načrta razvoja raziskovalne infrastrukture v Sloveniji* (spletni vir 1), ki sledi evropski pobudi ESFRI (spletni vir 2). Njegova značilnost je, da temelji na pobudah raziskovalne skupnosti, saj je nastal na podlagi javnega poziva za podajanje utemeljenih predlogov. Druga značilnost je ideja graditve panevropske infrastrukture, kar operativno pomeni članstvo in finančno udeležbo države pri konzorcijih, ki upravljajo enovite evropske infrastrukturne enote. Končni izbor vsebuje množico predlogov infrastrukture za naravoslovje.¹ Med

predlogi imajo po številu v *Načrtu* pomemben delež tudi programi s področja družboslovja in humanistike, ki so na ta način priznani kot enakopravni sestavni del povezane znanstvene infrastrukture. Družboslovno-humanistični sklop infrastrukturnih programov sestavljajo SHARE (*Survey of Health, Ageing and Retirement in Europe*) in ESS (*European Social Survey*) kot primera centralno koordinirane mednarodne anketne raziskave, ki periodično, na nekaj let, ustvarja kontinuirano podatkovno serijo, odprto dostopno za analizo vsakomur takoj ob nastanku. CESS-DA je podatkovna storitev za dostop do raznovrstnih podatkov s področja družboslovja, DARIAH pa podobno vlogo opravlja za humanistiko. Projekta bomo podrobneje predstavili v nadaljevanju skupaj s tretjim projektom, CLARIN, ki sicer še ni vključen v *Načrt*, si pa njegovi zastopniki prizadevajo za oblikovanje nacionalnega konzorcija izvajalcev in naknadno priključitev v panevropsko enoto za jezikovne vire in tehnologije. Na področju družboslovja v Sloveniji velja za najpomembnejši infrastrukturni projekt CESSDA - *Council of European Social Science Data Archives*. Družboslovje sodi med področja, ki se ponajajo s tradicijo podatkovnih centrov oziroma arhivov, saj so se ti začeli razvijati že po drugi svetovni vojni, vzporedno z razvojem računalniških tehnologij.² CESSDA kot krovna orga-

ropean Infrastructure for Transnational Medicine), biologije v ELIXIR (pan-European research infrastructure for biological information). Med pomembnejšimi evropskimi raziskovalnimi infrastrukturami so še LifeWatch, GBIF (*Global Biodiversity Information Facility*), EPOS (*European Plate Observing System*), European Research Infrastructure on Earthquakes, Volcanoes, Surface Dynamics and Tectonics, PRACE Research Infrastructure, C-ERIC (*Central European Research Infrastructure Consortium*), CERN (*European Organization for Nuclear Research*), ILL (*Institut Laue-Langevin*).

2 Med pomembnejšimi podatkovnimi centri, ki raziskovalcem omogočajo dostop do zbirk z različnih področij družboslovnega raziskovanja, so nemški GESIS – ZACAT, ki sodi pod Leibniz Institute for the Social Sciences, ameriški Inter-university Consortium for Political and Soci-

1 Slovenski raziskovalci s področja medicine so vključeni v EATRIS (*Eu-*

* Doc. dr. Janez Štebe, sociolog, predstojnik Arhiva družboslovnih podatkov, raziskovalec in predavatelj, Fakulteta za družbene vede, Univerza v Ljubljani. 1000 Ljubljana, Kardeljeva ploščad 5, janez.stebe@fdv.uni-lj.si; dr. Sonja Bezjak, sociologinja, samostojna strokovna sodelavka v Arhivu družboslovnih podatkov, Fakulteta za družbene vede, Univerza v Ljubljani. 1000 Ljubljana, Kardeljeva ploščad 5, sonja.bezjak@fdv.uni-lj.si.

nizacija evropskih arhivov družboslovnih podatkov že zdaj prek skupnega portala uporabnikom podatkov zagotavlja dostop do gradiva evropskih podatkovnih arhivov.³ Slovenija po stopnji razvitosti podatkovne infrastrukture ne sodi med razvite države, kot so Velika Britanija, Nemčija ali Nizozemska, pa vendarle je za družboslovje mogoče govoriti o napredni infrastrukturi. Predstavlja jo Arhiv družboslovnih podatkov (ADP), ki obstaja od druge polovice devetdesetih let, njegovo temeljno poslanstvo pa je izbor, hramba in dostop do podatkov pomembnejših in kakovostnih raziskav, povezanih s slovensko družbo. Kot aktivni član ADP zastopa Slovenijo v CESSDA. Članstvo v nadgrajeni panevropski javni infrastrukturni enoti, naslednici CESSDA pod istim imenom, ki je imela junija 2013 ustanovno skupščino, zahteva sodelovanje pri oblikovanju skupnih integriranih podatkovnih storitev (npr. AAI - *Authentication and Authorization Infrastructure* (spletni vir 3), katalogi, procesi), uporabo in vzdrževanje skupnih tezavrov (ELLST - *Multilingual European Language Social Science Thesaurus* (spletni vir 4)) in standardov (DDI - *Data Documentation Initiative* (spletni vir 5)), DSA - *Data Seal of Approval* (spletni vir 6)) ter orodij (NESSTAR (spletni vir 7), Easy (spletni vir 8)). Predvidena je tudi integracija storitev varnega ravnanja z občutljivimi podatki in nadzorovanega dostopa v mednarodnem okolju. Pri nadgraditvi trenutnih storitev na tem področju, ki jih zagotavlja omrežje dobro vzpostavljenih nacionalnih področnih podatkovnih centrov, pomaga projekt *Data Without Boundaries - DWB* (spletni vir 9) iz 7. okvirnega programa, katerega cilj je za raziskovalce urediti enostaven dostop do podatkov uradne statistike, tako na nacionalni kot tudi na mednarodni ravni. Sicer pa je pomembno področje delovanja CESSDA in njenih nacionalnih izvajalcev storitev tudi svetovanje na različnih področjih ravnanja s podatki, usposabljanje in izobraževanje za zahtevnejšo uporabo podatkov in širitev podatkovnih storitev in povezovanje z drugimi področji znanosti. Članice CESSDA so aktivne tudi kot mentorice v državah, v katerih je raziskovalna podatkovna storitvena infrastruktura še v zametkih. Tudi pri tem je slovenski nosilec dejavnosti ADP aktiven v zanimivem projektu SERSCIDA (januar 2012–junij 2014); v njem Slovenija nudi podporo pri izmenjavi znanja pri arhiviranju podatkov družboslovnih znanosti (spletni vir 10) med državami EU, vključenimi v združenje CESSDA, in državami Zahodnega Balkana (Hrvaška, Srbija in Bosna in Hercegovina). Dejavnosti druge od omenjenih infrastrukturnih enot, **DARIAH** (spletni vir 11), so se v Sloveniji okrepile leta 2011, in sicer v programu raziskovalnih infrastruktur Inštituta za novejšo zgodovino, ki ga predstavlja portal in v njem zajeti viri Zgodovina Slovenije SIStory (spletni vir 12) in Znanstvenoraziskovalnega centra SAZU, ki ima infra-

strukturo sicer porazdeljeno po posamičnih inštitutih. Digitalna infrastruktura za umetnost in humanistiko pred uporabnike stopa v obliki spletnega vozlišča SIDIH, in sicer s ciljem »spodbuditi raziskovalce k uporabi novih tehnologij ter izkoriščanju možnosti, ki jih te ponujajo« (spletni vir 13). Kot nacionalna spletna stran bo SIDIH povezovala slovenske raziskovalce s področja umetnosti in humanistike, da se dogovorijo o minimalnih standardih, ki bodo omogočili izmenjavo znanj in orodij, podatkov in zbirk. Vključenost v infrastrukturni program DARIAH slovenskim raziskovalcem omogoča, da vplivajo na razvoj digitalne humanistike, njeno decentraliziranost ter ohranitev enkratnosti posameznih raziskovalnih disciplin in podatkov.

Cilj infrastrukturnega programa **CLARIN** - *Common Language Resources and Technology Infrastructure* je raziskovalcem iz družboslovja in humanistike zagotoviti dostop do digitalnih podatkov o jeziku. V tem smislu CLARIN pomeni postopek zgraditve združenja evropskih podatkovnih repozitorijev, podatkovnih storitev in centrov znanj, do katerih bodo lahko dostopali vsi člani akademske skupnosti držav članic. Slovenija je od leta 2009 v postopku pridruževanja CLARIN, koordinator je pripravljalni fazi je Slovensko društvo za jezikovne tehnologije. S članstvom v tej evropski podatkovni infrastrukturi bo slovenskim raziskovalcem zagotovljen dostop in hranjenje jezikovnih virov in jezikovno-tehnoloških aplikacij.

Ali obstaja izčrpen seznam ponudbe raziskovalnih podatkov iz družboslovja in humanistike?

Za čim boljši izkoristek oziroma izrabo raziskovalnih podatkov so osrednjega pomena podatkovna središča, ki strokovno opravljajo storitve prevzema, izbora, obdelave in hrambe podatkov ter poskrbijo za dostop do njih. Temelj delovanja je torej skrb za kakovostno podatkovno gradivo, ki je odprto dostopno za raziskovalce in širšo javnost, s tem pa sta zagotovljena prihranek pri nepotrebnem podvajanju novih podatkov in višja kakovost spoznanj. Ta temeljna zamisel je tudi v izhodišču dokumenta *Principles and Guidelines for Access to Research Data from Public Funding* (spletni vir 14), ki zavezuje države članice k večji aktivnosti pri dostopu do raziskovalnih podatkov iz javno podprtega raziskovanja. V nadaljevanju predstavljamo nekatere ugotovitve iz poročila »Opis stanja na področju raziskovalnih podatkov v Sloveniji«, ki je nastalo v okviru ciljnega raziskovalnega projekta *Odprti podatki – priprava akcijskega načrta za vzpostavitev sistema odprtega dostopa do podatkov iz javno financiranih raziskav v Sloveniji*. S poročilom smo skušali predstaviti okoliščine, v katerih deluje slovenska raziskovalna skupnost in v katerih nastajajo in se izmenjujejo raziskovalni podatki. V večjem delu poročila je predstavljena analiza stanja v Sloveniji, ta je med drugim utemeljena na 22 poglobljenih intervjujih s slovenskimi raziskovalci in izvajalci podatkovnih storitev. V izbor smo vključili predstavnike 17 različnih raziskovalnih področij, ki v raziskovalnem okolju opravljajo različne vloge (ustvarjalec podatkov, uporabnik podatkov, knjižničar, razvijalec podatkovne tehnologije, vodja, idr.) (spletni vir 15). V nadaljevanju podajamo nekatere probleme in nastave rešitev o vprašanih, povezanih z dostopom do raziskovalnih podatkov, ki smo jih razbrali ob aktivnostih pri projektu *Odprti podatki*, s posebnim poudarkom na družboslovju in humanistiki. Za sogovornike smo ciljno izbirali nosilce, ki so že vpeti v življenjski cikel podatkov, ali kot ustvarjalci, skrbniki, uporabniki ali kot predstavniki vodstev.

al Research (ICPSR), britanski UK Data Archives, nizozemski DANS (*Data Archiving and Networked Services*), norveški Norwegian Social Science Data Services (NSD). V ZDA je osrednji center ICPSR. Tam je še več svetovno pomembnih centrov, kot je Minnesota Population center za dostop do harmoniziranih mikropodatkov popisov, ODUM Institute, ki hrani tudi kvalitativne podatke, idr.

3 Avstrija (WISDOM), Češka (CSDA), Danska (DDA), Estonija (ESSDA), Finska (FSD), Francija (Reseau Quetelet), Grčija (GSDB-EKKE), Nemčija (GESIS), Madžarska (TARKI), Irska (ISSDA), Italija (ADPSS Sociodata), Litva (LiDA), Luksemburg (CEPS), Nizozemska (DANS), Norveška (NSD), Portugalska (APIS), Romunija (RODA), Slovaška (SASD), Slovenija (ADP), Španija (ARCES/CIS), Švedska (SND), Švica (FORS), Velika Britanija (UKDA).

Tako je bilo eno od vprašanj, povezano z poznavanjem tovrstnih aktivnosti, tako lokalno kot v širših mednarodnih povezavah.

Raziskovalci, vključeni v našo raziskavo, so navajali, da pri svojem delu pogosto uporabljajo različne tuje, tudi komercialne, podatkovne zbirke. Na področju ekonomije npr. podatkovne zbirke Amadeus (spletni vir 16), Datastream (spletni vir 17) in BVD – Company information and business intelligence (spletni vir 18). Med domačimi zbirkami sta za področje ekonomije, geografije in druge pomembna vira Statistični urad Republike Slovenije (SURS) (spletni vir 19) in AJ PES (spletni vir 20). Pozanimali smo se tudi, kakšna je bila uporabniška izkušnja pri dostopu do javnih podatkov, ki so jih omenili. Poleg pogosto izraženih želja med predstavniki različnih disciplin, od medicine do geodezije, geografije in ekonomije, da bi bil dostop do nacionalnih uradnih podatkov preglednejši, so bili pogosto omenjeni tudi problemi pri uporabi mednarodnih podatkovnih zbirk, ki jih organizacije, kot so Eurostat, Unesco, OECD ipd., včasih zaračunavajo.

Intervjuvanci s področja humanistike so v pogovorih omenili več zbirk. Različni inštituti v ZRC SAZU svoje storitve oblikujejo tako, da bodo podatki na voljo širšemu krogu uporabnikov (npr. Glasbenonarodopisni inštitut, Arheološki inštitut, Inštitut Frana Ramovša idr.), ZRC SAZU pa za svoje inštitute razvija tudi skupno spletno mesto ARZENAL, kjer lahko raziskovalci odložijo vse, kar bi želeli dati na voljo uporabnikom. Za humanistiko je dober zgled tudi že omenjeno sodelovanje med Inštitutom za novejšo zgodovino in ZRC SAZU v evropskem programu DARI-AH. Za pomembnejšo pobudo na tem področju velja Europeana. V njej sodelujejo strokovnjaki in ustanove s področja dediščine (spletni vir 21). Osrednji center za Europeano v Sloveniji je dLib, ki deluje kot eden od glavnih skrbnikov repozitorija digitaliziranega gradiva iz zbirke NUK (glasbeno gradivo, redki zemljevidi, slikovni viri idr.), skrbi za arhiviranje spleta in deluje tudi kot pooblaščen agregator vsebin za Europeano, deluje pa tudi strokovno na področju digitalne hrambe in podpore širitve storitev odprtega dostopa, kot je zastavljena v projektu OpenAire (plus) (spletni vir 22).

Za področje književnosti velja omeniti projekt Gutenberg, ki omogoča brezplačen dostop do 42.000 e-knjig. Nekaj tovrstnih aktivnosti poteka tudi v Sloveniji (npr. dLib), več o tem pa v nadaljevanju o slovenskih zbirkah. Za preučevanje jezika sta bili poudarjeni podatkovna zbirka Bank of English (spletni vir 23) in Digitalna knjižnica Intratext (spletni vir 24).

Za arheologijo velja omeniti britanski Archeology Data Service, ki je namenjen zbiranju, opisovanju, katalogiziranju in digitalni hrambi podatkov, ki so rezultat arheoloških raziskav. V arheološkem arhivu skrbijo tudi za razširjanje dobrih praks v obliki seminarjev, delavnic in podobno (spletni vir 25).

Pokazalo se je, da je poznavanje zbirk največkrat ozko vezano na raziskovalčevo strokovno področje. Tako ugotavljamo, da je lahko že samo nepoznavanje razpoložljive podatkovne ponudbe ovira tako za pripravljenost za večjo udeležbo pri ponudbi odprto dostopnih podatkov kot tudi pri izrabi v interdisciplinarnem zajemu zanimivih vsebin. Pri odkrivanju področnih ali splošnih podatkovnih repozitorijev so raziskovalcem lahko v pomoč splošni katalogi. Med bolj poznanimi so npr. Databib (spletni vir 26), Registry of Research Data Repositories (spletni vir 27) in DataCite (spletni vir 28). Podoben izčrpen pregled bi bil koristen tudi med nacionalnimi ponudniki raziskovalno zanimivega podatkovnega gradiva.

Pregled domačih zbirk s področja humanistike in družboslovja

V intervjujih se je pokazalo, da imajo slovenski raziskovalci raznovrstne izkušnje s poznavanjem in rabo domačih podatkovnih zbirk. Nekateri naravoslovci in družboslovci so navajali mednarodne podatkovne zbirke, v nekaj primerih niso vedeli, ali domače zbirke sploh obstajajo. Spet drugi z družboslovja in humanistike so navajali predvsem domače zbirke. V nadaljevanju naštevamo tiste, ki so jih omenili udeleženci naše raziskave. Za boljši pregled smo nekaj zbirk dodali po lastnem izboru.

Na področju jezikoslovja že vrsto let delujejo Slovarske in besedilne zbirke Inštituta za slovenski jezik Frana Ramovša ZRC SAZU: Slovar slovenskega knjižnega jezika, Slovenski pravopis 2001, Nova beseda, Besede slovenskega jezika, Terminološki slovarji idr. (spletni vir 29). Na strežniku nl.ijs.si, Odsek za tehnologije znanja Inštituta »Jožef Stefan«, so uporabnikom na voljo jezikovni viri starejše slovenščine (spletni vir 30). Poleg teh so raziskovalcem na voljo tudi referenčni korpus slovenščine Fi-daPLUS (spletni vir 31) in korpus pisanega jezika GIGAFIDA. Slednji vsebuje zbirko tiskanih besedil in besedil s spleta, ki so izšla med letoma 1990 in 2011 – zbirka obsega milijardo besed (spletni vir 32). Iz njega izhaja korpus KRES s sto milijoni besed (spletni vir 33). Omeniti velja še korpus govorne slovenščine, imenovan GOS (spletni vir 34), in korpus ŠOLAR, ki ponuja pogled v pisanje šolajoče se mladine (spletni vir 35).

Za književnost več zbirk ponuja digitalizirana literarna dela oziroma podatke o njih. V intervjujih so bili omenjeni: Slovensko leposlovje na spletu (spletni vir 36), Slovenski zgodovinski roman (spletni vir 37), Slovenska kmečka povest (spletni vir 38). Pri projektu Wikivir so nastale zbirke Slovenska leposlovna klasika, Zbirka slovenskih mladinskih leposlovnih besedil in Zbirka slovenskih pravljic (spletni vir 39). Na strežniku nl.ijs.si je uporabnikom na voljo Digitalna knjižnica IMP z deli, ki so bila izdana med 1584 in 1918 (spletni vir 40). Nekatera klasična besedila so dostopna tudi na dLib (spletni vir 41).

V zgodovino pisje je referenčna točka spletna stran Zgodovina Slovenije - SIStory, ki nastaja v programu *Raziskovalna infrastruktura slovenskega zgodovino pisja* v Inštitutu za novejšo zgodovino (RI INZ). Na spletni strani so poleg izvirnih digitalnih objavljene tudi digitalizirane vsebine, identične izvirniku. Gre za gradivo v obliki slik, tiskanega in ustnega gradiva s področja zgodovine (spletni vir 42). V rubriki »Viri« sta uporabnikom prosto dostopni podatkovni zbirki Smrtne žrtve med prebivalstvom RS med drugo svetovno vojno in neposredno po njej ter Judje na Slovenskem (spletni vir 43). SIStory s svojo dejavnostjo predstavlja dobre zametke strokovno podprtega disciplinarnega podatkovnega središča.

Za arheologijo je osnovna baza arheoloških najdišč Slovenije v zbirki ARKAS; v njej so povezane podatkovne, besedilne baze in kartografske podlage. Enota je arheološko najdišče, ki ga opišejo na različne načine, da ga lahko povežejo z drugimi bazami, ki niso njihove. Nastala je pod okriljem Inštituta za arheologijo ZRC SAZU in predstavlja eno infrastrukturo za celotno arheološko dejavnost. V intervjujih je bila omenjena tudi Zbiva, »arheološka baza podatkov za vzhodne Alpe in obrobje v zgodnjem srednjem veku« (spletni vir 44).

Primer specializiranega spletišča s področja humanistike je Etno-infolab (spletni vir 45), ki obsega slikovno, zvočno in filmsko

gradivo s področja etnologije in kulturne antropologije. Omeniti velja še Enciklopedijo naravne in kulturne dediščine – DEDI, ki nastaja v okviru projekta *Prva digitalna enciklopedija naravne in kulturne dediščine na Slovenskem* in obsega nepremično, premično in živo kulturno ter naravno dediščino v treh okoljih: v digitalni enciklopediji, v interaktivnem spletnem atlasu Geopedia in v tri-razsežnostnem geografskem informacijskem sistemu Gaea+ (spletni vir 46).

V Glasbenonarodopisnem inštitutu ZRC SAZU so si za eno temeljnih nalog postavili »sestavljanje čim popolnejše zbirke slovenske glasbene folklorne«. V arhivu poleg slovenskih ljudskih pesmi hranijo raznovrstne zapise ljudske glasbe drugih narodov in dokumentarne posnetke (spletni vir 47). Obsežna zbirka dokumentarnega gradiva je uporabna le ob dobrem sodelovanju skrbnikov, ki poznajo gradivo, in uporabnikov. V sodelovanju obojih nastajajo zgoščenke, monografije in drugo:

Seveda subjektivno, objektivno, to je stvar avtorja, ampak tu se poskuša na nek znanstveni ali pa poljudnoznanstveni način, če je namenjen širši publiki, predstaviti tovrstno gradivo javnosti. In tega je veliko. Tako da se mi zdi, da je bila v tem smislu dobra pot. Še danes je precej uspešna, kot primer naj navedem, pride nekdo iz nekega kraja ali pa iz neke pokrajine in reče: »Joj, vi imate toliko gradiva, ste snemali 50 let, ste tukaj raziskovali, ali bi nam lahko naredili kopijo tega gradiva? Na primer avdio, mi bi tudi poslušali te posnetke.« Pa je odgovor po navadi tak: »Veste, saj samo poslušanje vam veliko ne bo pomenilo, to ni publicirano gradivo, to ni zbirka plošč. Potrebujete tudi vse metapodatke, naše terenske zvezke in tako naprej, da boste lahko iz tega izžrli to, kar bi lahko. Poleg tega se pa priporoča še poznavanje širšega konteksta.« Pa ne, da bi hoteli zapirati to gradivo. Tako se ponavadi potem dogovorimo, da skupaj, v sodelovanju z njimi naredimo publikacijo, v kateri se to gradivo predstavi, na primer konkretno nek pesemski repertoar neke pokrajine, s komentarji in tako naprej (znanstvenik s področja zvočnega arhiviranja, NO21).

Širša pobuda, ki naj bi povezala dogajanje v humanistiki, se skriva v nastajanju spletne strani SI-DIH, o kateri smo na kratko že spregovorili. Eden od ciljev SI-DIH je k sodelovanju pritegniti vse zainteresirane ponudnike ali hranitelje digitalnih ali digitaliziranih vsebin s področja humanistike.

Na področju umetnosti velja omeniti Podatkovno in slikovno bazo Moderne galerije v Ljubljani o razstavah (Raz) in umetnikih (Ume) 20. in 21. stoletja na Slovenskem. Ta je rezultat zbiranja, vrednotenja in obdelovanja pridobljenih informacij, s katerimi gradijo bazo podatkov o dogajanju v likovni umetnosti 20./21. stoletja na Slovenskem, o njenih akterjih, spremljavalcih (pisci, kritiki, kustosi, naročniki) in odzivnosti v stroki in medijih (spletni vir 48). Postopki digitalizacije arhivskih vsebin potekajo tudi na RTV Slovenija. Končni uporabniki arhiva do vsebin dostopajo prek spletne aplikacije. Za dostop do podatkov se je treba prijaviti v sistem.⁴

Za področje družboslovja je za slovensko družbo pomembna zbirka podatkov, ki jo hrani ADP. V njej je približno 600 raziskav z različnih družboslovnih področij. V glavnem gre za kvantitativne raziskave, nekaj pa je tudi kvalitativnih. ADP poleg

podatkovne datoteke za posamično raziskavo hrani tudi drugo spremljajoče gradivo, potrebno za razumevanje podatkov. Podatki so namenjeni raziskovanju, za dostop do njih je potrebna registracija. ADP je lahko dober zgled za oblikovanje področnih disciplinarnih podatkovnih središč v Sloveniji.

Posebno poglavje v družboslovju so podatki kvalitativnih raziskav, večina teh v Sloveniji nastane v okviru raziskovalnih skupin na Fakulteti za družbene vede, Fakulteti za socialno delo in pri komercialnih ponudnikih raziskovalnih storitev (npr. ARAGON). Omeniti velja tudi Muzej novejših zgodovine, v katerem od leta 2003 ustvarjajo Zbirko ustnih virov, tj. zbirko ustnih pričevanj (Štebe idr. 2011: 51–53). Svoje institucionalne repozitorije ustanavljajo tudi posamične ustanove. V začetni fazi so ti nemalokrat osredinjeni predvsem na hrambo in dostop do publikacij. Tak primer je repozitorij Pedagoške fakultete Univerze v Ljubljani (spletni vir 49). Med dejavnosti repozitorija bi bilo smotno vključiti tudi hrambo in dostop do raziskovalnih podatkov, ustvarjenih v tej ustanovi. Pomemben korak k zgraditvi podatkovne infrastrukture ustanov predstavlja projekt ODUN, v okviru katerega poteka zgraditev repozitorijev univerz v Ljubljani, Mariboru, Novi Gorici in na Primorskem, ki se je končal avgusta 2013. Eden od namenov projekta je »omogočiti odprt dostop do intelektualne produkcije slovenskih univerz zainteresirani javnosti doma in v tujini« (Ojsteršek 2013).

Ocena stanja

Na osnovi predstavljenega pregleda je mogoče naštetati nekaj problemov, ki jih je potrebno upoštevati pri nadaljnjih načrtih razvoja podatkovne infrastrukture za družboslovje in humanistiko:

1. Opraviti imamo z veliko raznovrstnostjo in razdrobljenostjo ozko specializiranih področij, t. i. veliki preostanek (angl. *long-tail*), pri katerem za razložek od prav tako razvpitega problema t. i. velikih podatkov ni ene rešitve za vse. Nadaljnji razvoj infrastrukture zahteva fleksibilnost, ki bo zagotovila specializirano obravnavo glede na vsebino in tip podatkov. Skupna pa mora biti politika, ki bo določila način delovanja podatkovnih središč ter dolžnosti in pravice raziskovalcev pri izročanju podatkov in pri njihovi uporabi. Potrebno je poskrbeti za minimalne standarde pri metapodatkih in harmonizacijo bibliografskih informacij za vključitev v pregledne kataloge.
2. Ponujene in urejene so referenčne zbirke na ravni kataloga, medtem ko je dostop do primarnega gradiva omejen. Primarno gradivo, npr. popis najdb arheološkega najdišča in njihove slike, je tudi gradivo, ki bi bilo zanimivo za uporabnike, vendar je zaradi tradicionalne usmerjenosti v graditev pregledne referenčne zbirke podatkov o najdiščih ta vir potisnjen v ozadje. Potrebno bi bilo ustvariti boljše zavedanje o pomenu primarnega gradiva in načinih, kako nadzorovano zagotavljati njegovo večjo izrabo.
3. Pogosto so zbirke nastale kot stranski produkt, delo navdušencev in angažiranja študentov. Pomembno je spoznati pomen tovrstnih podatkov in omogočiti pilotske projekte za njihovo pripravo. Glede na oceno pomena za znanost in nacionalno dediščino pa bi se na tej podlagi oblikovale možnosti za morebitno trajno skrbništvo.
4. Nekateri problemi so seveda skupni različnim področjem in zadevajo tako razvitejši kot manj razvite ponudnike podatkov. Splošno je vsekakor pomanjkanje sredstev in ustre-

4 Za več o projektu glej: Kosi in Žvelc 2011; Lavrenčič 2011.

zno usposobljenih kadrov. Tako bi bili potrebni ciljni načrti usposabljanja ali pa tudi prekvalifikacije splošnih podatkovnih knjižničarjev, po drugi strani pa dodatno usposobiti ustvarjalce gradiv, da prevzamejo naloge novih profilov podatkovnih znanstvenikov.

Zadržki in ovire na poti vzpostavljanja sistema odprtega dostopa do raziskovalnih podatkov

Poleg tehničnih in finančnih izzivov, ki se porajajo ob ustvarjanju raziskovalne podatkovne infrastrukture in storitev, se raziskovalna skupnost spopada še z drugimi zadržki in ovirami, ko gre za sistem odprtega dostopa do podatkov. Zlasti v humanistiki so raziskovalci, vključeni v intervjuje, poudarili nerazrešene dileme avtorskega prava in varovanja zasebnosti. Za ponazorilo v nadaljevanju navajamo njihove dobesečne izjave.

Vprašanje avtorskih pravic zadeva vsaj dva vidika. Med ustvarjalci podatkov je priznanje avtorstva predvsem priznanje zaslug za ustvarjalni napor in drugič, za uporabnike podatkov lahko avtorske pravice na delih pomenijo omejitve rabe oziroma oviro pred prostim razširjanjem. Ustvarjalci in skrbniki zbirk s področja humanistike so v intervjujih večkrat izrazili zaskrbljenost in strah, da bi v želji po odpiranju raziskovalnih podatkov zaradi nejasnosti in nepoznavanja prestopili meje zakonitega ravnanja, kar je eden od razlogov za restriktivno ravnanje.

Nepoznavanje se nanaša na ravnanje s podatki oziroma publikacijami, ki so jih ustvarili raziskovalci sami – npr. razmerje med avtorskimi in uredniškimi pravicami ali avtorskimi pravicami in pravicami financerja.

Če pa gledam svojo analizo, bi pa jaz čim več mojega dela on-line, ampak si ne upam zaradi tega, ker ne vem, kje so tu uredniške in kje so moje avtorske pravice. Ker tukaj se to povezuje še s citiranostjo. Vedno več hočemo biti citirani in citiran si, če daješ stvari on-line. Da bi me čim več ljudi prebralo, sem naredila svojo stran, kamor sem dala svoje članke, sicer v angleščini, ker to mednarodni prostor bolj zanima, in čim več mojih zapisov, ker bi rada, da me ljudje opazijo (Nina Vodopivec, raziskovalka s področja socialne antropologije).

Delež negotovosti se nanaša tudi na dela, ki so jih ustvarili drugi, skrbniki teh del (podatkov, publikacij, fotografij, slik idr.). pa ne vedo, pod kakšnimi pogoji jih smejo deliti z drugimi in, ali jih sploh smejo.

To bo ovira, čim več bo ljudi prihajalo z digitaliziranim gradivom, s fotografijami, s filmi, z oralno, ustno zgodovino, z glasbo ... To je zelo veliko različnih avtorskih pravic. Mi določenih gradiv sploh ne moremo dati on-line zato, ker se bojimo zaradi avtorskih pravic (Nina Vodopivec, raziskovalka s področja socialne antropologije).

Še posebej težavno je to v primerih, ko uporabniki podatkom dodajo novo vrednost (npr. slovarji, zemljevidi). V tem primeru ostaja vprašanje »lastništva« nedorečeno, nedorečenost pa je velikokrat (neupravičeno) argument za omejitve dostopa do podatkov. Podobno velja za dela in dokumente, pri katerih ni jasno, ali zanje še velja omejitev avtorskih pravic ali ne, in katerih avtorji ostajajo neznani. Več intervjuvancev je pravne dileme označilo za izjemno pomembno oviro.

To je pa posebno poglavje, mislim da, od vsega še najbolj pomembno. To je tako pomembno vprašanje, da bi

morali po mojem mnenju že v projektu imeti kak razgovor s pravniki, ki se ukvarjajo prav s tem področjem. Ker avtorstvo mora biti zagotovljeno (znanstvenik s področja geografije, NO8).

Izrazili so željo, da bi za področje urejanja avtorskih pravic potrebovali strokovno pomoč oziroma svetovalno storitev. Raziskovalne ustanove si pravne službe praviloma ne morejo privoščiti, zato bi bilo smotno razmišljati o razvoju skupne svetovalne službe za celotno raziskovalno skupnost.

Mi nimamo pravnika, tako kot marsikdo drug ne, zato področje urejamo na tak in drugačen način, po evropskih standardih in zahtevah, da je vse, kar mora biti, avtorsko pravno razčiščeno in tako naprej. Toda mi nismo strokovnjaki za to, sredstev pa tudi ne dobimo toliko, da bi lahko vsako leto najeli pravnika, ki stane nekaj deset tisoč evrov. Zato bi bilo smiselno razmišljati o nekem skupnem centru, saj bi lahko bil skupen za tri, štiri ministrstva. Samo da je neka storitev (znanstvenik s področja zgodovine, NO10).

Problemiziranje avtorskopравnih vprašanj in iskanje rešitev za odprti dostop je med svojimi nalogami poudaril SIDIH /DARIAH. Priznanje avtorstva, kakor ga razumemo danes, prinese vrsto omejitev, zato ni primerna rešitev, če želimo, da podatki ostanejo dostopni čim širšemu krogu uporabnikov. Pri nastajanju novih del je smotno upajati uporabo odprtih licenc, ki čim manj omejujejo ponovno rabo (Ball 2012).

Nekateri intervjuvanci so opozorili na problem varovanja zasebnosti vpletenih v raziskave. To vprašanje je še posebej aktualno pri kvalitativnih raziskavah, ki vključujejo manjše število udeležencev, katerih identiteto je težje zakriti.

Pri nas je še ena težava, sploh takrat, ko gre za študij primera, je zelo težko zagotoviti anonimnost. Ker če si v nekem konkretnem okolju, se da že iz teh zgodb, ki so napletene, zelo hitro razbrati, za koga gre. Tako da ne moremo vsega dati v javnost. In tukaj se najbolj sprašujem, kako to zagotoviti. Če daš podatke, ki so širše dostopni, ne govorim zgolj znotraj hiše, ker tukaj smo tako ali tako vsi temu zavezani, ampak ko bi bili pa malo širše dostopni, je pa že vprašanje (znanstvenik s področja socialnega dela, NO7).

Z varovanjem zasebnosti imajo opraviti tudi večje in obsežnejše zbirke podatkov, pomembne za družboslovje, npr. podatki o posameznikih, ki jih zbira in hrani SURS ali IVZ. Da bi zadostili pravici o varovanju zasebnosti, so raziskovalcem na voljo podatki v anonimizirani obliki. V posebnem postopku so odstranjeni neposredni identifikatorji, s katerimi bi bilo mogoče ugotoviti identiteto posameznika, vključenega v raziskavo. V primerih, ko to ni dovolj, so za ravnanje z občutljivimi podatki predvidene tudi druge rešitve, npr. varna soba, reguliran dostop do podatkov. Tudi etične komisije lahko pripomorejo pri boljšem seznanjanju in zgodnjem odkrivanju problemov in rešitev glede z dostopom do podatkov povezanih dilem. Nekateri se sicer bojijo prevelike birokratizacije nadzora nad predlogi raziskovalnih projektov (Malnar 2010), vendar ocenjujemo, da bi v našem okolju tovrstna regulacija prinesla več koristi kot škode, tako za raziskovalno skupnost pri etično ustreznem pridobivanju in izrabi podatkov kot za državljane; ti se pogosto čutijo preobremenjeni s premalo utemeljenimi raziskavami, ki pogosto ostanejo zaprte v predalu.

Postopek anonimizacije kot način varovanja sodelujočih je posebej zahteven v primeru slikovnega, zvočnega in video gradiva. Ponovna uporaba in zagotavljanje dostopa do primarnega gradiva, ko gre za »kvalitativne« podatke, je še vedno prej izjema kot pravilo. Širjenje kulture razširjanja podatkov skupaj z usposabljanjem za izrabo in pripravo testnih primerov za učenje lahko pomembno oživi ponudbo na tem področju (Gray in Byrne 2012: 5). V primerih, ko raziskovalci - ustvarjalci podatkov ali skrbniki podatkov niso seznanjeni z metodami in postopki, ki zagotavljajo varovanje pravice do zasebnosti, razširjanju podatkov niso naklonjeni in jih, da bi se izognili kršitvam podatkov, raje ne delijo z drugimi. Zato je tudi pri tem nadvse pomembno izmenjevanje izkušenj in prenašanje dobrih praks med področji. Pri reguliranju in omejevanju rabe občutljivih podatkov za raziskovalne namene je nujno opozoriti na potrebe in cilje znanosti, ki so spoznavati družbo in pomagati pri njenem razvoju. Na te, »višje« cilje je treba misliti, ko se ob vse večjem zavedanju uveljavljanja pravice do zasebnosti sprejemajo odločitve o ponovni rabi občutljivih podatkov. Tudi za to je treba o rabi podatkov za raziskovalne namene misliti ločeno od komercialnih ali drugih namenov in postoriti vse, da bodo podatki opremljeni tako, da bodo lahko dosegli čim večje število uporabnikov, hkrati pa ne bodo oropani pomembnejših raziskovalnih vsebin. Pri tem bi morala znanstvena skupnost aktivneje spremljati sprejemanje nove zakonodaje in spodbujati upoštevanje interesov znanstvenega raziskovanja, ki morajo zaradi splošne koristi za skupnosti kot dejavnosti biti izvzeti iz splošnih določb zakonodaje na tem področju.⁵ V intervjujih, pa tudi na delavnicah in razpravah, ki so potekale v okviru projekta *Odprti podatki*, so udeleženci večkrat poudarili pomen ustreznega skrbništva v primerih, ko gre za varovanje podatkov o redkih primerkih, ki so velikega pomena bodisi za stroko bodisi za državo:

Podatki iz zbirke ARKAS (Arheološki kataster Slovenije) so za zdaj dostopni le za strokovne namene. Interesenti si dostop prek spleta lahko zagotovijo na podlagi prošnje za dovoljenje, ki je praviloma odobrena vsem iz arheološke stroke in študentom arheologije, kakor tudi tistim, ki sodelujejo z nami bodisi v skupnih projektih ali tako, da nam posredujejo nove podatke o arheoloških najdiščih. Za tak restriktiven dostop smo se odločili zato, ker si arheologi ne želijo nestrokovnih in nenadzorovanih posegov raznih ljubiteljev in zbirateljev v arheološka najdišča ter iskanja arheoloških predmetov iz nekih osebnih nagibov in materialnih koristi, kar se dogaja. Za namene splošnega izobraževanja smo pripravili druge aplikacije, ki so odprte širši zainteresirani javnosti (npr. Interaktivna karta Slovenije z zbirkami ZRC SAZU, ARZENAL, ZBIVA) in ne vključujejo občutljivih podatkov, npr. o natančni lokaciji najdišča (raziskovalec s področja arheologije, NO14).

Za premoščanje ovir in zadržkov pri avtorskih pravicah, varovanju zasebnosti in zaščiti redkih primerov so potrebni ustrezna priprava in hramba podatkov ter dostop pod ustreznimi pogoji. Raziskovalci poročajo tudi o tem, da bi bilo treba surove podatke še dodatno prečistiti in pregledati ter opremiti z več metapodatki. V nasprotnem primeru se neizkušen uporabnik ne more znajti in lahko napačno interpretira vsebino. Kot rešitev za tovrstne

primere se ponuja večja vključenost primarnih raziskovalcev pri uvajanju novih uporabnikov. Tako bi pomembni raziskovalci - ustvarjalci kot skrbniki in dobri poznavalci kompleksnih podatkovnih virov opravljali tudi infrastrukturno storitev, tj. skrb, zagotavljanje dostopa, usposabljanje in svetovanje za uporabo podatkov. Takšne primere si lahko predstavljamo tako v družboslovju, npr., da ustvarjalci anketnih podatkov s kompleksno longitudinalno zasnovijo svetujejo o posebnosti načrta in pristopov k analizi, kot v humanistiki, kjer bi npr. centri, kot je Glasbeno-narodopisni inštitut, ki že zdaj razvijajo dobre prakse in poznajo najvišje standarde pri ustvarjanju in izrabi podatkov, lahko svojo dejavnost širili navzven skupaj z nadzorovanim dostopom do podatkov. Na področjih, kjer ne poznajo druge rabe podatkov ali je področje nacionalno omejeno, so razprave o odpiranju podatkov trčile na nemalo odpora.

Mi smo zelo nepovezani, kar je značilno za humanistiko v celoti. Osnovna pomanjkljivost humanistike je, da je nacionalno omejena. To ne velja samo za slovensko, ampak tudi za druge. Zato so nacionalne filologije označene kot konzervativne stroke, kar nam seveda ni prav in želimo našo podobo spremeniti. Sam opozarjam na podobne projekte v tujini v zadnjih 10 letih (Miran Hladnik, znanstvenik s področja slovenistične literarne vede).

V pogovorih z različnimi raziskovalci ugotovimo, da je zaprtost v veliki meri stvar nepoznavanja področja podatkovne infrastrukture in storitev, odklonilen odnos pa stvar predsodkov v smislu, da gre pri drugi oziroma ponovni rabi podatkov za neznanstven pristop ipd. Tako npr. raziskovalec slovenist opaža spor med t. i. tradicionalisti, ki svoje analize razvijejo na osnovi natančnega branja besedil, in drugimi, ki analize gradijo na podatkovnih zbirkah.

Gre za opozicijo med dvema tipoma dela v stroki. Tradicionalno literarnovedno pisanje nima z velikimi podatkovnimi zbirkami nobenih opravkov in je kljub temu še vedno deležno določenega družbenega ugleda, saj kreira družbena pričakovanja v zvezi s slovensko književnostjo in jim ustreza. Drugo je pa to, čemur se reče literarna znanost in si jo predstavljamo samo na osnovi čim boljših, čim bolj natančno in popolno zbranih podatkov (Miran Hladnik, znanstvenik s področja slovenistične literarne vede).

Naklonjenost podatkovnim zbirkam je pogosto povezana tudi s poznavanjem računalniške tehnologije.

Bioinformatika, genomika ima tako imenovane antologije in tam se da napisati marsikaj. To je zato, ker so v tem delo udeleženi tudi računalničarji in informatiki, ki to vidijo na ta način, da se da to nekako avtomatično preoblikovati. Tukaj je ta zavest najbolj prisotna, druge pa niti ne (znanstvenik s področja biologije, NO6).

Pozitivna pa je tudi izkušnja medgeneracijskega sodelovanja.

Jaz vidim, da res potrebujem mlajše kolegice, prav zaradi tega, ker ste že na začetku rekli. Starejši kolegi so navajeni na svoj način dela in tudi če bomo nekaj vzpostavili, jih je ponavadi zelo težko prepričati o teh stvareh, pokazati, da so smiselne, kaj so koristi ... Pri mlajših se mi pa zdi, da pravzaprav šele vstopajo v to polje in je nekoliko lažje. Tudi za izobraževanja in za vse te stvari so še zelo odprti. Čeprav po drugi strani pa potem včasih naletimo na teža-

⁵ Glej npr. odgovor Economic and Social Research Council iz Velike Britanije na osnutke sprememb evropske zakonodaje (spletni vir 50).

vo, ko si nekdo misli, kaj mi bo sedaj tam en mladi pamet solil, da moram jaz tukaj neko stvar izpolnit (znanstvenik s področja socialnega dela, NO7).

Sklepi

Eden od ciljev prispevka je bil domačo javnost seznaniti z razpoložljivo in nastajajočo podatkovno infrastrukturo v Sloveniji ter s povezavami z mednarodnim podatkovnim okoljem. Ugotavljamo, da je celovit pregled stanja ponudbe na tem področju težko doseči zaradi premajhne povezanosti in razdrobljenosti. Pregled projektov, programov, zbirk, pobud, arhivov in centrov s področja humanistike in družboslovja, ki se posredno ali neposredno ukvarjajo z raziskovalnimi podatki, je poskus, kako prispevati k večji in boljši ozaveščenosti o odpiranju raziskovalnih podatkov v Sloveniji in s tem tudi k večji izrabi obstoječega gradiva in možnosti, ki jih to prinaša družboslovju in humanistiki.

Že kratek pregled pokaže, da je to področje tudi v Sloveniji, čeprav nekoliko manj kot v mednarodnih krogih, precej živahno in prinaša široko paleto problemov kakor tudi rešitev, ki zahtevajo novo organiziranje raziskovalne infrastrukture, raziskovanja in financiranja raziskovalne dejavnosti. Vse to pa seveda prinaša vrsto zadržkov, predsodkov in zahteva uskladitev tako na politični (konsenzualni) ravni, na ravni kulture (tekmovalnost, konkurenčnost) in tudi na ravni samih metod raziskovanja. V družboslovju in humanistiki so morda bolj kot na drugih področjih znanosti opazne velike razlike med različno usmerjenimi raziskovalci. Gre za prepletanje med viri, podatki, in njihovo interpretacijo. Na nekaterih področjih gre tudi za vprašanje nacionalno in regionalno omejenih terminologij in klasifikacij (npr. arheološke klasifikacije), spet drugod je celostna izkušnja, pridobljena z navzočnostjo *in situ* videti nenadomestljiva.

Medpodročno sodelovanje ter izmenjava znanj in izkušenj bosta prispevali k razvoju podatkovne infrastrukture in storitev ter k odpiranju novih raziskovalnih poti. Omenjene uspešne izkušnje so bile pogosto povezane z uspelim srečanjem med naravoslovno-tehnično izobraženimi sodelavci in družboslovno-humanistično podkovanimi vsebinci. S tem v zvezi velja omeniti poročilo o izkušnjah projektov iz razpisa One Culture, ki se spopada s problemi velikih količin digitaliziranega gradiva s področja humanistike in družboslovja ter uporabe novih metod raziskovanj, ki jih prinaša razvoj računalniških tehnologij (spletni vir 51). Podobno opaža poročilo iz delavnice projekta, ki išče odgovore na skupne izzive različnih, prej omenjenih evropskih infrastrukturnih enot s področja družboslovja in humanistike; to je projekt **DASISH** (spletni vir 52). Tradicionalni epistemološki zidovi so se začeli rušiti s širjenjem novih podatkovnih virov, npr. sistematičnim zapisovanjem osebnih zgodovín istih posameznikov skozi čas, kar je prineslo strokovnejšo organizacijo ravnanja s podatki in več skupinskega dela (Gray in Byrne 2011: 1).

Ti in drugi izzivi, povezani s pripravo in dostopom do raziskovalnih podatkov, delajo to področje zanimivo in inovativno. Zato ne preseneča velik delež posameznikov s področja humanistike in družboslovja, ki se poleg svoje raziskovalne dejavnosti posvečajo delovanju pri podatkovni infrastrukturi. Navdušenje je tu opaznejše kakor v naravoslovju. Velike izzive prinaša tudi povezovanje različnih podatkovnih virov in analitičnih pristopov ter s tem povezano interdisciplinarno povezovanje. Koristi prihranka stroškov in prednosti pri kakovosti spoznanj, ki ga kot motiv odprte znanosti in ožje, odprtega dostopa do podatkov, omenjamo

na začetku, se na otipljiv način kaže v posamičnih primerih. Tak primer bi lahko bila dodana vrednost ustvarjenih podatkov v povezavi med prostorom in literaturo, ki nastajajo v raziskovalnem projektu *Prostor slovenske literarne kulture. Literarna zgodovina in prostorska analiza z geografskim informacijskim sistemom*. Tematika kar sama vabi k širšim interdisciplinarnim pristopom pri ustvarjanju podatkov in pri analizah ter prikazih izven kroga sodelavcev projekta (Perenič 2012).

Podobne možnosti širše izrabe se nakazujejo pri projektu digitalizacije zgodovinskih popisov na območju RS, ki poteka pod okriljem SIStory (Šorn idr. 2012). Značilnosti tega projekta so izboljšanje dostopa do gradiva v skrbništvu javnih arhivov, sodelovanje uporabnikov in študentov pri prepoznavanju besedila zapisov in pri vnašanju prepoznanih besedil v podatkovno bazo, zagotavljanje dostopa do surovih podatkov za širšo javnost (če kdo išče podatke o svojih prednikih) in za znanstvene in izobraževalne namene ter vključenost v povezano infrastrukturo za humanistiko SI-DIH (Pančur 2013). Sodelovanje javnosti pri ustvarjanju in izrabi (angl. *crowd sourcing*, *citizen science*), skrb za kakovost in dostop, vključevanje v širše infrastrukturne programe, kot jih razberemo iz naštetega, vse to so tudi značilne prednosti odprte znanosti (Lyon 2009).

Sklenemo lahko s tem, da je potencialov odprte znanosti, ki jih predstavljajo dostopnejši raziskovalni podatki, prav v družboslovju in humanistiki še zelo veliko, in da se na nekaterih področjih šele odkrivajo prave koristi odprtejšega dostopa. Razveseljivo je, da se ob tem krepi tudi institucionalizacija dejavnosti, ki se kaže tako v pripravljenosti nosilcev politike, da sledijo razvitim raziskovalnim okoljem z večjim poudarkom na razvoju podatkovne infrastrukture, tako pri vključevanju v programe ESFRI kot pri sodelovanju pri načrtovanju bodočih aktivnosti graditve mreže podatkovnih središč in pripravi ustreznih zavezujočih politik.

Literatura

BALL, Alex: *How to License Research Data*. Edinburgh: Digital Curation Centre, 2012 (*DCC How-to Guides*), 1–16 (dostopno na: <http://www.dcc.ac.uk/resources/how-guides>).

GRAY, Jane in Claire Byrne (ur.): *Data Service Infrastructure for the Social Sciences and Humanities: Deliverable Report*. DASHIS, 2012, 1–12 (dostopno na: http://dasish.eu/publications/projectreports/D8.7_-_Qualitative_Data_Workshop_Report__1_.pdf).

KOSI, Bojan in Martin Žvelc: Mediateka – digitalizacija in trajno hranjenje zvokovnih vsebin RTV Slovenija. *Knjižnica* 55/1, 2011, 115–128.

LAVRENČIČ, Aleksander: Hramba digitalnega gradiva v televizijskem arhivu. *Knjižnica* 55, 1, 2011, 129–138.

LYON, Liz: *Open Science at Web-Scale: Optimising Participation and Predictive Potential*. Bath: Digital Curation Centre, 2009 (JISC Consultative Report), 1–45 (dostopno na: <http://www.jisc.ac.uk/media/documents/publications/research/2009/open-science-report-6nov09-final-sentojisc.pdf>).

MALNAR, Brina: Raziskovalna etika med znanostjo, zasebnostjo in birokracijo. *Družboslovne razprave* XXVI/64, 2010, 7–24 (dostopno na: <http://druzboslovnerazprave.org/clanek/pdf/2010/64/1/>).

OJSTERŠEK, Milan: *Predstavitev možnosti vključitve raziskovalnih podatkov v nacionalno infrastrukturo odprtega dostopa*. Predstavitev v okviru Delavnice »Napredne tehnologije za vzpostavitev raziskovalne podatkovne infrastrukture v Sloveniji«, FDV, 22. maj 2013. Ljubljana: Arhiv družboslovnih podatkov, Fakulteta za družbene vede, Univerza v Ljubljani 2013 (dostopno na: http://www.adp.fdv.uni-lj.si/odpp10D3/presentations/Milan_Ojstersek.pdf).

PANČUR, Andrej: Osebna korespondenca, 2013.

PERENIČ, Urška: Prostor v literaturi in literatura v prostoru. *Slavistična revija* 60/3, 2012, 259–264 (dostopno na: http://www.srl.si/sql_pdf/SRL_2012_3_01.pdf).

ŠORN, Mojca idr.: Spletni portal Zgodovina Slovenije - Sistory: Prost in odprt dostop do digitaliziranih in digitalnih zgodovinskih in zgodovinsopisnih vsebin. *Knjižničarske novice* 3, 2012, 17–20.

ŠTEBE, Janez idr.: *Opis stanja na področju raziskovalnih podatkov v Sloveniji: Poročilo projekta Odprti podatki – priprava akcijskega načrta za vzpostavitev sistema odprtega dostopa do podatkov iz javno financiranih raziskav v Sloveniji*. Ljubljana: Arhiv družboslovnih podatkov, Fakulteta za družbene vede, Univerza v Ljubljani, 2012, 1–131.

ŠTEBE, Janez idr.: Archiving and Re-using Qualitative and Qualitative Longitudinal Data in Slovenia. *IASSIST Quarterly 2010/2011*, 2011, 50–59.

Spletni viri

Spletni vir 1: <http://www.arhiv.mvzt.gov.si/fileadmin/mvzt.gov.si/pageuploads/pdf/znanost/nacrt-RI.pdf>, 29. 5. 2013.

Spletni vir 2: http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri, 29. 5. 2013.

Spletni vir 3: <http://aai.arnes.si/>, 29. 5. 2013.

Spletni vir 4: <http://elsst.esds.ac.uk/>, 29. 5. 2013.

Spletni vir 5: <http://www.ddalliance.org/>, 29. 5. 2013.

Spletni vir 6: <http://dataealofapproval.org/>, 29. 5. 2013.

Spletni vir 7: <http://www.nesstar.com/>, 29. 5. 2013.

Spletni vir 8: <https://easy.dans.knaw.nl/ui/home>, 29. 5. 2013.

Spletni vir 9: <http://www.dwbproject.org/>, 29. 5. 2013.

Spletni vir 10: <http://www.iassist2013.org/program/sessions/session-f2/#c248>, 29. 5. 2013.

Spletni vir 11: <http://www.dariah.eu/>, 29. 5. 2013.

Spletni vir 12: <http://www.sistory.si/>, 29. 5. 2013.

Spletni vir 13: <http://www.sidih.si/>, 29. 5. 2013.

Spletni vir 14: <http://www.oecd.org/sti/sci-tech/38500813.pdf>, 29. 5. 2013.

Spletni vir 15: http://www.adp.fdv.uni-lj.si/o_arhivu/publikacije/odpp10_opis_stanja/, 29. 5. 2013.

Spletni vir 16: <http://biblio.iue.it/record=b1381738>, 29. 5. 2013.

Spletni vir 17: http://thomsonreuters.com/products_services/financial/financial_products/a-z/datastream/, 29. 5. 2013.

Spletni vir 18: <http://www.bvdinfo.com/home.aspx>, 29. 5. 2013.

Spletni vir 19: http://www.stat.si/drz_stat.asp, 29. 5. 2013.

Spletni vir 20: http://www.ajpes.si/Statistike/Druge_statistike?id=534, 29. 5. 2013.

Spletni vir 21: <http://pro.europeana.eu/>, 29. 5. 2013.

Spletni vir 22: <http://www.openaire.eu/en/component/content/article/326-openaireplus-release>, 29. 5. 2013.

Spletni vir 23: <http://www.mycobuild.com/about-collins-corpus.aspx>, 29. 5. 2013.

Spletni vir 24: <http://www.intratext.com/>, 29. 5. 2013.

Spletni vir 25: <http://archaeologydataservice.ac.uk/>, 29. 5. 2013.

Spletni vir 26: <http://databib.org/>, 29. 5. 2013.

Spletni vir 27: <http://www.re3data.org/>, 29. 5. 2013.

Spletni vir 28: <http://www.datacite.org/members>, 29. 5. 2013.

Spletni vir 29: <http://bos.zrc-sazu.si/>, 29. 5. 2013.

Spletni vir 30: <http://nl.ijs.si/imp/index-sl.html>, 29. 5. 2013.

Spletni vir 31: <http://www.fidaplus.net/Index.htm>, 29. 5. 2013.

Spletni vir 32: <http://www.slovenscina.eu/korpusi/gigafida>, 29. 5. 2013.

Spletni vir 33: <http://www.slovenscina.eu/korpusi/kres>, 29. 5. 2013.

Spletni vir 34: <http://www.slovenscina.eu/korpusi/gos>, 29. 5. 2013.

Spletni vir 35: <http://www.slovenscina.eu/korpusi/solar>, 29. 5. 2013.

Spletni vir 36: <http://slovenskaliteratura.ff.uni-lj.si/sl.html>, 29. 5. 2013.

Spletni vir 37: http://lit.ijs.si/zgr_dons.html, 29. 5. 2013.

Spletni vir 38: <http://lit.ijs.si/skp1.html>, 29. 5. 2013.

Spletni vir 39: <http://sl.wikisource.org/wiki/Wikivir:WikiProjekt>, 29. 5. 2013.

Spletni vir 40: <http://nl.ijs.si/imp/dl/index-date.html>, 29. 5. 2013.

Spletni vir 41: <http://www.dlib.si/>, 29. 5. 2013.

Spletni vir 42: <http://www.sistory.si/>, 29. 5. 2013.

Spletni vir 43: <http://www.sistory.si/publikacije/?menu=10>, 29. 5. 2013.

Spletni vir 44: <http://zrcalo1.zrc-sazu.si/zbiva/>, 29. 5. 2013.

Spletni vir 45: <http://www.etnoinfolab.org/>, 29. 5. 2013.

Spletni vir 46: <http://www.dedi.si/>, 29. 5. 2013.

Spletni vir 47: <http://gni.zrc-sazu.si/sl#v>, 29. 5. 2013.

Spletni vir 48: <http://razume.mg-lj.si/?lang=si>, 29. 5. 2013.

Spletni vir 49: <http://pefprints.pef.uni-lj.si/>, 29. 5. 2013.

Spletni vir 50: <http://www.esrc.ac.uk/about-esrc/information/eu-data-protection.aspx>, 29. 5. 2013.

Spletni vir 51: <http://www.clir.org/pubs/reports/pub151>, 29. 5. 2013.

Spletni vir 52: http://dasish.eu/about_dasish/, 29. 5. 2013.

Rudiments of Open Databases as a Research Basis for Social Sciences and the Humanities

Increasing needs for the development of data infrastructure are felt virtually in all research spheres. The paper provides an overview of domestic and related international projects on the development of data infrastructure in the humanities and in social sciences. Especially important are the following international centers: CESSDA (Consortium of European Social Science Data Archives) for social sciences, DARIAH (Digital Research Infrastructure for the Arts and Humanities) for the humanities, and CLARIN (Common Language Resources and Technology Infrastructure) for language resources. They are linked to the relevant national units whose long-term importance is confirmed in the national document titled *Development Plan of Research Infrastructure in Slovenia*. The Plan presents emerging and current databases designed for research in social sciences and the humanities, as well as data centers specialized in the reception, storage, and dissemination of research data. Due to vigorous development, innovative approaches, and new solutions the creation of data infrastructures is interesting and appealing. Particularly challenging are the use of metadata standards for long-term digital storage and inclusion in common, transparent catalogs in order to facilitate user searching and access, and thus achieve better utilization of available sources. However, this development also brings considerable concern and a number of obstacles that will have to be overcome in order to set up in Slovenia a system of universal open access to research data supported by public funding. For this purpose we need umbrella national policies on open access to research data. Researchers, who are the creators and guardians of such data, which in a broader sense includes different types of digital sources, need to have clearly defined responsibilities in ensuring access for other users. It is also necessary to provide adequate support in the form of data services by specialized data centers, as well as other incentives. Uncertainty regarding copyright, for instance, frequently leads to excessive restrictions that may be avoided if appropriate advice and the use of commonly-established solutions for broader fields were provided. A similar solution applies to the ever-topical issue of the protection of privacy.

The paper further analyses the current situation in the treatment of research data and the ensuing opportunities for the research community. These opportunities are primarily in overcoming the traditional paradigmatic division into individual disciplines; in the multi-method and multi-data approach; and in the multidisciplinary use of data, which can yield new insights.

