

Z občansko znanostjo do validacije podatkov o COVID-19 v Sloveniji – primer skupnosti Sledilnik

Abstract

Validating Data on the COVID-19 Pandemic in Slovenia Through Citizen Science – The Sledilnik Example

In early 2020, the world found itself in the middle of the COVID-19 pandemic. In Slovenia, the first confirmed infection was recorded on the 4th of March 2020, and the numbers increased in the following weeks. The National Institute of Public Health (NIPH) published aggregated data on the number of infected individuals and later also the number of deaths caused by the disease. However, data users, including a number of leading Slovenian researchers, noticed that this data was often inconsistent in comparison with publications in other sources. They created an informal community called *COVID-19 Sledilnik* (COVID-19 Tracker) that established a direct connection with the NIPH and health care institutions and published, validated and transformed structured data into a format suitable for visualisations and further analysis, and also the development of models and forecasts. The article presents the *Sledilnik* community as an example of a community science initiative, as well as its origin, structure and results. In the second part, the authors present guidelines for COVID-19 data management that were prepared by a working group at the Research Data Alliance. The article also evaluates the data management in the *Sledilnik* community. We find that the community, although unaware of these guidelines, has largely followed principles of openness and transparency. The authors provide suggestions on how their work could be upgraded and implemented within public institutions and how it could leverage Slovenia's existing data services infrastructure.

Keywords: citizen science, data management, COVID-19 pandemic, COVID-19 Tracker Slovenia project

Ana Slavec is a sociologist and holds a PhD in statistics. She is a member of the Research Data Alliance and previously served as its ambassador for the field of technology, more

specifically renewable materials. As part of the COVID-19 Sledilnik team, she participated in the collection and analysis of data from public opinion polls related to the coronavirus epidemic. She is employed as a postdoctoral researcher and consulting statistician at the InnoRenew CoE Centre of Excellence (ana.slavec@innorenew.eu).

Andrej Srakar holds a PhD in economics and is a doctoral student of mathematical statistics. He is a founding member of the Tracker scientific society and one of the earliest members of the COVID-19 Sledilnik team, playing an active role in modeling the epidemic and presenting it to the scientific community. He is the initiator and coordinator of the international project YoungStatS of FENStatS. He is employed as a research associate at the Institute for Economic Research (IER), and is also an assistant professor of economics and quantitative methods at the Faculty of Economics, University of Ljubljana (andrej.srakar@ier.si).

Povzetek

V začetku leta 2020 se je svet znašel sredi pandemije bolezni COVID-19. V Sloveniji smo prvo potrjeno okužbo zabeležili 4. 3. 2020, v naslednjih tednih pa je število okuženih naraščalo. Nacionalni inštitut za javno zdravje (NIJZ) je objavljala agregirane podatke o številu okuženih in kasneje tudi umrlih, ko so se ti začeli pojavljati. Vendar so uporabniki podatkov, med njimi vrsta vodilnih slovenskih raziskovalcev, opazili, da so ti velikokrat nekonsistentni z objavami v drugih virih. Povezali so se v neformalno skupnost COVID-19 Sledilnik, ki je vzpostavila neposredno povezavo z NIJZ in zdravstvenimi zavodi ter strukturirane podatke objavljala, validirala in preoblikovala v format, primeren za vizualizacije, nadaljnjo analizo ter razvijanje modelov in napovedi. Prispevek skupnost Sledilnik predstavi kot primer pobude občanske znanosti, njen nastanek, strukturo in rezultate. V drugem delu predstavimo smernice za ravnanje s podatki COVID-19, ki jih je pripravila delovna skupina Združenja za raziskovalne podatke, ter primerjavo in ovrednotenje načina ravnanja s podatki v skupnosti Sledilnik. Ugotovimo, da je skupnost, čeprav teh smernic ni poznala, v veliki meri sledila podobnim načelom odprtosti in transparentnosti. Podamo predloge, kako bi lahko njeno delo nadgradili, implementirali v okviru javnih institucij ter uporabili obstoječo infrastrukturo podatkovnih storitev, ki je na voljo v Sloveniji.

Ključne besede: občanska znanost, ravnanje s podatki, pandemija COVID-19, COVID-19 Sledilnik

Ana Slavec je sociologinja in doktorica statistike. Je članica Združenja za raziskovalne podatke RDA in njegova nekdanja ambasadorica za področje tehnike – obnovljivih materialov. V okviru projekta COVID-19 Sledilnik sodeluje pri zbiranju in analizi podatkov javnomnenjskih raziskav, povezanih z epidemijo koronavirusne bolezni. Zaposlena je kot podoktorska raziskovalka in svetovalka za statistiko v centru odličnosti InnoRenew CoE (ana.slavec@innorenew.eu).

Andrej Srakar je doktor ekonomskih znanosti in doktorski študent matematične statistike. Je ustanovni član Znanstvenega društva Sledilnik in skoraj od samih začetkov eden izmed članov ekipe COVID-19 Sledilnik (aktiven pri modeliranju epidemije in predstavljanju Sledilnika znanstveni skupnosti). Je pobudnik in koordinator mednarodnega projekta YoungStatS pri združenju FENStatS. Zaposlen je kot znanstveni sodelavec na Inštitutu za ekonomska raziskovanja (IER) ter docent za ekonomijo in kvantitativne metode na Ekonomski fakulteti Univerze v Ljubljani (andrej.srakar@ier.si).

Uvod

Čeprav smo sredi maja 2020 v Sloveniji kot prvi v Evropi razglasili konec epidemije bolezni COVID-19, smo v drugi polovici leta 2020 doživeli drugi val, ki je bil veliko hujši od prvega. Tudi ponovna uvedba ukrepov in razglasitev epidemije sredi oktobra nista bistveno pomagali pri zajeitvi epidemije. V začetku leta 2021, ko je nastal ta prispevek, število primerov še vedno stagnira, do bistvenega upada torej ni prišlo.

Prvemu in drugemu valu je skupno, da so ju odgovorne javne institucije pričakale nepripravljene. Čeprav imamo organizacije, ki zbirajo podatke o zdravju prebivalcev Slovenije, te niso bile zmožne pripraviti javnih in zanesljivih podatkov o prisotnosti in širjenju te bolezni ter jih na jassen in transparenten način predstaviti odločevalcem in javnosti. Informacije iz različnih virov namreč pogosto niso bile konsistentne, zato so se nekateri uporabniki podatkov iz različnih okolij, od strokovnjakov za analizo podatkov do raziskovalcev z različnih področij znanosti, povezali v neformalni skupnosti COVID-19 Sledilnik (v nadaljevanju Sledilnik). Skupnost je vzpostavila neposredno povezavo z Nacionalnim inštitutom za javno zdravje (NIJZ) in zdravstvenimi zavodi. Pridobljene podatke validira¹ in objavlja v formatu, primerem za nadaljnjo analizo in modeliranje. V predstavitev rezultatov kot eno osnovnih načel svojega delovanja navaja podatkovno odličnost (Jakulin in Renko, 2020), kar je v statutu novoustanovljenega Znanstvenega društva Sledilnik (2020) zapisano tudi kot njegov namen in cilj.

Z izzivi glede dostopa do podatkov in ravnanja z njimi pa se ne soočamo samo pri nas, ampak je to globalni problem. S tem področjem se posebej aktivno ukvarja Združenje za raziskovalne podatke (Research Data Alliance – RDA), leta 2013 ustanovljena mednarodna organizacija, ki z različnimi aktivnostmi skrbi za zmanjšanje socialnih in tehničnih ovir pri izmenjavi raziskovalnih podatkov. Člani so raziskovalci, znanstveniki in drugi strokovnjaki za podatke iz različnih držav po svetu, ki so organizirani v interesne in delovne skupine. Marca 2020 se je v okviru združenja vzpostavila interdisciplinarna delovna skupina, ki je aprila objavila prvo različico priporočil in smernic za ravnanje s podatki o COVID-19 za različne skupine deležnikov, od raziskovalcev in podatkovnih analitikov do oblikovalcev politik, pa tudi državljanov (RDA, 2020). Slednji v teh procesih nastopajo bodisi le kot subjekti analize bodisi kot prostovoljci v okviru pobude občanske znanosti (ang. *citizen science*).

¹ Postopki validacije podatkov so predstavljeni v okviru pogosto zastavljenih vprašanj na spletni strani projekta (<https://covid-19.sledilnik.org/>), nekaj besed pa jim namenimo tudi v nadaljevanju prispevka.

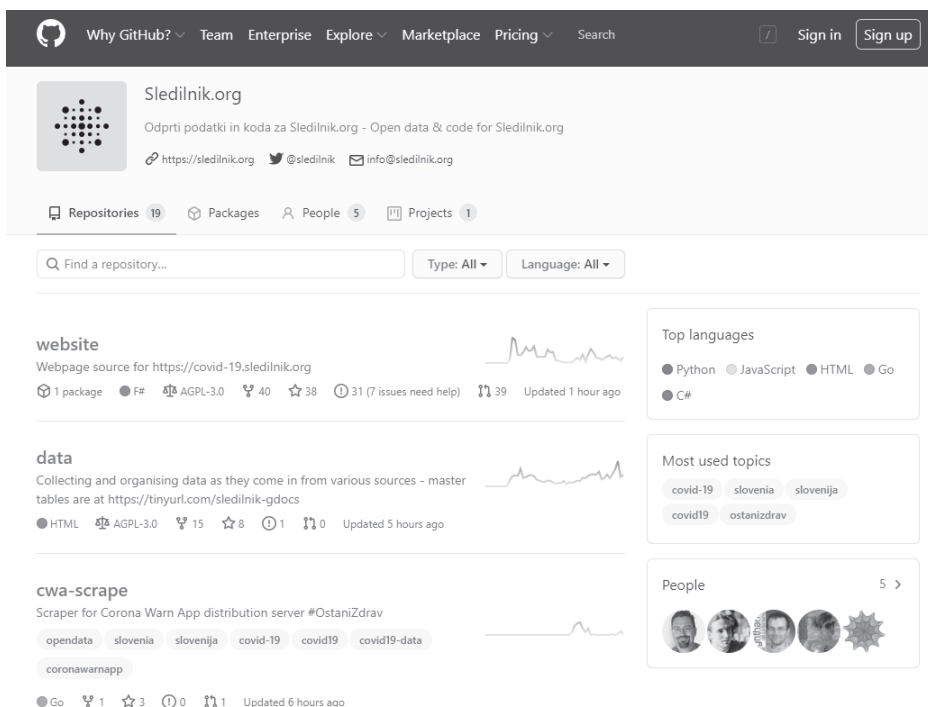
Skupnost Sledilnik je v okviru proučevanja odprte znanosti v Sloveniji zanimiva tako kot primer dobrih praks ravnanja s podatki kot tudi primer občanske znanosti. V članku predstavimo nastanek in delovanje skupnosti, nato pa jo ovrednotimo, najprej s pomočjo načel občanske znanosti, kot jih je oblikovalo Evropsko združenje za občansko znanost (European Citizen Science Association – ECSA), in nato v luči smernic Združenja za raziskovalne podatke. V tem delu se osredotočimo predvsem na koncept načel FAIR (Wilkinson idr., 2016), ki presega razumevanje odprtosti podatkov² kot preprosto odprtodostopne objave podatkovnih datotek na spletu, ampak skuša podati smernice, kako izboljšati najdljivost (ang. *Findability*), dostopnost (ang. *Accessibility*), interoperabilnost (ang. *Interoperability*) in ponovno uporabnost (ang. *Reusability*) podatkov prek ustreznega ravnanja z njimi. V kolikšni meri se delo Sledilnika, ki se ponaša s spodbujanjem razvoja podatkovne odličnosti, sklada s temi načeli? Kaj lahko Sledilnik stori, da podatke naredi bolj najdljive, dostopne, interoperabilne in predvsem ponovno uporabne?

Nastanek in delovanje skupnosti Sledilnik

Zametke skupnosti COVID-19 Sledilnik lahko najdemo v preglednici v orodju Google Sheets, ki jo je 11. marca, teden po prvem primeru potrjene okužbe v Sloveniji, ustvaril Luka Renko, v njej zbral do takrat znane podatke o epidemiji ter jo odprto delil na družabnih omrežjih Twitter in Facebook (Renko, 2020a). Že v prvih nekaj dneh mu je pri dopolnjevanju preglednice na pomoč priskočilo več ljudi (ibid.). Med njimi je bil Andraž Vrhovec, ki je naslednji dan, 12. marca, dodal skripto za avtomatski izvoz v format z vejico ločenih vrednosti (ang. *Comma-separated values* ali CSV) ter objavo v repozitoriju, ki gostuje v okviru storitve GitHub (Slika 1).

Pet dni pozneje, 17. marca, je bil postavljen delovni prostor v spletnem orodju Slack (Renko, 2020a), ki je po skoraj desetih mesecih delovanja še vedno glavna platforma za organizacijo dela in interno komunikacijo članov ekipe. Pri tem ne gre le za komunikacijo v sklopu delovnih procesov, ampak

2 Pri odprtih podatkih gre za idejo o prosti dostopnosti (nekaterih) podatkov vsem za ponovno uporabo in objavo, brez omejitev avtorskih pravic ter prepovedi kopiranja in objavljanja. Podatki so lahko odprti, a ne sledijo načelom FAIR (najdljivost, dostopnost, interoperabilnost in ponovna uporabnost). In obratno, načela FAIR ne zahtevajo popolne odprtosti podatkov, ampak zadostuje odprtost metapodatkov. Za neagregirane mikropodatke pogosto obstajajo utemeljeni razlogi, zakaj ne morejo biti dostopni vsakomur. Velja načelo »toliko odprti, kot je mogoče, toliko zaprti, kot je potrebno« (ang. *as open as possible, as closed as necessary*). Z vidika odprte znanosti je skladnost z načeli FAIR pomembnejša od same odprtosti.



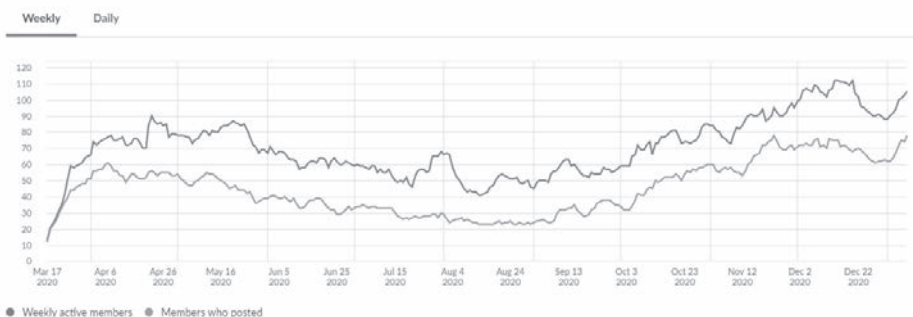
Slika 1: Podatki in koda Sledilnika v repozitoriju GitHub, <https://github.com/sledilnik>. Vir: Zaslonska slika, 10. januar 2021.

je Slack za člane tudi prostor druženja, kjer razpravljajo o različnih temah, povezanih z epidemijo COVID-19 v Sloveniji. Pogovori potekajo na 54 različnih kanalih, od katerih so po številu članov, ki so karkoli objavili, največji #team-intro (uvodne predstavitve), #general (za obvestila in komunikacijo na ravni celotne skupnosti), #random (namenjen razpravam, ki se ne tičejo neposredno dela), #modeling (modeliranje) in #dataviz (vizualizacije podatkov) (Slack, 2021a).

V začetku januarja, ko je bil ta članek zaključen, je bilo v orodju Slack registriranih kar 276 članov. Skoraj tretjina se jih je pridružila že marca (84), do konca aprila pa je imel že več kot polovico članstva celotnega leta (145). Kot prikazuje Slika 2, je bilo v delovnem okolju Slack v času od konca marca do konca maja (konec prvega vala epidemije) tedensko aktivnih 60 do 90 uporabnikov. Sporočila je objavljalo manj uporabnikov, od 40 do 60. V poletnih mesecih je prišlo do upada aktivnosti, z drugim valom pa je število uporabnikov naraslo in decembra doseglo vrh pri okrog 110 tedensko aktivnih uporabnikih ter od 60 do 70 takih, ki so objavili vsaj eno sporočilo (Slack, 2021b).

Active members

See how many people are active – meaning they posted a message or read at least one channel or direct message.

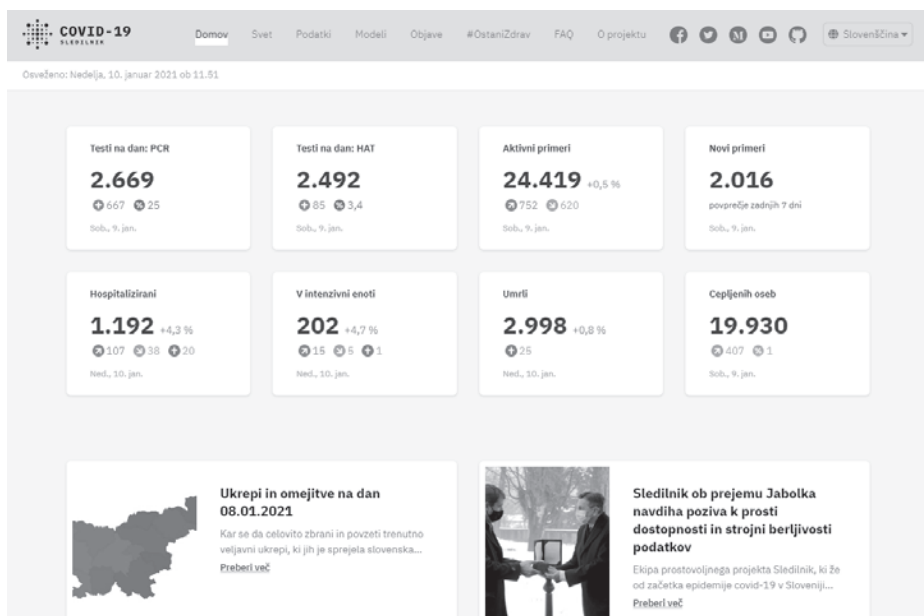


Slika 2: Tedenska aktivnost članov Sledilnika po mesecih. Vir: Zaslonska slika analitike v orodju Slack, 10. januar 2021.

Vendar vsi člani skupnosti niso enako aktivni. Podrobnejši podatki analitike delovnega okolja Slack v celoti gledano kažejo, da kar 73 članov (26 %) nikoli ni objavilo nobenega sporočila, 120 (43 %) jih je objavilo vsaj eno sporočilo, manj kot 50, tj. 43 (16 %), jih je objavilo vsaj 50, a manj kot 500 sporočil, 33 (12 %) jih je objavilo vsaj 500, a manj kot 5000 sporočil, najaktivnejših 7 (3 %) pa je v skoraj desetih mesecih delovanja skupnosti objavilo 5000 ali več sporočil (Slack, 2021b). Poleg tega objavljanje sporočil ne pomeni dejanske aktivnosti pri nalogah Sledilnika, kjer je aktivnih 55 članov, katerih delovanje predstavljamo v nadaljevanju.

Eksterna komunikacija skupnosti Sledilnik poteka predvsem prek spletne strani. Prva verzija spletne strani, kjer so predstavljeni rezultati dela Sledilnika, je bila pripravljena 18. marca in 25. marca objavljena na domeni sledilnik.org (Renko, 2020a). Vzporedno s spletno stranjo je nastala grafična podoba, ki jo je ustvaril Jure Sobočan. Aprila je spletna stran dobila svoj strežnik, maja pa še strežnik aplikacijskega programskega vmesnika (ang. *Application Programming Interface* ali API), ki podatke posreduje v obliki JSON (ang. *JavaScript Object Notation*).

Z dodajanjem vsebin je bila spletna stran večkrat spremenjena in prilagojena – njeno podobo v začetku januarja 2021 prikazuje Slika 3. Na prvi strani so prikazi stanja COVID-19 v Sloveniji, ki se osvežujejo dnevno. V drugih zavihkih vsebuje tudi primerjavo s stanjem v drugih državah, opis virov podatkov, opis modeliranja, seznam objav, analizo delovanja aplikacije #OstaniZdrav, pogosto zastavljena vprašanja ter opis projekta. Desno zgoraj so povezave do drugih storitev, kjer ekipa Sledilnika objavlja svoje vsebine (družabni omrežji Facebook in Twitter, založniška platforma Medium, kanal YouTube in stran GitHub).



Slika 3: Spletna stran <https://covid-19.sledilnik.org/>. Vir: Zaslonska slika, 10. januar 2021.

Decembra je Sledilnik doživel precej medijskega odmeva, saj je Luka Renko prejel naziv ime tedna na Valu202 in bil nominiran za Delovo osebnost leta, Sledilnik pa je prejel priznanje Statističnega društva Slovenije za odličnost statističnega poročanja v medijih ter jabolko navdih predsednika države. Del tega so bile številne objave in zapisi v različnih slovenskih tiskanih in spletnih medijih ter na radiu in televiziji. Spletno stran Sledilnik je decembra 2020 obiskalo skoraj milijon različnih uporabnikov. Prispevki na Mediumu so bili prebrani že več kot 50.000-krat. Sledilnikovo Facebook stran spremlja več kot 3.000 oseb, na Twitterju pa ima več kot 6.000 sledilcev.

Konec leta je Sledilnik svoje delovanje formaliziral z ustanovitvijo Znanstvenega društva Sledilnik, ki je bilo 28. decembra vpisano v register in katerega predsednik je Luka Renko. V petem členu društvenega statuta (Znanstveno društvo Sledilnik, 2020) je kot njegov namen in cilj zapisano spodbujanje razvoja »podatkovne odličnosti, kroženja in uporabne vrednosti podatkov ter podatkovne znanosti v dobro življenja«. Poudarjeno je tudi, da društvo »goji strokovnost, razsodnost in rahločutnost« ter »sledi znanstveni metodi razmišljanja in znanstvenemu pristopu analize družbenih in naravnih pojavov«.

V nadaljevanju predstavljamo področja delovanja Sledilnika, podatke, modeliranje in vsebino, ter analiziramo področja delovanja in demografske značilnosti 47 najaktivnejših članov Sledilnika.

Področja delovanja Sledilnika

Med tremi sklopi dela v skupnosti najprej omenimo delo s *podatki*, ki ga koordinirata Luka Renko in dr. Maja Založnik. Zajema delo na podatkovnem modelu, virih in njihovi validaciji ter na vizualizacijah. To je osrednja dejavnost, s katero se je Sledilnik tudi pričel. Viri podatkov so navedeni na podstrani Podatki (<https://covid-19.sledilnik.org/sl/data>), kako poteka delo z njimi, pa opisuje podstran s pogosto zastavljenimi vprašanji (<https://covid-19.sledilnik.org/sl/faq>), vendar gre za splošen opis, brez podrobnosti, ki bi omogočale sledljivost – te so dostopne le interno v okviru dokumentacije, povezane z dnevnim posodabljanjem podatkov in komunikacijo znotraj orodja Slack.

Za vnos in obdelavo podatkov Sledilnik uporablja javno dostopne podatke iz uradnih virov, kot so Nacionalni inštitut za javno zdravje (NIJZ), Vlada RS, Ministrstvo za zdravje itn., podatke iz administrativnih virov zdravstvenega sistema, kot so UKC Ljubljana, UKC Maribor, UK Golnik in drugi, virov Civilne zaščite ter iz nacionalnih in lokalnih medijev. Po odprtju šol ob koncu prvega vala je Sledilnik pridobil tudi podatke Ministrstva za izobraževanje, znanost in šport o okužbah in karantenah v šolah. Od poletja spremlja delovanje aplikacije Ministrstva za javno upravo #OstaniZdrav in objavlja število objavljenih dnevnih ključev, ki služi za oceno števila uporabnikov, ki so vnesli potrditev okužbe. Sledilnik dobiva podatke enot intenzivne nege v vseh 13 bolnišnicah, zaradi njihove obremenitve pa je prevzel tudi pripravo zbirnega poročila za deležnike (Renko, 2020b). Od jeseni Sledilnik na podlagi podatkov Ministrstva za notranje zadeve in Statističnega urada RS pripravlja prikaz presežnih smrti v letu 2020.

Rezultati Sledilnika, ki so prikazani na spletni strani, so vsebinsko omejeni samo na epidemiološke podatke o številu testiranih (po različnih metodah), številu novih in številu aktivnih primerov, številu hospitaliziranih, številu pacientov v enotah intenzivne terapije in številu umrlih. Ekipa Sledilnika ne nadzoruje točnosti izvirnih podatkov in ne objavlja podatkov, ki niso pridobljeni iz uradnih virov ali sredstev javnega obveščanja, zato pa vse podatke navzkrižno preverja, da so pravilni in skladni z izvirnimi. V primeru neskladnosti podatkov ne objavi, ampak skuša v sodelovanju z viri podatkov ugotoviti razlog za neujemanje.

Zbrani in preverjeni podatki so poleg preglednice v orodju Google Sheet na voljo v obliki CSV-datotek, kjer gre za nelastniški format, zato uporabnika ne omejujejo na lastniško programsko opremo, in aplikacijskega vmesnika REST (ang. *RESTful Web Services*), ki za razliko od standardnega protokola SOAP (ang. *Simple Object Access Protocol*) nima tako strogih pravil in omogoča

bolj prilagodljivo arhitekturo ter je zato bolj dostopen za uporabnike. Vse skozi spodbujamo nadaljnjo uporabo podatkov in grafov ter sodelovanje pri zbiranju, obdelavi in prikazu. Sledilnikovi podatki so bili uporabljeni v številnih znanstvenih prispevkih slovenskih avtorjev z različnih področij (denimo medicine, statistike, matematike, fizike, kemije, biologije in ekonomije). Podatke torej pridobivamo od različnih institucij, 25. decembra pa je Sledilnik prešel na avtomatizirano obdelavo dnevni in tedenskih poročil NIJZ, kar je korak k prehodu na povsem avtomatizirane procese zbiranja, urejanja in prikaza podatkov (Renko, 2020b).

Ekipa se pri delu sooča s številnimi izzivi, ki so največkrat posledica tega, da podatki niso v želenem formatu (Renko, 2020c). Idealno bi bilo imeti dezagregirane podatke, ki pa dolgo niso bili na voljo niti NIJZ kot osrednji instituciji na tem področju. Ta jih sedaj ima, vendar so zaprte narave, medtem ko so Sledilniku in drugim uporabnikom na voljo le agregirani podatki. Po eni strani je to razumljivo, saj gre pri zdravstvenih podatkih za občutljive osebne podatke, ki ne morejo biti odprto objavljeni, ampak morajo biti ustrezno anonimizirani, za njihovo obdelavo pa morajo biti uporabniki ustrezno usposobljeni ter zavezani k zaupnosti. Po drugi strani pa obstajajo tehnologije, izdelane z upoštevanjem koncepta vgrajene zasebnosti (ang. *privacy by design*), ki omogočajo varno analitiko tudi najbolj občutljivejših osebnih podatkov, katerih obdelavo varuje Splošna uredba o varstvu podatkov (GDPR). Poleg tega Sledilnik nima dostopa do vseh vrst podatkov, ki bi jih želel – tako denimo manjkajo demografski podatki tako o testiranih kot o hospitaliziranih in tistih v enotah intenzivne terapije (Renko, 2020c). Koristni bi bili tudi podatki o njihovih pridruženih boleznih.

Pomemben del Sledilnika je tudi *modeliranje*, ki ga je do novembra 2020 koordiniral dr. Aleks Jakulin, sedaj pa je to delo prof. dr. Janeza Žiberta. V delovnem okolju Slack ima svoj kanal #modeliranje, v katerem je vsaj eno sporočilo objavilo 71 oseb (Slack, 2021a). Pri Sledilniku sodeluje več skupin, ki se ukvarjajo z različnimi vrstami modeliranja (Renko, 2020a). Eden od modelov je nadgrajena verzija standardnega oddelčnega modela SEIR (dovzetni, izpostavljeni, okuženi, ozdravljeni; ang. *susceptible, exposed, infected, recovered*) iz epidemiološke statistike, ki jo pripravlja izr. prof. dr. Janez Žibert z Zdravstvene fakultete Univerze v Ljubljani. Inštitut za biostatistiko in medicinsko informatiko Medicinske fakultete Univerze v Ljubljani pod vodstvom prof. dr. Maje Pohar Perme ocenjuje stopnjo reprodukcije virusa na osnovi semimehanskega semiparametričnega Bayesovega modela. Dr. Žiga Zaplotnik s Fakultete za matematiko in fiziko (FMF) Univerze v Ljubljani pa modelira širjenje virusa po socialnem omrežju prebivalcev Slovenije s pomočjo modela, temelječega na agentih (pri čemer kombinira modelska

spoznanja s področij podnebnih sprememb in epidemiologije). V drugem valu je bila skupina manj aktivna, kar pripisujemo manjšemu zanimanju odločevalcev za modeliranje.

Tretji večji sklop je delo z *vsebinami*, ki ga koordinira Marko Brumen; vključuje tekstopisje, spletno stran, prevode, komunikacijo, grafično oblikovanje in družabna omrežja. Sledilnik intenzivno objavlja na Facebooku in Twitterju, daljše objave pa na platformi Medium. Dinamične vizualizacije podatkov so objavljene na Sledilnikovem kanalu YouTube. Zbrane podatke je skupnost poskušala postaviti v pravi kontekst, jih ustrezno interpretirati ter pretvoriti v znanje za informiranje o ukrepih na področju javnega zdravja. Vsebine so posredovane tudi medijem. Člani v sklopu te dejavnosti odgovarjajo na novinarska vprašanja in nastopajo v medijih.

Sledilnik ima močno razvojno komponento, pri kateri ne gre le za tehnične nadgradnje, temveč za aktivno dodajanje novih funkcionalnosti in podatkov. V zadnjem času je tako zanimivo vzorčenje odpadnih voda za hitrejšo informacijo o razširjenosti virusa, kjer poteka sodelovanje z raziskovalci na Nacionalnem inštitutu za biologijo, ter podatki mobilnih operaterjev, ki jih je pri modeliranju uporabljal dr. Žiga Zaplotnik s Fakultete za matematiko in fiziko Univerze v Ljubljani, za analizo poteka epidemije pa so jih uporabili tudi raziskovalci Inštituta za biostatistiko in medicinsko informatiko Medicinske fakultete Univerze v Ljubljani, med katerimi je nekaj članov Sledilnika. Novembra je ekipa Sledilnika sodelovala pri izvedbi ankete Mladih zdravnikov o spoštovanju zaščitnih ukrepov ter ankete študentov Medicinske fakultete Univerze v Mariboru o odnosu do cepljenja.

Vzporedno s temi aktivnostmi potekajo upravljanje s skupnostjo, koordiniranje različnih aktivnosti in sestanki. Še posebej veliko dodatnih aktivnosti je potekalo v zvezi z ustanovitvijo Znanstvenega društva Sledilnik, še več pa jih bo povezanih z njegovim delovanjem.

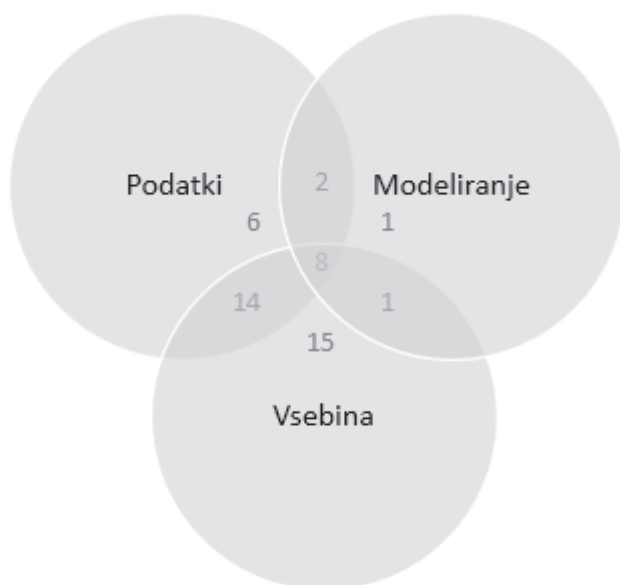
Analiza aktivnih sodelavcev Sledilnika

Na spletni strani Sledilnika je na podstrani O projektu (<https://covid-19.sledilnik.org/sl/about>) seznam njegovih sodelavcev ter aktivnosti, pri katerih sodelujejo. V začetku januarja, ko je ta članek nastal, je bilo na seznamu 55 oseb. Od tega trije nimajo računa v orodju Slack, ampak komunikacija z njimi poteka prek elektronske pošte (ena validatorka podatkov in dva prevajalca).

Vseh 55 sva z namenom pridobivanja podatkov za ta članek povabila k izpolnitvi kratkega obrazca, s katerim sva zbrala njihove podatke o aktivnosti, izobrazbi, zaposlitvenem statusu, starosti in spolu (Slavec in Srakar,

2021). Odzvalo se jih je le 47 (85 %). Skoraj polovica vprašanih (23) se je skupnosti Sledilnik pridružila že marca. V tipičnem tednu jih 17 Sledilniku nameni manj kot 5 ur, 9 jih nameni do 10 ur, 7 do 15 ur, naslednjih 7 do 20 ur, preostalih 7 pa 20 ur in več.

Pri vprašanju o aktivnostih pri Sledilniku je bilo mogoče izbrati več kategorij s seznama. 38 sodelavcev je izbralo eno ali več aktivnosti na področju vsebin (družabna omrežja, grafično oblikovanje, komunikacija, prevodi, spletna stran, tekstopisje in vsebine), 30 jih je izbralo vsaj eno izmed aktivnosti na področju dela s podatki (podatkovni model, validacija, viri, vizualizacije), 12 pa jih je izbralo modeliranje. Kot prikazuje Slika 4, je 8 sodelavcev aktivnih na vseh treh področjih, 14 je aktivnih pri podatkih in vsebini, dva pri podatkih in modeliranju ter eden pri modeliranju in vsebini.



Slika 4: Področja delovanja 47 sodelavcev Sledilnika.

V ekipi je več kot dvakrat več moških (33) kot žensk (14). Le 7 jih je mlajših od 30 let, 18 jih je v tridesetih, 16 v štiridesetih, 6 pa jih je starejših od 50 let, od tega je eden starejši od 65. Največ jih je zaposlenih v zasebnem sektorju (27), skoraj tretjina pa v javnem (15). Med tistimi, ki so odgovarjali, sta tudi dva upokojenca, dva brezposelna in en študent. Vsi imajo končano vsaj štiri-letno srednjo šolo, večina (39) pa jih ima tudi visokošolsko izobrazbo. Med njimi je 12 doktorjev in štirje magistri znanosti.

Ena izmed značilnosti Sledilnika je njegova interdisciplinarnost, kar se vidi tudi pri področjih izobrazbe njegovih sodelavcev. Kar 19 vprašanih se je

izobraževalo ali usposabljaljo na več kot enem področju (po klasifikaciji KLASIUS-P-16), nekateri tudi na treh ali štirih. Med področji prevladujejo informacijske in komunikacijske tehnologije (20), naravoslovje, matematika in statistika (18) ter družbene vede, novinarstvo in informacijska znanost (13). Nekaj se jih je izobraževalo tudi na področju umetnosti in humanistike (7), poslovnih in upravnih ved ali prava (6), izobraževalnih znanosti in izobraževanja učiteljev (5) ter po eden na področju tehnike, proizvodne tehnologije in gradbeništva ter zdravstva in socialne varnosti.

Med vsemi respondenti jih je le 9 evidentiranih v informacijskem sistemu o raziskovalni dejavnosti v Sloveniji SICRIS. Med njimi sta dva raziskovalca na področju inteligentnih sistemov oziroma programske opreme, dva na področju sociologije, po eden pa prihaja s področij verjetnostni račun in statistika, meteorologija in oceanografija, fizika srednjih in nizkih energij, javno zdravstvo in komunikologija.

V takšnem zlivanju in srečevanju področij prihaja do številnih razprav, pogosto zelo polemičnih, ki pa so verjetno dober prikaz trenutnega stanja v slovenski znanosti in povzamejo razprave, tudi sicer prisotne v znanstveni skupnosti (denimo o statistiki in delu s podatki, o delu raziskovalcev in znanstvenikov, o vlogi znanosti v družbi ter o splošno družbenih vprašanjih, politiki in ideologiji).

Sledilnik kot primer občanske znanosti

Občanska znanost je eden od prevodov³ angleškega termina *citizen science*, ki pomeni aktivno participacijo javnosti v znanstveno-raziskovalnih projektih.⁴ Slovar Cambridge jo definira kot znanstveno delo, na primer zbiranje podatkov, ki ga opravljajo običajni ljudje brez posebnih kvalifikacij, z namenom pomagati znanstvenikom. To je ozka opredelitev, saj sodelujoči v tovrstnih projektih niso le laiki, ampak gre pogosto za znanstvenike, kar se kaže tudi v primeru Sledilnika.⁵

3 V slovenščino se pogosto prevaja tudi dobesedno kot državljanska znanost, kar po mnenju terminološke sekcije Inštituta za slovenski jezik Frana Ramovša termina ne opredeli dovolj natančno, zato predlaga prevod ljubiteljska (ZRC SAZU, 2017). V praksi se uporabljata tudi prevoda skupnostna znanost (na primer <https://covid.si/>) in participativna znanost (ECSA, 2016). Prevod občanska smo izbrali, ker ga uporablja prispevek Zdravka Mlinarja v tej številki ČKZ (Mlinar, 2021).

4 Izraz je bil prvič uporabljen za opis sodelovanja prostovoljcev pri zbiranju vzorcev dežja za oceno njegove kislosti (Kerson, 1989). Šlo je za člane okoljskega društva Audubon, ki ima že od leta 1900 tradicijo opazovanja in štetja ptic, pri čemer gre prav tako za občansko znanost.

5 Poleg Sledilnika se je v času epidemije COVID-19 v Sloveniji pojavilo še nekaj skupin, ki so se ali poskusile vzpostaviti kot primer občanske znanosti, na primer projekt Covid.si, ali

Onkraj slovarske definicije obstaja več drugih, natančnejših definicij občanske znanosti, ki jih je težko poenotiti (Haklay idr., 2010). Težko je tudi oceniti obseg tovrstnih dejavnosti, saj ni enotne podatkovne baze, ki bi zbiral podatke o projektih. Kot približek si lahko ogledamo podatke o aktivnih in preteklih projektih občanske znanosti, ki jih beležita repozitorija SciStarter (<https://scistarter.org/>) in Zooniverse (<https://www.zooniverse.org/>). Največ jih je s področja narave, biologije, ekologije in okolja, medtem ko je na področju zdravja in medicine, kamor uvrščamo projekt Sledilnik, tovrstnih projektov precej manj. Od 1356 projektov občanske znanosti v repozitoriju SciStarter jih je le 107 s področja zdravja in medicine, med njimi pa jih je le 5 na temo epidemije COVID-19 (od tega gre v treh primerih za anketo, v enem primeru za označevanje tvitov in v enem za sodelovanje v igri). Med 315 projekti v repozitoriju Zooniverse jih je s področja medicine le 18, pri čemer pa nobeden ni povezan s temo pandemije.

Ena od zelo pomembnih dimenzij projektov občanske znanosti je krepitev demokratičnosti znanosti in zaupanja vanjo, zato je velikega pomena, da pri tem sledimo minimalnim standardom kakovosti, odprtosti in etičnosti, vključno z ustreznim ravnanjem s podatki (Heigl idr., 2019). Delovna skupina Evropskega združenja za občansko znanost ECSA je oblikovala deset načel občanske znanosti, ki so jih na Zavodu Kersnikova prevedli v slovenščino (European Citizen Science Association, 2016). V Tabeli 1 delovanje skupnosti Sledilnik vrednotimo skozi teh deset načel in ugotovimo, da že sledi vsem načelom, razen desetemu, saj še nima sporazumov o etični in pravni zaščiti podatkov in intelektualne lastnine. To je še toliko bolj pomembno pri projektu, ki se ukvarja z obdelavo zdravstvenih podatkov, kjer so zaradi občutljivosti vsebine standardi varnosti še toliko višji. Prvi korak v tej smeri je formalizacija delovanja, kar se je zgodilo konec decembra z ustanovitvijo društva. Možnosti za izboljšave so tudi pri drugih načelih, predvsem pri devetem prek vzpostavitve sistema nadzora kakovosti.

so poskusile organizirati podatke o epidemiji, na primer COVID-19 Alpaka (<https://covid19.alpaka.si/>) in Ustavimo korono (<https://ustavimokorono.si/>). Tudi pri njih je šlo za sodelovanje raziskovalcev in drugih državljanov prostovoljcev – kar je, kot bomo videli v nadaljevanju, ena ključnih značilnosti občanske znanosti –, vendar jim ni uspelo pridobiti toliko prostovoljcev, njihove aktivnosti pa v javnosti zaenkrat niso bile tako odmevne. Poleg tega ni bilo vidnega prispevka na področju ravnanja s podatki. Alpaka COVID-19 in Ustavimo korono uporabljata Sledilnikove podatke, ki jih le prikazujeta na malo drugačen način. Pri projektu Covid.si pa ne gre za epidemiološke podatke, ampak za podatke o molekulskem sidranju. Zato se v prispevku osredotočamo le na študijo primera skupnosti Sledilnik.

Tabela 1: Vrednotenje Sledilnika skozi deset načel občanske znanosti⁶

Načelo ECSA	Vrednotenje Sledilnika skozi načelo
1. Projekti občanske znanosti aktivno vključujejo pripadnike družbe v znanstvena prizadevanja, ki ustvarjajo nova znanja in razumevanja.	Člani skupnosti Sledilnik, ki niso znanstveniki, imajo možnost biti aktivno vključeni v proces znanstvenih raziskav kot sodelavci projekta.
2. Projekti občanske znanosti imajo pristne znanstvene rezultate.	Sledilnik odločevalce redno obvešča o rezultatih raziskav ter jim svetuje glede ukrepanja. Nekateri člani na podlagi rezultatov pripravljajo tudi znanstvene objave.
3. Profesionalni znanstveniki in občanski znanstveniki imajo koristi od sodelovanja pri projektih.	Poleg osebnega zadovoljstva ob prispevanju rezultatov, družbeno koristnem delovanju in vplivu na proces oblikovanja politik je sodelovanje pri Sledilniku za nekatere člane priložnost za izobraževanje. Člani, ki so zaposleni kot raziskovalci, imajo korist tudi zaradi objav rezultatov raziskav.
4. Občanski znanstveniki lahko sodelujejo pri več fazah znanstvenega procesa.	Člani skupnosti imajo možnost sodelovati v vseh fazah raziskave, od njenega načrtovanja in pridobivanja podatkov do analize in predstavitve rezultatov. Podatke o okužbah in karantenah po šolah smo na primer zbirali z množičnim izvajanjem (ang. <i>crowdsourcing</i>) (Renko 2020b).
5. Občanski znanstveniki prejmejo povratne informacije o projektu.	Kako se podatki uporabljajo in kakšni so rezultati projekta, je transparentno in dostopno vsem članom skupnosti.
6. Občanska znanost je kot katerikoli drugi raziskovalni pristop, ki ima svoje omejitve in pristranskosti, ki se jih moramo zavedati in jih nadzorovati.	Omejitve in pristranskosti uporabljenih raziskovalnih pristopov so ustrezno izpostavljene.
7. Podatki in metapodatki, ⁷ pridobljeni v projektih občanske znanosti, so javno dostopni in, kjer je mogoče, objavljeni v odprtokodnih formatih.	Večina podatkov in metapodatkov, zbranih v okviru projektov Sledilnika, je odprto dostopna. Podatki o nekaj najnovejših projektih bodo prav tako pripravljeni za objavo.
8. Občanski znanstveniki so navedeni v projektnih rezultatih in publikacijah.	V rezultatih in publikacijah so navedeni vsi, ki so k izdelku prispevali.

6 Vir (ECSA, 2016) uporablja izraza »participativna znanost« in »participativni znanstvenik«, ki smo ju tu prevedli v »občanska znanost« in »občanski znanstvenik«.

7 Metapodatki so podatki o podatkih. Podrobneje jih obravnavamo v naslednjem poglavju.

9. Programe občanske znanosti ocenjujemo glede na kakovost njihovih znanstvenih rezultatov, kakovost zbranih in obdelanih podatkov, izkušenj vključenih posameznikov ter širše družbene in politične učinke.	O kakovosti znanstvenih rezultatov je še prezgodaj govoriti, saj večina del še ni bila objavljena v recenziranih publikacijah. Prav tako je težko oceniti kakovost obdelanih podatkov, saj je odvisna od kakovosti izvornih podatkov. Izkušenj vključenih posameznikov projekt ne meri sistematično, lahko pa spremljamo družbene in politične učinke Sledilnikove aktivnosti na področju osveščanja prebivalstva, o čemer pričajo obiski spletnih strani, številne medijske objave ter sodelovanje z vladno strokovno skupino.
10. Vodje projektov občanske znanosti morajo upoštevati etične in pravne vidike zaščite podatkov, intelektualne lastnine, sporazumov o izmenjavi podatkov, zaupnosti, pripisovanja zaslug in vpliva na okolje pri izvajanju aktivnosti.	Podatki, ki so do sedaj javno objavljeni, ne vsebujejo nobenih osebnih podatkov. Obstaja nekaj podatkovnih datotek, ki so rezultat anket; te pa še niso ustrezno pripravljene, tako da bo treba poskrbeti tudi za te vidike.

Vrednotenje ravnanja s podatki v skladu s priporočili Združenja za raziskovalne podatke (RDA)

Skupina za COVID-19, ki deluje v okviru Združenja za raziskovalne podatke, je imela dva cilja: 1. jasno opredeliti podrobne smernice za izmenjavo podatkov v okviru okoliščin pandemije COVID-19, z namenom deležnikom pomagati pri sledenju najboljšim praksam za povečanje učinkovitosti njihovega dela; in 2. razviti smernice za oblikovalce politik, da čim bolj povečajo pravočasno izmenjavo podatkov in ustrezne odzive v izrednih zdravstvenih razmerah. Delovna skupina se je razdelila na osem podskupin, ki so se osredotočile na eno od osmih področij: klinični podatki, sodelovanje skupnosti, epidemiologija, podatki avtohtonih skupnosti, pravni in etični vidiki, genomika, družboslovje in programska oprema. Za Sledilnik je vsebinsko ustrezno predvsem področje epidemiologije, kjer je v smernicah izpostavljeno, naj podatkovni modeli vključujejo klinične podatke, mejnike bolezni, kazalnike in podatkovna poročila, sledenje kontaktom in dejavnike osebne tveganja (RDA, 2020). Vendar je pri tem omejen s podatki, ki so dostopni, zato ga težko vrednotimo skozi smernice, specifične za to področje. Kvečje-

mu bi lahko skozi ta priporočila ovrednotili NIJZ in druge vire podatkov, kar pa presega namene tega članka.

Za vrednotenje Sledilnika se zdijo bolj relevantna nekatera splošna priporočila⁸ delovne skupine, in sicer tretje (načela FAIR za ravnanje s podatki), peto (metapodatki, tj. podatki o podatkih), šesto (dokumentacija raziskovalnega konteksta in metodologij), sedmo (uporaba zaupanja vrednih repozitorijev za deljenje raziskovalnih podatkov) in osmo (objava podatkov takoj ob objavi rezultatov). Čeprav le tretje priporočilo neposredno omenja načela FAIR, so ta povezana tudi z ostalimi štirimi. Metapodatki in dokumentacija so namreč ključnega pomena za zagotavljanje najdljivosti in ponovno uporabnost podatkov, repozitoriji in objava pa za njihovo najdljivost in dostopnost. Zato podatke Sledilnika spodaj vrednotimo skozi načela najdljivosti, dostopnosti, interoperabilnosti in ponovne uporabnosti. Posamezne smernice so predstavljene v prilogi, v besedilu članka pa so navedene njihove oznake (na primer F1).

Najdljivost podatkov Sledilnika

Podatki, ki jih Sledilnik pridobi iz različnih virov, so na spletni strani Sledilnika na voljo v obliki CSV-datoteke in aplikacijskega vmesnika REST. Hranimo jih na GitHubu, spletnem gostiteljskem servisu za repozitorije v sistemu za upravljanje z izvorno kodo Git. Vendar bo raziskovalec, ki ne pozna ne spletne strani covid-19.sledilnik.org ne GitHuba, podatke Sledilnika težko našel, saj niso indeksirani v nobenem od repozitorijev, ki jih najdemo v registru repozitorijev re3data.org. Za epidemiološke podatke o COVID-19 je na voljo kar nekaj repozitorijev, vendar bi bilo treba poiskati takega, ki bi omogočil tudi povezavo med podatki ter kodo za prikaz in analizo, kar je glavna dodana vrednost podatkov Sledilnika v primerjavi z izvirnimi podatki, ki jih ustvarjajo NIJZ in druge javne ustanove.

Pri izbiri repozitorija bi bilo treba izbrati tudi ustrezno metapodatkovno shemo za predstavitev podatkov. Podatki na Sledilnikovi spletni strani in Githubu sedaj niso opisani z bogatimi metapodatki (F2) in niso indeksirani v iskalnem viru. Podatkovne datoteke imajo imena, vendar niso enolična. S ključnimi besedami so opremljene le nekatere. Sledilnik za podatke in metapodatke ne uporablja nobenih globalno edinstvenih trajnih identifikatorjev (F1).

8 Prvo (usklajena prizadevanja za spodbujanje odprte znanosti s pomočjo oblikovanja politik in vlaganje v poenostavitev pretoka podatkov med lokalnimi subjekti in mednarodnimi sistemi), drugo (vlaganje v infrastrukturo, ki podpira ravnanje s podatki) in četrto (zgodnje načrtovanje, vključno z opredelitvijo odgovornosti in zadolžitev v obliki načrtov ravnanja s podatki) priporočilo delovne skupine RDA zadevajo predvsem odločevalce.

Dostopnost podatkov Sledilnika

Podatki, ki jih Sledilnik objavlja na spletni strani in Githubu, so prosto dostopni vsem obiskovalcem. Do njih je mogoče dostopati z uporabo standardiziranega komunikacijskega protokola (A1). Protokol je odprt, brezplačen in splošno izvedljiv (A1.1), zato ni potrebe po omogočanju postopka preverjanja pristnosti in avtorizacije (A1.2). Ni pa znano, ali bodo metapodatki dostopni tudi po tem, ko podatki ne bodo več na voljo (A2).

Poteka tudi postopek predaje raziskovalnih podatkov anketnih raziskav, ki so jih partnerske organizacije (Mladi zdravniki Slovenije, Medicinska fakulteta Univerze v Mariboru in Valicon) izvedle v sodelovanju s Sledilnikom, v trajno hrambo Arhivu družboslovnih podatkov, kar bo še en korak v smeri večje dostopnosti.

Interoperabilnost podatkov Sledilnika

Na začetku delovanja Sledilnika je bila jedro podatkovnega modela Google Sheet preglednica, ki so jo kasneje zamenjale CSV-datoteke. Ta preglednica se od sredine decembra, ko je ekipi uspelo avtomatizirati večino procesov avtomatizacije podatkov, ne osvežuje več in je bila zamenjana s posameznimi CSV-datotekami. V uporabi je sicer še nekaj Google Sheet preglednic, ki so osnova za CSV-datoteke, nekatere pa so strojno obdelane neposredno v CSV-formatu. Kot že omenjeno, so vsi podatki v strojno berljivi obliki in skupaj z izvorno kodo za njihovo obdelavo na voljo na GitHubu.

Pri interoperabilnosti pa ne gre le za povezljivost formatov, ampak tudi vsebine na ravni konceptov in izvedbe merjenja, ki bi omogočili harmonizacijo in primerjave s podatki drugih držav. Za reprezentacijo znanja je v podatkih uporabljen formalen, dostopen in široko uporabljen jezik (I1), vendar za metapodatke niso uporabljeni slovarji, ki sledijo načelom FAIR (I2). Prav tako manjkajo kvalificirane reference na druge metapodatke in podatke (I3).

Ponovna uporabnost podatkov Sledilnika

Podatki na Githubu imajo dokumentacijo v obliki readme.me datotek, vendar ni zelo izčrpna, zato ne moremo reči, da so podatki bogato opisani z natančnimi in ustreznimi lastnostmi (R1). Metapodatki niso povezani s podrobnim poreklom (R3) in ne ustrezajo skupnostnim standardom relevantnega področja (R4). Na področju medicine sta uveljavljena dva metapodatkovna standarda, UMLS (<https://www.nlm.nih.gov/research/umls/index.html>) in LOINC (<https://loinc.org/>), ki bi ju lahko uporabili. Razvoj

standardov na tem področju poteka tudi v okviru delovne skupine Združenja za raziskovalne podatke, imenovane Epidemiološki skupni standard za nadzor nad poročanjem o podatkih (<https://www.rd-alliance.org/group/epidemiology-common-standard-surveillance-data-reporting-wg/case-statement/rdacodata>).

Vsebine na spletni strani covid-19.sledilnik.org so zaščitene s simbolom avtorskih pravic »vse pravice zadržane«, vendar je na podstrani »O projektu« (<https://covid-19.sledilnik.org/sl/about>) pod pogoji uporabe zapisano, da so podatki v javni domeni in se jih lahko prosto uporablja, ureja in predeluje v netržne vsebine ob navedbi vira. Velja torej CC-BY-SA 3.0 licenca Creative Commons, ki dovoljuje deljenje in predelovanje vsebin pod pogojem, da navedemo avtorja, označimo nastale spremembe ter vsebino delimo pod enakimi pogoji. Vendar bi lahko bila ta na strani jasneje in vidneje predstavljena (R2).

Kljub temu imajo podatki Sledilnika kar nekaj uporabnikov, od drugih spletnih strani za analizo in vizualizacijo podatkov⁹ do medijev in vladnih služb. Predvsem slednje licenco pogosto kršijo, saj pri uporabi rezultatov Sledilnika tega ne navajajo kot vira in vsebine ne razširjajo pod isto licenco. Razmisliti bi veljalo, ali bi lahko vsaj za nekatere rezultate Sledilnika uporabili bolj odprto in manj omejujočo licenco CC0.

Poleg ponovne uporabnosti podatkov velja izpostaviti tudi ponovno uporabnost programske kode, saj je bila uporabljena pri vzpostavitvi podobnega projekta v Severni Makedoniji, ki se imenuje COVID-19 Treker (<https://covid-19.treker.mk/>) in ga koordinira eden izmed članov skupnosti Sledilnik ter ima še od 5 do 10 drugih aktivnih članov.

Razprava in sklep

Sledilnik je skupnost prostovoljcev, ki je s pravočasnim in natančnim zbiranjem, poročanjem in izmenjavo podatkov z raziskovalno in zdravstveno skupnostjo ter oblikovalci politik omogočila spremljanje razvoja pandemije v Sloveniji. Gre za interdisciplinarno skupnost, kot je v Sloveniji še nismo videli ne po obsegu ne po raznolikosti profilov sodelavcev. Združuje tako aktivne raziskovalce, znanstvenike z različnih področij znanosti, od naravoslovja do družboslovja, kot tudi strokovnjake na področjih programi-

⁹ Poleg prej omenjenih slovenskih spletnih strani Alpaka in Ustavimo korono podatke Sledilnika prikazujejo tudi nekatere druge strani, na primer Worldometer (<https://www.worldometers.info/coronavirus/>) in Our World in Data (<https://ourworldindata.org/coronavirus-testing>).

ranja, modeliranja, tekstopisja in odnosov z javnostmi. Zanimivo je, da je najmanj sodelujočih s področja medicine in javnega zdravja.

V okviru procesov odpiranja znanosti v Sloveniji je Sledilnik zanimiv kot primer občanske znanosti in transparentnih praks ravnanja s podatki. Skupnost Sledilnik ustreza skoraj vsem od desetih načel Evropskega združenja za občansko znanost. Prostor za izboljšave je v okviru vzpostavitve sistema nadzora kakovosti rezultatov ter v ureditvi etičnih in pravnih vidikov ravnanja s podatki. Formalizacija delovanja z ustanovitvijo Znanstvenega društva Sledilnik je lahko eden od temeljev nadaljnjih pogovorov o tovrstnih vidikih ravnanja s podatki.

Čeprav se skupnost Sledilnik rada ponaša s podatkovno odličnostjo, je naše vrednotenje skladnosti s priporočili in smernicami Združenja za raziskovalne podatke pokazalo, da na tem področju še ni bilo narejeno vse, kar bi lahko bilo. Nedvomno je zelo dobro poskrbljeno za dostopnost, a do vseh zbranih podatkov zaenkrat še ni mogoče dostopati. Ena glavnih dodanih vrednosti Sledilnikovega dela je bistveno izboljšana interoperabilnost podatkov v primerjavi s podatki, pridobljenimi iz uradnih in administrativnih virov, vendar bi bilo treba za korak dlje aplicirati še ustrezne metapodatkovne slovarje ter kvalificirane reference na druge podatke. Urejene podatke, ki jih objavlja skupnost Sledilnik, uporabljajo tudi druge spletne strani doma in po svetu. Za izboljšanje njihove ponovne uporabnosti je treba metapodatke obogatiti v skladu z ustreznimi metapodatkovnimi standardi. S tem bi izboljšali tudi najdljivost podatkov, ki bi jih bilo smiselno indeksirati tudi v primernem področnem repozitoriju. Pri tem bi lahko uporabili tudi obstoječo javno infrastrukturo podatkovnih storitev, ki je na voljo v Sloveniji.

Vseh nalog, ki jih je treba izvesti za boljšo skladnost z načeli FAIR, Sledilnik kot prostovoljna skupnost sam ne more opraviti, ampak bi to morala biti naloga pristojnih institucij. Ministrstvo za zdravje RS se je v okviru pobude OECD za reforme v zdravstvu med drugim zavezalo, da bo izboljšalo informatizacijo zdravstva in reorganiziralo podatkovne tokove tako v smeri večje varnosti kot povezljivosti in interoperabilnosti podatkov ter njihove dostopnosti za podatkovne analize v javnem interesu (OECD, 2017). Z implementacijo tehnologij, temelječih na konceptu vgrajene zasebnosti, bi lahko podatke na varnejši način odprto delili in zainteresiranim javnostim omogočili boljše razumevanje poteka in širjenja bolezni, kar bi bila osnova za učinkovitejšo strategijo zajeziitve epidemije.

Poleg obstoječih zavez bi lahko Slovenija v ravnanju s podatki o epidemiji šla še korak dlje in se pridružila državam, kot so Kitajska, Nizozemska, Švedska in Velika Britanija, ki so se na tem področju povezale v omrežje VODAN (Virus Outbreak Data Network). Gre za aktivnost organizacij CODATA, RDA,

WDS in GO FAIR (<https://www.go-fair.org/implementation-networks/overview/vodan/>), ki pa se zaenkrat osredotoča na klinične podatke. Ni znano, ali kaj podobnega obstaja tudi za epidemiološke podatke, vsekakor pa je to aktivnost, ki se ji lahko pridružijo institucije v zdravstvenem sistemu, ne pa prostovoljna skupnost, kot je Sledilnik.

Kljub delovanju Sledilnika, ki odločevalcem v Sloveniji zagotavlja konsistentno in transparentno predstavitev podatkov, tem v drugem valu še ni uspelo zajezi epidemije. Poraja se vprašanje, ali je to posledica tega, da so pri odločanju o ukrepih premalo upoštevali podatke in analize, ki jih je omogočila skupnost Sledilnik, oziroma da so ukrepe za zajezi epidemije velikokrat sprejemali na podlagi premalo resnih analiz. Primer Sledilnika spodbuja razmišljanje o bodočem delu s podatki in o njihovi uporabi, tako na ravni odločevalcev (države in lokalnih skupnosti) in organizacij kot posameznikov. Tudi v bodoče potrebujemo tovrstne pobude, kjer v odprtem dialogu različnih področij poteka izmenjava zamisli, izračunov in predlogov, kar se prenaša v vsakdanje življenje.

Omejitev pričujočega prispevka je, da v okviru Sledilnikovega podatkovnega modela nismo podrobneje obravnavali vidika kakovosti vhodnih podatkov, ki jih Sledilnik uporablja. Tudi ravnanje s podatki v institucijah, ki jih zbirajo, bi lahko bilo zanimiv predmet analize, na podlagi katere bi ugotovili, kje so potrebne izboljšave. Prav tako se tu osredotočamo le na podatke, ki so del Sledilnikovega podatkovnega modela, medtem ko so za spremljanje epidemije pomembni tudi drugi viri (rezultati Nacionalne raziskave o razširjenosti COVID-19 in različnih javnomnenjskih anket). Nekaj teh virov je bilo uporabljenih tudi v okviru modeliranja epidemije, vendar jih v članku ne predstavimo. Zanimarili smo tudi vidik mednarodne primerljivosti podatkov. Postavlja se namreč vprašanje, ali bomo lahko kot globalna skupnost zaradi tako različnih vstopnih podatkov sploh zares ocenili razsežnosti epidemije. Na podlagi primerjalnega ravnanja s podatki v različnih državah bi verjetno lahko prišli do pomembnih zaključkov.

Poleg podatkovnih vidikov bi veljalo v prihodnje več pozornosti nameniti podrobnejšemu raziskovanju interakcije druženja v interdisciplinarnih skupnostih, ki nastajajo v okviru projektov občanske znanosti, ter ugotoviti, katere značilnosti skupnosti najbolj pripomorejo k njihovemu uspehu. Zlasti bi bilo zanimivo primere skupnosti, kot je Sledilnik, podrobneje preučiti z vidika dinamike njihovega delovanja, saj je bila ta v času obeh valov pandemije drugačna kot v vmesnem obdobju. Preučiti bi veljalo tudi stališča članov skupnosti do vprašanj, povezanih z lovljenjem ravnotežja med odprtostjo in zaščito podatkov. Tudi mnoge druge značilnosti skupnosti Sledilnik, kot so umeščenost v siceršnji družbeni kontekst, vezanost

na obdobje, ki pomeni eno najhujših kriz človeštva po drugi svetovni vojni (in s tem velikokrat zelo čustveno pogojena odzivanja ter potrebo po hitrem delovanju), prepleti človeških in strokovnih vidikov, ter zelo redki podobni poskusi povezovanja strok pri nas bodo raziskovalni izziv tudi v prihodnje.

Literatura

- ECSA (2016): *Deset načel participativne znanosti*. Dostopno na: https://ecsa.citizen-science.net/wp-content/uploads/2020/02/ecsa_ten_principles_of_cs_slovenian.pdf (10. januar 2021).
- Haklay, Mordechai, Daniel Dörler, Florian Heigl, Marina Manzoni, Susane Hecker in Katrin Vohland (2021): What Is Citizen Science? The Challenges of Definition. V *The Science of Citizen Science*, K. Vohland idr. (ur.), 13–33. Springer, Cham. Dostopno na DOI: 10.1007/978-3-030-58278-4_2.
- Heigl, Florian, Barbara Kieslinger, Katharina T. Paul, Julia Hulik in Daniel Dörler (2019): Opinion: Toward an International Definition of Citizen Science. *Proceedings of the National Academy of Sciences of the United States of America* 116(17): 8089–8092. Dostopno na DOI: 10.1073/pnas.1903393116.
- Jakulin, Aleks in Luka Renko (2020): Data Excellence and Experience of the Covid-19 Tracker Project. Spletna konferenca *Open Data – Digital Economy Accelerator*. Dostopno na: <https://www.youtube.com/watch?v=T6tS29XuEk> (13. januar 2021).
- Kerson, Raymond (1989): Lab for the Environment. *MIT Technol Rev* 92(1): 11–12.
- Mlinar, Zdravko (2021): Kaj nam prinašata koncept in gibanje občanska znanost/*Citizen Science*? Uveljavljanje raziskovanja kot sestavine vsakdanjega življenja. *Časopis za kritiko znanosti* 282: 23–63.
- OECD (2017): *The Next Generation of Health Reforms: Ministerial Statement. Health Ministerial Meeting*. Dostopno na: <https://www.oecd.org/health/ministerial/ministerial-statement-2017.pdf> (30. december 2020).
- RDA (2020): *RDA COVID-19: Recommendations and Guidelines*. Peta edicija, končni osnutek, 28. maj. Dostopno na: <https://www.rd-alliance.org/group/rda-covid19-rda-covid19-omics-rda-covid19-epidemiology-rda-covid19-clinical-rda-covid19-0> (9. november 2020).
- Renko, Luka (2020a): V slovo eni GSheet tabeli. Kako je nastal Sledilnik, v 14 čivkih. *COVID-19 Sledilnik Medium*, 26. december. Dostopno na: <https://medium.com/sledilnik/v-slovo-eni-gsheet-tabeli-323d98b2203c> (28. december 2020).
- Renko, Luka (2020b): Pretakanja podatkovja nekega Sledilnika ... V slovo eni GSheet tabeli, 2. del. *COVID-19 Sledilnik Medium*, 26. december. Dostopno na: <https://medium.com/sledilnik/pretakanje-podatkovja-nekega-sledilnika-c74df9e0801e> (28. december 2020).

- Renko, Luka (2020c): Kaj hočemo jutri? V slovo eni GSheet tabeli, 3. del. *COVID-19 Sledilnik Medium*, 27. december. Dostopno na: <https://medium.com/sledilnik/kaj-ho%C4%8Demo-jutri-23a3fef5d5cbe> (28. december 2020).
- Slack (2021a): *sledilnik.org Channel Analytics All time – Jan 10, 2021*. CSV datoteka.
- Slack (2021b): *sledilnik.org Member Analytics All time – Jan 10, 2021*. CSV datoteka.
- Slavec, Ana in Andrej Srakar (2021): *Anketa 47 aktivnih sodelavcev Sledilnika*. EnKlikAnketa.
- Znanstveno društvo Sledilnik, Statut. Dostopno na: https://drive.google.com/file/d/1h-SerCtHW7eL25N_wMYVrwzLFikvwxVSR/view (15. januar 2021).
- ZRC SAZU (2017): Ljubiteljska znanost. *Terminološka svetovalnica*. Dostopno na: <https://isjfr.zrc-sazu.si/sl/terminologisce/svetovanje/ljubiteljska-znanost#v> (9. november 2020).
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg idr. (2016): The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci Data* 3: 160018. Dostopno na DOI: 10.1038/sdata.2016.18.

Priloga: Načela FAIR

Najdljivost

- F1. (Meta)podatkom se dodeli globalno edinstven trajni identifikator.
- F2. Podatki so opisani z bogatimi metapodatki.
- F3. (Meta)podatki jasno in izrecno vključujejo identifikatorje podatkov, ki jih opisujejo.
- F4. (Meta)podatki so registrirani ali indeksirani v iskalnem viru.

Dostopnost

- A1. Na podlagi njihovega identifikatorja je do (meta)podatkov mogoče dostopati z uporabo standardiziranega komunikacijskega protokola.
 - A1.1. Protokol je odprt, brezplačen in splošno izvedljiv.
 - A1.2. Protokol po potrebi omogoča postopek preverjanja prisotnosti in avtorizacije.
- A2. (Meta)podatki so dostopni tudi, ko podatki niso (več) na voljo.

Interoperabilnost

- I1. (Meta)podatki za reprezentacijo znanja uporabljajo formalen, dostopen in široko uporabljen jezik.
- I2. (Meta)podatki uporabljajo slovarje, ki sledijo načelom FAIR.
- I3. (Meta)podatki vključujejo kvalificirane reference na druge (meta)podatke.

Ponovna uporabnost

- R1. (Meta)podatki so bogato opisani z množico natančnih in ustreznih lastnosti.
- R2. (Meta)podatki se izdajo z jasno in dostopno licenco za uporabo podatkov.
- R3. (Meta)podatki so povezani s podrobnim poreklom.
- R4. (Meta)podatki ustrezajo skupnostnim standardom relevantnega področja.