

To je recenzirani rokopis članka, objavljenega v Časopisu za kritiko znanosti, 2021, letnik 49, št. 282, str. 170–192.



Maša Kerstein in Marko Lovec

Odpiranje javnih podatkov v podporo odločanju v slovenskem kmetijstvu¹

Abstract

Opening Up Public Data in Support of Decision-Making in Slovenian Agriculture

The article is focused on the (non) use of data in public debate and decision-making processes, using the (common) agricultural policy in Slovenia as an example. Studies conducted within the scope of the Research Data Alliance show that descriptions of data sources and structures for the discovery, access, interoperability and re-use of data play an important role from a technical aspect, as well as in terms of content. The actual demand for data is also an important factor to consider. Studying public administrative data in the field of agriculture in Slovenia reveals a limited “supply” in terms of documentation, inclusion in catalogues, and equipping data to ensure connectivity and access solutions. At the same time, it appears that changes in “demand” at the EU level such as the strengthening of societal policy objectives, the increased flexibility of measures, and the justification of decisions with “facts”, could be leveraged to strengthen the use of data in decision-making processes.

Keywords: open data, administrative sources, FAIR, semantics, agriculture, Common Agricultural Policy

Maša Kerstein is a PhD student at the Faculty of Social Sciences, University of Ljubljana. She is employed at the Ministry of Agriculture, Forestry and Food. The topic of her doctoral dissertation is the introduction of fact-based decision making in Slovenia within the framework of the EU’s Common Agricultural Policy (masa.kerstein@gmail.com).

Marko Lovec is an Associate Professor at the Faculty of Social Sciences, University of Ljubljana. He conducts research on the EU’s Common Agricultural Policy. (marko.lovec@fdv.uni-lj.si).

Povzetek

Članek izhaja iz (ne)uporabe podatkov v javni razpravi in procesih odločanja na primeru (Skupne) kmetijske politike v Sloveniji. Iz študij, opravljenih v okviru mednarodnega Združenja za raziskovalne podatke, sledi, da pomembno vlogo igrajo opisi podatkovnih virov in struktur za odkrivanje, dostop, povezljivost in ponovno uporabo podatkov, tako v vsebinskem kot tehničnem (informacijskem) smislu, pa tudi dejansko povpraševanje po podatkih. Raziskava javnih administrativnih podatkov na področju kmetijstva v Sloveniji kaže na omejitve »ponudbe« v smislu dokumentacije, vključenosti v kataloge, opremljanja podatkov za zagotavljanje povezljivosti ter rešitev za dostop. Hkrati kaže, da bi bilo spremembe »povpraševanja« na ravni EU, kot so krepitev različnih družbenih ciljev politike, večja

¹ Raziskavo, na kateri temelji članek, sta financirala MKGP in ARRS v okviru Ciljnega raziskovalnega projekta Zasnova informacijskih rešitev v podporo izvajanju SKP EU na osnovi podatkov (CRP ITzaSKP).

prilagodljivost ukrepov in utemeljevanje odločitev z »dejstvi«, mogoče izkoristiti za krepitev uporabe podatkov v procesih odločanja, pa tudi kakovostne, širše javne razprave.

Ključne besede: odprti podatki, administrativni viri, FAIR, semantika, kmetijstvo, Skupna kmetijska politika

Maša Kerstein je doktorska študentka na Fakulteti za družbene vede Univerze v Ljubljani. Zaposlena je na Ministrstvu za kmetijstvo, gozdarstvo in prehrano. Tema njene doktorske naloge je uvajanje odločanja na podlagi dejstev v Sloveniji v okviru Skupne kmetijske politike EU (masa.kerstein@gmail.com).

Marko Lovec je izredni profesor na Fakulteti za družbene vede Univerze v Ljubljani. Raziskovalno se ukvarja s Skupno kmetijsko politiko EU (marko.lovac@fdv.uni-lj.si).

Uvod: javna razprava brez javnih podatkov

Javna razprava, utemeljena na dejstvih, je ključ transparentne, s tem pa odgovorne in uspešne javne politike. Podatki omogočajo postavljanje smiselnih ciljev, določanje učinkovitih ukrepov ter učenje s pomočjo preverjanja doseženega in primerjav. To velja tako za javne politike kot za državno in javno upravo, raziskovalno-razvojne institucije, zasebne ponudnike blaga in storitev ter slehernega posameznika. Digitalizacija procesov širi možnosti izmenjave in uporabe velikega števila podatkov, s tem pa priložnosti za izboljševanje kakovosti odločanja in povečevanje blaginje. Vendar pa je slednja odvisna od ustreznih javnih politik. Podatki so namreč javna dobrina, saj so individualni stroški za ustrezno razpoložljivost običajno (pre)visoki. Tudi ko so podatki t. i. klubska dobrina, ki je na voljo tistim, ki plačajo licenčnino, razmerje med stroški in koristmi pogosto ne odtehta vse potencialne javne vrednosti podatkov. Podatki imajo kot dobrina namreč to značilnost, da se z razširjanjem uporabe njihova vrednost ne zmanjšuje, temveč povečuje. Javne politike in ukrepi na področju podatkov vključujejo izobraževanje in opolnomočenje akterjev in deležnikov po celotnem ciklu od ustvarjanja do uporabe, pravila in prakse glede hrambe, dostopa, varovanja zasebnosti in licenc, skupno digitalno infrastrukturo, protokole, standarde, knjižnice, usmerjeno podatkovno raziskovalno-razvojno dejavnost ter programe podatkovnega državljanstva.

Namen tega članka je analizirati uporabo podatkov v procesih odločanja in izvajanja kmetijske politike v Sloveniji. Kmetijska politika v smislu širše javne razprave velja za eno bolj zaprtih politik z bolj notranjim krogom informiranih deležnikov, od predstavniških do raziskovalnih institucij na tem področju. Obenem je kmetijska politika deležna vedno večje pozornosti javnosti; kmetje opozarjajo na vse nižje odkupne cene hrane, potrošniki pa postajajo občutljivi na izvor hrane, načine pridelave in predelave ter njihove posledice za zdravje ljudi, okolje (vključno z naravo) in dobrobit živali.

Pri odločanju in izvajanju kmetijske politike v Sloveniji podatki ne igrajo skoraj nobene vloge. Namesto logiki dejstev kmetijska politika pogosto sledi logiki vplivnih lobističnih skupin; te z

argumenti, ki ne temeljijo na podatkih ali so celo v nasprotju z njimi, vplivajo na javno mnenje. Kmetijska politika je v veliki meri sestavljena iz ukrepov Skupne kmetijske politike (SKP), prenesenih v nacionalno zakonodajo.² Kot je pokazala raziskava strateškega načrtovanja ukrepov v Sloveniji (Erjavec idr., 2018; Šumrada idr., 2020; Lovec idr., 2020a), je relevantnost ukrepov nizka, zasledujejo preveč ciljev brez upoštevanja intervencijske logike na podlagi podatkov in študij, njihovi učinki pa so temu primerno razpršeni. Čeprav ima politika določene pozitivne učinke, na primer na dohodkovni položaj prejemnikov, so ti doseženi na neučinkovit način. Ob tem ostajajo odprta vprašanja o različnih razvojnih, okoljskih in socialnih vplivih politike, od vplivov na konkurenčnost, okolje in naravo do dobrobiti živali, razvitosti podeželskih območij ter demografskih vprašanj. Za odgovore na ta vprašanja pogosto niso na voljo sistematično urejeni podatki in študije.

Ta članek problem uporabe podatkov pri odločanju in izvajanju kmetijske politike naslavlja z vidika odkrivanja, dostopa, povezljivosti in (ponovne) uporabe podatkov v slovenskem kmetijstvu (t. i. standardi FAIR, glej Wilkinson idr., 2016). Osredinja se predvsem na javne administrativne podatke, torej tiste, ki jih zbirajo na podlagi izvajanja zakonskih nalog (v okviru SKP) in so tako premalo izkoriščena javna dobrina. Ugotavlja, v kolikšni meri podatki izpolnjujejo zahteve FAIR ter s katerimi javnimi ukrepi bi bilo mogoče spodbuditi razvoj standardov za smiselno uporabo podatkov v procesih odločanja in izvajanja politike, in sicer tako, da bi krepili odprto javno razpravo in zagotavljali koristi vsem deležnikom politike.

Članek je sestavljen iz dveh delov. V konceptualnem delu pregledujemo aktivnosti v okviru mednarodnega Združenja za raziskovalne podatke (*Research Data Alliance – RDA*) na področju kmetijstva za razvoj standardov za čim širšo uporabo podatkov. V empirično-raziskovalnem delu predstavljamo rezultate raziskave za Slovenijo, ki temelji na projektu CRP ITzaSKP, v okviru katerega so bili opravljeni intervjuji in fokusne skupine z deležniki v zvezi s podatkovnimi potrebami in možnostmi, pregled nacionalnih administrativnih virov ter eksperiment uporabe administrativnih podatkov za izboljšanje procesov odločanja, načrtovanja in izvajanja politike. V zaključku članka obravnavamo pomen rezultatov za uporabo podatkov v slovenskem kmetijstvu ter predlagamo ukrepe in nadaljnje raziskave.

Konceptualni okvir: semantični viri podatkov v kmetijstvu

² Tudi krog informirane razprave o SKP na ravni EU je razmeroma zaprt (Lovec, 2016). Hkrati dejstvo, da je kmetijska politika v veliki meri določena na ravni EU, ob »permisivnem konsenzu« in položaju neto prejemnice negativno vpliva na obseg javne razprave o ukrepih v Sloveniji.

V okviru skupine RDA za agrisemantiko, ki se ukvarja s standardi podatkov v kmetijstvu, so v zadnjih letih aktivnosti potekale v dveh fazah: najprej je bil pripravljen konceptualni okvir za raziskovanje in opravljen pregled virov podatkov (zbirk) (Aubin idr., 2017), v drugi fazi pa so se raziskovalci ukvarjali s povpraševanjem po kmetijskih podatkih na posameznih področjih (Whitehead idr., 2018) ter skušali na podlagi »ponudbe« in »povpraševanja« oblikovati predloge za različne akterje, vključene v proces načrtovanja, zbiranja in (ponovne) uporabe podatkov (Whitehead idr., 2019).

Ponudba podatkov

Raziskovalci RDA so izhajali iz vloge digitalizacije in spleta, ki sta omogočila neposredni dostop do naraščajočega obsega podatkov v različnih oblikah na vseh področjih, vključno s kmetijstvom. Pri tem so opozorili na pomen metapodatkov – podatkov o podatkih, ki služijo (avtomatiziranemu) odkrivanju in povezovanju podatkov za različne uporabnike in namene, od javnih in strokovnih služb ter raziskovalne dejavnosti do kmetov, potrošnikov in podjetij. Po mnenju raziskovalne skupine metapodatki prinašajo različne izzive, od terminologije ter ravni in podrobnosti opazovanj do različnih izhodišč in potencialnih konfliktnih interesov.³ Osnova kakovostnih metapodatkov so dokumentacija, upoštevanje standardov, trajni identifikatorji ter skupna informacijska infrastruktura.

Raziskovalci so svoj pristop opredelili kot »semantičen«. Semantika se ukvarja s pomenom jezika, bodisi »naravnega« (področje lingvistike) ali formalnega (računalništvo) (Aubin idr., 2017: 3). Nanaša se na splošno berljivost podatkovnih struktur za različne namene uporabe, tako z vsebinskega kot računalniškega vidika. Z besedami raziskovalcev je semantika definirana kot »opis pomena podatkov, ki ga omogočajo semantični viri (strukture), s ciljem ustvariti informacijo, ki bi lahko eksplicitno pomagala najti, razumeti in ponovno uporabiti podatke (zbirke)« (Aubin idr., 2017: 1).

Semantika omogoča prepoznavanje t. i. entitet, ki povezujejo posamezna področja zbiranja in uporabe podatkov (Aubin idr., 2017: 4–5). Entiteta je opredeljiv subjekt ali objekt, na katerega se lahko nanašajo različna opazovanja (primere za kmetijstvo predstavljamo v nadaljevanju). Entiteta je ključna z vidika povezljivosti – značilnosti sistemov, ki omogoča obdelavo

³ Konfliktni interesi se nanašajo na potencialne konflikte med interesi tistih, na katere se podatki nanašajo, ter interesi ustvarjanja in uporabe; na vpliv uporabe podatkov na nasprotujoče si interese; na konflikte med različnimi nameni ustvarjanja, ki pogojujejo možnosti uporabe; ter na konflikte med različnimi nameni ustvarjanja in uporabe podatkov. Podatki o podatkih so pomembni za razumevanje omejitev in pristranskosti ter s tem za odgovorno uporabo. Pomembno pa je tudi vprašanje strukturnih neenakosti v okviru dostopa, razumevanja, možnosti in koristi uporabe, ki izhajajo iz razlik v interesih, znanju, finančnih virih ter organiziranosti različnih akterjev in interesnih skupin. Za etične vidike ravnanja s podatki v kmetijstvu glej GODAN, 2020.

informacij, ustvarjenih s strani nekoga drugega. Semantika na račun povezljivosti vpliva na (ponovno) uporabo podatkov ter preprečuje podvajanje. Pri tem so ključni deli metapodatki: vsebujejo elemente, ki razlagajo pomene in možne vrednosti podatkov, v različnih oblikah, od slovarjev, seznamov, kod in klasifikacijskih shem do taksonomij idr. Za te elemente v zvezi s standardom podatkov so Aubin idr. (2017: 6–8) uporabili skupen izraz »semantični viri ali strukture«.

Kmetijske podatke lahko definiramo kot podatke, ki so ustvarjeni ali uporabljeni v kmetijstvu oziroma v povezavi s kmetijstvom (primarna pridelava hrane in proizvodnja kmetijskih izdelkov). S področjem kmetijstva se ukvarja več disciplin, od (agrarne) ekonomije do biosistemskih ved in družboslovja. Včasih v okviru kmetijstva obravnavamo tudi področje hrane (Aubin idr., 2017: 8). Semantika omogoča vpogled v podatkovne strukture, ki so pogosto vezane na posamezno disciplino ali raziskovalno področje, na način, ki širi možnosti uporabe podatkov na različnih ravneh, vključno z interdisciplinarnimi raziskavami, ki jih narekujejo trendi v sodobnem kmetijstvu. Uporaba podatkov, ustvarjenih s (povsem) drugim namenom, ne krepi zgolj vrednosti teh podatkov, temveč tudi znanstveno vrednost raziskave.

Primer delitve vsebinskih področij ali domen je bil oblikovan v okviru pobude GODAN (*Global Open Data for Agriculture and Nutrition*). Poznamo še delitve glede na vrste podatkov, statističnih, podatkov merilnih mest/senzorjev, opazovanj zemlje in drugih, s katerimi so povezani različni viri, tipične vsebine ter ravni opazovanja (enota, splošnost). Različni so tudi formati podajanja podatkov. Nazadnje je tu še vprašanje odprtosti podatkov, ki se ob predpostavki urejenega lastništva nanaša na pravne in tehnične vidike za pripravljanje podatkov za uporabo ali predelavo v metapodatke.

Semantika je kot pristop pomembna zato, ker je v praksi poznavanje struktur v ozadju podatkov zelo različno – od uporabnikov, ki običajno vedo najmanj, prek razvijalcev aplikacij in knjižničarjev, ki jih semantični viri zanimajo posredno, do razvijalcev in vzdrževalcev (skrbnikov), ki združujejo različna znanja in podatke najboljše poznajo.

Pomen semantike se kaže v različnih procesih, povezanih z načeli FAIR (Aubin idr., 2017: 9–19). Ko gre za iskanje podatkov (*Findable*), je pomembna kontrolirana terminologija, ne le jezikovno, temveč tudi v smislu logičnih definicij, ki omogočajo avtomatizirano prepoznavanje. Primer globalnega standarda so definicije in kazalniki v okviru Svetovne banke ter Organizacije za kmetijstvo in prehrano (*Food and Agriculture Organisation – FAO*) (t. i. kmetijski slovar ali *Agrovoc*) (glej tudi VEST, 2020). V fazi pridobivanja in urejanja podatkov (*Accessible, Interoperable*) so pomembni skupni identifikatorji za posamezne entitete, povezani prek spletnih orodij (odprti povezani podatki – *Open Linked Data*), v standardnih formatih,

konsistentno uporabljeni v času in odprti javni uporabi (Caracciolo idr., 2012). Na ravni programske opreme so pomembni vmesniki za povezovanje aplikacij – *Application Programming Interface* ali API. Organizacije, kot so FAO, Mednarodni center za kmetijske in bioznanosti CABI (*Centre for Agricultural and Biosciences International*) ter različne nacionalne institucije, že več let delajo na poenotenju najpogostejših konceptov, uporabljenih v kmetijstvu. Na razumevanje in ponovno uporabo (*Reusable*) vplivajo tudi logični opisi skladno z ontologijami, ki se nanašajo na tipe subjektov, razmerja med njimi in podobno ter morajo biti dostopne in povezljive. To je še posebej pomembno pri sistemih za podporo odločanju v kompleksnih situacijah, ki zajemajo alternativne možnosti ter upoštevanje tveganj in negotovosti (Lindblom, 2017). Takšni sistemi, ki zajemajo na primer povezovanje podatkov iz statistike, senzorjev in opazovanj zemlje, potrebujejo knjižnico modelov in ontologij z nedvoumnimi definicijami. Pojavljajo se tako v okviru preciznega kmetijstva (Kamilaris idr., 2016) kot ekosistemskih storitev (na primer aplikacija umetne inteligence za ekosistemske storitve ARIES – *Artificial Intelligence for Ecosystem Services*), odločanja na kmetiji in javnih politik (Veerle idr., 2015; Rose idr., 2016; 2018).

Nato je raziskovalna skupina za agrisemantiko opravila pregled in analizo obstoječih semantičnih virov podatkov (uporabila je zemljevid podatkovnih standardov za kmetijstvo in hrano v okviru GODAN, povezan z registrom FAO, ter repozitorij ontologij na AgroPortalu Univerze Montpellier). Rezultati pregleda so pokazali, da je bila tudi v primeru katalogiziranih virov pogosta hkratna uporaba globalnih in lokalnih standardov, pogost je bil razkorak med navedenim na spletu ter stanjem na podlagi dejanske dokumentacije, ki je bila sicer prisotna v le slabih tretjini primerov. Dobra polovica standardov podatkov je bila strojno bralna, licence pa je imela petina podatkov, od tega je večji del sicer omogočal prosto uporabo. Pokazale so se tudi razlike med domenami: največ standardov v obliki ontologij je bilo na voljo na področju znanosti o rastlinah. Sledili so naravni viri, kjer je bil delež visok predvsem na račun geoprostorskih podatkov, čeprav so se na ravni podrobnosti pokazale razlike. Na področju prsti so bili standardi dobri, vendar pogosto premalo formalizirani. Na področju rabe površin so manjkale splošne klasifikacije. Na področju administrativnih in zakonodajnih podatkov je bil zaznan velik potencial hkrati s šibko standardizacijo. Ponekod so bili na voljo generični statistični podatki z malo povezavami na domensko specifično semantiko. Na področju verige vrednosti je bil v zadnjih letih zaznan napredek, vendar je večinoma šlo za regulatorne ali sintaktične vire z malo referencami na skupno semantiko. Do določene standardizacije je prišlo na področju hrane, na primer v povezavi s sestavo (hranili) (Aubin idr., 2017: 4. poglavje; za raziskave po domenskih področjih, ki so potekale v nadaljevanju, glej Whitehead idr., 2018).

Povpraševanje in priporočila

Po zaključku prve faze raziskovanja je raziskovalna skupina ugotovila, da na področju stanja semantičnih virov od začetka aktivnosti do bistvenega napredka ni prišlo – večina virov ni omogočala strojnega branja in ni imela urejenih ontologij, viri so bili bodisi poduporabljeni ali pa so se uporabljali onkraj namena uporabe. Zato so se raziskovalci odločili, da se s pozivom obrnejo na ustvarjalce in uporabnike podatkov ter tako identificirajo ozka grla, ki bi jih lahko naslovili s predlogi. Na podlagi prejetih odgovorov so ugotovili, da je problem semantičnih virov v širokem naboru namenov zbiranja in obdelave ter povezanih strokovnih profilov in kompetenc. Orodij je bilo skoraj toliko kot namenov zbiranja. Opredeljeni so bili trije izzivi: (a) potreba po dostopnosti ontologij neontologom, (b) skupna orodja in platforme za zniževanje stroškov lokalnih in *ad hoc* rešitev ter (c) integrirana orodja za prevzem in urejanje, ki bi zmanjšala potreben nabor kompetenc za delo s semantičnimi viri (Whitehead idr., 2019: 1–2). Nato je raziskovalna skupina oblikovala več priporočil za izboljšanje semantičnih virov za povezljivost podatkov (Whitehead idr., 2019):

- K semantičnim virom je treba pristopiti s perspektive življenjskega cikla podatkov, od načrtovanja, vrednosti in možnosti uporabe do poslovnih modelov upravljanja, hrambe in zagotavljanja dostopa, vključno s tehničnimi rešitvami in infrastrukturo.
- Ustvarjalci in upravljavci podatkov naj v čim večji meri uporabljajo skupne modele in globalne identifikatorje, merila enot, parametre opazovanj ter dokumentacijo procesov in rezultatov.
- Knjižničarji ter razvijalci orodij in platform naj zagotavljajo integrirana orodja za prepis, formate, orodja v oblaku, najbolj razpoložljiva odprtokodna orodja, povezave do repozitorija za ponovno uporabo, vključitev v kataloge, posodabljanje, funkcionalnosti za iskanje in vizualizacijo.
- Odločevalci naj skrbijo za usmerjanje na posamezna strateška področja uporabe podatkov, razmišljajo naj dolgoročno in globalno.

V nadaljevanju predstavljamo izvirno raziskavo semantičnih virov administrativnih kmetijskih podatkov v Sloveniji. Raziskava se namenoma ukvarja z administrativnimi podatki kot najobsežnejšim virom (mikro)podatkov o slovenskem kmetijstvu (z vidika kmetijske politike), ki se uporablja za namene različnih raziskovanj (nacionalna in mednarodna statistika, delo raziskovalno-razvojnih institucij). Čeprav so administrativni podatki skladno s pregledom dela raziskovalne skupine za agrisemantiko v okviru RDA zgolj eno izmed področij podatkov, gre za področje z velikim potencialom. Poleg tega razpoložljivi podatki posegajo na številna druga

področja, od že omenjenih statistik, geoprostorskih podatkov in rabe površin do podatkov o živalih in drugih. Skladno z nameni zbiranja in uporabe ter manjšim obsegom standardizacije so pričakovani problemi semantike bolj osnovni, na primer prisotnost dokumentacije, popis virov in podobno. K prilagojenim priporočilom za Slovenijo skladno s splošnimi priporočili skupine za agrisemantiko se vračamo v razpravi po predstavitvi rezultatov.

Metodologija raziskave za Slovenijo

Z javnimi administrativnimi viri podatkov v slovenskem kmetijstvu se je ukvarjalo že več raziskav. V okviru projekta Volk idr. (2017) sta bili opravljeni analiza uporabnosti administrativnih virov za strukturne raziskave na področju standardnega prihodka (Rednak idr., 2017) ter raziskava kakovosti podatkov mreže računovodskih podatkov na kmetiji FADN za Slovenijo (Kožar, 2017). V okviru projekta Erjavec idr. (2018) je bil opravljen pregled razpoložljivih statistik in (administrativnih) podatkov za strateško načrtovanje v slovenskem kmetijstvu. Na kmetijskem inštitutu uporabljajo administrativne izbirke za dopolnjevanje namenskih strokovnih zbirk, kot je CPZ Govedo (Jeretina, 2019).

Vendar pa so se omenjene raziskave ukvarjale le z delom administrativnih podatkov, relevantnih za posamezne namene, vprašanje semantičnih virov in struktur pa so naslavljale le posredno. V nasprotju s tem je projekt CRP ITzaSKP izhajal iz potencialne uporabe administrativnih podatkov v slovenskem kmetijstvu kot takšnih, in to za različne, še posebej takšne namene uporabe, ki ob ustvarjanju podatkov niso bili predvideni (Lovec idr., 2020b). Projekt je bil sestavljen iz treh sklopov: v okviru prvega sklopa sta bila opravljena pregled in vrednotenje administrativnih registrov, evidenc in zbirk; v okviru drugega sklopa je bil opravljen pregled povpraševanja po podatkih in podatkovnih storitvah za preverjanje možnosti razvoja slednjih na podlagi prepoznavanja ozkih grl; v okviru tretjega sklopa je potekal eksperiment uporabe administrativnih zbirk za izvajanje konkretnih nalog v kontekstu odločanja in izvajanja kmetijske politike (SKP).

Projekt je sledil aktivnostim skupine za agrisemantiko v okviru RDA, tako da se je usmeril k ponudbi in povpraševanju po podatkih, hkrati s tem pa se je osredotočil na vidike semantike, ki so ustrezali administrativnim virom ter obsegu standardov in uporabe (potencialna ustreznost entitet, metapodatki in viri, zakonodajne, raziskovalne in druge potrebe deležnikov).

V okviru prvega raziskovalnega sklopa, tj. pregleda in vrednotenja, je bil cilj opredeliti vsebinsko relevantnost (vrednost) podatkov z vidika širših možnosti uporabe. Pri tem smo se

oprli na točkovanik, ki ga za prevzem raziskav uporablja ADP.⁴ Poleg tega smo ocenili statistično kakovost zbirk v smislu zanesljivosti, točnosti, natančnosti in ročnosti, skladno s standardi, ki veljajo za administrativne vire (SURS, 2017). Statistike so trenutno najpogostejši način uporabe administrativnih virov in s tem njihove standardizacije ter ocenjevanja kakovosti. Nazadnje smo ocenili še organizacijsko zrelost dajalca podatkov. Pri tem smo se dotaknili standardov FAIR z vidika javnih organizacij (OECD, 2014), ki se nanašajo na dokumentacijo in metapodatke, katalogizacijo, protokole dostopa, licence in namenska orodja. Bolj kot dejanska ocena stanja je bil namen krepiti razumevanje vrednosti, kakovosti in odprtih podatkov.

Tabela 1: Ovrednotenje podatkov glede na uporabljene globalne/lokalne standarde.

Kategorija	Oprelitev	Vir	Lestvica
Vsebinska relevantnost	Objekt, raven in obseg opazovanja	Prevzemni točkovanik ADP (2011; NERC, 2020)	1 – neustrezno/ni na voljo, 2 – lokalni standard,
Statistična kakovost	Zanesljivost, točnost, natančnost, ročnost, trajnost	Standardi SURS (2017; Eurostat, 2014) za administrativne vire	3 – potencialno globalni standard, 4 – delno globalni standard,
Organizacijska zrelost	Katalogi, opisi in oprema podatkov, licence, vmesniki	FAIR (CESSDA, 2017) in OECD (2014)	5 – globalni standard v pomembnem delu

Vir: lasten prikaz.

Pregled in ocenjevanje sta potekala tako, da smo najprej pridobili t. i. skrbniške priročnike izbranih administrativnih registrov, evidenc in zbirk, nato pa smo na podlagi gručenja (področja, organizacijske enote, razvitost zbirk) opravili podatkovne intervjuje s skrbniki, informatiki in drugimi strokovnimi ter vodstvenimi delavci, pristojnimi za posamezno področje dela.

V okviru drugega sklopa projekta, ki se je nanašal na povpraševanje po podatkih (Lovec idr., 2020c), smo opravili pregled referenčnih raziskav ter strateških in zakonodajnih dokumentov EU in v Sloveniji na področju SKP. Opravili smo pogovore z različnimi deležniki in potencialnimi uporabniki podatkov, od zaposlenih na ministrstvu za kmetijstvo in v organih v njegovi sestavi do kmetijske zbornice, svetovalne službe, nevladnih organizacij, raziskovalnih organizacij in zasebnega sektorja. Med drugim smo pregledali rezultate projekta FLINT (*Farm Level Indicators for New Topics in Policy Evaluation*), ki je oral ledino na področju globalnih

⁴ Točkovanik zajema več kriterijev, ki so razdeljeni v štiri skupine: področna ustreznost; zgodovinska, kulturna vrednost in enkratnost; znanstvena relevantnost, na primer večrazsežnost, izčrpnost zajema in primerljivost; relevantnost za izobraževalne in druge namene.

standardov mikropodatkovnih virov v zvezi z odločanjem in izvajanjem SKP na ravni EU z vidika vključevanja ekonomskih, okoljskih in družbenih dimenzij (Poppe, 2016). Pregledali smo predlog nove SKP za obdobje 2021–2027, ki je predvidela spremenjen izvedbeni model v smislu večje nacionalne prilagodljivosti, hkrati pa obveznega utemeljevanja odločitev in spremljanja izvajanja na podlagi enotnih kazalnikov in podatkov (Evropska komisija, 2018). Pregledali smo tudi podatkovne potrebe, opredeljene v nacionalni resoluciji (MKGP, 2019) in raziskavah v zvezi s strateškim načrtovanjem v kmetijstvu (Erjavec idr., 2018).

Na podlagi pregleda administrativnih virov (prvi sklop) in izraženega povpraševanja (drugi sklop) sta bila v okviru tretjega sklopa izvedena eksperimentalni prevzem administrativnih podatkov ter prikaz izvedbe analitičnih nalog (Zupan idr., 2020). Namen je bil prepoznati potencialne težave in stroške, povezane z opisi podatkov, formati, človeškimi viri in infrastrukturo, v odnosu do možnosti in koristi uporabe s pomočjo orodij za analitično obdelavo in vizualizacijo.

Ugotovitve 1: Ovrednotenje ponudbe javnih administrativnih podatkov z vidika metapodatkov ter ustreznosti in kakovosti virov

Pregled in vrednotenje javnih administrativnih virov v slovenskem kmetijstvu sta pokazala, da je težava že samo odkrivanje podatkov (*Findable*). Posamezne zbirke podatkov so sicer navedene na portalu Ministrstva za javno upravo Odprti podatki Slovenije (OPSI), vendar je seznam nepopoln, opisi pa zelo skromni. Na podlagi pogovorov s skrbniki in uradniki na Ministrstvu za kmetijstvo, gozdarstvo in prehrano (MKGP) ter v organih v njegovi sestavi (Agencija za kmetijske trge in razvoj podeželja – AKTRP – ter Uprava za varnost hrane, veterinarstvo in varstvo rastlin – UVHVVR) smo identificirali več kot 50 naslovov registrov, evidenc in zbirk, ki so bile zelo različne po namenu zbiranja, obsegu, vrsti in količini podatkov, ravni povezanosti z drugimi zbirkami, formatih, uporabniških orodjih in drugem. Izkazalo se je, da je bila že priprava seznama velik zalogaj, saj so zaposleni na MKGP in v organih v njegovi sestavi poznali predvsem tiste zbirke, s katerimi so delali. Podobno so tudi informirani zunanji uporabniki (na primer raziskovalci s področja agrarne ekonomije) poznali predvsem posamezne zbirke ali dele zbirk, ki so jih uporabljali pri svojem delu. Pregleda nad celoto podatkovnih zbirk ni imel nihče, prav tako zbirke niso bile enotno zavedene v nobenem internem, kaj šele zunanjem katalogu.

Dostop do podatkov (*Accessible*) je bil prav tako omejen predvsem na posamezne zbirke, informirane zunanje uporabnike in znane načine uporabe. Poleg prevzemanja in navzkrižnega preverjanja posameznih podatkovnih sklopov med MKGP in organi v njegovi sestavi je imel

izvoz iz posameznih zbirk urejen SURS, za nekatere druge uporabnike, kot so občine, pa so bili predpripravljeni posamezni podatki. Do posameznih podatkov, povezanih s strukturo kmetijstva, rabo površin, proizvodnjo (na primer vina), izvajanjem ukrepov idr., so dostopali tudi raziskovalci, študentje, novinarji in drugi. Vendar pa je dostop temeljil na njihovem lastnem zavedanju o obstoju vira oziroma poznavanju znotraj internih krogov uporabnikov.

Povezljivost (*Interoperability*) je bila omejena z več vidikov. Prvi vidik je bil (ne)obstoj podatkov o podatkih oziroma metapodatkov. Edina dokumentacija o zbirkah, ki je bila na voljo, so bili t. i. skrbniški priročniki, ki so bili prvenstveno namenjeni upravnim procesom, ki pa so vključevali nekatere informacije o namenih zbiranja, zakonodajni podlagi, virih podatkov, načinih obdelave, logičnih in drugih kontrolah ter povezanih zbirkah. Informacije so bile vodene na neenoten način, pogosto pa so bile netočne in neažurne. Na podlagi informacij v skrbniških priročnikih smo opredelili sedem podatkovnih domen: subjekte (KMG oziroma kmetija, vključno s površinami v rabi in dejavnostmi), živali v reji, ukrepe kmetijske politike, okolje, naravo, proizvode, kakovost in podeželje. Kot osrednjo entiteto smo opredelili kmetijsko gospodarstvo (KMG), ki je področja povezovalo s pomočjo enoličnega identifikatorja KMG-MID. Za preverjanje in pridobivanje dodatnih informacij smo nato opravili intervjuje z informatiki in vzorcem skrbnikov po sedmih domenah. Pokazalo se je, da je zbiranje primarnih podatkov vezano predvsem na ozko zakonsko opredeljene namene, iz česar so izhajale tako opazovane enote, ravni in splošnost opazovanj. Spremembe v zakonodaji, enonamenske lokalne zbirke (pogosto ločene za vsako časovno obdobje izvajanja) in šifranti so povzročali nepovezanost in podvajanje že znotraj podatkovno-informacijskih sistemov MKGP in organov v njegovi sestavi, pri čemer so bili nekateri deli, predvsem podatki o subjektih in rabi zemljišč, bolj povezani, drugi pa manj. Posamezne zbirke so bile vodene v zastarelih programskih okoljih, nekatere so temeljile celo na ročnih vpisih v Excelove tabele, hranjene v lokalnih mapah posameznih skrbnikov. Ne glede na to so bili javni administrativni podatki primarni in najkakovostnejši (natančen, točen, ročen) vir številnih podatkov, pomembnih z vidika strukturnih statističnih raziskovanj, kot so podatki o KMG (nosilec, naslov), rabi površin (travniki, njive, trajni nasadi), posameznih proizvodih (vino) ter živalih v reji (predvsem za govedo). Številne podatke so zbirali v okviru t. i. zbirne vloge v povezavi z zahtevami za izplačilo javnih podpor v okviru ukrepov SKP. Do teh podatkov so najpogosteje dostopali SURS, raziskovalci in drugi. Vendar pa je bil prenos surovih podatkov iz živih registrov v uporabne zbirke tudi zanje pogosto velik izziv. Kakovost posameznih virov je temeljila na navzkrižnem preverjanju, kontrolah ter visokih zakonskih sankcijah v primeru neskladnosti. Tako so v primeru goveda zbirke z veliko natančnostjo omogočale vpogled v dogodke in

premike posamezne živali. Znotraj posameznih domen in zbirk so bile sicer (skladno z različnimi zakonskimi cilji in zahtevami zbiranja – za področje identifikacije parcel in nekaterih vrst živali denimo veljajo zakonsko urejeni standardi kakovosti) precejšnje razlike v kakovosti posameznih podatkov.

Potencialno relevantni so bili tudi podatki s področij okolja, narave in podeželja. V prvem primeru je šlo večinoma za t. i. grafične sloje v zvezi s prostorskimi značilnostmi ter območji varovanja narave, melioracijskimi napravami in komasacijami, v drugem primeru pa za podatke o izobraževanjih in usposabljanjih v okviru ukrepov SKP ter prostovoljno vključenih v mrežo za podeželje.

Glej shemo 1: Ocene vsebinske relevantnosti, statistične kakovosti in organizacijske zrelosti po domenah. Vir: Lovec idr., 2020b.

Shema 1 zaradi višjih ocen relevantnosti podatkov nasproti statistični kakovosti, predvsem pa organizacijski zrelosti, nakazuje na podizkoriščenost podatkov. Med vsebinsko relevantnostjo zbirk in formalnimi vidiki semantičnih virov (statistična in organizacijska kakovost) je hkrati nakazana korelacija z nekaj izjemami, kot sta visoka statistična kakovost posameznih zbirk rejenih živali ter razmeroma nizka vsebinska relevantnost okoljskih in prostorskih slojev (z vidika »globalnih potreb« po podatkih na tem področju). Oboje je povezano s specifičnimi zakonskimi zahtevami v odnosu do širših družbenih potreb oziroma relevantnosti podatkov (večje ali manjše).

Največja omejitev z vidika ponovne uporabe podatkov (*Reusable*) je bilo dejstvo, da je strukturo posamezne zbirke in ozadja, pomembna za uporabo in interpretacijo podatkov, najpogosteje poznal le skrbnik posamezne zbirke, za katerega je posredovanje podatkov pomenilo dodatno delovno obremenitev. Na posameznih področjih so bili sicer na voljo javni vmesniki – pregledovalniki podatkov (na primer javni pregledovalnik površin v rabi) in surovi podatki (o rabi površin), vendar so se opisi podatkov sklicevali predvsem na zakonodajo, manjkali pa so številni elementi, potrebni za smiselno uporabo in interpretacijo podatkov (posodabljanje, opis sprememb v kvaliteti skozi čas). Številne druge informacije so bile skupaj z opisi podatkov na voljo na podlagi individualnih povpraševanj in priprave. S tehničnega vidika so omejitve ponovne uporabe s pomočjo vzpostavitve stalnih uporabniških vmesnikov (API) pomenili namenski zaprtokodni informacijski sistemi ter odvisnost od zunanjih

izvajalcev. MKGP in organi v njegovi sestavi niso imeli razvite podatkovne analitike, kompetence in viri za to pa so bili zelo omejeni tudi na ravni države.⁵

Ugotovitve 2: Podatkovne zahteve/potrebe kot priložnost za razvoj (semantičnih) standardov podatkov

Drugi in tretji sklop raziskave sta se ukvarjala s povpraševanjem po podatkih na podlagi strateških in zakonskih potreb ter eksperimentalnem preverjanju možnih tehničnih rešitev.

Povpraševanje po podatkih

V okviru projekta FLINT je bilo za vrednotenje trajnostnih vidikov kmetovanja na ravni kmetij v EU oblikovanih 30 kazalnikov, po domenah razdeljenih v tri skupine (ekonomski, okoljski in družbeni) (Poppe, 2016). Izvorna raziskava se je ukvarjala predvsem s FADN, vendar je opozarjala na vključevanje administrativnih virov. Primerjava s kazalniki FLINT kaže, da je na podlagi identifikatorja KMG-MID mogoče v Sloveniji s pomočjo javnih administrativnih virov zagotoviti (vsaj delne) podatke za kar tretjino kazalnikov. To velja predvsem za ekonomsko področje ter kazalnike o naložbah in zavarovanjih (na podlagi ukrepov), trajanju kmetij ter shemah kakovosti in izpostavljenosti tveganjem (nekmetijske dejavnosti), deloma tudi za učinkovito rabo površin (standardni prihodek). Na okoljskem področju je teh kazalnikov manj (trajno travinje, površine z ekološkim pomenom, potencialno izračuni TGP s pomočjo standardnih faktorjev), podobno je na družbenem področju (vključenost v izobraževanje in usposabljanje). Vključevanje javnih administrativnih virov v pripravo globalnih kazalnikov za Slovenijo bi zaradi zahtev skladnosti, metapodatkov in dostopanja pozitivno vplivalo na razvoj semantičnih virov, zaradi splošnosti in primerljivosti kazalnikov pa tudi na informirano javno razpravo in odločanje.

Uredba o strateškem načrtovanju SKP v obdobju 2021–2027 prinaša okoli 100 različnih vrst kazalnikov v podporo odločanju (Evropska Komisija, 2019: priloga 1). Poleg kazalnikov stanja in vpliva (*impact*), ki so v domeni Evropske komisije in se nanašajo na obstoječe statistične vire, so novost predvsem kazalniki izvedbe (*output*) (št. kmetij, površin, glav velike živine, vključenih v ukrepe) in rezultata (*result*) (zaveze pri vključenih iz naslova izboljševanja stanja, ki izhajajo iz definicije ukrepov SKP); ti so odgovornost držav članic ter zahtevajo postavljanje letnih ciljev, spremljanje in poročanje ter lahko v primeru nedoseganja vodijo v sankcije. Poleg

⁵ Poleg omejenih kadrovskih in finančnih virov so intervjuvanci kot razlog za stanje najpogosteje navajali premajhno uporabnost podatkov, omejitve z vidika varovanja zasebnosti ter potencialne negativne posledice prostega dostopa do sicer javnih podatkov (na primer uporaba podatkov v profitne namene).

tega naj bi države članice s podatki utemeljevale tudi druge odločitve, pri katerih jim je omogočena večja fleksibilnost, denimo prilagoditve pogojev za upravičenost do ukrepa glede na specifične lastnosti države. Uredba še posebej zahteva spremljanje podatkov o nekaterih okoljskih in družbenih učinkih (obseg vključenih v ukrepe, pomlajevanje), celoten proces pa naj bi bil odprt deležnikom in širši javnosti.

Nov izvedbeni model SKP je bil kritiziran zaradi šibke povezave med kazalniki vpliva in drugimi kazalniki (vrednotenje bo potekalo šele po zaključku izvedbenega cikla in ne bo vplivalo niti na prihodnji cikel politike po letu 2027) ter zaradi značilnosti novih kazalnikov, ki so usmerjeni predvsem k povečevanju vključenosti kmetij v posamezen ukrep, kar vodi v maksimizacijo porabe sredstev namesto doseganja dejanskega vpliva ukrepov (Lovec idr., 2020a). Ne glede na to pa uredba prinaša nekatere pozitivne vplive na razvoj podatkov: prvič, ker teži k povezovanju administrativnih in statističnih podatkov v fazi načrtovanja in spremljanja ter s tem različnih virov in struktur; drugič, ker zahteva kontinuirano in standardizirano pridobivanje podatkov za vodenje vrednosti kazalnikov na podlagi administrativnih zbirk; in tretjič, ker zahteva načrtovanje in spremljanje v odprtem procesu z deležniki, s tem pa tudi razvoj spletnih orodij za dostop, pregledovanje in vizualizacijo podatkov.

Nacionalna resolucija v zvezi s kmetijsko politiko (MKGP, 2019) in raziskovalni projekti so dodali še nekaj specifičnih podatkovnih potreb, ki jim je prav tako mogoče zadostiti z administrativnimi viri. Na ekonomskem področju je denimo mogoče factorske izračune izboljšati z administrativnimi podatki o investicijskih ukrepih in standardnem prihodku – najmanj na ravni preverjanja vzorcev in populacij FADN in SURS (Rednak idr., 2017). Na okoljskem področju je mogoče ukrepe v zvezi z uporabo zavarovanj in zaščitnih sredstev uporabiti za oblikovanje kazalnika. Na družbenem področju je podatke o številu vključenih v ukrep dobrobiti živali prav tako mogoče uporabiti kot kazalnik (Erjavec idr., 2018).

Omeniti velja še Evropsko podatkovno strategijo (Evropska komisija, 2020a), ki med drugim predvideva javno odpiranje vseh podatkov, ki temeljijo na financiranju iz virov EU (torej skoraj vsi podatki, povezani s kmetijsko politiko v Sloveniji), in Strategijo od vil do vilic (Evropska komisija, 2020b), ki v povezavi z novo SKP in Evropskim zelenim dogovorom predvideva dopolnitev FADN s podatki s področij okoljske in družbene trajnosti ter nove, ambiciozne kvantitativne cilje na področjih gnojenja, tveganja uporabe pesticidov in ekološke proizvodnje, ki bodo zahtevali sistematično uporabo obstoječih in zbiranje ter zagotavljanje dostopa do nekaterih novih podatkov. Omenimo lahko še spodbude za razvoj satelitskih podatkov (opazovanja zemlje) in podatkov/opazovanj v okviru razvoja preciznega kmetijstva, ki so prav

tako del »horizontalne« strategije digitalizacije, zahteve in spodbude za razvoj podatkov v okviru inventarjev biomase, snovnih tokov in trajnostnega potenciala (biogospodarstvo) ter zaščite okolja in narave (izpusti TGP, energijska bilanca, poraba vode, stanje biološke raznovrstnosti, klimatski modeli).

Eksperiment

V okviru eksperimentalnih aktivnosti smo na podlagi javnega poziva deležnikom najprej oblikovali seznam zelenih podatkovnih nalog. Te so se večinoma nanašale na osnovne vpogled v strukturne značilnosti KMG in na prostorske spremembe v povezavi z vključenostjo v ukrepe, v manjši meri pa na možnosti podatkovne analitike (modeliranje profilov in vzročnih zvez ter predvidevanje). Nato smo od AKTRP pridobili podatke za vse kmetije, vključene v sistem kmetijskih subvencij (približno 57.000 kmetij), za obdobje 2015–2019 (trenutni cikel SKP), ki so vključevali do okoli 400 spremenljivk na letni ravni. Od MKGP (register kmetijskih gospodarstev) so bili pridobljeni še podatki o zemljiščih (z rabo) površin. Podatke smo s pomočjo KMG-MID združili na ravni enote KMG ter izvedli nekaj testnih nalog v prosto dostopnem in odprtokodnem programskem okolju (Orange). Tako smo denimo skušali preveriti, ali je mogoč hiter vpogled v trend in strukturo KMG, vključenih v ukrep ekološkega kmetijstva. Skladno s Strategijo od vil do vilic naj bi se število vključenih kmetij in kmetijskih površin pod ekološkim kmetijstvom v Sloveniji bistveno povečalo. Vpogled je pokazal, da v Sloveniji starostna struktura ne vpliva na vključevanje v ekološko kmetijstvo, hkrati s tem pa velja obratna korelacija z vključevanjem v nekatere druge ukrepe na področju okolja in narave. To nakazuje, da brez celovitih sprememb ciljev na področju ekološkega kmetijstva ne bomo dosegli. V okviru druge demonstracije smo skušali ugotoviti, ali obstoječi podatki omogočajo določanje dejavnikov in napovedovanje zaraščanja površin na posameznih KMG z zadostno statistično verjetnostjo po različnih metodah (glede na posamezno kmetijo ali generalno verjetnost). Test je pokazal, da je to mogoče. Med drugim je preizkus nakazal, da vključenost v posamezen ukrep (zaradi neciljnosti večine ukrepov) ni bistven dejavnik, temveč pomembnejšo vlogo igrata občina (tj. razvitost okolja) ter struktura kmetije (velikost in proizvodna usmerjenost) (Zupan idr., 2020).

Treba je opozoriti, da namen eksperimenta ni bil ugotavljanje dejanskih vzročno-posledičnih zvez, ki bi zahtevale podrobnejšo namensko analizo v sodelovanju domenskih ekspertov, temveč prikaz potenciala podatkov za analitične naloge na relevantnih področjih, ki jih lahko ob ustreznem opremljanju podatkov in dostopu izvaja kdorkoli. Ne glede na to sta primera analitično potrdila že omenjene ocene vrednotenja v zvezi z vplivom ukrepov SKP.

Eksperiment je potrdil nekatere ugotovitve pregleda in vrednotenja administrativnih virov: podatki imajo izjemen potencial, oviri pa sta poznavanje strukture semantičnih virov ter potrebna različna znanja in kompetence za smiselno uporabo in interpretacijo. Pridobivanje, združevanje in urejanje podatkov za posamezne namene je zaradi kompleksne strukture, fragmentiranosti na področja in v času, pomanjkanja dokumentacije, tudi v zvezi s kakovostjo in načini uporabe, različnih enonamenskih sistemov, odsotnosti standardnih vmesnikov in drugega zahtevno z vidika časa in človeških virov. Pogosto je odvisno od nekaj ključnih posameznikov, ki imajo dovolj potrebnih znanj (na primer skrbnikov) in osebnih povezav.

Sklepna razprava s predlogi

Delo na globalni ravni v okviru skupine za agrisemantiko in raziskava javnih administrativnih virov v Sloveniji omogočata oblikovanje nekaterih zaključkov in predlogov, ki bi lahko v okviru aktivnosti javnih nosilcev in različnih deležnikov prispevali k razvoju in uporabi podatkov v podporo odločanju in javni razpravi v slovenskem kmetijstvu.

Prvič, posameznike moramo opolnomočiti za (s)poznavanje podatkov, priložnosti in tveganj, ki jih prinašajo. Raziskava je imela socializacijski učinek (tako kot tudi druge omenjene). Opozorili smo na različne možnosti (brezplačnih) izobraževanj in usposabljanj ustvarjalcev podatkov, knjižničarjev in drugih, ki so na voljo v okviru mednarodnih organizacij, in nekatere celo spodbudili k udeležbi. Žal pa dosedanje aktivnosti niso dovolj naslavljale (malih) kmetov in potrošnikov kot najštevilčnejše skupine. Njihovo vključenost bi lahko zagotovili z razvojem uporabniških aplikacij, ki bi črpale iz administrativnih virov, kot so samokontrola sprejetih obveznosti v okviru ukrepov (na primer QR kode na pesticidih, lokacijsko pogojene zahteve) za kmete, podatki o proizvodih (QR kode na izdelkih) in lokacijski podatki o kmetijah (ekološka in podobne oblike proizvodnje, standardi). Aplikacije so lahko decentralizirane ali pa omogočajo različne ravni dostopa do podatkov in statistik. Uporaba povratnih informacij bi omogočala koristi tudi dajalcem, na primer optimizacijo svetovanja in kontrol ali spremljanje masnih tokov in trženjske analize.

Drugič, razviti moramo odgovorno javno podatkovno kulturo. Trenutno ustvarjanje podatkovnih zbirk vodijo ozke zahteve, ki izhajajo iz zakonodaje in predpisov, dostop pa kljub »javnosti« podatkov ovirajo izgovori, povezani z visokimi stroški, omejenimi koristmi in zlorabami. V nasprotju s tem bi širša uporaba podatkov okrepila njihovo vrednost, kakovost in zaupanje. Evropska podatkovna strategija navaja primere evropskih držav, v katerih morajo zasebni lastniki zagotoviti javni dostop do podatkov iz aktivnosti: tako kot je gozd v določeni meri javno dobro, so javno dobro tudi podatki o upravljanju z njim. Dajalci so hkrati tisti, ki

lahko imajo od dajanja podatkov največjo korist, tako neposredno (analize in primerjave) kot posredno (vrednost podatkov se prelije v vrednost dajanja). V smislu širše javne razprave lahko ključno vlogo odigrajo deležniške in raziskovalne organizacije z močnimi primeri potenciala podatkov za družbeno blaginjo. Lahko gre za preproste statistike, kot je razdeljevanje sredstev med velikimi in malimi KMG, odsotnost podatkov o učinkih na okolje in naravo, neuravnoteženo demografsko strukturo nosilcev KMG ali za kompleksnejše analize dohodkovnih, okoljskih in socialnih učinkov ukrepov. Možnosti pa se kažejo tudi za odločevalce. Politike na ravni EU so okvir za spremembe, vendar jih je treba ponotranjiti kot priložnost. Zakaj ne bi na primer kazalnika vključenosti v ukrep dobrobiti živali naredili za blagovno znamko?

Tretjič, treba je sistematično delati na razvoju podatkov in upravljanju z njimi po celotnem življenjskem ciklu. Povezovanje ter izkoriščanje obstoječih možnosti in sinergij je ključno za znižanje stroškov in širjenje možnosti uporabe (vrednosti). Vsako ustvarjanje podatkov naj se začne s pregledom obstoječih virov in predhodno oceno vrednosti za različne uporabnike. V čim večji meri je treba upoštevati skladnost z obstoječimi globalnimi in lokalnimi standardi. Na voljo so brezplačne spletne platforme ter orodja za metapodatke in arhiviranje (v okviru projekta smo poskrbeli za opis ključnih administrativnih zbirk v DATAVERSE, 2020). Spodbujati je treba različne poslovne modele priprave, zagotavljanja, delitve in uporabe podatkov s strani različnih organizacij, tudi profitno usmerjenih, ob ustreznih nadomestilih tistim, ki podatke delijo. Javne oblasti lahko v okviru javno financiranih raziskav zahtevajo uporabo in ustrezno deljenje ustvarjenih zbirk. V okviru projekta smo pokazali na konkretne možnosti izkoriščanja obstoječih virov in sinergij z novimi potrebami (nadgradnja zbirk, modelni izračuni na kmetijskem inštitutu z različnimi kazalniki SKP in e-tal z izboljšanimi podatki o založenosti tal, dostop KGZS do posameznih podatkov in avtomatizirane aplikacije, na primer za gnojilne načrte, uporaba komplementarnih zbirk za navzkrižno oceno kakovosti ter dopolnjevanje vzorcev in populacij, glej Lovec idr., 2020b; 2020c). Uporabo kmetijskih podatkovnih zbirk je treba vključiti v izobraževalni proces. Izpostaviti je treba pozitivne primere uporabe zbirk (razpisane zaključne naloge, nagrade) ter spodbujati izmenjavo med ustvarjalci in uporabniki podatkov, razvijalci aplikacij in knjižničarji (skupni projekti, paneli na konferencah).

Četrtič, podatkovna znanost/analitika je najkompleksnejša raven razvoja in uporabe podatkov, ki pa lahko ima največje učinke na blaginjo. Potrebne investicije, na primer posegi v informacijske sisteme za zagotavljanje centralizirane analitike, so lahko glede na obstoječe stanje visoko tvegane z vidika stroškov in potrebnega časa za zagotavljanje ustreznih rezultatov.

Vodijo lahko v odvisnost od posameznih dobaviteljev licenčne opreme. Alternativa temu so lahko vsaj sprva manjši projekti, usmerjeni h konkretnim ciljem, kjer je tveganje razpršeno, uporaba (obstojećih) odprtokodnih orodij za podatkovno analitiko in učenje s poskušanjem. Preprosta enotna statična zbirka, ustvarjena v okviru eksperimenta, kaže na številne možnosti analitičnih vpogledov za različne namene, od načrtovanja do spremljanja in vrednotenja. Omogoča denimo ugotavljanje in testiranje ciljnih skupin in mejnih vrednosti, ugotavljanje vzročno-posledičnih zvez in napovedovanje, in sicer tako na agregatni ravni kot na ravni KMG (Zupan idr., 2020c). Čeprav je bilo vse to mogoče tudi do sedaj, je vsaka posamezna naloga zahtevala ustvarjanje ali uporabo namenske zbirke in orodja ter dolgotrajno delo skrbnika, domenskega strokovnjaka, statistika in drugih. To je bil tudi razlog za majhen dejanski obseg podatkovne analitike v kmetijstvu. Eksperiment je medtem pokazal, da je mogoče z razumnim vložkom in tveganji bistveno skrajšati potreben čas in vire ter demokratizirati dostop do podatkov in vpoglede.

Literatura

- ADP (2011): *Odprti podatki. Intervju – Stanje, problemi, odgovornost in skrb nad podatki*. Dostopno na: https://www.adp.fdv.uni-lj.si/podatki/odpp11/odpp11_vp1_sl_v1_rl.pdf (6. januar 2021).
- Aubin, Sophie, Caterina Caracciolo in Panagiotis Zervas (2017): *Landscaping the Use of Semantics to Enhance the Interoperability of Agricultural Data*. *RDA Agrisemantics Working Group*. Poročilo. Dostopno na: <https://www.rd-alliance.org/system/files/documents/Deliverable1%20-%20Landscaping.pdf> (6. januar 2021).
- Caracciolo, Caterina, Armando Stellato, Ahsan Morshed, Gudrun Johannsen, Sachit in Johannes Keizer (2012): The AGROVOC Linked Dataset. *Semantic Web Journal* 4(3): 341–348.
- CESSDA (2017): *How FAIR Are Our Data*. Dostopno na: https://www.cessda.eu/content/download/3845/35038/file/20170707_How_FAIR_are_your_data_Jones.pdf (6. januar 2021).
- DATAVERSE (2020): *Kmetijski podatki (ADP)*. Dostopno na: https://dataverse.adp.fdv.uni-lj.si/dataverse/kmetijski_podatki (25. januar 2021).
- Erjavec, Emil, Tanja Šumrada, Luka Juvančič, Ilona Rac, Tomaž Cunder, Matej Bedrač in Marko Lovec (2018): *Vrednotenje slovenske kmetijske politike v obdobju 2015–2020. Raziskovalna podpora za strateško načrtovanje po letu 2020*. Ljubljana: Kmetijski inštitut Slovenije.
- Evropska komisija (2018): *Proposal for a Regulation Establishing Rules on Support for Strategic Plans to Be Drawn Up by Member States under the Common Agricultural Policy (CAP Strategic Plans)*. COM (2018) 392 final 2018/0216 (COD), Bruselj, 1. junij 2018.
- Evropska komisija (2020a): *A European Strategy for Data*. COM(2020) 66 final. Bruselj, 19. februar 2020.
- Evropska komisija (2020b): *Farm to Fork Strategy for a Fair, Healthy and Environmentally-Friendly Food System*. COM/2020/381 final. Bruselj, 20. maj 2020.
- EUROSTAT (2014): *ESS Handbook for Quality Reports. Methodologies and Working Papers*. Luxembourg: Eurostat.
- GODAN (Global Open Data for Agriculture and Nutrition) (2020): *Data Ethics Workshop. Posnetek delavnice*. Dostopno na: <https://www.godan.info/news/godan-data-ethics-workshop> (6. januar 2021).

- Jeretina, Janez (2019): *Predstavitev CPZ Govedo. Odprti raziskovalni podatki v Sloveniji*. 14. november 2019.
- Kožar, Maja (2017): Analiza kakovosti rezultatov FADN za Slovenijo. V *Razvoj celovitega modela kmetijskih gospodarstev in povezanih podatkovnih zbirk za podporo pri odločanju v slovenskem kmetijstvu, zaključno poročilo*, T. Volk idr. (ur.), 53–142. Ljubljana: BF, KIS in KGZS.
- Lindblom, Jessica, Magnus Ljung in Anders Jonsson (2017): Promoting Sustainable Intensification in Precision Agriculture: Review of Decision Support Systems Development And Strategies. *Precision Agriculture* 18(3): 309–331.
- Lovec, Marko (2016): *The European Union's Common Agricultural Policy Reforms: Towards a Critical Realist Approach*. Basingstoke: Palgrave Macmillan.
- Lovec, Marko, Tanja Šumrada in Emil Erjavec (2020a): New CAP Delivery Model, Old Issues. *Intereconomics* 55(2): 112–119.
- Lovec, Marko, Janez Štebe, Maja Dolinar, Brigita Bočkaj in Maša Kerstein (2020b): *Pregled zbirk podatkov na področju kmetijstva (RI.1) in predlogi za izboljšave (RI.2)*. Ljubljana: Fakulteta za družbene vede.
- Lovec, Marko, Janez Štebe, Brigita Bočkaj, Maša Kerstein, Gregor Kramberger in Andrej Jamšek (2020c): *Poročilo v zvezi z delovnim sklopom 3: strateški pristop k podatkom : rezultati Ciljnega raziskovalnega projekta R3.1 (analiza podatkovnih potreb za izvajanje prihodnje SKP), R3.2 (povzetek posvetovanja), R3.3. (zakonske zahteve in dobre prakse), R3.4 in R3.5 (predlog strategije in načrt)*. Ljubljana: Fakulteta za družbene vede.
- MKGP (2019): *Resolucija Naša hrana, podeželje in naravni viri po 2021. Strateški okvir razvoja slovenskega kmetijstva, predelave hrane in podeželja*. Dostopno na: <https://e-uprava.gov.si/download/edemokracija/datotekaVsebinsa/382689?disposition=inline> (6. januar 2021).
- NERC (2020): *Data Value Checklist*. Dostopno na: <https://nerc.ukri.org/research/sites/data/policy/data-value-checklist/> (6. januar 2021).
- OECD (2014): *OECD Expert Group for International Collaboration on Microdata Access. Final Report*. Dostopno na: <http://www.oecd.org/sdd/microdata-access-final-report-OECD-2014.pdf> (6. januar 2021).
- Poppe, Krijn (2016): *Farm Level Indicators for New Topics in Policy Evaluation. Intro and Objectives Meeting. 199th FADN Committee*. Bruselj, november. Dostopno na: <https://www.flint-fp7.eu/downloads/presentations/flint%20in%20fadn%20committee%20nov%202016.pdf> (6. januar 2021).
- Rednak, Miroslav, Tina Volk, Barbara Zagorc, Ben Moljk in Maja Kožar (2017): Ocena standardnega prihodka KMG in njegova uporaba za presojo ukrepov in učinkov kmetijske politike. V *Razvoj celovitega modela kmetijskih gospodarstev in povezanih podatkovnih zbirk za podporo pri odločanju v slovenskem kmetijstvu, zaključno poročilo*, T. Volk idr. (ur.), 7–52. Ljubljana: BF, KIS in KGZS.
- Rose, David C., William J. Sutherland, Caroline Parker, Matt Loble, Michael Winter, Carol Morris, Susan Twining, Charles Ffoulkes, Tatsuya Amano in Lynn Dicks (2016): Decision Support Tools for Agriculture: Towards Effective Design and Delivery. *Agricultural Systems* 149: 165–174.
- Rose, David C., Joe Fodey, Caroline Park., William J. Sutherland in Lynn Dicks (2018): Involving Stakeholders in Agricultural Decision Support Systems: Improving User-Centred Design. *International Journal of Agricultural Management* 6: 80–89.
- SURS (2017): *Metodološki priročniki. Kazalniki kakovosti 2017, št. 2*. Dostopno na: <https://www.stat.si/StatWeb/File/DocSysFile/9690/kazalniki-kakovosti-2017.pdf> (6. januar 2021).
- SURS (2019): *Metodološko pojasnilo. Struktura kmetijskih gospodarstev*. Pripravil: Aleš Krajnc. Zadnjič osveženo: 20. december 2017. Dostopno na: <https://www.stat.si/StatWeb/File/DocSysFile/8225> (6. januar 2021).

Veerle, van den Eynden, Kate Schreckenber in Louise Corti (2015): Sharing Social Data in Multidisciplinary, Multi-Stakeholder Research. Best Practice Guide for Researchers. *ESPA Working Paper Series*, št. 1, december. Dostopno na: <http://www.espa.ac.uk/files/espa/Sharing%20social%20data.pdf> (6. januar 2021).

VEST (2020): *Agroportal. Map of Standards by Domain*. Dostopno na: http://vest.agrisemantics.org/about/structure?qt-content_organization_tabs=3#qt-content_org (6. januar 2021).

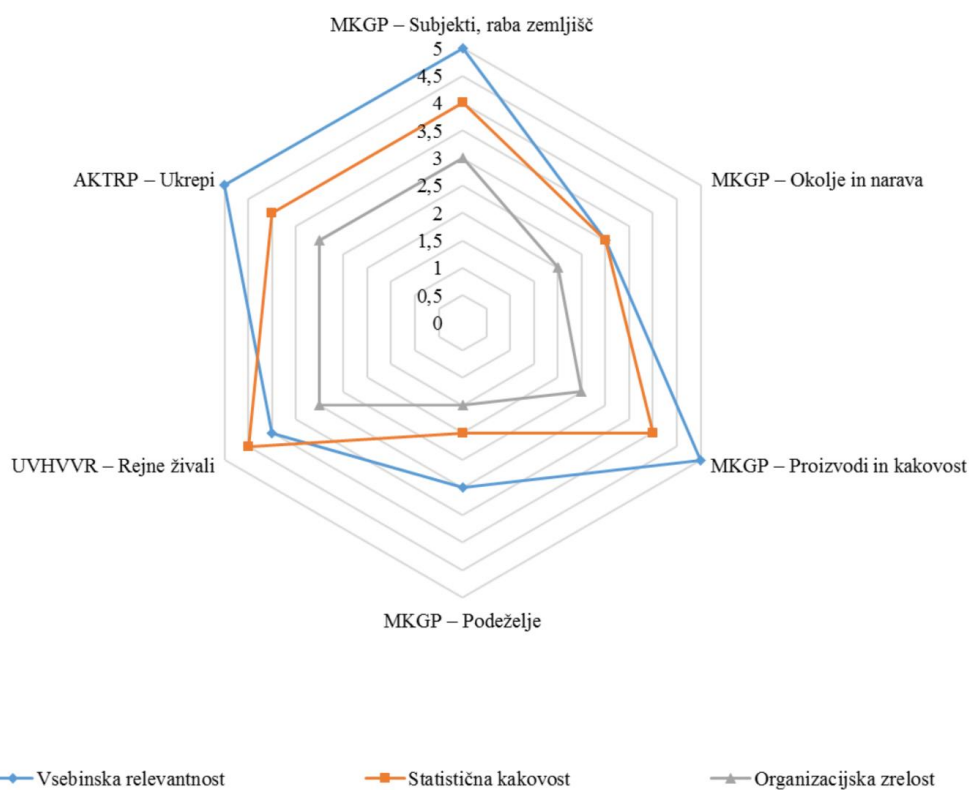
Volk, Tina, Jure Brečko, Emil Erjavec, Damjan Jerič, Stane Kavčič, Maja Kožar, Ben Moljk, Miroslav Rednak, Barbara Zagorc in Jaka Žgajnar (2017): Uvod: Sklepne projektne ugotovitve in priporočila. V *Razvoj celovitega modela kmetijskih gospodarstev in povezanih podatkovnih zbirk za podporo pri odločanju v slovenskem kmetijstvu*, Volk idr. (ur.), 5–7, 291–295. Ljubljana: BF, KIS in KGZS.

Whitehead, Brandon, Caterina Caracciolo in Sophie Aubin (2018): *A Collection of Use Cases in Agriculture and Nutrition Collected by the Agrisemantics Working Group of the Research Data Alliance*. Dostopno na: <https://www.rd-alliance.org/groups/agrisemantics-wg.html> (6. januar 2021).

Whitehead, Brandon, Caterina Caracciolo in Sophie Aubin (2019): *39 Hints to Facilitate the Use of Semantics for Data on Agriculture and Nutrition. Recommendations from the RDA Agrisemantics Working Group*. Dostopno na: <https://www.rd-alliance.org/group/agrisemantics-wg/outcomes/39-hints-facilitate-use-semantics-data-agriculture-and-nutrition> (6. januar 2021).

Wilkinson, Mark D. idr. (2016): The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Scientific data* 3: 160018.

Zupan, Blaž, Maša Kerstein, Vesna Tanko, Andreja Kovačič, Janez Štebe, Brigita Bočkaj in Marko Lovec (2020): *Poročilo delovnega sklopa 1 (tehnični pregled zbirk), delovnega sklopa 2 (metodologija izgradnje analitičnih zbirk), delovnega sklopa 4 (demonstracijski primeri) in delovnega sklopa 5 (zbirka podatkov)*. Ljubljana: Fakulteta za družbene vede.



Shema 1: Ocene vsebinske relevantnosti, statistične kakovosti in organizacijske zrelosti po domenah.
Vir: Lovec idr., 2020b.