

METODOLOGIJA ZA KVALITATIVNO VREDNOTENJE ODPRTIH PODATKOV IN PREGLED PODATKOVNIH ZBIRK NA PORTALU OPSI

JURE JUVAN, KLARA ŽNIDERŠIČ, MATIJA MAROLT,
MATEVŽ PESEK

Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Ljubljana, Slovenija
jure.juvan@fri.uni-lj.si, klara.znidersic@fri.uni-lj.si, matija.marolt@fri.uni-lj.si,
matevz.pesek@fri.uni-lj.si

Odpri podatki se iz dneva v dan pojavljajo v vse večjem številu, zato je naslovitev njihove uporabe ključnega pomena, ker postajajo tudi vse dostopnejši, pa jih je smiselno obravnavati tudi njihovo kakovost. Da bi poskrbeli za uspešno (ponovno) uporabo razpoložljivih odprtih podatkov, moramo najprej ugotoviti, kakšni ti podatki so. V prejšnji študiji smo v sklopu ciljnega raziskovalnega projekta zato na podlagi našega pregleda obstoječih raziskav in pristopov razvili metodologijo, ki je namenjena kvalitativnemu vrednotenju odprtih podatkov. S pomočjo ročnega pregleda smo tako na portalu Odprtih podatkov Slovenije (OPSI) ocenili kvalitativno stopnjo odprtih podatkov in podatkovnih zbirk, znotraj katerih se slednji nahajajo. Namen naše metodologije ni, da bi do potankosti zadostila specifikam različnih področij, temveč da bi zaobjela karseda široko in raznoliko paleto podatkovnih zbirk. S predlagano metodologijo in izdelanim orodjem za avtomatizirano preverjanje kakovosti odprtih podatkov si tako želimo povečati vsesplošno uporabo slovenskih in evropskih odprtih podatkov z najrazličnejših področij.

DOI
[https://doi.org/
10.18690/um.fov.2.2026.7](https://doi.org/10.18690/um.fov.2.2026.7)

ISBN
978-961-299-114-2

Ključne besede:
odprti podatki,
kvalitativno vrednotenje
podatkov,
metodologija,
ročni pregled,
OPSI

DOI
[https://doi.org/
10.18690/um.fov.2.2026.7](https://doi.org/10.18690/um.fov.2.2026.7)

ISBN
978-961-299-114-2

METHODOLOGY FOR QUALITATIVE EVALUATION OF OPEN DATA AND DATASETS REVIEW ON THE OPSI PORTAL

JURE JUVAN, KLARA ŽNIDERŠIČ, MATIJA MAROLT,
MATEVŽ PESEK

University of Ljubljana, Faculty of Computer and Information Science, Ljubljana,
Slovenija
jure.juvan@fri.uni-lj.si, klara.znidersic@fri.uni-lj.si, matija.marolt@fri.uni-lj.si,
matevz.pesek@fri.uni-lj.si

Keywords:

open data,
qualitative data assessment,
methodology,
manual review,
OPSI

Open data emerge in greater numbers on a daily basis, therefore, it is crucial to assess their use. They are also becoming more and more accessible, therefore their qualitative assessment is of utmost importance. In order to ensure successful (renewed) use of the available open data, we must first assess their current state. In prior research we have developed a methodology for qualitative open data assessment and in manually reviewing the datasets on the OPSI portal, we have assessed the qualitative degree of open data and datasets where the first are stored. The goal of our methodology is not to adapt its use to specific datasets and fields but to make it generally applicable to different datasets. With the proposed methodology and developed tool for automatic qualitative open data assessment we strive to widen the use of Slovenian and European open data in various fields.



University of Maribor Press

1 Uvod

Ministrstva in javne agencije vse bolj sledijo trendu odprtih podatkov, katerega glavna ideja je poskrbeti, da odprti podatki postanejo na voljo javnosti preko spletnih portalov, storitev in uporabniških vmesnikov (Braunschweig idr., 2012). S tem, ko odprte podatke predstavimo javnosti, pa nastopijo raznolike problematike. Te se nanašajo predvsem na vrednotenje njihove kakovosti oziroma kvalitativnosti in ponovno uporabljivost (Vetrò idr., 2016), ki pa v raziskavah nista visoko zastopana, saj se te večinoma usmerjajo v preučevanje metapodatkov. Problematike vrednotenja kakovosti in ponovne uporabljivosti odprtih podatkov lahko tako najdemo v vsebini odprtih podatkov in podatkovnih zbirk, iz katerih izhajajo: pomankljivi, nenatančni, nepopolni, zastareli, neunikatni in neveljavni podatki, pa tudi nejasne vrednosti, prevelike količine informacij za procesiranje, pomanjkanje bistvenih podatkov, nato nesmiselni, nedosledni, nestrukturirani podatki in povezljivost oziroma uparljivost podatkovnih zbirk ter njihovih podatkov, neustrezno opredeljeni ali težko dostopni formati, odsotnost standardov in programske opreme za obdelavo odprtih podatkov ter razdrobljenost slednje in aplikacij (Pesek idr., 2025).

V študiji Braunschweig idr. (2012) so ugotovili, da problemi z odprtimi podatki nastajajo že ob dejstvu, da so ti raziskani zgolj na specifičnih primerih, zato problem posledično terja celostno študijo o odprtih podatkih. Analiza v študiji je pokazala, da večina petdesetih obravnavanih spletnih portalov nima ustreznih standardov in vmesnikov uporabniškega programa (API), pri čemer njihovi odprti podatki niso strojno berljivi ali v ustreznem formatu. To dejstvo vztrajno preprečuje ponovno uporabljivost in možnost uporabe avtomatiziranih orodij, kar v praksi pomeni, da mnoge odprte podatkovne zbirke niso zares odprte. Hkrati nastajajo problemi v zvezi z metapodatki, ki so prisotni v različnih merah, pri čemer ustvarjajo probleme povezljivosti oziroma interoperabilnosti z drugimi odprtimi podatki in podatkovnimi zbirkami na drugih spletnih podatkovnih platformah oziroma portalih. Lahko rečemo, da so spletne platforme z odprtimi podatki odprte na različnih nivojih, kar pomeni, da skupnost odprtih podatkov ni zadostno organizirana in potrebuje močnejše smernice za uspešnejše medsebojno delovanje.

V študiji je govora tudi o tehnikah soočanja s podatkovnimi zbirkami odprtih podatkov, pri čemer avtorji poudarjajo, da je integracija podatkov v podatkovne zbirke zahtevna in da za svojo uspešno realizacijo potrebuje vztrajno in redno ročno pregledovanje. Tako kot avtorji študije tudi mi s svojim pristopom predlagamo uporabo in razvoj avtomatiziranih orodij za preverjanje odprtih podatkov na širši ravni in ne zgolj na specifičnih primerih. Na ta način zagotovimo boljšo povezljivost in ponovno uporabljivost podatkov in njihovih podatkovnih zbirk, s čimer neposredno izboljšamo zanesljivost celotnega podatkovnega portala. Ker je naša študija neposredno vezana na vrednotenje kvalitete odprtih podatkov, moramo tudi definirati, kaj predstavlja »dobro kvaliteto podatkov« (Kitchin, 2014). Ta se uresniči z diskretnostjo, razumljivostjo – pri čemer mora biti vsakršen podatek individualen, samostojen, ločljiv ter jasno definiran – in agregacijo podatkov – kar pomeni, da morajo biti podatki združljivi v podatkovne zbirke – ter z ustreznostjo metapodatkov, ki v resnici predstavljajo podatke o podatkih. Pri vsem tem pa moramo upoštevati tudi, da morajo biti podatki, kolikor jim je pripisana dobra kvaliteta, prisotni v več različnih podatkovnih zbirkah in se med seboj povezovati, saj le na tovrsten način lahko ustvarijo koheziven in koherenten pregled stanja podatkov v različnih zbirkah, ki ga ne moremo ustvariti zgolj na podlagi stanja znotraj ene same podatkovne zbirke. Z vsemi naštetimi značilnostmi lahko tako rišemo tudi vzporednice s portalom OPSI, ki predstavlja osrednji predmet obravnave naše študije, saj smo na tem portalu predhodno opravili ročni pregled podatkovnih zbirk (Pesek idr., 2025).

Ročni pregled je pokazal, da so na portalu prisotne podobne problematike kot v pregledu literature (Žnideršič idr., 2025), pri čemer velja izpostaviti sintaktične napake v obliki manjkajočih črk, šumnikov, posebnih znakov, potem raznolikost formatov ter vprašljivo povezljivost podatkov in ocenjevanje njihovih podatkovnih zbirk (Pesek idr., 2025). Z metodologijo (Pesek idr. 2025) si tako nadalje prizadevamo ustvariti enovito metodo za ocenjevanje podatkovnih zbirk in odprtih podatkov, ki nima vrednosti zgolj v lokalnem, slovenskem okolju, temveč tudi v širšem, evropskem podatkovnem prostoru.

V sledečem poglavju se bomo najprej poglobili v spletni portal OPSI, kjer podajamo njegov oris in navajamo svoje ugotovitve ročnega pregleda ter izpostavljamo pomanjkljivosti portala, ki jih bomo naslovili z metodologijo, predstavljeno v tretjem poglavju, v zadnjem poglavju pa bomo povzeli obravnavane problematike in se

usmerili v prihodnost odprtih podatkov, evropskih podatkovnih portalov in podatkovne krajine nasploh.

2 Ročni pregled na portalu OPSI

2.1 Portal OPSI

Nacionalni portal odprtih podatkov Slovenije (OPSI)¹, ki ga lahko vidimo na Slika 1, predstavlja enotno nacionalno spletno točko za objavo odprtih podatkov za celotni javni sektor, portal pa je vzpostavilo Ministrstvo za javno upravo (O portalu OPSI). Tako kot angleški portal odprtih podatkov (Data GOV UK)² ter evropski portal odprtih podatkov (Data Europa EU)³ je tudi OPSI zgrajen na odprtokodni programski opremi (O portalu OPSI). Pred vzpostavitvijo portala so se določene odprte zbirke objavljale na portalu Nacionalnega interoperabilnostnega okvira (NIO)⁴, zaradi specifične strukture objav in predvsem usmerjenosti na uporabnike pa je bilo po vzoru večine držav potrebno vzpostaviti namenski portal odprtih podatkov, kjer so na voljo metapodatkovni opisi vseh zbirk javnega sektorja ter zbirke odprtih podatkov (O portalu OPSI). Tako lahko opredelimo vlogo portala OPSI z dvema funkcijama (ZDIJZ-LP, 2017; O portalu OPSI): 1) gre za centralni katalog evidenc in zbirk podatkov v državi oziroma za centralni popis metapodatkov o vseh evidencah in podatkovnih zbirkah, ki jih vodijo državni organi, občine in drugi organi javnega sektorja; 2) gre za enotno spletno mesto za objavo podatkov iz zbirk v odprtih in strojno berljivih formatih, v primeru, da je specifična zbirka v odprtih formatih že objavljena na drugem spletnem mestu, pa je na portalu OPSI zagotovljena spletna povezava do te strani. Sočasno je tudi v skladu z EU Direktivo o ponovni uporabi podatkov javnega sektorja in zakonodaje o dostopu do informacij javnega značaja (ZDIJZ-E, 2015; ZDIJZ-G, 2022; Pesek idr., 2025).

Portal OPSI zastopa Ministrstvo za javno upravo, ki je sistemsko odgovorno za področje dostopa do informacij javnega značaja, pri čemer za tehnični vidik portala skrbi Ministrstvo za digitalno preobrazbo (O portalu OPSI). Ekipo OPSI sestavljajo sodelavci oziroma zaposleni z obeh navedenih ministrstev, ki skrbijo za delovanje

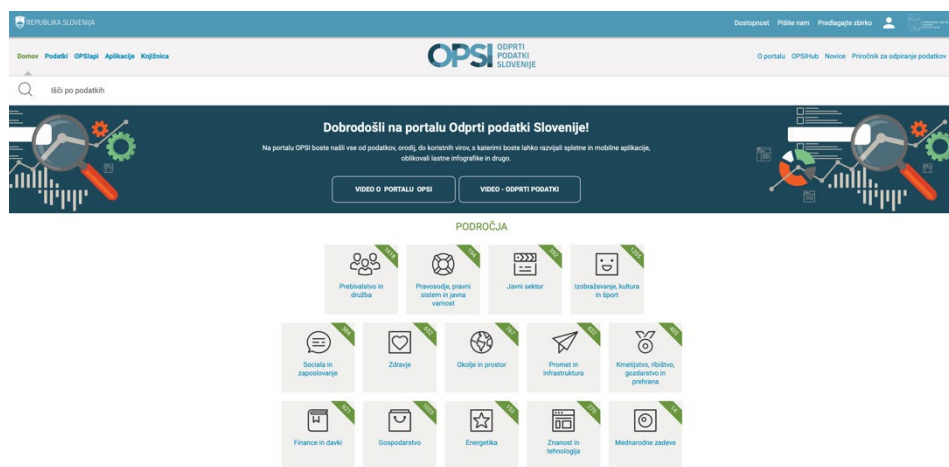
¹ <https://podatki.gov.si/>.

² <https://www.data.gov.uk/>.

³ <https://data.europa.eu/>.

⁴ <https://nio.gov.si>.

portala, kakovost objavljenih podatkov na portalu, mrežo urednikov, komuniciranje in sodelovanje z uporabniki in obiskovalci portala (O portalu OPSI). Ob tem ekipa komunicira z državljani in podjetji, organizira in sodeluje na številnih dogodkih, kot so FOP⁵, konference, predavanja, hackatoni, ti pa so povezani z odprtimi podatki javnega sektorja v Sloveniji in širše na področju Evropske unije (O portalu OPSI). Cilj portala OPSI je, da Slovenija postane primer dobre prakse na področju odpiranja podatkov in uporabe javnih podatkov, hkrati pa si prizadeva za okrepljeno sodelovanje z vsemi deležniki (občine, javni sektor, gospodarstvo, nevladne organizacije, raziskovalci, start-up podjetja, univerze in srednje šole) (O portalu OPSI). Namen sodelovanja z različnimi deležniki pa je, da se odprte podatke v čim večji meri uporabi za izdelavo aplikacij, pri čemer skrbniki portala tako uporabnike močno spodbujajo, naj podatke pridobijo in jih ponovno uporabijo pri izdelavi novih produktov ali storitev z dodano vrednostjo, ki bodo služili dobremu servisu in storitvam za državljane in podjetja (Prenova OPSI, 2016; O portalu OPSI).



Slika 1: Začetna stran spletnega portala OPSI in podatkovna področja, 17. marca 2025.

Vir: zajem zaslona, <https://podatki.gov.si/>.

Podatki na portalu OPSI so dosegljivi v strojno berljivi obliki, ki omogoča izdelavo novih analiz, aplikacij in vizualizacij, so dostopni z odprto licenco in se lahko prosto ponovno uporabljajo, kar pomeni, da jih lahko vsakdo brez posebne zahteve ali

⁵ Festival odprtih podatkov javnega sektorja.

registracije pridobi in ponovno uporabi ter na podlagi le-teh izdelava že omenjene produkte ali storitve z dodano vrednostjo (O portalu OPSI).

Na portalu po posamičnih področjih objavljajo podatke vsi organi javnega sektorja, sistem objav pa deluje na podlagi uredniškega sistema – ki sestoji iz glavnih urednikov, področnih urednikov in vsebinskih urednikov – pri čemer podatke oziroma zbirke ali nove izdelke, ki nastajajo v procesu ponovne uporabe, lahko objavijo tudi zunanji deležniki oziroma uporabniki zbirk podatkov (O portalu OPSI).

2.2 Ročni pregled

Naš ročni pregled je potekal od oktobra 2023 do februarja 2024 (Pesek idr., 2025). Primer ene izmed ročno pregledanih zbirk lahko najdemo na sledečih slikah, ki so razdeljene po izbranih kategorijah. Spodnje podatkovne zbirke, prikazane na slikah, denimo izhajajo s podatkovnega področja Izobraževanja, kulture in športa. Ob koncu pregleda je bilo na portalu objavljenih okoli 3.640 zbirk, v času pisanja članka, 4. marca 2025, pa jih je 5.338.

Na Slika 2 lahko vidimo navedena imena podatkovnih zbirk, ki smo jih ročno pregledali, z dodano hiperpovezavo do spletne strani, kjer se nahaja, in datumom ročnega pregleda, ko so bile zbirke ročno pregledane. V samem Excel dokumentu, znotraj katerega je potekal vpis podatkov ročnega pregleda, smo podatkovne zbirke porazdelili po posamičnih kategorijah, ki ustrezajo njihovemu področju.

Na Slika 3 lahko najprej vidimo navedene formate podatkovnih zbirk, ki smo jih ročno pregledali, in popolnost podatkovnih zbirk, pri čemer smo za merilo vzeli subjektivno oceno stanja podatkovne zbirke: če smo označili, da je popolnost podatkovne zbirke navedena z odgovorom: 1) DA, pomeni, da so prisotni vsi podatki in zato podatkovna zbirka velja kot popolna; 2) POMANKLJIVA, pomeni, da so manjkali podatki v manjšem številu (ena ali več); 3) NE, pomeni, da je manjkalo veliko podatkov in zato podatkovna zbirka velja kot nepopolna; 4) znak poševnice »/«, pomeni, da zbirka ne vsebuje podatkov oziroma so prisotne zgolj kategorizacije oziroma identifikatorji podatkov, ki sami na sebi ne povedo ničesar, in zato podatkovna zbirka velja kot nepopolna.

IME ZBIRKE	SPLETNA POVEZAVA	DATUM
Register nepremične kulturne dediščine	https://podatki.gov.si/dataset/register-nepremicne-kulturne-dediscine	1.12.23
Register nepremične kulturne dediščine	https://podatki.gov.si/dataset/register-nepremicne-kulturne-dediscine	1.12.23
Register nepremične kulturne dediščine	https://podatki.gov.si/dataset/register-nepremicne-kulturne-dediscine	1.12.23
Register nepremične kulturne dediščine	https://podatki.gov.si/dataset/register-nepremicne-kulturne-dediscine	1.12.23
Razvid športnih objektov	https://podatki.gov.si/data/search?open_data=True&all_podrocje=Izobrazevanje.%20kultura%20in%20sport	1.12.23
Izbor iz registra nepremične kulturne dediščine	https://podatki.gov.si/dataset/izbor-iz-registra-nepremicne-kulturne-dediscine	1.12.23
Izbor iz registra nepremične kulturne dediščine	https://podatki.gov.si/dataset/izbor-iz-registra-nepremicne-kulturne-dediscine	1.12.23
Število novih doktorjev in doktoric znanosti, povprečna starost ob doktoriranju in povprečen čas za dokončanje študija po področju znanosti doktorata ter spolu, Slovenija, večletno	https://podatki.gov.si/dataset/surs2367004s	1.12.23
Evidenca javnih zavodov na področju kulture	https://podatki.gov.si/dataset/evidenca-javnih-zavodov-s-podrocja-kulture	1.12.23
Evidenca javnih zavodov na področju kulture	https://podatki.gov.si/dataset/evidenca-javnih-zavodov-s-podrocja-kulture	1.12.23
POHODNIŠKE POTI OBČINA GROSUPLJE	https://podatki.gov.si/dataset/pohodniske-poti-obcina-grosuplje	1.12.23
POHODNIŠKE POTI OBČINA GROSUPLJE	https://podatki.gov.si/dataset/pohodniske-poti-obcina-grosuplje	1.12.23
Vrtci in osnovne šole v Mestni občini Koper	https://podatki.gov.si/dataset/vrtci-osnovne-sole-v-mestni-obcini-koper	1.12.23
Vrtci in osnovne šole v Mestni občini Koper	https://podatki.gov.si/dataset/vrtci-osnovne-sole-v-mestni-obcini-koper	1.12.23
Seznam osnovnih šol v Mestni občini Novo mesto	https://podatki.gov.si/dataset/seznam-osnovnih-sol-v-mestni-obcini-novo-mesto	1.12.23
Seznam osnovnih šol v Mestni občini Novo mesto	https://podatki.gov.si/dataset/seznam-osnovnih-sol-v-mestni-obcini-novo-mesto	1.12.23
POHODNE POTI PO OBČINI IVANČNA GORICA	https://podatki.gov.si/dataset/pohodne-poti-po-obcini-ivancna-gorica	1.12.23
POHODNE POTI PO OBČINI IVANČNA GORICA	https://podatki.gov.si/dataset/pohodne-poti-po-obcini-ivancna-gorica	1.12.23
NAMESTITVENE ZMOGLJIVOSTI OBČINA IVANČNA GORICA	https://podatki.gov.si/dataset/namestitvene-zmogljivosti-obcina-ivancna-gorica	1.12.23
NAMESTITVENE ZMOGLJIVOSTI OBČINA IVANČNA GORICA	https://podatki.gov.si/dataset/namestitvene-zmogljivosti-obcina-ivancna-gorica	1.12.23
Občina Rogaška Slatina - pohodniške poti	https://podatki.gov.si/dataset/obcina-rogaska-slatina-pohodniske-poti	1.12.23
Občina Rogaška Slatina - pohodniške poti	https://podatki.gov.si/dataset/obcina-rogaska-slatina-pohodniske-poti	1.12.23
Pohodniške poti v Občini Škofljica	https://podatki.gov.si/dataset/pohodniske-poti-v-obcini-skofljica	1.12.23
Pohodniške poti v Občini Škofljica	https://podatki.gov.si/dataset/pohodniske-poti-v-obcini-skofljica	1.12.23
Register kulturne dediščine v Občini Žalec	https://podatki.gov.si/dataset/register-kulturne-dediscine-v-obcini-zalec3	1.12.23

Slika 2: Ime, spletna povezava in datum dostopa do podatkovnih zbirk v Excel dokumentu s podatki ročnega pregleda.

Vir: lasten.

Naslednji navedbi v stolpcih vključujeta pravilnost kategorizacije – ki se nanaša na pravilnost, jasnost in smiselnost zapisa ter pomena poimenovanja stolpcev – in zanesljivost, pri čemer smo za merilo prav tako vzeli subjektivno oceno stanja podatkovne zbirke. Če smo označili, da je zanesljivost podatkovne zbirke navedena z odgovorom: 1) DA, pomeni, da je (vsaj približno) popolna, razumljiva in jasna v zadostni meri in zato podatkovna zbirka velja kot zanesljiva; 2) NE, pomeni, da zbirka ni (niti približno) popolna, razumljiva in jasna v zadostni meri in zato

podatkovna zbirka velja kot nezanesljiva; 3) znak poševnice »/«, pomeni, da podatkovna zbirka ne vsebuje podatkov oziroma so prisotne zgolj kategorizacije oziroma identifikatorji podatkov, ki sami na sebi ne povedo ničesar, in zato podatkovna zbirka velja kot nezanesljiva.

FORMAT	POPOLNOST	PRAVLNOST KATEGORIZACIJE	ZANESLJIVOST
XLS	POMANJKLJIVO	DA	NE
GPKG	POMANJKLJIVO	DA	NE
GML	POMANJKLJIVO	DA	NE
ZIP	POMANJKLJIVO	DA	NE
CSV	POMANJKLJIVO	DA	NE
ZIP	POMANJKLJIVO	DA	NE
CSV	POMANJKLJIVO	DA	NE
PCAXIS	POMANJKLJIVO	DA	NE
DOCX	POMANJKLJIVO	DA	DA
CSV	POMANJKLJIVO	DA	DA
CSV	DA	DA	DA
XLSX	DA	DA	DA
GEOJSON	/	/	/
CSV	/	/	/
CSV	DA	DA	DA
XLSX	DA	DA	DA
CSV	DA	DA	DA
XLSX	DA	DA	DA
CSV	POMANJKLJIVO	DA	DA
XLSX	POMANJKLJIVO	DA	DA
CSV	POMANJKLJIVO	DA	NE
XLS	POMANJKLJIVO	DA	NE
CSV	POMANJKLJIVO	DA	NE
XLSX	POMANJKLJIVO	DA	NE
CSV	POMANJKLJIVO	DA	NE
XLSX	POMANJKLJIVO	NE: podatkov sploh ni, je zelo neurejena tabela	NE

Slika 3: Format, popolnost, pravilnost kategorizacije in zanesljivost podatkov v podatkovnih zbirkah v Excel dokumentu s podatki ročnega pregleda.

Vir: lasten.

Na Slika 4 lahko vidimo kategorije uparljivosti ali spremenljivosti ali povezljivosti podatkov, pri kateri smo ocenjevali, v kolikor imajo podatki znotraj danega stolpca podatke, ki so v obliki: 1) šifranta (npr. šifrant občine Domžale je 1230); 2) besedila (npr. besedilo občine Domžale je Domžale); 3) obojega (npr. šifrant in besedilo občine Domžale se pojavlja v obliki 1230 Domžale); 4) ne obstajajo ali 5) niso navedeni, pri čemer smo uporabili znak poševnice »/«.

Sledi kategorija potrebe po naknadnem urejanju oziroma popravljanju podatkov podatkovne zbirke, pri kateri smo ocenjevali, v kolikor je potrebno podatkovno zbirko ponovno oceniti in spreminjati njene značilnosti, pri čemer smo podali odgovore: 1) DA; 2) MOGOČE; 3) DA in 4) znak poševnice »/«, če podatkovna zbirka ni vsebovala podatkov oziroma so prisotne zgolj kategorizacije oziroma identifikatorji podatkov, ki sami na sebi ne povedo ničesar, in zato podatkovna zbirka velja kot nepopravljiva.

Predzadnja kategorija se nanaša na podatke o obstoju, pri čemer smo obravnavali časovni obstoj podatkov, ki je bil bodisi naveden bodisi ne. Zadnja obravnavana kategorija pa je vezana na druge opombe oziroma komentarje, ki so se pojavili tekom pregledovanja podatkovnih zbirk in njihovih podatkov, ti vključujejo: 1) navedbo avtorjev, 2) zaznane formate, 3) pojavljanje posebnih znakov v tabelah in 4) druge posebnosti.

Daleč največ podatkovnih zbirk (O portalu OPSI) je prisotnih na področju Prebivalstva in družbe (1.237), takoj zatem na področju Gospodarstva (923), Izobraževanja, kulture in športa (774), Financ in davkov (454), Okolja in prostora (391), Prometa in infrastrukture (370), Kmetijstva, ribištva, gozdarstva in prehrane (352), Znanosti in tehnologije (246), Sociale in zaposlovanja (241), Zdravja (127), Javnega sektorja (118), Energetike (59), Pravosodja, pravnega sistema in javne varnosti (40) ter Mednarodnih zadev (6).

V spodnjem podpoglavju navajamo številne problematike, na katere smo naleteli tekom ročnega pregleda podatkovnih zbirk na portalu OPSI, utemeljene na podlagi specifičnih primerov.

UPARLJIVOST/SPREMENLJIVOST/POVEZLJIVOST (ŠIFRANT/BESEDILO)	POTREBA PO NAKNADNEM UREJANJU/POPRAVLJIVOSTU ANJU?	PODATKI O OBSTOJU	OPOMBE
BESEDILO (Občina)	MOGOČE	/	Niso navedeni vsi avtorji
BESEDILO (Občina)	MOGOČE	/	Niso navedeni vsi avtorji
BESEDILO (Občina)	MOGOČE	/	Niso navedeni vsi avtorji
BESEDILO (Občina)	MOGOČE	/	Niso navedeni vsi avtorji
BESEDILO (Občina)	MOGOČE	/	Obstajata 2 CSV datoteki.
BESEDILO (Občina)	MOGOČE	/	Niso navedeni vsi avtorji
BESEDILO (Občina)	MOGOČE	/	Niso navedeni vsi avtorji
NE	DA	Podatki obstajajo za 2009, 2012 in 2020	Podatki manjkajo; pojavlja se "-" in neutemeljeni črki M ter N
BESEDILO + ŠIFRANT (Občina)	MOGOČE	Podatki obstajajo za 2023	En podatek manjka; pojavlja se "-"
BESEDILO + ŠIFRANT (Občina)	MOGOČE	Podatki obstajajo za 2023	En podatek manjka; pojavlja se "-"
NE	NE	/	
NE	NE	/	Zaznani format: XLS
/	/	/	Zaznani format: JSON
/	/	/	
NE	NE	/	
NE	NE	/	Zaznani format: XLS
NE	NE	/	
NE	NE	/	Zaznani format: XLS
ŠIFRANT (Občina)	NE	/	Niso navedeni vsi izvajalci in kontaktni podatki
ŠIFRANT (Občina)	NE	/	Zaznani format: XLS
NE	DA	/	Niso navedene vse dolžine poti
NE	DA	/	Niso navedene vse dolžine poti
NE	DA	/	Niso navedena naselja in potek poti
NE	DA	/	Niso navedena naselja in potek poti
BESEDILO (Občina)	DA	/	Manjkajo podvrsti in avtorji

Slika 4: Uparljivost ali spremenljivost ali povezljivost, potreba po naknadnem urejanju oziroma popravljanju, podatki o obstoju in opombe za podatkovne zbirke v Excel dokumentu s podatki ročnega pregleda.

Vir: lasten.

2.3 Ugotovitve ročnega pregleda

2.3.1 Nestandardiziranost

Prvotni problem, s katerim smo se srečali pri ročnem pregledu, razvidnem s Slika 2, je raznolikost formatov, v katerih se podatkovne zbirke pojavljajo. V idealni situaciji si želimo enoten standardiziran format, ki bi nam omogočal enostavno obdelavo podatkov, srečujemo pa se z nekaterimi standardnimi formati: PCAXIS, CSV, XLS, HTML, ZIP, XML, KML, GeoJSON, (PDF), JSON, GML, GPKG (tudi gpkg), KMZ, DOC, TXT, SHP, ODT, RDF ter CVS. Načeloma se ti med seboj razlikujejo predvsem po tipu vsebovanih podatkov (npr. geografske koordinate zahtevajo poseben tip shranjevanja), na spletni strani OPSI pa je trenutno največja zastopanost formata PCAXIS (4.438 podatkovnih zbirk), ki je specifičen format, uporabljen zgolj v Sloveniji na SURS-u (Statistični urad Republike Slovenije) ter še v nekaterih drugih državah v Evropi in izven nje⁶. Za interpretiranje formata PCAXIS obstaja omejen nabor orodij, vendar ta niso dostopna na vseh platformah. Program je tako zasnovan zgolj za operacijski sistem Windows, kar pomeni, da ni dostopen na drugih operacijskih platformah, kot so macOS, Linux itd., kar ne predstavlja idealnih pogojev za pregledovanje podatkovnih zbirk na napravah, ki nimajo nameščenega operacijskega sistema Windows.

Umrli po dnevu smrti, Slovenija, mesečno

V tej zbirki so samodejno zajeti metapodatki, katerih vir je VLADA REPUBLIKE SLOVENIJE STATISTIČNI URAD REPUBLIKE SLOVENIJE in ustrezajo izvorni zbirki z naslovom "Umrli po dnevu smrti, Slovenija, mesečno".

Dejanski podatki so na voljo v formatu PC-Axis (.px). Med dodatnimi povezavami lahko dostopate do strani izvirnega portala za vpogled in izbor podatkov, na voljo pa je tudi program PX-Win, ki si ga lahko brezplačno prenesete. Oba omogočata izbor podatkov za prikaz, spreminjanje oblike izpisa in shranjevanje v različne formate, poleg tega pa tudi pregledovanje in izpis tabel neomejene velikosti ter nekaj osnovnih statističnih analiz in grafičnih prikazov.

Podatkovni viri (1)


Umrli po dnevu smrti, Slovenija, mesečno

 Več ...

 Prenos (49.1 KB)

Dodatne povezave (4)


Izbor podatkov

 Več ...

 Povezava

Slika 5: PCAXIS format podatkovne zbirke »Umrli po dnevu smrti, Slovenija, mesečno« in dostop do nje na spletnem portalu OPSI, 17. marec 2025.

Vir: <https://podatki.gov.si/dataset/surs051018s>.

⁶ <https://www.unecce.org/fileadmin/DAM/stats/documents/ece/ces/ac.61/2004/wp.36.s.e.pdf>.

Na Slika 5 vidimo primer zbirke, ki obstaja zgolj v formatu PCAXIS. V primeru da naša naprava nima nameščenega operacijskega sistema Windows, moramo pri dodatnih povezavah pritisniti na gumb »Povezava«, nakar izberemo podatke različnih kategorij za prikaz.

2.3.2 Nepopolnost, nejasnost in nerazumljivost podatkov

Pri obdelavi in analizi podatkov si želimo predvsem, da bi bili podatki čim popolnejši, podatkovni tipi čim natančneje določeni in celotna podatkovna struktura karseda smiselno zastavljena. V mislih imamo predvsem to, da bi v vsaki podatkovni tabeli bilo čim manj manjkajočih vrednosti ter da so vsi stolpci in vrstice informativni, to pa pomeni, da ne smemo podvajati vrstic, stolpci pa ne smejo biti v preveliki korelaciji (npr. leto rojstva in starost). Nasproten primer, torej ko je korelacija dobrodošla, predstavljajo podatki, ki so med seboj nujno povezani, a ne predstavljajo istih podatkov, zato potrebujemo več vrstic ali stolpcev. Želimo si tudi, da so podatki enostavno označeni, razumljivi in hkrati dovolj natančno opišejo vsebino, ki jo predstavljajo. Kot primer si lahko vzamemo kategorijo EMŠO, ki je navadno navedena na vrhu stolpca, pri čemer je zgoraj napisana identifikacijska številka, kar je nejasno, preširoko, posledično pa ima lahko več kot zgolj enojni pomen. Zapis mora biti enostaven in zadostno specifičen na način, da jasno opisuje podatke, ki jih naslavlja. V vsakem stolpcu mora biti določen tip podatkov, vsi podatki v stolpcu pa se morajo ujemati z njim.

2.3.3 Nesmiselnost podatkov

Želimo tudi, da je podatkovni tip smiselno in ustrezno široko zastavljen. Kot primer je »datum rojstva«, ki se denimo navadno zapisuje s številkami, ne pa z nizi črk, tako pa posledično prihaja do napak pri obdelavi podatkov. Če dovoljujemo zapis z vsemi znaki (string), je rezultat ponovno neuporaben, pri čemer dobimo nestandardiziran in nedosleden podatkovni tip. V stolpcih, kjer se vsaki vrstici priredi kategorično oznako, na primer »spok«, morajo biti vse vrednosti iz določene končne množice vrednosti, na primer »moški«, »ženska«, »neopredeljen«, ne pa nedoločene vrednosti, ki ne spadajo v to končno množico vrednosti, na primer »neznano«. Pod doslednost podatkovnih tipov spada tudi problem praznih polj v tabeli, saj je pri določenih stolpcih nesprejemljivo, da vsebujejo prazna polja, kar s težavo določimo, saj se pogosto zgodi, da ne vemo, če je polje dejansko prazno ali se v polju nahaja zgolj

napaka v zapisu. Ker ne moremo vseh praznih številskih polj nadomestiti s številom 0 zaradi neposrednega vpliva na interpretacijo podatkov, je potrebno, da se problematike zavedamo in jo poskusimo razrešiti, kjer je na podlagi lastne presoje to smiselno.

2.3.4 Neveljavnost podatkov

Pri številskih podatkih želimo, kjer je to smiselno, da vrednosti obstajajo znotraj določenega intervala, na primer »prihodek gospodinjstva« ne sme biti negativna vrednost, »število otrok« pa ne sme biti decimalno število.

2.3.5 Neunikatnost podatkov

Določeni stolpci v tabelah vsebujejo unikatne podatke, kot so »EMŠO«, »davčna številka« in tako dalje, zato si želimo, da se podatki ne bi po nepotrebnem podvajali oziroma ponavljali. Problematika je neposredno vezana tudi na nerazumljivost podatkov, saj je ta znatno povečana z navajanjem dvojnih podatkov – ki predstavljajo isti pojem – oziroma njihovim podvajanjem.

2.3.6 Nedoslednost in nestrukturiranost podatkovnih zbirk in podatkov

Da bi podatkovne zbirke postale doslednejše oziroma med seboj po strukturi bolj podobne, potrebujejo jasno določen okvir in merila. Znotraj podatkovnih zbirk povzročajo težave predvsem stolpci, kjer poimenovanja in ključi za identifikacijo v posamičnih stolpcih pogosto nejasno označujejo to, kar predstavljajo. S tem mislimo predvsem na napake v zapisu (manjkajoče črke in posledično napačne črke, manjkajoči šumniki, številčni znaki in razni drugi posebni znaki (npr. črka »è« kot nadomestek za črko »č«; uporaba znaka podčrtaja »_« med besedami namesto presledkov; »_id« namesto drugega poimenovanja, npr. »št.«), neuporaba velikih začetnic oziroma uporaba zgolj malih črk (npr. opomba) ali uporaba zgolj velikih črk (npr. DATUM) ter samo zaporedje posamičnih stolpcev. Nedoslednost se pojavi tudi na ravni naslovov podatkovnih zbirk in datotek, pri čemer pri naslovih opuščajo velike začetnice (npr. turistična društva) ali pa uporabljajo zgolj velike črke (npr. VOLITVE POSLANCEV IZ RS V EVROPSKI PARLAMENT). Idealna zbirka naj bi tako na zunanji ravni vključevala jasno kategorizacijo – znotraj katere najdemo posamične podatkovne zbirke, razdeljene denimo po letnicah – jasen in ustrezen

naslov, delujoče, razumljive in strojno berljive datoteke, metapodatke in omogočen prost dostop, na notranji ravni oziroma znotraj tabel pa jasno oznako podatkov brez manjkajočih črk, ki jih nadomestijo drugi, posebni znaki, ter zapolnitev manjkajočih podatkov.

2.3.7 Povezljivost, uparljivost ali interoperabilnost podatkovnih zbirk in podatkov

Pereča je tudi problematika v zvezi z uparljivostjo posamičnih podatkovnih zbirk in njihovih podatkov, ki so neposredno vezani na doslednost in strukturiranost. Pogosto smo ob pregledu naleteli na napake v zapisu, pri čemer podatki izgubijo rdečo nit medsebojnega povezovanja oziroma uparljivosti, s čimer je tudi okrnjena. Da bi med seboj povezovali notranje ravni podatkovnih zbirk (podatke znotraj tabel), moramo najprej zagotoviti povezljivost podatkov na zunanji ravni (podatkovne zbirke znotraj posamičnih gesel oziroma naslovov), predlogi za izboljšanje pa so predstavljeni v prejšnji problematiki. Način, kako bi dodatno izboljšali povezljivost podatkovnih zbirk, je, da bi več podatkovnih zbirk združili znotraj enega gesla oziroma naslova. Z implementacijo predlogov na ravni povezljivosti oziroma uparljivosti bi tako pripomogli k enostavnejšemu in učinkovitejšemu iskanju po samem podatkovnem portalu in podatkovnih zbirkah ter njihovih podatkih.

2.3.8 Petzvezdični sistem za ocenjevanje

Aktualna je tudi problematika, ki se nanaša na Petzvezdični sistem za ocenjevanje, ki je poglobljen pri povezovanju odprtih podatkov, saj vrednoti formate ter dostopnost podatkov. Sistem vrednoti izključno odprtost podatkovnih zbirk, posledično pa vsaka podatkovna zbirka, kot je razvidno s Slika 6, samodejno prejme zgolj tri izmed pet zvezdic (z izjemo ene petzvezdične podatkovne zbirke), saj izpolnjuje zgolj prve tri pogoje, kar močno omeji ponovno uporabljivost za bodoče uporabnike. Da bi zvišali samodejno oceno vseh podatkovnih zbirk, moramo k vsaki zbirki dodati ustrezne manjkajoče informacije, da bi uspešno izpolnjevale vseh pet pogojev (Žnideršič idr., 2024):

1. podatki so dostopni na spletu v formatu z odprto licenco,

2. podatki so na voljo v strojno berljivi obliki (npr. tekstovna datoteka namesto digitalno preslikane razpredelnice),
3. podatki so dostopni v nelastniškem formatu (npr. CSV namesto Excel),
4. za identificiranje so uporabljeni odprti standardi organizacije W3C (RDF in SPARQL) ter URI,
5. podatki so za namen kontekstualizacije povezani z ostalimi dostopnimi podatki.

OCENA ODPRTOSTI
★★★★★ (1)
★★★★☆ (5320)

Slika 6: Ocena odprtosti vseh zbirk v razdelku »Podatki« na spletnem portalu OPSI, 17. marec 2025.

Vir: zajem zaslona, https://podatki.gov.si/data/search?open_data=true.

3 Metodologija

Proces izdelave, zasnova metodologije in definicije dimenzij ter njihovih metrik so bili že opisani v naši prejšnji študiji (Pesek idr., 2025), v tej študiji pa bomo definicije nadgradili z natančnejšimi usmeritvami in predstavitvijo tabel za posamične dimenzije z njihovimi metrikami.

3.1 Definicije dimenzij in njihovih metrik

Kot smo nakazali že v prejšnji študiji (Pesek idr., 2025), naša tabela, ki jo najdete v prilogi, sestoji iz gradnikov, kot so dimenzije, metrike, spremenljivke, formule s spremenljivkami metrik ter merila rezultatov metrik – prilagojena posamičnim spremenljivkam. Tako bomo predstavili več posamičnih tabel, ki prikazujejo metrike posamičnih dimenzij.

3.11 Popolnost

Dimenzija, ki določa, do katere stopnje podatki vsebujejo vrednosti za vse pričakovane značilnosti, je razvidna iz Tabela 1.

3.1.1.1 Odstotek popolnih celic

Metrika je izračunana na podlagi števila vrstic (v), števila stolpcev (s), števila nepopolnih celic (cn) in števila vseh celic (c).

3.1.1.2 Odstotek popolnih vrstic

Metrika je izračunana na podlagi števila vrstic (v) in števila nepopolnih vrstic (vn).

3.1.1.3 Odstotek popolnih stolpcev

Metrika je izračunana na podlagi števila stolpcev (s) in števila nepopolnih stolpcev (sn).

Tabela 1: Metrike dimenzije Popolnost.

Dimenzija	Metrika (m)	Spremenljivke	Formula	Merilo
Popolnost	Odstotek popolnih celic (Vetrò idr., 2016)	v : število vrstic, s : število stolpcev, cn : število nepopolnih celic, c : število celic	$c = v \cdot s;$ $m_1 = \left(1 - \frac{cn}{c}\right) \cdot 100$	[0 %, 100 %]
	Odstotek popolnih vrstic (Vetrò idr., 2016)	v : število vrstic, vn : število nepopolnih vrstic	$m_2 = \left(1 - \frac{vn}{v}\right) \cdot 100$	[0 %, 100 %]
	Odstotek popolnih stolpcev	s : število stolpcev, sn : število nepopolnih stolpcev	$m_3 = \left(1 - \frac{sn}{s}\right) \cdot 100$	[0 %, 100 %]
	Odstotek popolnosti metapodatkov v podatkovnih zbirkah	m : število metapodatkov, mn : število nepopolnih metapodatkov	$m_4 = \left(1 - \frac{mn}{m}\right) \cdot 100$	[0 %, 100 %]
	Odstotek predvidenih vrednosti v celicah	vr : število vrednosti, vrn : število naključnih vrednosti (črke, številke, simboli itd.)	$m_5 = \left(1 - \frac{vrn}{vr}\right) \cdot 100$	[0 %, 100 %]

Vir: lasten; osnovna različica je objavljena v študiji Pesek idr., 2025.

3.1.1.4 Odstotek popolnosti metapodatkov v podatkovnih zbirkah

Metrika je izračunana na podlagi števila metapodatkov (m) in števila nepopolnih metapodatkov (mn).

3.1.1.5 Odstotek predvidenih vrednosti v celicah

Metrika označuje odsotnost naključnih vrednosti v celicah znotraj podatkovnih zbirk. Spremenljivki za izračun metrike sta število vrednosti (vr) in število naključnih vrednosti (vrn), kot so naključne črke, številke, posebni znaki itd., pri čemer vsakršno odstopanje znotraj celice velja kot naključna vrednost.

3.1.2 Unikatnost

Dimenzija, ki določa, do katere stopnje so podatki edinstveni, je razvidna iz Tabela 2.

3.1.2.1 Odstotek nepodvojenih vrstic

Metrika je izračunana na podlagi števila podvojenih vrstic (vp) in števila vrstic (v).

3.1.2.2 Odstotek nepodvojenih vrednosti v celicah

Metrika je izračunana na podlagi števila podvojenih vrednosti (pvr) in števila vrednosti (vr).

3.1.2.3 Pearsonov korelacijski koeficient

Metrika opisuje, do katere stopnje sta v podatkovnih zbirkah povezana dva različna stolpca. Spremenljivki za izračun metrike sta izbrani stolpec (x) in stolpec za primerjavo (y). Da bi izračunali Pearsonov korelacijski koeficient, kovarianco stolpcev delimo s standardnima odklonoma stolpcev. Merilo je izraženo z vrednostjo od -1 do vključno 1.

Tabela 2: Metrike dimenzije Unikatnost.

Dimenzija	Metrika (m)	Spremenljivke	Formula	Merilo
Unikatnost	Odstotek nepodvojenih vrstic (Vetrò idr., 2016; Neumaier idr., 2016)	vp: število podvojenih vrstic, v: število vrstic	$m_6 = \left(1 - \frac{vp}{v}\right) \cdot 100$	[0 %, 100 %]
	Odstotek nepodvojenih vrednosti v celicah	pvr: število podvojenih vrednosti, vr: število vrednosti	$m_7 = \left(1 - \frac{pvr}{vr}\right) \cdot 100$	[0 %, 100 %]
	Pearsonov korelacijski koeficient	x: izbrani stolpec, y: stolpec za primerjavo	$\rho_{x,y} = \frac{cov(X,Y)}{\sigma_x \sigma_y}$	[-1, 1]

Vir: lasten; osnovna različica je objavljena v študiji Pesek idr., 2025.

3.1.3 Doslednost

Dimenzija, ki določa, do katere stopnje se podatki pojavljajo v enaki obliki, je razvidna iz Tabela 3.

3.1.3.1 Odstotek pravilno poimenovanih stolpcev

Metrika predstavlja stolpce, ki so poimenovani brez slovničnih napak in s šifrantom oziroma besedilom ter je izračunana na podlagi števila nepričakovano poimenovanih stolpcev (spn) in števila stolpcev (s).

3.1.3.2 Odstotek doslednosti podatkovnih tipov

Metrika v procentih označuje doslednost uporabe podatkovnih tipov v podatkovnih zbirkah. Število nepričakovanih podatkovnih tipov, ki je poleg števila vseh uporabljenih podatkovnih tipov spremenljivka za izračun metrike, predstavlja podatkovne tipe, ki niso predvideni za področje določene podatkovne zbirke. Metriko tako izračunamo na podlagi števila podatkovnih tipov (pt) in števila nepričakovanih podatkovnih tipov (ptn).

Tabela 3: Metriki dimenzije Doslednost.

Dimenzija	Metrika (<i>m</i>)	Spremenljivke	Formula	Merilo
Doslednost	Odstotek pravilno poimenovanih stolpcev	s: število stolpcev, spn: število nepričakovano poimenovanih stolpcev	$m_8 = \left(1 - \frac{spn}{s}\right) \cdot 100$	[0 %, 100 %]
	Odstotek doslednosti podatkovnih tipov	pt: število podatkovnih tipov, ptn: število nepričakovanih podatkovnih tipov	$m_9 = \left(1 - \frac{ptn}{pt}\right) \cdot 100$	[0 %, 100 %]

Vir: lasten; osnovna različica je objavljena v študiji Pesek idr., 2025.

3.1.4 Uparljivost

Dimenzija, ki določa, do katere stopnje se podatki med seboj povezujejo, je razvidna iz Tabela 4.

3.1.4.1 Odstotek navedenih podatkovnih tipov

Metrika v procentih označuje prisotnost navedenih podatkovnih tipov v podatkovnih zbirkah. Število nenavedenih podatkovnih tipov označuje podatkovne tipe, kot so zaporedna številka, datum, ime občine itd., ki ne obstajajo znotraj podatkovne zbirke. Metriko tako izračunamo na podlagi števila podatkovnih tipov (pt) in števila nenavedenih podatkovnih tipov (ptnn).

3.1.4.2 Odstotek veljavnih podatkovnih tipov

Metrika v procentih označuje prisotnost veljavnih podatkovnih tipov v podatkovnih zbirkah. Število neveljavnih podatkovnih tipov označuje podatkovne tipe, kot so zaporedna številka, datum, ime občine itd., ki vsebujejo napake v zapisu znotraj podatkovne zbirke. Metriko tako izračunamo na podlagi števila podatkovnih tipov (pt) in števila neveljavnih podatkovnih tipov (ptnv).

3.1.4.3 Odstotek strojno berljivih podatkovnih zbirk

Metriko izračunamo na podlagi števila podatkovnih zbirk (pz) in števila strojno neberljivih podatkovnih zbirk (pzsn).

Tabela 4: Metrike dimenzije Uparljivost.

Dimenzija	Metrika (m)	Spremenljivke	Formula	Merilo
Uparljivost	Odstotek navedenih podatkovnih tipov	pt: število podatkovnih tipov, ptnn: število navedenih podatkovnih tipov	$m_{10} = \left(1 - \frac{ptnn}{pt}\right) \cdot 100$	[0 %, 100 %]
	Odstotek veljavnih podatkovnih tipov	pt: število podatkovnih tipov, ptnv: število neveljavnih podatkovnih tipov	$m_{11} = \left(1 - \frac{ptnv}{pt}\right) \cdot 100$	[0 %, 100 %]
	Odstotek strojno berljivih podatkovnih zbirk	pz: število podatkovnih zbirk, pzsn: število strojno neberljivih podatkovnih zbirk	$m_{12} = \left(1 - \frac{pzsn}{pz}\right) \cdot 100$	[0 %, 100 %]

Vir: lasten; osnovna različica je objavljena v študiji Pesek idr., 2025.

3.1.5 Ponovna uporabljivost

Dimenzija, ki določa, do katere stopnje so podatki bodočim uporabnikom na voljo za ponovno uporabo, je razvidna iz Tabela 5. Navedene metrike so zasnovane za uporabo na nivoju portala in ne za posamezne podatkovne zbirke.

3.1.5.1 Odstotek podatkovnih zbirk z licenco

Metriko izračunamo na podlagi števila podatkovnih zbirk (pz) in števila podatkovnih zbirk brez licence (pzbl).

3.1.5.2 Odstotek podatkovnih zbirk s kontaktno točko

Metrika v procentih označuje, do katere stopnje so v podatkovnih zbirkah prisotni kontaktni podatki v primeru, da ima uporabnik vprašanja glede podatkov in se želi obrniti na lastnika. Metriko tako izračunamo na podlagi števila podatkovnih zbirk (pz) in števila podatkovnih zbirk brez kontaktne točke (pzbkt).

3.1.5.3 Odstotek podatkovnih zbirk z informacijo o izdajatelju

Metriko izračunamo na podlagi števila podatkovnih zbirk (pz) in števila podatkovnih zbirk brez informacije o izdajatelju (pzbii).

Tabela 5: Metrike dimenzije Ponovna uporabljivost.

Dimenzija	Metrika (<i>m</i>)	Spremenljivke	Formula	Merilo
Ponovna uporabljivost (MQA)	Odstotek podatkovnih zbirk z licenco	pz: število podatkovnih zbirk, pzbil: število podatkovnih zbirk brez licence	$m_{13} = \left(1 - \frac{pzbil}{pz}\right) \cdot 100$	[0 %, 100 %]
	Odstotek podatkovnih zbirk s kontaktno točko	pz: število podatkovnih zbirk, pzbkt: število podatkovnih zbirk brez kontaktne točke	$m_{14} = \left(1 - \frac{pzbkt}{pz}\right) \cdot 100$	[0 %, 100 %]
	Odstotek podatkovnih zbirk z informacijo o izdajatelju	pz: število podatkovnih zbirk, pzbii: število podatkovnih zbirk brez informacije o izdajatelju	$m_{15} = \left(1 - \frac{pzbii}{pz}\right) \cdot 100$	[0 %, 100 %]

Vir: lasten; osnovna različica je objavljena v študiji Pesek idr., 2025.

3.1.6 Natančnost

Dimenzija, ki določa, do katere stopnje se podatki zanesljivo pojavljajo v pravilni obliki in ne dopuščajo napak, je razvidna iz Tabela 6.

3.1.6.1 Odstotek sintaktično natančnih celic

Metrika je izračunana na podlagi števila celic z napakami pri kodiranju (cnk), kot so pretvorba šumnikov in posebnih znakov, številke itd., in števila celic (c).

Tabela 6: Metrika dimenzije Natančnost.

Dimenzija	Metrika (<i>m</i>)	Spremenljivke	Formula	Merilo
Natančnost	Odstotek sintaktično natančnih celic (Vetrò idr., 2016)	cnk: število celic z napakami (šumniki, posebni znaki, številke itd.), c: število celic	$m_{16} = \left(1 - \frac{cnk}{c}\right) \cdot 100$	[0 %, 100 %]

Vir: lasten; osnovna različica je objavljena v študiji Pesek idr., 2025.

3.1.7 Razumljivost

Dimenzija, ki določa, do katere stopnje so podatki razumljivi in enostavni za uporabo, je razvidna iz Tabela 7.

3.1.7.1 Jasnost poimenovanja

Metrika v procentih ocenjuje, do katere subjektivne stopnje je v podatkovnih zbirkah prisotna jasnost poimenovanja. Spremenljivke, ki predstavljajo predpogoj oziroma so upoštevane za sistem točkovanja, so podatki (p), podatkovni tipi (pt) ter sintaktično natančna in semantično ustrezna pravila za poimenovanje (sspp). Da bi izračunali jasnost poimenovanja, moramo pri podatkih in podatkovnih tipih upoštevati sintaktično natančna in semantično ustrezna pravila za poimenovanje.

3.1.7.2 Odstotek stolpcev v razumljivem podatkovnem tipu

Metrika v procentih označuje prisotnost stolpcev v razumljivem podatkovnem tipu v podatkovnih zbirkah. Število nerazumljivih podatkovnih tipov, predstavlja podatkovne tipe, pri katerih na podlagi danega imena ni razumljivo označeno, katere podatke vključujejo. Metrika je tako izračunana na podlagi števila stolpcev v nerazumljivem podatkovnem tipu (snpt) in števila stolpcev (s).

Tabela 7: Metriki dimenzije Razumljivost.

Dimenzija	Metrika (<i>m</i>)	Spremenljivke	Formula	Merilo
Razumljivost	Jasnost poimenovanja	p: podatki, pt: podatkovni tipi, sspp: sintaktično natančna in semantično ustrezna pravila za poimenovanje	Uporaba NLP (Natural Language Processing) (Chowdhary idr., 2020) ali ročno preverjanje podatkov za dodelitev točk jasnosti.	[0 %, 100 %]
	Odstotek stolpcev v razumljivem podatkovnem tipu	snpt: število stolpcev v nerazumljivem podatkovnem tipu, s: število stolpcev	$m_{17} = \left(1 - \frac{snpt}{s}\right) \cdot 100$	[0 %, 100 %]

Vir: lasten; osnovna različica je objavljena v študiji Pesek idr., 2025.

3.1.8 Aktualnost

Dimenzija, ki določa, do katere stopnje so podatki objavljeni pravočasno, je razvidna iz Tabela 8.

3.1.8.1 Zamuda v objavi

Metrika opisuje, od kdaj do kdaj so bili določeni podatki dostopni, in je izračunana na podlagi datuma dostopnosti informacij (*ddi*), datuma objave (*do*), začetnega datuma obdobja obstoja podatkovne zbirke (*zd*) ter končnega datuma obdobja obstoja podatkovne zbirke (*kd*). Merilo je izraženo z vrednostjo od minus neskončno do vključno 1.

Tabela 8: Metrika dimenzije Aktualnost.

Dimenzija	Metrika (<i>m</i>)	Spremenljivke	Formula	Merilo
Aktualnost	Zamuda v objavi (Vetrò idr., 2016)	<i>ddi</i> : datum dostopnosti informacij, <i>do</i> : datum objave, <i>zd</i> : začetni datum obdobja obstoja podatkovne zbirke, <i>kd</i> : končni datum obdobja obstoja podatkovne zbirke	$ddi = kd + 1;$ $m_{18} = 1 - \frac{do - kd}{kd - zd}$	$(-\infty, 1]$

Vir: lasten; osnovna različica je objavljena v študiji Pesek idr., 2025.

3.1.9 Skladnost

Dimenzija, ki določa, do katere stopnje so podatki skladni z določenimi standardi, je razvidna iz Tabela 9.

3.1.9.1 Skladnost z eGMS:

Metrika opisuje, do katere mere so podatkovne zbirke skladne z e-GMS, in je izračunana na podlagi vira (*v*), datuma izdelave (*di*), kategorije (*k*), naslova (*n*), opisa (*o*), identifikatorja (*id*), izdajatelja (*iz*) in jezika (*j*). Merilo je izraženo z vrednostjo od 0 do vključno 8.

3.1.9.2 Petzvezdični odprti podatki

Metrika ocenjuje skladnost podatkovnih zbirk s pogoji petzvezdičnega sistema za ocenjevanje, pri čemer je vsaka podatkovna zbirka vrednotena s toliko zvezdicami, kot izpolnjuje pogojev. Pogoji so dostopnost podatkov v formatu z odprto licenco (dpol), podatki v strojno berljivi obliki (psb), dostopnost podatkov v nelastniškem formatu (dpnf), uporaba odprtih standardov organizacije W3C in URI za identifikacijo (uosi) ter kontekstualizacija podatkov in ostalih dostopnih podatkov (kp). Merilo je izraženo z vrednostjo od 0 do vključno 5.

Tabela 9: Metriki dimenzije Skladnost.

Dimenzija	Metrika (<i>m</i>)	Spremenljivke	Formula	Merilo
Skladnost	Skladnost z eGMS (Vetro idr., 2016)	v: vir, di: datum izdelave, k: kategorija, n: naslov, o: opis, id: identifikator, iz: izdajatelj, j: jezik	$m_{19} =$ $= v + di + k + n$ $+ o + id + iz + j$	[0, 8]
	Petzvezdični odprti podatki (Berners-Lee, 2006)	dpol: dostopnost podatkov v formatu z odprto licenco, psb: podatki v strojno berljivi obliki, dpnf: dostopnost podatkov v nelastniškem formatu, uosi: uporaba odprtih standardov organizacije W3C in URI za identifikacijo, kp: kontekstualizacija podatkov in ostalih dostopnih podatkov	$m_{20} =$ $= dpol + psb$ $+ dpnf + uosi + kp$	[0, 5]

Vir: lasten; osnovna različica je objavljena v študiji Pesek idr., 2025.

3.1.10 Sledljivost

Dimenzija, ki določa, do katere stopnje je mogoče spremljati sledi podatkovnih zbirk, je razvidna iz Tabela 10.

3.1.10.1 Sled izdelave

Metrika spremlja število sledi o izdelavi podatkovne zbirke in je izračunana na podlagi vira (v) in datuma izdelave (di). Merilo je izraženo z vrednostjo od 0 do vključno 2.

3.1.10.2 Sled posodobitev

Metrika spremlja število sledi o posodobitvah podatkovne zbirke in je izračunana na podlagi seznama posodobitev (sp) in datuma posodobitev (dp). Merilo je izraženo z vrednostjo od 0 do vključno 2.

Tabela 10: Metriki dimenzije Sledljivost.

Dimenzija	Metrika (<i>m</i>)	Spremenljivke	Formula	Merilo
Sledljivost	Sled izdelave (Vetrò idr., 2016)	v: vir, di: datum izdelave	$m_{21} = v + di$	[0, 2]
	Sled posodobitev (Vetrò idr., 2016)	sp: seznam posodobitev, dp: datum posodobitev	$m_{22} = sp + dp$	[0, 2]

Vir: lasten; osnovna različica je objavljena v študiji Pesek idr., 2025.

4 Zaključek

Vse od začetka trenda odprtih podatkov do danes ni bilo slišati veliko kritičnih glasov, ki bi si prizadevali za izboljšanje kvalitete odprtih podatkov na podatkovnih portalih (Umbrich idr., 2015), vendar so se, kot je bilo tudi pričakovano, ob prvih kritičnih pristopih pojavile problematike. Kljub problematikam so denimo v študiji Vetrò idr. (2016) predstavili model, katerega rešitve ne delujejo zgolj na njihovem podatkovnem portalu, temveč tudi na naključnih, s čimer presegajo omejitve večine raziskav. Ko smo zasnovali našo metodologijo (Pesek idr. 2025), smo upoštevali širok nabor predlogov iz pregleda literature (Žnideršič idr., 2025). Zato velja trditi, da je metodologija zasnovana široko in tako na področju odprtih podatkov naslavlja mnogo težavnih vidikov, ki se nanašajo zlasti na problematike, naslovljene v uvodu študije. Tako kot Moraga idr. (2009) smo predloge za rešitve udeležili ne samo na ravni spletne strani, temveč tudi v obliki izdelanega avtomatiziranega orodja, ki bo predmet obravnave v prihodnji študiji. Cilj avtomatiziranega orodja je olajšati in

pospešiti delo programerjev, ki ustvarjajo spletne strani, hkrati pa z zmanjšanim časovnim vnosom razrešiti problematike, na katere smo naleteli v ročnem pregledu, s tem pa zagotavljati višjo kvaliteto podatkov. Dodatno spodbujamo tudi k raziskavam, ki bi obravnavale druženje tako splošnih rešitev kot tudi tistih, ki so usmerjene zgolj v specifične portale, saj bi na ta način lahko ustvarili približek vseuporabne rešitve za dvig in ohranjanje kvalitete odprtih podatkov.

Velja omeniti tudi pojma *Uredba o podatkih* (Data Act)⁷ in *Evropska uredba o upravljanju podatkov* (European Data Governance Act)⁸ – ki sta usklajena z vrednotami in principi Evropske unije in sta med seboj zato nujno povezana – katerih cilj je nasloviti izzive in priložnosti podatkov v Evropski uniji, pri čemer poudarjata pomembnost in nujnost prostega dostopa in pravic uporabnikov, hkrati pa zagotavljata varnost njihovih osebnih podatkov. Njun cilj je zagotoviti tudi zaupanje v deljenje podatkov, okrepiti mehanizme za voljo ponovne uporabljivosti podatkov in razrešiti tehnične težave, ki se ob tem pojavljajo. Z omenjeno Uredbo o podatkih pa Evropska komisija posplošuje urejanje odnosa med uporabniki in ponudniki storitev ter uvaja nove smernice za obdelavo podatkov in predvsem deljenje podatkov s tretjimi osebami. Narava Uredbe se posledično dotika tudi ustvarjanja nove vrednosti, inovacij in poslovnih modelov, za kar so pomembni predvsem podatki z visokim učinkom za razvoj novih izdelkov, kar je tudi eden izmed glavnih ciljev nacionalnega portala OPSI, ki so razdelani v drugem poglavju.

Opaziti je, da odprti podatki v splošnem tendirajo k boljši povezljivosti oziroma interoperabilnosti, ponovni uporabljivosti in posledično k višji kvaliteti. Pomembni dejavniki, ki so pogosto v obravnavi, pa so tudi njihova varnost, dostopnost in odprtost, ki so tako vezani neposredno na podjetje kot tudi na uporabnike same. Z vedno večjim zavedanjem pomena in širjenjem raziskovalne dejavnosti o odprtih podatkih in njihovih evropskih podatkovnih portalih ter prizadevanjem za uresničevanje predvsem zgoraj navedenih kategorij skrbimo, da bi ustvarili podatkovno krajino, ki vključuje zares odprte, povezljive oziroma interoperabilne in ponovno uporabne podatke z visoko, nespreminjajočo in dosledno kvalitativno vrednostjo.

⁷ <https://digital-strategy.ec.europa.eu/en/policies/data-act>.

⁸ <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act>.

Priznanje

Projekt št. V2-2388 je sofinancirala Javna agencija za raziskovalno dejavnost Republike Slovenije iz državnega proračuna.

Viri in literatura

- Berners-Lee, T. (2006). Linked data-design issues. <http://www.w3.org/DesignIssues/LinkedData.html>
- Braunschweig, K., Eberius, J., Thiele, M. & Lehner, W. (2012). The State of Open Data – Limits of Current Open Data Platforms. V *WWW '12: Proceedings of the 21st International Conference on World Wide Web*.
- Chowdhary, K., Chowdhary, K. R. (2020). Natural language processing. *Fundamentals of artificial intelligence* (str. 603-649). https://doi.org/10.1007/978-81-322-3972-7_19
- Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage. <https://doi.org/10.4135/9781473909472>
- Metadata Quality Assessment (MQA). <https://data.europa.eu/mqa/methodology?locale=sl>. Dostop: 16. 3. 2025
- Moraga, C., Moraga, M. Á., Caro, A., & Calero, C. (2009). SPDQM: SQuaRE-aligned portal data quality model. V Ninth International Conference on Quality Software, QSIC (str. 24-25). <https://doi.org/10.1109/QSIC.2009.23>
- Neumaier, S., Umbrich, J., Polleres, A. (2016). Automated quality assessment of metadata across open data portals. *Journal of Data and Information Quality (JDIQ)*, 8(1), 1-29. <https://doi.org/10.1145/2964909>
- O portalu OPSI. <https://podatki.gov.si/o-portalu>. Dostop: 15. 3. 2025
- Pesek, M., Juvan, J., Žnideršič, K., & Marolt, M. (2025). Methodology for Qualitative Evaluation of Open Data. V *Conference on Organizational Science Development Human Being, Artificial Intelligence and Organization* (str. 730).
- Prenova informacijskega sistema za objavljanje odprtih podatkov javnega sektorja – OPSI (Prenova OPSI). (2016). <https://www.gov.si/zbirke/projekti-in-programi/prenova-informacijskega-sistema-za-objavljanje-odprtih-podatkov-javnega-sektorja-opsi/>. Dostop: 4. 2. 2026.
- Skupno poročilo o izvajanju zakona o dostopu do informacij javnega značaja za leti 2015 in 2016 (ZDIJZ-LP). (2017). https://podatki.gov.si/sites/default/files/reports/LP_ZDIJZ_2015_2016.pdf. Dostop: 4. 2. 2026.
- Umbrich, J., Neumaier, S., Polleres, A. (2015). Quality Assessment and Evolution of Open Data Portals. 2015 3rd International Conference on Future Internet of Things and Cloud (str. 404-411). <https://doi.org/10.11>
- Vetrò, A., Canova, L., Torchiano, M., Minotas, C. O., Iemma, R., & Morando, F. (2016). Open data quality measurement framework: Definition and application to Open Government Data. *Government Information Quarterly*, 33(2), 325-337. <https://doi.org/10.1016/j.giq.2016.02.001>
- Zakon o spremembah in dopolnitvah zakona o dostopu do informacij javnega značaja (ZDIJZ-E). (2015). <https://pisrs.si/pregledPredpisa?id=ZAKO7087>. Dostop: 14. 3. 2025
- Zakon o spremembah in dopolnitvah zakona o dostopu do informacij javnega značaja (ZDIJZ-G). (2022). <https://pisrs.si/pregledPredpisa?id=ZAKO8367>. Dostop: 14. 3. 2025
- Žnideršič, K., Marolt, M., & Pesek, M. (2025). Overview of Recent Methodologies for Open Data Quality Assessment. *Informatica* (str. 1-24). <https://doi.org/10.15388/25-INFOR614>
- Žnideršič, K., Marolt, M., Veršič, A., & Pesek, M. (2024). Metodologije za kvalitativno vrednotenje kakovosti odprtih podatkov. *Uporabna informatika*, 32(1). <https://doi.org/10.31449/upinf.224>

O avtorjih

Jure Juvan zaključuje magistrski študij muzikologije na Filozofski fakulteti v Ljubljani, kjer je diplomiral leta 2024. Od leta 2024 je član Laboratorija za računalniško grafiko in multimedije na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Ukvarja se z raziskovanjem vzorcev v glasbi, igrificiranega e-učenja glasbe ter metodološkimi pristopi za izboljšanje kvalitativnega vrednotenja zbirk odprtih podatkov. *Jure Juvan is completing his master's degree in musicology at the Faculty of Arts in Ljubljana where he graduated in 2024. Since 2024, he has been a member of the Laboratory for Computer Graphics and Multimedia at Faculty of Computer and Information Science of the University of Ljubljana. His research interests include patterns in music, gamified e-learning of music and methodologies for improving the qualitative evaluation of open data collections.*

Klara Žnideršič se je Laboratoriju za računalniško grafiko in multimedije na Fakulteti za računalništvo in informatiko pridružila leta 2023. Njeno raziskovalno delo je usmerjeno v pridobivanje informacij iz glasbe, sodobne pristope h glasbeni pedagogiki ter uporabo razširjene resničnosti v učnih procesih. Ukvarja se tudi z zbirkami odprtih podatkov in razvojem podpornih rešitev za sledenje regulativam na tem področju. *Klara Žnideršič has been a member of the Laboratory for Computer Graphics and Multimedia at Faculty of Computer and Information Science at the University of Ljubljana since 2023. Her research focuses on music information retrieval, contemporary approaches to music pedagogy, and the use of XR in educational processes. Her work also includes open data collections and the development of support solutions for regulatory compliance in this field.*

Dr. **Matija Marolt** je redni profesor in predstojnik Katedre za multimedijo ter vodja Laboratorija za računalniško grafiko in multimedije na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Njegova glavna raziskovalna področja so analiza in obdelava multimedijskih signalov ter razvoj kompleksnih informacijskih sistemov. Njegova skupina trenutno sodeluje pri projektih, ki raziskujejo uporabo umetne inteligence za razvoj pisnih in ustnih arhivov, biomedicinskih elektronskih mikroskopskih zvezkov, evolucijo glasbe in vibracijsko komunikacijo živali. *Dr. Matija Marolt is a full professor, chair of the Department of Multimedia, and head of the Laboratory for Computer Graphics and Multimedia at the Faculty of Computer and Information Science, University of Ljubljana. His main research areas are the analysis and processing of multimedia signals and the development of complex information systems. His group is currently involved in projects exploring the application of AI for analysis of written and oral archives, biomedical electron microscopy volumes, evolution of music, and vibrational animal communication.*

Dr. **Matevž Pesek** je izredni profesor in raziskovalec na Fakulteti za računalništvo in informatiko Univerze v Ljubljani, kjer je diplomiral (2012) in doktoriral (2018). Od leta 2009 je član Laboratorija za računalniško grafiko in multimedije. Njegovo raziskovalno delo je usmerjeno v pridobivanje informacij iz glasbe, e-učenje ter razvoj aplikacij z namenom ustvarjanja družbenega vpliva. Od leta 2014 se aktivno ukvarja z raziskovanjem odprtih podatkov in je soavtor več spletnih storitev, med drugim Avtolog.si, Tocen.si in Vodostaji.si. *Dr. Matevž Pesek is an associate professor and researcher at the Faculty of Computer and Information Science, University of Ljubljana, where he received his B.Sc. (2012) and his PhD (2018). He has been a member of the Laboratory for Computer Graphics and Multimedia since 2009. His research interests are e-learning, music information retrieval, and the development of applications aimed at creating societal impact. Since 2014, he has been actively involved in open data research and has co-developed several online services, including Avtolog.si, Tocen.si, and Vodostaji.si.*

Tabela 11: Metodologija za kvalitativno vrednotenje odprtih podatkov na spletnem portalu OPSI – dimenzije, metrike, spremenljivke, formule in merila.

Dimenzija	Metrika (<i>m</i>)	Spremenljivke	Formula	Merilo
Popolnost	Odstotek popolnih celic (Vetrò idr., 2016)	v: število vrstic, s: število stolpcev, cn: število nepopolnih celic, c: število celic	$c = v \cdot s; m_1 = \left(1 - \frac{cn}{c}\right) \cdot 100$	[0 %, 100 %]
	Odstotek popolnih vrstic (Vetrò idr., 2016)	v: število vrstic, vn: število nepopolnih vrstic	$m_2 = \left(1 - \frac{vn}{v}\right) \cdot 100$	[0 %, 100 %]
	Odstotek popolnih stolpcev	s: število stolpcev, sn: število nepopolnih stolpcev	$m_3 = \left(1 - \frac{sn}{s}\right) \cdot 100$	[0 %, 100 %]
	Odstotek popolnosti metapodatkov v podatkovnih zbirkah	m: število metapodatkov, mn: število nepopolnih metapodatkov	$m_4 = \left(1 - \frac{mn}{m}\right) \cdot 100$	[0 %, 100 %]
	Odstotek predvidenih vrednosti v celicah	vr: število vrednosti, vrn: število naključnih vrednosti (črke, številke, simboli itd.)	$m_5 = \left(1 - \frac{vrn}{vr}\right) \cdot 100$	[0 %, 100 %]
Unikatnost	Odstotek nepodvojenih vrstic (Vetrò idr., 2016; Neumaier idr., 2016)	vp: število podvojenih vrstic, v: število vrstic	$m_6 = \left(1 - \frac{vp}{v}\right) \cdot 100$	[0 %, 100 %]
	Odstotek nepodvojenih vrednosti v celicah	pvr: število podvojenih vrednosti, vr: število vrednosti	$m_7 = \left(1 - \frac{pvr}{vr}\right) \cdot 100$	[0 %, 100 %]
	Pearsonov korelacijski koeficient	x: izbrani stolpec, y: stolpec za primerjavo	$\rho_{x,y} = \frac{cov(X,Y)}{\sigma_x \sigma_y}$	[-1, 1]

Dimenzija	Metrika (m)	Spremenljivke	Formula	Merilo
Doslednost	Odstotek pravilno poimenovanih stolpcev	s: število stolpcev, spn: število nepričakovano poimenovanih stolpcev	$m_8 = \left(1 - \frac{spn}{s}\right) \cdot 100$	[0 %, 100 %]
	Odstotek doslednosti podatkovnih tipov	pt: število podatkovnih tipov, ptn: število nepričakovanih podatkovnih tipov	$m_9 = \left(1 - \frac{ptn}{pt}\right) \cdot 100$	[0 %, 100 %]
Uparljivost	Odstotek navedenih podatkovnih tipov	pt: število podatkovnih tipov, ptnn: število navedenih podatkovnih tipov	$m_{10} = \left(1 - \frac{ptnn}{pt}\right) \cdot 100$	[0 %, 100 %]
	Odstotek veljavnih podatkovnih tipov	pt: število podatkovnih tipov, ptnv: število neveljavnih podatkovnih tipov	$m_{11} = \left(1 - \frac{ptnv}{pt}\right) \cdot 100$	[0 %, 100 %]
	Odstotek strojno berljivih podatkovnih zbirk	pz: število podatkovnih zbirk, pzsn: število strojno neberljivih podatkovnih zbirk	$m_{12} = \left(1 - \frac{pzsn}{pz}\right) \cdot 100$	[0 %, 100 %]
Ponovna uporabljivost (MQA)	Odstotek podatkovnih zbirk z licenco	pz: število podatkovnih zbirk, pzbl: število podatkovnih zbirk brez licence	$m_{13} = \left(1 - \frac{pzbl}{pz}\right) \cdot 100$	[0 %, 100 %]
	Odstotek podatkovnih zbirk s kontaktno točko	pz: število podatkovnih zbirk, pzbkt: število podatkovnih zbirk brez kontaktne točke	$m_{14} = \left(1 - \frac{pzbkt}{pz}\right) \cdot 100$	[0 %, 100 %]
	Odstotek podatkovnih zbirk z informacijo o izdajatelju	pz: število podatkovnih zbirk, pzbi: število podatkovnih zbirk brez informacije o izdajatelju	$m_{15} = \left(1 - \frac{pzbi}{pz}\right) \cdot 100$	[0 %, 100 %]
Natančnost	Odstotek sintaktično natančnih celic (Vetro idr., 2016)	cnk: število celic z napakami (šumniki, posebni znaki, številke itd.), c: število celic	$m_{16} = \left(1 - \frac{cnk}{c}\right) \cdot 100$	[0 %, 100 %]
Razumljivost	Jasnost poimenovanja	p: podatki, pt: podatkovni tipi, sspp: sintaktično natančna in semantično ustrezna pravila za poimenovanje	Uporaba NLP (Natural Language Processing) (Chowdhary idr., 2020) ali ročno preverjanje podatkov za dodelitev točk jasnosti.	Subjektivno [0, 5]
	Odstotek stolpcev v razumljivem podatkovnem tipu	snpt: število stolpcev v nerazumljivem podatkovnem tipu, s: število stolpcev	$m_{17} = \left(1 - \frac{snpt}{s}\right) \cdot 100$	[0 %, 100 %]

Dimenzija	Metrika (<i>m</i>)	Spremenljivke	Formula	Merilo
Aktualnost	Zamuda v objavi (Vetrò idr., 2016)	ddi: datum dostopnosti informacij, do: datum objave, zd: začetni datum obdobja obstoja podatkovne zbirke, kd: končni datum obdobja obstoja podatkovne zbirke	$ddi = kd + 1;$ $m_{18} = 1 - \frac{do - kd}{kd - zd}$	$(-\infty, 1]$
Skladnost	Skladnost z eGMS (Vetrò idr., 2016)	v: vir, di: datum izdelave, k: kategorija, n: naslov, o: opis, id: identifikator, iz: izdajatelj, j: jezik	$m_{19} = v + di + k + n + o + id + iz + j$	$[0, 8]$
	Petzvezdični odprti podatki (Berners-Lee, 2006)	dpol: dostopnost podatkov v formatu z odprto licenco, psb: podatki v strojno berljivi obliki, dpnf: dostopnost podatkov v nelastniškem formatu, uosi: uporaba odprtih standardov organizacije W3C in URI za identifikacijo, kp: kontekstualizacija podatkov in ostalih dostopnih podatkov	$m_{20} = dpol + psb + dpnf + uosi + kp$	$[0, 5]$
Sledljivost	Sled izdelave (Vetrò idr., 2016)	v: vir, di: datum izdelave	$m_{21} = v + di$	$[0, 2]$
	Sled posodobitev (Vetrò idr., 2016)	sp: seznam posodobitev, dp: datum posodobitev	$m_{22} = sp + dp$	$[0, 2]$

Vir: lasten; osnovna različica je objavljena v študiji Pesek idr., 2025.