

PREGLED OBSTOJEČIH PRISTOPOV K VREDNOTENJU KAKOVOSTI ODPRTIH PODATKOV

KLARA ŽNIDERŠIČ, MATEVŽ PESEK

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

Poglavje obravnava pregled obstoječih pristopov k vrednotenju kakovosti odprtih podatkov, pri čemer se osredotoča na metodologije, standarde in praktične primere uporabe. Avtorja navedeta ključne pravne podlage in izpostavljata pomen zagotavljanja kakovosti (meta)podatkov za njihovo uporabnost. Predstavljeni so različni standardi in prakse, med njimi DCAT, standardi ISO, načela FAIR in petzvezdični sistem za ocenjevanje odprtosti podatkov, ter uveljavljeni modeli in orodja za ocenjevanje kakovosti. Analiza vključuje tudi pregled dobrih praks v različnih državah in na portalih, kar ponuja vpogled v učinkovite pristope k zagotavljanju kakovostnih odprtih podatkov. Prispevek tako služi kot vodilo za razumevanje kakovosti odprtih podatkov ter možnosti njihovega učinkovitejšega vrednotenja in uporabe.

DOI
[https://doi.org/
10.18690/um.fov.2.2026.6](https://doi.org/10.18690/um.fov.2.2026.6)

ISBN
978-961-299-114-2

Ključne besede:
odprti podatki,
kvalitativno vrednotenje,
pregled metodologij,
zakonodaja,
implementacije
metodologij

DOI
[https://doi.org/
10.18690/um.fov.2.2026.6](https://doi.org/10.18690/um.fov.2.2026.6)

ISBN
978-961-299-114-2

Keywords:

open data,
qualitative assessment,
methodology review,
legislation,
methodology
implementation

OVERVIEW OF EXISTING APPROACHES TO ASSESSING THE QUALITY OF OPEN DATA

KLARA ŽNIDERSIČ, MATEVŽ PESEK

University of Ljubljana, Faculty of Computer Science and Informatics, Ljubljana,
Slovenia

klara.znidersic@fri.uni-lj.si, matevz.pesek@fri.uni-lj.si

The chapter reviews current approaches to assessing the quality of open data, focusing on methodologies, standards, and practical use cases. The authors present key legal foundations and emphasize the importance of ensuring the quality of (meta)data for usability. Various standards and practices are discussed, including DCAT, ISO standards, FAIR principles, and the five-star system for evaluating data openness, as well as established models and tools for quality assessment. The analysis also provides an overview of good practices in different countries and portals, offering insight into effective methods for ensuring high-quality open data. This paper thus serves as a guide to understanding open data quality and the possibilities for more effective assessment and use.



University of Maribor Press

1 Uvod

Tehnologija, ki je dandanes vseprisotna v vsakdanjem življenju, omogoča številne storitve, ki poenostavljajo in pospešujejo številne procese, obenem pa z vse intenzivnejšo rabo teh storitev kot glavni ali stranski produkt nastaja veliko število podatkov. V slednjih so raziskovalci in odločevalci prepoznali sredstvo za spodbujanje nadaljnjega razvoja in z namenom, da bi spodbudili njihovo čim širšo (ponovno) uporabo, vzpostavili številne smernice in regulative. Osrednji pojem teh dokumentov so odprti podatki, ki so definirani kot podatki, do katerih lahko vsak svobodno dostopa, jih uporablja in deli za kakršenkoli namen (Open Knowledge Foundation, 2015). Izraz odprti podatki se ne navezuje zgolj na vladne podatke, temveč tudi na podatke iz zasebnega sektorja, ki imajo v kombinaciji z (neodprtimi) deljenimi ali osebnimi podatki še večji potencial. Poudariti velja, da prava vrednost podatkov ni v podatkih samih, temveč v izdelkih, storitvah in vsebini, ki so izboljšani ali omogočeni z odprtimi podatki. Zato se ponovna uporaba podatkov ne spodbuja le v javnih službah, ampak tudi v zasebnih podjetjih.

Na podlagi poročil Evropske komisije ocena skupne vrednosti trga odprtih podatkov v Evropski uniji izrazito raste – leta 2019 je bila ocenjena na 185 milijard evrov, do leta 2025 pa naj bi dosegla vsaj 199 milijard evrov oziroma, po optimističnem scenariju, celo 334 milijard evrov (Huyer & Knippenberg, 2020). S kopičenjem podatkov na evropskih podatkovnih portalih, ki vpliva na rast vrednosti trga, hkrati rastejo tudi pričakovanja in zahteve uporabnikov ponujenih podatkov. Toda komercialna podjetja, ki bi z rabo odprtih podatkov lahko spodbudila velike inovacije in ekonomsko rast, se za uporabo podatkov pogosto ne odločajo zaradi razlogov, povezanih z ekonomskimi aspekti in pomanjkljivim znanjem (Zuiderwijk idr., 2015). Kljub navidez preprosti uporabi brezplačno dostopnih odprtih podatkov se namreč podjetja srečujejo z ovirami pri identificiranju primernih naborov podatkov in pripravi le-teh za uporabo, pri čemer se ovire neredko navezujejo na kakovost podatkov (Krasikov & Legner, 2023).

Dolgoletno zbiranje in vse bolj zgoščeno kopičenje podatkov izpostavlja vedno bolj perečo problematiko vzdrževanja zadostne stopnje kakovosti zbranih podatkov. Pri objavljanju odprtih podatkov namreč pogosto ni prioriteta zagotavljanje najvišje kakovosti, temveč omogočanje dostopnosti in objava čimvečjega nabora podatkov. Toda prav pomanjkljivi, nepravilni, neurejeni ali nejasni podatki izrazito vplivajo na

uporabno vrednost zbirke. Le zanesljivi vhodni podatki namreč lahko zagotavljajo zanesljive rezultate njihove analize in nadaljnje uporabe. Če torej proces objave zbirk odprtih podatkov ne vključuje tudi ocene pripravljenosti podatkov za nadaljnjo rabo z vidika kakovosti, je težko vnaprej določiti, ali bo izbrani nabor podatkov sploh uporaben ali pa je morda potreben obsežnega dodatnega urejanja. Prav tako včasih ni jasno, ali so podatki povezljivi in interoperabilni, kar je pri vzpostavljanju novih storitev eden ključnih faktorjev. Identifikacija ustreznih podatkovnih zbirk lahko torej za potencialne uporabnike predstavlja preveliko breme z vidika časovnih, kadrovskih in materialnih virov, zaradi česar se pogosto odločijo za iskanje alternativnih virov podatkov, ki so primerno urejeni in formatirani ter zbirk odprtih podatkov ne uporabljajo ali pa jih uporabljajo le v omejenem obsegu.

Raziskovalci in razvijalci, ki so se lotili identifikacije primernih podatkovnih zbirk in so v procesu že v zgodnjih fazah vključili oceno kakovosti podatkov, so svoje pristope predstavili v različnih metodoloških opisih. V začetnem delu se poglavje usmerja v evropsko in slovensko zakonodajo, ki ureja področje odprtih podatkov. Ta zakonodaja lahko služi kot pravna podlaga za ocenjevanje kakovosti podatkovnih zbirk v Sloveniji ter nudi okvir za njihovo izboljšanje in uporabo v različnih kontekstih. Poglavje nato prinaša pregled ključnih lestvic in standardov, ki so raziskovalcem predstavljali usmeritev pri implementaciji metodologij, ter se osredotoča na predstavitev najvidnejših metodologij za vrednotenje kakovosti zbirk podatkov.

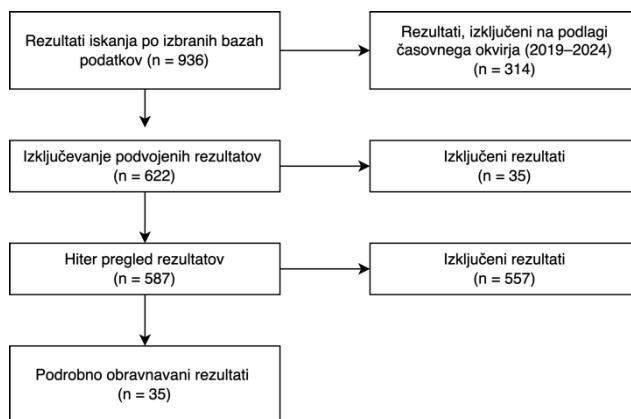
Metodologija pregleda

Našo raziskavo smo osnovali na podlagi treh raziskovalnih vprašanj:

1. Katere so obstoječe metodologije za kvalitativno vrednotenje (meta)podatkov?
2. Kateri standardi so v teh metodologijah najpogosteje uporabljeni in upoštevani?
3. Katere evropske in državne regulative urejajo področje odprtih podatkov v Sloveniji?

Pri pregledu literature smo sledili smernicam modela PRISMA (M. J. Page idr., 2021) za sistematično iskanje virov v bibliografskih zbirkah. Pregled je zajemal tri faze – identifikacijo, pregledovanje in vključevanje.

Fazo identifikacije smo izvedli z iskanjem v bazah Science Direct, Scopus, IEEE Xplore, ACM Digital Library in Web of Science na podlagi ključnih besed: »open data« AND (»quality assessment« OR »quality evaluation«) AND (metrics OR framework OR guidelines OR methodology OR measurement). Proces identifikacije je shematično prikazan na Sliki 1.



Slika 1: Identifikacija literature o kvalitativnem vrednotenju odprtih podatkov

Vir: lasten

Filtriranje smo osnovali na naslednjih kriterijih:

- rezultati iz petletnega obdobja 2019–2024, ki sovpadajo z implementacijo evropske direktive o odprtih podatkih,
- odstranjeni podvojeni rezultati,
- rezultati so dostopni,
- rezultati so v angleškem jeziku.

Nabor rezultatov je zajemal 35 člankov, ki smo jih podrobneje pregledali in na podlagi prebranega izločili manj relevantno gradivo. V primerih, ko so avtorji članka navajali gradivo izjemnega pomena za naš pregled, smo ga naknadno dodali v nabor literature. Pregledali smo tudi letna poročila o stanju odprtih podatkov v Evropski uniji, ki so dostopni na portalu Evropske unije, saj ta poleg vpogleda v stanje pogosto vsebuje smernice in primere dobrih praks za zagotavljanje kvalitete. Del našega dela je zajemal tudi pregled zakonodaje, zato smo nabor razširili še s pravnimi dokumenti,

dostopnimi na vladnih straneh Republike Slovenije in Evropske unije. Zbrano gradivo smo razvrstili v kategorije – zakonodaja in analitična poročila, standardi in formati ter metodologije. Tako so tudi predstavljena v nadaljnjem besedilu.

3 Zakonodaja

3.1 V Sloveniji

V slovenski zakonodaji se na naslovno tematiko odprtih podatkov navezujeta Zakon o dostopu do informacij javnega značaja (v nadaljevanju ZDIJZ) (Uradni list RS, št. 24/2003) ter Uredba o posredovanju in ponovni uporabi informacij javnega značaja (*Uredba o posredovanju in ponovni uporabi informacij javnega značaja*, 2016; *Uredba o spremembah in dopolnitvah Uredbe o posredovanju in ponovni uporabi informacij javnega značaja*, 2022), ki v pravni red Republike Slovenije prenašata evropsko direktivo, podrobneje opisano v nadaljevanju. Dopolnitve in spremembe so sledile z ZDIJZ-E (Uradni list RS, št. 102/2015) in ZDIJZ-G (Uradni list RS, št. 141/2022) (*Zakon o spremembah in dopolnitvah Zakona o dostopu do informacij javnega značaja (ZDIJZ-E)*, 2015; *Zakon o spremembah in dopolnitvah Zakona o dostopu do informacij javnega značaja (ZDIJZ-G)*, 2022). Vsebinsko podlago za dopolnitve in spremembe ZDIJZ so predstavljale prenovljene evropske direktive.

Z omenjenim zakonom je urejen postopek, ki vsakomur omogoča prost dostop in ponovno uporabo informacij javnega značaja. ZDIJZ-E definira relevantne pogosto uporabljene izraze strojno berljivo obliko, odprti podatek, formalni odprti standard in izraz metapodatek, ZDIJZ-G pa dodaja še definicije izrazov raziskovalni podatki, dinamični podatki, nabor podatkov velike vrednosti, informacija javnega značaja in ponovna uporaba informacij javnega značaja.

- **Strojno berljiva oblika** pomeni, da so podatki ali metapodatki strukturirani tako, da lahko programske aplikacije z lahkoto določajo, prepoznajo in izločajo določene podatke, vključno s posameznimi dejanskimi navedbami in njihovo notranjo strukturo.
- Izraz **formalni odprti podatek** pomeni standard, ki je javno dostopen, brez omejitev uporabe in natančno opredeljuje zahteve za zagotavljanje interoperabilnosti programske opreme in je določen v pisni obliki ter ga

ministrstvo, pristojno za upravo, objavi na nacionalnem portalu odprtih podatkov.

- **Formalni odprti standard** pomeni standard, ki je javno dostopen, brez omejitev uporabe in natančno opredeljuje zahteve za zagotavljanje interoperabilnosti programske opreme, je določen v pisni obliki ter ga ministrstvo, pristojno za upravo, objavi na nacionalnem portalu odprtih podatkov.
- **Metapodatek** pomeni informacije, ki opisujejo zbirko podatkov in storitve v zvezi s podatki ter omogočajo njihovo iskanje, evidentiranje in uporabo.
- **Dinamični podatki** pomenijo podatke v digitalni obliki, ki se pogosto ali v realnem času posodablajo, zlasti zaradi svoje spremenljivosti ali hitre zastarelosti. Med dinamične podatke se ne uvrščajo podatki, ki se obdelujejo v zbirkah, ki vsebujejo posebne vrste osebnih podatkov.
- **Nabori podatkov velike vrednosti** pomenijo podatke s področij geoprostora, opazovanja Zemlje, okolja, meteorologije, statistike, gospodarskih družb in lastništva gospodarskih družb ter mobilnosti, katerih ponovna uporaba je povezana s pomembnimi koristmi za družbo, okolje in gospodarstvo, zlasti zaradi njihove primernosti za ustvarjanje storitev in aplikacij z dodano vrednostjo ter novih, visokokakovostnih in dostojnih delovnih mest ter števila potencialnih prejemnikov storitev in aplikacij z dodano vrednostjo, ki temeljijo na teh naborih podatkov, in so določeni v predpisu iz 11. člena tega zakona.
- **Informacija javnega značaja** je informacija, ki izvira iz delovnega področja organa, nahaja pa se v obliki dokumenta, zadeve, dosjeja, registra, evidence ali drugega dokumentarnega gradiva (v nadaljnjem besedilu: dokument), ki ga je organ izdelal sam ali v sodelovanju z drugim organom ali pridobil od drugih oseb.
- **Ponovna uporaba informacij javnega značaja** pomeni uporabo s strani fizičnih oseb ali pravnih oseb za pridobitne ali nepridobitne namene razen za prvotni namen v okviru javne naloge, zaradi katerega so bili dokumenti izdelani. Uporaba informacij za izvajanje javnih nalog v organu ali izmenjava informacij med organi za izvajanje javnih nalog se ne šteje za ponovno uporabo informacij.

S področjem zbirke odprtih podatkov je najbolj neposredno povezan 10. člen zakona, ki ureja posredovanje informacij javnega značaja v svetovni splet. Zakon obvezuje vsak organ k posredovanju določenih informacij javnega značaja v splet brezplačno in na namenskem portalu, če ta za določeno področje obstaja. Po

navodilih za zagotavljanje odprtih podatkov za ponovno uporabo v členu 10.b ZDIJZ-E organi praviloma omogočajo objavo v odprtih formatih, ki upoštevajo formalne odprte standarde, v strojno berljivi obliki, skupaj z metapodatki. Seznam vseh zbirk podatkov z metapodatki ter zbirke odprtih podatkov ali povezave na spletne strani, kjer so objavljene zbirke odprtih podatkov, morajo biti po zakonu dostopni na nacionalnem portalu odprtih podatkov javnega sektorja, ki ga vodi pristojno ministrstvo. Objavljeni podatki morajo biti brezplačno na voljo za rabo v pridobitne ali druge namene pod pogoji, da se ponovna uporaba izvaja v skladu z Zakonom o varstvu osebnih podatkov in da se ob vsaki uporabi navede vir podatkov.

Uredba o posredovanju in ponovni uporabi informacij javnega značaja v 2. členu specificira vsebino kataloga informacij javnega značaja, ki mora vključevati osnovne podatke o katalogu (naziv organa in odgovorne uradne osebe, datum objave oz. zadnje spremembe), splošne podatke o organu in informacijah javnega značaja, s katerimi razpolaga, opis načina dostopa do informacij javnega značaja, relevantne stroškovnike in seznam najpogostejše zahtevanih informacij javnega značaja. Javne evidence in druge informatizirane zbirke morajo biti opremljene z naslednjimi metapodatki:

1. naziv institucije;
2. naziv zbirke podatkov;
3. področje;
4. opis vsebine;
5. pravne podlage;
6. ključne besede;
7. območje, na katero se vsebovani podatki nanašajo;
8. dostopnost podatkov;
9. morebitne posebne pogoje za uporabo oziroma ponovno uporabo;
10. morebitne pogoje glede zaračunavanja cene za ponovno uporabo;
11. morebitni internetni naslov za dostop oziroma vpogled podatkov;
12. morebitni internetni naslov do odprtih podatkov (strojno berljiva oblika);
13. datum nastanka;
14. pogostost osveževanja podatkov;
15. obliko zapisa;
16. jezik zapisa.

Ob tem Uredba opredeljuje obravnavanje dokumentov, ki so v postopku izdelave, definira vrste informacij javnega značaja, ki se posredujejo v svetovni splet, določa načine in natančen stroškovnik posredovanja teh informacij ter vsebino poročil, ki jih morajo po zakonu izdajati državni organi in organi lokalnih skupnosti na letni ravni.

3.2 V Evropski uniji

Javni sektor je v Evropski uniji eden najbolj podatkovno zasičenih sektorjev. Podatki, ki se v tem okolju zbirajo, proizvajajo, reproducirajo in razširjajo, vključujejo podatke s področja geografije, vremena, sociale, turizma, podjetništva, izobraževanja, podatke iz javno financiranih raziskovalnih projektov in digitalizirane knjige iz knjižnic. Med njimi so številni podatki »odprti«, kar pomeni informacije javnega sektorja, do katerih je mogoče enostavno dostopati ter jih ponovno uporabiti, pogosto pod neomejenimi pogoji. Neposredni vpliv ponovne uporabe odprtih podatkov na višanje njihove gospodarske vrednosti je bil sorazmerno zgodaj prepoznan za nezanemarljivega, zato je Evropska komisija že leta 2003 izdala prve smernice na tem področju. V naslednjih letih je zaradi hitrega tehnološkega razvoja kot rezultat predlogov Evropske komisije prišlo do nadaljnjih sprememb in nadgradenj predhodno sprejetih direktiv, zadnja med njimi pa je izšla leta 2019 in predstavlja najbolj aktualne smernice, med drugim relevantne za izdelavo metodologije, ki bo namenjena določanju kakovosti podatkov.

Direktiva 2003/98/ES (Direktiva 2003/98/ES Evropskega parlamenta in Sveta z dne 17. novembra 2003 o ponovni uporabi informacij javnega sektorja, 2003) je določevala splošni zakonodajni okvir na evropski ravni, minimalni nabor pravil za ponovno rabo podatkov v državah članicah EU in praktična sredstva za olajšanje ponovne uporabe obstoječih dokumentov. Njen namen je bil med drugim omogočiti izkoriščanje potenciala zbranih informacij za vse države članice ter s tem prispevati h gospodarski rasti. Predstavljeni so bili princip zahtevkov za ponovno uporabo podatkov ter načela obračunavanja dostopa, upoštevanja razpoložljivih formatov in izogibanja diskriminatornim razlikam rabe podatkov v javnem in zasebnem sektorju.

Spremembe v naslednjem desetletju je nasloвила **Direktiva 2013/37/EU** (Direktiva 2013/37/EU Evropskega parlamenta in Sveta z dne 26. junija 2013 o spremembi Direktive 2003/98/ES o ponovni uporabi informacij javnega sektorja, 2013). V njej

so zaradi ambicioznih pristopov nekaterih držav članic sprejeli minimalno uskladitev na področju dela z odprtimi podatki, s čimer so poskušali premostiti večje ovire čezmejnega informacijskega prometa. Osrednji cilj direktive je bil sprostitvev gospodarskega potenciala ponovne uporabe podatkov, ki jih ustvarja javni sektor, prek oblikovanja posodobljene skupne podlage za ponovno rabo podatkov po vsej EU. Ta je hkrati predstavljala spodbudo javnim organom, da zagotovijo brezplačen elektronski dostop do podatkov. V okviru direktive so bile na novo dodane oz. posodobljene definicije pojmov (strojno berljivi format, odprti format, formalni odprti format), natančneje razčlenjen postopek in višina obračunavanja dostopa ter definiran položaj digitaliziranih kulturnih virov in posebni pogoji pri uporabi podatkov univerzitetnih knjižnic, muzejev in arhivov, najbolj neposredno vezane na objavo podatkov pa so bile posodobljene praktične ureditve. Poleg primerne strojno berljivega formata, seznama virov glavnih dokumentov in portalov, ki so povezani s seznamami virov, je poudarek usmerjen na opremljanje dokumentov z ustreznimi metapodatki, doseganje največje stopnje natančnosti in ločljivosti, zagotavljanje interoperabilnosti, predlagana pa je tudi funkcionalnost medjezikovnega iskanja dokumentov. Kot vir dodatnih smernic pri obdelavi podatkov je izpostavljena Direktiva 2007/2/ES (Direktiva 2007/2/ES Evropskega parlamenta in Sveta z dne 14. marca 2007 o vzpostavitvi infrastrukture za prostorske informacije v Evropski skupnosti (INSPIRE), 2007).

Najaktualnejša direktiva za delo z odprtimi podatki in ponovno uporabo informacij javnega sektorja (Direktiva (EU) 2019/1024 Evropskega parlamenta in Sveta z dne 20. junija 2019 o odprtih podatkih in ponovni uporabi informacij javnega sektorja, 2019) podaja predlog prenove in dopolnitve področja. Bistvena vodila in točke iz prejšnjih direktiv so bile redefinirane in vključene med prenovljene smernice. Novosti in največje spremembe so bile predlagane na področjih zagotavljanja dostopa do dinamičnih podatkov v realnem času z ustreznimi tehničnimi sredstvi, medsebojnega povezovanja vzpostavljenih zbirk podatkov na več nivojih EU in povečanja ponudbe dragocenih javnih podatkov za ponovno uporabo – tudi podatkov javnih podjetij, raziskovalnih organizacij in organizacij, ki financirajo raziskave.

Pred razčlenitvijo predlogov na naštetih področjih je smotrno natančneje definirati najpogostejše uporabljene izraze pri delu z odprtimi podatki, izmed katerih so v direktivi med drugim navedene opredelitve naslednjih pojmov:

- **Strojno berljivi format** predstavlja datotečni format, strukturiran tako, da lahko programske aplikacije z lahkoto določajo, prepoznajo in izločajo določene podatke, vključno s posameznimi navedbami dejstev in njihovo notranjo strukturo.
- **Odprti format** je datotečni format, neodvisen od uporabljene platforme in dostopen javnosti brez omejitev, ki bi preprečevale ponovno uporabo dokumentov.
- **Formalni odprti standard** pomeni standard, ki je določen v pisni obliki in natančno opredeljuje zahteve za zagotavljanje interoperabilnosti programske opreme.
- **Nabori podatkov velike vrednosti** pomenijo dokumente, katerih ponovna uporaba je povezana s pomembnimi koristmi za družbo, okolje in gospodarstvo, zlasti zaradi njihove primernosti za ustvarjanje storitev in aplikacij z dodano vrednostjo ter novih, visokokakovostnih in dostojnih delovnih mest ter zaradi števila potencialnih prejemnikov storitev in aplikacij z dodano vrednostjo, ki temeljijo na teh naborih podatkov.

5. člen aktualne direktive poziva organe javnega sektorja in javna podjetja, da dajejo na voljo dokumente v katerem koli obstoječem formatu ali jeziku ter, kadar je to možno in primerno, z elektronskimi sredstvi v formatih, ki so odprti, strojno berljivi in dostopni ter jih je mogoče najti in ponovno uporabiti, skupaj z njihovimi metapodatki. Kadar je možno, je treba zagotoviti skladnost formata in metapodatkov s formalnimi odprtimi standardi. Obenem direktiva organe spodbuja k pripravi in objavi dokumentov v skladu z načelom »vgrajena in privzeta odprtost«, vendar pa organi niso obvezani k pripravi ali prilagajanju dokumentov, kadar bi to pomenilo nesorazmeren napor, ki bi presegal enostaven postopek.

Kot novost v direktivi so bili obravnavani dinamični podatki, ki zaradi svoje narave za polno izkoriščanje svojega gospodarskega potenciala zahtevajo čimbolj ažuren dostop. Ključno vlogo pri omogočanju dostopa do dinamičnih podatkov igrajo razumni roki. Idealni rok za zagotovitev dostopa do tovrstnih podatkov bi bilo posodabljanje v realnem času, kadar pa to zaradi tehničnih ali finančnih omejitev ni mogoče, je pomembno, da se dinamični podatki objavijo v časovnem okvirju ali z začasnimi tehničnimi omejitvami, ki neupravičeno ne ovirajo izkoriščanja njihovega gospodarskega in družbenega potenciala. Predvsem v primerih, ko je preverjanje

podatkov ključno zaradi upravičenih razlogov javnega interesa (zlasti na področju javnega zdravja in varnosti), je treba v primeru tehničnih in finančnih omejitev sprejeti konkretne ukrepe za izboljšanje ažurnosti. Direktiva organe javnega sektorja spodbuja k omogočanju dostopa do dinamičnih podatkov za ponovno uporabo takoj, ko so podatki zbrani, in prek ustreznih vmesnikov za aplikacijsko programiranje (API).

Več pozornosti v novi direktivi se pri postavitvi zbirk usmerja v medsebojno povezovanje, tako znotraj države kot čezmejno. Odprti podatki, ki predstavljajo pomemben primarni material za proizvode in storitve z digitalno vsebino, postajajo z razvojem naprednih digitalnih tehnologij (umetna inteligenca, internet stvari ...) še pomembnejši vir vsebin, pri čemer je vse bolj bistvena široka čezmejna geografska pokritost. Smernice za praktično ureditev v 9. členu predvidevajo poenostavitev medjezikovnega iskanja dokumentov, kar bi omogočilo tudi združevanje metapodatkov na ravni Unije, in hkrati dostopnost praktičnih ureditev za poenostavitev iskanja dokumentov (npr. s seznamami virov glavnih dokumentov z ustreznimi podatki) prek spleta in v strojno berljivem formatu, ter vzpostavitev enotne točke dostopa do podatkov v formatih, prilagojenih rabi z elektronskimi sredstvi. Ponovna uporaba dokumentov je v skladu s 6. členom direktive brezplačna, vendar je ob tem sprejemljivo zaračunavanje povračila mejnih stroškov, ki nastanejo zaradi reprodukcije, zagotavljanja in razširjanja dokumentov, pa tudi anonimizacije osebnih podatkov in ukrepanja za varovanje poslovnih informacij.

S povečanjem ponudbe javno dostopnih podatkov za ponovno uporabo je neposredno povezano javno financiranje področje raziskovalnih podatkov. Direktiva nagovarja k podpori razpoložljivosti raziskovalnih podatkov prek sprejemanja nacionalnih politik in ustreznih ukrepov, ki bi zagotavljali, da bodo podatki in metapodatki iz raziskav v skladu z načelom »vgrajene odprtosti« prosto dostopni in združljivi z načeli FAIR. Ponovna uporaba teh podatkov je dovoljena, če so financirani iz javnih sredstev ter so jih raziskovalci, raziskovalne organizacije ali organizacije, ki raziskave financirajo, že objavili prek institucionalnih ali tematskih repozitorijev.

Aktualna direktiva obravnava tudi posebne nabore podatkov velike vrednosti, ki so tematsko kategorizirani v skupine *Geoprostorski podatki*, *Opazovanje zemlje in okolje*, *Meteorološki podatki*, *Statistični podatki*, *Družbe in lastništvo družb* ter *Mobilnost*.

Opredelitev teh naborov temelji na oceni njihovega potenciala za ustvarjanje pomembnih družbeno-gospodarskih, okoljskih ali inovativnih koristi, koristi za veliko število uporabnikov, doprinosu k ustvarjanju prihodka ter za kombiniranje z drugimi nabori podatkov. Za posebne nabore aktualna direktiva narekuje, da so na voljo brezplačno, so strojno berljivi in dostopni preko API ter, kadar je to ustrezno, omogočajo masovni prenos (bulk download). Načini objave in ponovne uporabe lahko vključujejo pogoje za ponovno uporabo, formate podatkov in metapodatkov ter tehnične ureditve za razširjanje, vendar morajo biti združljivi z dovoljenji odprtega standarda.

Sprejeta direktiva naj bi povečala možnosti uporabe podatkov, s čimer bi družbe EU lahko izkoristile potencial teh informacij ter prispevale h gospodarskemu razvoju, ustvarjanju visokokakovostnih delovnih mest in zaščiti, zlasti v korist lokalnih skupnosti. Hkrati ostajajo nekateri tehnološki izzivi, ki so povezani z neusklajenostjo nacionalnih zakonodaj, zato bo nadaljnji razvoj informacijske družbe sledil vsaj minimalno harmoniziranim nacionalnim pravilom in praksam glede ponovne uporabe dokumentov javnih sektorjev.

3.3 Ostali EU dokumenti

Poleg opisanih direktiv je v Evropski uniji izšlo še več drugih pravnih dokumentov, do katerih je dostop omogočen na spletišču EUR-Lex. Spletišče upravlja Urad za publikacije Evropske unije in je redno posodablano ter na voljo v vseh 24 uradnih jezikih EU.

Ključno spodbujevalno vlogo IKT v smislu podpore EU pri doseganju njenih ciljev opredeljuje desetletna Evropska digitalna agenda (Evropska komisija, 2010). Njen glavni cilj je bil načrtovati pot, po kateri bo najbolj izkoriščen družbeni in gospodarski potencial IKT, zlasti internet, ki je nepogrešljiv medij gospodarske in družbene aktivnosti. Strategija za enotni digitalni trg za Evropo (Evropska komisija, 2015), ki je dopolnila obstoječo Evropsko digitalno agendo, je bila sprejeta leta 2015. Izoblikovala je cilj zagotovitve pravičnega, odprtega in varnega digitalnega okolja in 16 usmerjenih ukrepov, utemeljila pa je na treh glavnih stebrih:

1. Boljši dostop za potrošnike in podjetja do digitalnega blaga in storitev po vsej Evropi.

2. Oblikovanje ustreznega okolja in enakih konkurenčnih pogojev za omogočanje razvoja digitalnih omrežij in inovativnih storitev.
3. Čim boljše izkoriščanje potenciala za rast digitalnega gospodarstva.

S področjem odprtih podatkov se najtesneje povezuje tretji steber, ki se osredotoča na prost pretok podatkov v EU in računalništvo v oblaku, opredelitev prednostnih nalog glede standardov in interoperabilnosti naprav, programov, odložišč podatkov, storitev in omrežij, ki so ključna za enotni digitalni trg, ter podporo vključujoči digitalni družbi, v kateri imajo državljani potrebne spretnosti za izkoriščanje interneta in izboljšanje možnosti za zaposlitev.

Strategija za izvajanje Evropskega okvira interoperabilnosti (EIF) (Evropska komisija, 2017) opredeljuje interoperabilnost kot ključni dejavnik, ki omogoča digitalno preobrazbo in glavni pogoj za omogočanje elektronske komunikacije in izmenjave informacij med javnimi upravami ter s tem uresničujev enotnega digitalnega trga. Vpliv interoperabilnosti sega od pravne do tehnične ravni in ni omejen zgolj na informacijsko in komunikacijsko tehnologijo (IKT).

EIF predstavlja skupno dogovorjen pristop k izvajanju javnih storitev na interoperabilen način in opredeljuje osnovne smernice za interoperabilnost v obliki skupnih načel, modelov in priporočil. Izmed 47 priporočil za interoperabilnost je 24 povezanih z 12 temeljnimi načeli evropskih javnih storitev, ki zajemajo naslednja področja:

1. Subsidiarnost in sorazmernost
2. Odprtost
3. Preglednost
4. Možnost ponovne uporabe
5. Tehnološka nevtralnost in prenosljivost podatkov
6. Osredotočenost na uporabnika
7. Vključevanje in dostopnost
8. Varnost in zasebnost
9. Večjezičnost
10. Upravna poenostavitev
11. Hramba informacij
12. Ocena učinkovitosti in uspešnosti

Konceptualni model za zagotavljanje integriranih javnih storitev države članice usmerja pri načrtovanju, razvoju, izvajanju in vzdrževanju integriranih javnih storitev na vseh ravneh upravljanja, od lokalne do ravni EU. Modularen model zajema ohlapno povezane sestavne dele storitev, ki so med seboj povezani prek skupne infrastrukture in v obliki sklopa uporabnih priporočil predstavlja praktična orodja.

Ambicija vzpostavitve podatkovno okretnega gospodarstva in možnosti aktivnega usmerjanja digitalne preobrazbe ter postavljanje globalnih standardov tehnološkega razvoja je spodbudila oblikovanje digitalne prihodnosti Evrope (Evropska komisija, 2020). Poglavitni cilji strategije so:

- Tehnologija, ki deluje za ljudi
- Pravično in konkurenčno gospodarstvo
- Odprta, demokratična in trajnostna družba

Komisija je med drugim napovedala pospeševanje naložb v gigabitno povezljivost Evrope, digitalno izobraževanje za spodbujanje digitalne pismenosti in kompetenc ter okrepljen program za krepitev digitalnih znanj in spretnosti. Kot ključni ukrep za doseg drugega cilja je komisija predvidela izdelavo Evropske strategije za podatke, ob tretji točki pa spremembe na področju pravil za poglobitev notranjega trga za digitalne storitve.

Uredba o vzpostavitvi programa Digitalna Evropa (Uredba (EU) 2021/694 Evropskega parlamenta in Sveta z dne 29. aprila 2021 o vzpostavitvi programa Digitalna Evropa in razveljavitvi Sklepa (EU) 2015/2240 (Besedilo velja za EGP), 2021) dopolnjuje druge programe EU, ki podpirajo digitalni prehod (npr. Obzorje Evropa), ter področje digitalizacije v Instrumentu za povezovanje Evrope. Digitalna Evropa je program financiranja EU, zasnovan za približevanje digitalne tehnologije podjetjem, državljanom in javni upravi. Njegovi glavni cilji so izboljšanje konkurenčnosti EU v svetovnem digitalnem gospodarstvu, premostitev digitalnega razkoraka v EU ter krepitev evropske zmogljivosti na ključnih področjih digitalne tehnologije. Program se osredotoča na visokozmogljivo računalništvo, umetno inteligenco, kibernetsko varnost in zaupanje, napredne digitalne veščine ter uvajanje in optimalno uporabo digitalnih zmogljivosti ter interoperabilnost.

Leta 2021 je bil pod naslovom Digitalni kompas: evropska pot v digitalno desetletje (Evropska komisija, 2021) vzpostavljen še en večletni program, v katerem so zastavljene digitalne ambicije za leto 2030 ter opredeljeni ključni mejniki in sredstva za uresničitev teh ambicij. Glavna vizija digitalnega kompasa za leto 2030 je opolnomočenje državljanov in podjetij, k čemur so usmerjene štiri točke:

1. Veščine: Digitalno usposobljeno prebivalstvo in visoko usposobljeni strokovnjaki na digitalnem področju
2. Infrastruktura: Varne in učinkovite trajnostne digitalne infrastrukture
3. Poslovanje: Digitalna preobrazba podjetij
4. Vlada: Digitalizacija javnih storitev

V okviru programa so konkretni cilji merjeni na nacionalni ravni in na ravni EU s ključnimi kazalniki uspešnosti, ki so bili napovedani kot del digitalnega kompasa.

4 Podatkovni portali in analitična poročila

Pred vzpostavitvijo Uradnega portala za evropske podatke (EDP) je v okviru EU delovala Evropska platforma za informacije javnega sektorja oz. European Public Sector Information Platform (EPSI), ki je podobno kot EDP letno izdajala tematska poročila o raziskavah podatkov. Uradni portal za evropske podatke (European Data Portal)¹ je bil vzpostavljen novembra 2015 in predstavlja enotno točko dostopa do odprtih podatkov držav Evropske unije, institucij, agencij in organov EU ter drugih evropskih držav. Zaradi velike količine podatkov iz omenjenih virov so v okviru portala pogosto izvedene študije, ki predstavljajo podlago za izboljšavo, popravilo in razvoj portala, rezultati izvedenih raziskav in analiz pa poleg tega služijo tudi razvoju ostalih podobnih portalov in zbirk z odprtimi podatki. Od leta 2015 je bilo vsako leto izdelano poročilo o zrelosti odprtih podatkov, med njimi najaktualnejše je izšlo za leto 2024 in je zajelo informacije o napredku pri delu z odprtimi podatki v posameznih državah članicah EU in nekaterih drugih državah, izbranih na letni ravni (kandidatke za članice, države v Evropskem združenju za prosto trgovino (EFTA), države vzhodnega partnerstva (EaP) ter druge). Poročilo o zrelosti odprtih podatkov je osredotočeno na štiri glavna področja – državno zakonodajo, meritve vpliva ponovne uporabe podatkov, oceno nacionalnih portalov in mehanizme za kvaliteto

¹ <https://data.europa.eu>

(meta)podatkov. Namen tovrstnih poročil je pomagati vsem vključenim državam pri razumevanju stopnje zrelosti svojih odprtih podatkov in opazovanju ter primerjanju lastnega napredka z drugimi, ob tem pa poročila izpostavljajo najboljše prakse, ki bi jih veljalo prenesti v druge nacionalne in lokalne okvire.

Zadnja revizija metodologije za izdelavo poročila je bila izvedena leta 2024 in je ohranila glavne štiri dimenzije ter njihove pripadajoče kazalnike, spremenila pa je nekatere zahteve za podrobnejša poročila in vključila obravnavo naborov podatkov velike vrednosti. Poročilo iz leta 2024 (M. Page idr., 2024) ocenjuje napredek 34 evropskih držav – vključno z 27 državami članicami EU, tremi državami EFTA (Islandija, Norveška, Švica) in štirimi državami kandidatkami (Albanija, Bosna in Hercegovina, Srbija, Ukrajina) – pri spodbujanju dostopnosti in ponovne uporabe podatkov javnega sektorja. Leta 2024 je povprečna stopnja zrelosti odprtih podatkov v državah EU-27 ohranila povprečje preteklega leta (83 %), med državami pa so najvišje uvrščene Francija (100 %), Poljska (98 %) in Slovaška (96 %). Slovenija je dosegla 89,4 %, kar je nad povprečjem EU, njeni rezultati po posameznih dimenzijah pa so bili:

- Državna zakonodaja 93 % (povprečje EU: 91 %)
- Portal: 86,6 % (povprečje EU: 82 %)
- Vpliv ponovne uporabe: 88,8 % (povprečje EU: 80,5 %)
- Kakovost: 89,4 % (povprečje EU: 79,7 %)

Ti rezultati potrjujejo, da Slovenija uspešno krepi svojo strategijo za odprte podatke in se med evropskimi državami uvršča na 14. mesto na tem področju.

EDP redno skrbi tudi za izvajanje poglobljenih raziskav specifičnih področij odprtih podatkov, izsledke raziskav pa objavlja v analitičnih poročilih na portalu. Najnovejša poročila iz leta 2024 med drugim obravnavajo:

1. najboljše prakse naborov podatkov velike vrednosti (Moro & Page, 2024),
2. vpliv umetne inteligence v procesiranju in ustvarjanju novih podatkov (Graux idr., 2024),
3. potrebe po odprtih podatkih za raziskovalce in akademike (Corcho, 2024),
4. granularne sisteme upravljanja odprtih podatkov (Graux, 2024a),
5. lastništvo podatkov pod evropsko podatkovno zakonodajo (Graux, 2024c),

6. poslovne modele za vire javnih odprtih podatkov (Pizzamiglio, 2024),
7. simbiozo varovanja podatkov in odprtih podatkov (Graux, 2024b),
8. geoprostorske trende (Fritze idr., 2024).

V enem od analitičnih poročil iz leta 2022 je bila zrelost odprtih podatkov v Sloveniji podrobneje predstavljena kot primer dobre prakse zaradi strme rasti ocene zrelosti v preteklih letih. Slovenija je tedaj dosegla skoraj popolno oceno (99 %) v kategoriji zakonodaje, vzpostavila je nacionalni portal odprtih podatkov (OPSI), ki služi kot enotna spletna točka za objavo podatkov javnega sektorja, sočasno pa poudarja pomen izobraževanja izdajateljev podatkov. Slovenski pristop poudarja pomen vključevanja različnih deležnikov, vključno z javnimi institucijami, podjetji in raziskovalci, za čim večji izkoristek odprtih podatkov, ter gradnjo aktivnih skupnosti, ki vključujejo tako ponudnike kot uporabnike podatkov. Slovenija je v letu 2022 dosegla pomemben napredek na področju odprtih podatkov tudi s pristopom k Mednarodni listini o odprtih podatkih,² kar potrjuje njeno zavezanost k odprtosti in ponovni uporabi podatkov.

5 Temelji za zagotavljanje kakovosti podatkov

Dolgoletno zbiranje podatkov in vse bolj zgoščeno kopičenje informacij izpostavlja velik pomen ohranjanja kar najvišje stopnje kakovosti zbranih podatkov. V ta namen so bili razviti različni standardi in pristopi, ki ponudnikom podatkov predstavljajo oporo in smernice za zagotavljanje kakovosti, v procesih razvoja metodologij za vrednotenje pa so uporabljeni za prepoznavanje podatkov ali indikacijo stopnje kakovosti.

5.1 Besednjak podatkovnih katalogov DCAT

Besednjak *Data Catalog Vocabulary (DCAT)* (Albertoni idr., 2024) za podatkovni model RDF je bil razvit v okviru organizacije W3C (Konzorcij za svetovni splet) ter ponuja smernice in specifikacije za opisovanje zbirk podatkov (tudi v katalogih podatkov) z namenom izboljšanja prepoznavnosti in interoperabilnosti naborov podatkov na spletu. Gre za standard, ki se lahko uporablja globalno in ni povezan z nobenim regulativnim okvirjem. DCAT je lahko za specifične potrebe različnih

² <https://opendatacharter.org>

podatkovnih portalov dopolnjen z aplikacijskimi profili. Kot specifikacija za opis povezanih javnih podatkov v Evropi je v uporabi aplikacijski profil za evropske podatkovne portale.³ Profil, ki je bil posebej zasnovan za evropske javne uprave, ponuja dodatna navodila in omejitve, s katerimi je zagotovljeno izpolnjevanje evropskih zahtev in potreb. Prav tako je aplikacijski profil usklajen z direktivo INSPIRE, ki določa pravne okvire za vzpostavitev in delovanje evropske infrastrukture za prostorske informacije, s čimer zagotavlja doslednost pri opisovanju geoprostorskih naborov podatkov. DCAT-AP vključuje tudi posebne smernice za zajemanje pravnih in licenčnih informacij v zvezi z nabori podatkov v skladu z evropskimi pravnimi zahtevami za odprte podatke ter zaradi svoje specializacije pripomore k večji interoperabilnosti med državami članicami EU.

Metoda RDF (Resource Description Framework) je namenjena opisovanju podatkov z definiranjem odnosov med podatkovnimi objekti (RDF Working Group, 2014). Za predstavitev informacij uporablja trojice subjekt – predikat – objekt, za opis podatkovnih katalogov pa se pogosto uporablja nabor razredov in lastnosti, ki jih definira DCAT. Sedem glavnih razredov besednjaka je:

1. **dcatalog:Catalog**: zajema metapodatke o zbirkah podatkov in podatkovnih storitvah;
2. **dcatalog:Dataset**: predstavlja zbirko podatkov, ki so lahko na voljo v različnih oblikah (števila, besedilo, slikovne pike, slike, zvok in druge večpredstavnostne vsebine);
3. **dcatalog:Distribution**: predstavlja dostopno obliko zbirke podatkov (datoteka za prenos);
4. **dcatalog:DataService**: predstavlja zbirko operacij, dostopnih prek API vmesnika;
5. **dcatalog:Resource**: je starševski razred razredov `dcatalog:Dataset`, `dcatalog:DataService` in `dcatalog:Catalog` in predstavlja razširitveno točko za definicijo katerekoli vrste kataloga;
6. **dcatalog:DatasetSeries**: je zbirka podatkov za predstavitev ločenih zbirk s skupnimi značilnostmi;
7. **dcatalog:CatalogRecord**: je zapis v katalogu, ki se nanaša na informacije o registraciji podatkov.

³ <https://semiceu.github.io/DCAT-AP/releases/3.0.0/>

Za skladnost z aplikacijskim profilom DCAT-AP je važno upoštevanje smernic, ki so določene v dokumentaciji (Van Nuffelen, 2024), vendar profil kljub omejitvam uporabnikom pušča veliko svobode, da ga prilagodijo svojim potrebam. Primer opisa podatkovne zbirke z uporabo DCAT-AP v sintaksi Turtle:

```
@prefix dcat: <http://www.w3.org/ns/dcat#> .  
@prefix dct: <http://purl.org/dc/terms/> .  
@prefix example-ds: <https://data.gov.gr/id/dataset/> .  
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
  
example-ds:BeePopulation a dcat:Dataset;  
dct:title "Bee population"@en;  
dct:description "A dataset about bee population in Greece"@en;  
dct:publisher <https://agencies.gov.gr/id/GreekEnvironmentAgency> .
```

5.2 Standard ISO 25012

Mednarodna organizacija za standardizacijo (ISO) je razvila serijo standardov, katerih glavni cilj je usmerjati razvoj programskih izdelkov s specifikacijo kakovostnih zahtev in kriterijev za evalvacijo, združeno pod kratico SQuaRE (System and Software Requirements and Evaluation) (Mednarodna organizacija za standardizacijo, 2008). V tej družini standardov so zbrani:

- ISO 2500n - standardi za upravljanje kakovosti,
- ISO 2501n - standardi za opis modela kakovosti,
- ISO 2502n - standardi za merjenje kakovosti,
- ISO 2503n - standardi za kakovostne zahteve,
- ISO 2504n - standardi za preverjanje kakovosti.

Za podatke, shranjene v strukturirani obliki v računalniškem sistemu, je splošni model kakovosti podatkov opredeljen s standardom ISO/IEC 25012:2008. Uporablja se lahko za opredeljevanje kvalitativnih meril ali načrtovanje in izvajanje ocenjevanja kakovosti podatkov. Standard definira 15 karakteristik modela, te pa so razvrščene v dve glavni kategoriji – inherentna in sistemsko odvisna kakovost podatkov. Inherentna kakovost podatkov se nanaša na same podatke in njihov notranji potencial, da zadostujejo izraženim in impliciranim zahtevam, ko so uporabljeni pod določenimi pogoji. Kategorija se nanaša zlasti na podatkovne vrednosti, razmerja med temi vrednostmi ter metapodatke. Kategorija kakovosti

podatkov, ki je odvisna od sistema, pa določa stopnjo, do katere je kakovost dosežena in ohranjena v računalniškem sistemu, ko so podatki uporabljeni pod določenimi pogoji. V tej kategoriji na kakovost vplivajo zmogljivosti komponent računalniških sistemov (strojna in programska oprema).

Tabela 1: Karakteristike kvalitete podatkov po ISO/IEC 25012

Inherentne kakovosti			
Natančnost			
Popolnost	Dostopnost	Natančnost	Dostopnost
Doslednost	Skladnost	Sledljivost	Prenosljivost
Verodostojnost	Zaupnost	Razumljivost	Obnovljivost
Aktualnost	Učinkovitost		
Sistemske odvisne kakovosti			

Vir: lasten

5.2.1 Inherentne karakteristike

- Natančnost: Stopnja, do katere imajo podatki attribute, ki pravilno predstavljajo resnično vrednost predvidenega atributa.
 - Semantična natančnost: Bližina vrednosti podatkov naboru vrednosti, ki po opredelitvah področja velja za semantično pravilnega.
 - Sintaktična natančnost: Bližina vrednosti podatkov naboru vrednosti, ki po opredelitvah področja velja za sintaktično pravilnega.
- Popolnost: Stopnja, do katere podatki vsebujejo vrednosti za vse pričakovane attribute in sorodne primere entitete (ang. related entity instances).
- Doslednost: Stopnja, do katere imajo podatki attribute, ki si ne nasprotujejo in so skladni z drugimi podatki. To velja tako med podatki o eni entiteti kot med podobnimi podatki o primerljivih entitetah.
- Verodostojnost: Stopnja, do katere imajo podatki attribute, ki jih uporabniki obravnavajo kot resnične in verodostojne, kar vključuje principe avtentičnosti (ang. the truthfulness of origins, attributions, commitments).
- Aktualnost: Stopnja, do katere imajo podatki attribute, ki so primerne starosti.

5.2.1 Inherentne in sistemsko odvisne karakteristike

- Dostopnost: Stopnja, do katere je mogoče dostopati do podatkov, zlasti za ljudi, ki zaradi invalidnosti potrebujejo podporno tehnologijo ali posebno konfiguracijo.
- Skladnost: Stopnja, do katere imajo podatki attribute, ki ustrezajo veljavnim standardom, konvencijam ali predpisom in podobnim pravilom v zvezi s kakovostjo podatkov.
- Zaupnost: Stopnja, do katere imajo podatki attribute, ki zagotavljajo, da so dostopni in razpoložljivi samo pooblaščenim uporabnikom v določenem kontekstu uporabe.
- Učinkovitost: Stopnja, do katere imajo podatki attribute, ki jih je mogoče obdelati in zagotoviti pričakovane ravni zmogljivosti z uporabo ustreznih količin in vrst virov.
- Natančnost: Stopnja, do katere imajo podatki attribute, ki so natančni ali omogočajo razlikovanje.
- Sledljivost: Stopnja, do katere imajo podatki attribute, ki zagotavljajo revizijsko sled dostopa do podatkov in vseh sprememb podatkov.
- Razumljivost: Stopnja, do katere imajo podatki lastnosti, ki uporabnikom omogočajo, da jih berejo in razlagajo, ter so izraženi v ustreznih jezikih, simbolih in enotah.

5.2.3 Sistemsko odvisne karakteristike

- Dostopnost: Stopnja, do katere imajo podatki attribute, ki pooblaščenim uporabnikom in/ali aplikacijam omogočajo, da jih pridobijo.
- Prenosljivost: Stopnja, do katere imajo podatki lastnosti, ki omogočajo njihovo namestitve, zamenjavo ali prenos iz enega sistema v drugega z ohranitvijo obstoječe kakovosti.
- Obnovljivost: Stopnja, do katere imajo podatki lastnosti, ki jim omogočajo, da vzdržujejo in ohranjajo določeno raven delovanja in kakovosti tudi v primeru okvare.

5.3 Petzvezdični sistem za ocenjevanje

Leta 2006 je Tim Berners-Lee predstavil svoj sistem za ocenjevanje povezanih podatkov, s katerim je želel spodbuditi predvsem upravljavce vladnih podatkov k dvigu kakovosti podatkov in boljši poveztljivosti (Berners-Lee, 2006). Sistem temelji na standardnih spletnih tehnologijah, kot so HTTP, RDF in URI, ki jih uporablja za deljenje informacij na strojno berljiv način. S tem omogoča povezovanje podatkov iz različnih virov in razširja možnosti semantičnih poizvedb.

Za vpeljavo odprtih podatkov je Berners-Lee (2012) predlagal petzvezdično shemo, pri kateri je zadostovanje pogojem vsake predhodne stopnje predpogoj za izpolnjevanje zahtev naslednje stopnje:

1. Podatki so dostopni na spletu v formatu z odprto licenco.
2. Podatki so na voljo v strojno berljivi obliki (npr. Excel datoteka namesto digitalno preslikane).
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.

5.4 Načela FAIR

Potreba po izboljšavi infrastrukture, ki podpira ponovno uporabo znanstvenih podatkov, je spodbudila razvoj principov za izboljšavo ponovne uporabe zbirk podatkov (Wilkinson idr., 2016). Načela FAIR poleg podpore posameznikom pri ponovni uporabi podatkov dajejo poseben poudarek na izboljšanje strojnih zmognosti za samodejno iskanje in uporabo podatkov. Utemeljena so na štirih osrednjih kriterijih – najdljivost (F – findability), dosegljivost (A – accessibility), interoperabilnost (I – interoperability) in ponovno uporabnost (R – reusability). FAIR opisuje jedrnata, od področja neodvisna načela na visoki ravni, posamezni elementi principov pa so sorodni, vendar neodvisni in ločljivi. Načela, ki uporabljajo oznako »(meta)podatki« so aplikativna tako na nivoju podatkov kot metapodatkov. V nadaljevanju so naštetna načela in posamezni elementi principov:

- Najdljivost
 - F1 (Meta)podatkom je dodeljen globalno unikaten in trajen identifikator.
 - F2 Podatki so opisani z bogatimi metapodatki (definiranimi v R1).
 - F3 Metapodatki jasno in nedvoumno navajajo identifikator podatkov, ki jih opisujejo.
 - F4 (Meta)podatki so registrirani ali indeksirani v viru, ki omogoča iskanje.
- Dostopnost
 - A1 (Meta)podatke je mogoče pridobiti na podlagi identifikatorja z uporabo standardiziranega komunikacijskega protokola.
 - A1.1 Protokol je odprt, brezplačen in univerzalno izvedljiv.
 - A1.2 Protokol po potrebi omogoča postopek avtentikacije in avtorizacije.
 - A2 Metapodatki so dostopni, tudi če podatki niso več na voljo.
- Interoperabilnost
 - I1 (Meta)podatki uporabljajo formalni, dostopni, skupni in široko aplikativni jezik za predstavitev gradiva.
 - I2 (Meta)podatki uporabljajo slovarje, ki sledijo načelom FAIR.
 - I3 (Meta)podatki vključujejo kvalificirane reference na druge (meta)podatke.
- Ponovna uporabnost
 - R1 (Meta)podatki so bogato opisani s številnimi natančnimi in ustreznimi atributi.
 - R1.1 (Meta)podatki so objavljeni z jasno in dostopno licenco za uporabo podatkov.
 - R1.2 (Meta)podatki so povezani s podrobnim izvorom.
 - R1.3 (Meta)podatki izpolnjujejo standarde skupnosti, ki so relevantni za posamezno področje.

6 **Pristopi h kvalitativnim meritvam podatkov**

Razvoj standardov, smernic in zakonodaje na področju kakovosti podatkov in metapodatkov je spodbudil raziskave in oblikovanje novih metodologij in pristopov k ocenjevanju kvalitete.

Šestnajst izbranih člankov predlaga različne metodologije, okvire vrednotenja in/ali metrike za ocenjevanje kakovosti podatkov. Medtem ko se večina teh metod osredotoča izključno na ocenjevanje podatkov, nekatere vključujejo dodatne faze, ki prispevajo k pripravi in nadzoru podatkov. Na primer, Krasikov & Legner (2023) ne predstavlja le metodologije za ocenjevanje podatkov, temveč tudi metodo za pregledovanje in pripravo ocenjenih podatkov za nadaljnjo uporabo, kar procesu dodaja znatno vrednost. Poleg tega tudi pristopa Fadlallah idr. (2023) in Raca idr. (2021) vključujeta fazo priprave, da zagotavljata ustrezno pripravljenost podatkov za ocenjevanje.

Metode se večinoma posvečajo ocenjevanju metapodatkov in podatkov, s čimer je zagotovljeno, da se pri ocenjevanju upoštevata tako struktura kot vsebina podatkovnih zbirk. Izstopa prispevek Molodtsov & Nikiforova (2024) s poudarkom na zasnovi podatkovnih portalov. Slednja lahko pomembno vpliva na dostopnost, uporabnost in celotno uporabniško izkušnjo pri delu s podatki, zato zagotavljanje, da portal izpolnjuje visoke oblikovne in dostopnostne standarde, lahko bistveno izboljša interakcijo uporabnikov s podatki. Poleg tega avtorji Krasikov & Legner (2023) in Kusnirakova idr. (2022) predlagajo specifične metrike za ocenjevanje shem podatkovnih zbirk, ki lahko pripomorejo k strukturni celovitosti podatkov.

V metodologijah se pogosto pojavljajo tri ravni abstrakcije: kategorija, dimenzija in metrika. Po definicijah Debattista idr. (2016) kategorija predstavlja skupino kvalitativnih dimenzij, v katerih je kot indikator uporabljena skupna vrsta informacije. Takšno kategoriziranje omogoča boljšo preglednost vseh vidikov kakovosti podatkov, zlasti pri velikem številu dimenzij. Metrika je konkretno merilo kakovosti za posamezni kazalnik in je običajno povezana z merilnim postopkom, ki vrne oceno oz. vrednost metrike. Posamezna dimenzija ima običajno več metrik, njihova vrednost pa je lahko numerična ali logična (boolean).

Čeprav vse metodologije stremijo k zagotavljanju bolj oprijemljivih informacij o stanju podatkovnih portalov, večinoma v obliki numeričnih ali opisnih ocen, se med seboj razlikujejo v načinu definiranja in poimenovanja dimenzij, ki jih uporabljajo za ocenjevanje. Pomanjkanje standardizacije predstavlja izziv pri primerjanju rezultatov različnih metod in otežuje integracijo ugotovitev iz več študij. Nekateri avtorji poskušajo premagati te izzive z izboljšanjem obstoječih pristopov – za primer, Ferradji & Benchikha (2022) predlagata izboljšano različico dveh formul, predstavljenih v starejših metodah.

Tabela 2: Osnovni podatki o izbranih prispevkih

Referenca	Metoda	Področje	Št. dimenzij	Primer uporabe
Wentzel idr. (2023)	ocenjuje	metapodatki	5	portal EU data.europa.eu
Lämmel idr. (2020)	zbira, ocenjuje	metapodatki	3	odprti vladni podatki (OVP) v Nemčiji
Šlibar & Mu (2022)	ocenjuje	metapodatki	2	OVP v Kanadi, ZDA, Novi Zelandiji
Nogueras-Iso idr. (2021)	ocenjuje	metapodatki	6	OVP v Španiji
Hafidz idr. (2023)	ocenjuje	metapodatki	3	OVP indonezijskih lokalnih oblasti
Krasikov & Legner (2023)	pregleduje, ocenjuje, pripravlja	metapodatki, shema, podatki	7	/
Fadlallah idr. (2023)	pripravlja, ocenjuje	podatki	8	podatki o sevanju libanonske komisije za atomsko energijo
Yan idr. (2023)	ocenjuje	podatki, spremljanje, obdelani podatki	5	/
Wang idr. (2020)	ocenjuje	podatki	6	OVP na Kitajskem in v ZDA
Álvarez Sánchez idr. (2019)	ocenjuje	podatki	8	zdravstveni podatki iz Severne Irske
Kusnirakova idr. (2022)	ocenjuje	podatki, shema	5	češki odprti podatki
Alogaiel & Alrwais (2023)	ocenjuje	podatki	9	OVP v Savdski Arabiji
Bouchelouche idr. (2022)	ocenjuje	podatki	1	OVP v ZDA
Raca idr. (2021)	pripravlja, ocenjuje	podatki	1	6 portalov OVP v Zahodnem Balkanu
Ferradji & Benchikha (2022)	ocenjuje	podatki	2	Wikidata
Molodtsov & Nikiforova (2024)	ocenjuje	portal	9	33 nacionalnih portalov v EU in GCC

Vir: lasten

6.1 Kakovost metapodatkov

Pet člankov je bilo usmerjenih v metapodatke in je poudarilo njihov pomen pri zagotavljanju visoke uporabne vrednosti podatkovnih zbirk. Wentzel idr. (2023) so ocenili kakovost metapodatkov v okviru standarda DCAT-AP. Svoje dimenzije ocenjevanja so opredelili na podlagi štirih načel FAIR in petzvezdične lestvice odprtosti podatkov ter dodali peto dimenzijo, kontekstualnost. Razvili so tudi razširljiv postopek meritve in implementirali svojo metodologijo v obliki storitve Piveau Metrics. Metodologija *Metadata Quality Assessment (MQA)* za preučevanje kakovosti metapodatkov, zbranih na portalu EDP, je neposredno povezana z delom Wentzel idr. Temelji namreč na enakih dimenzijah – štirih, ki izhajajo iz načel FAIR, ter dimenziji kontekstualnosti – ter preverja stanje izbranih indikatorjev na podlagi enakih gesel DCAT-AP. V večini primerov metodologija *MQA* ocenjuje zgolj prisotnost oz. odsotnost indikatorja med metapodatki, končno oceno pa tvori seštevek delnih ocen opisanih petih kategorij. Najvišja možna vsota za posamezno zbirko podatkov znaša 405 točk, točke pa so opredeljene tudi z opisno oceno po vnaprej določenem ključu (odlično, dobro, zadostno ali slabo).

Lämmel idr. (2020) opisujejo pristop k zbiranju metapodatkov in vrednotenju zbranih metapodatkov. Njihova ocena je bila osredotočena na popolnost zahtevanih informacij, razpoložljivost povezave URL in splošno skladnost s predvideno shemo. Poudarili so, da je mogoče najdljivost, pomembno načelo FAIR, bistveno izboljšati s popolnostjo in kakovostjo metapodatkov.

Šlibar & Mu (2022) predstavljata postopek pregleda skladnosti metapodatkov s smernicami za objavljanje na portalih odprtih vladnih podatkov in izračunov ocen popolnosti in konsistentnosti metapodatkov.

Nogueras-Iso idr. (2021) so ocenjevali kakovost geografskih metapodatkov na dva načina. Prva metoda je temeljila na standardu ISO 19157 in izbranih šestih dimenzijah – popolnost, položajna točnost, tematska točnost, logična usklajenost, časovna točnost in kakovost prostega besedila –, za drugo pa so izbrali *MQA*. Kot najprimernejši način za predstavitev rezultatov so označili RDF besednjak *Data Quality Vocabulary*.

Ogrodje *Open Data Portal Quality* (Hafidz idr., 2023) se osredotoča na dva aspekta kakovosti: odprtost podatkov in transparentnost. V sklopu teh dveh kategorij so se osredotočili na dimenzije obstoj, skladnost in odprti podatki ter dve, ki so jih naknadno izključili (natančnost in dostopnost). Poddimenzije so računali na podlagi prisotnosti ključev DCAT, za utežitev dimenzij in določitev končnih rezultatov pa so avtorji uporabili analitični hierarhični proces.

6.2 Kakovost podatkov

Model *BIGQA (Big data Quality Assessment)* (Fadlallah idr., 2023) so razvili za ocenjevanje kakovosti velikih naborov podatkov na podlagi izbranih 8 karakteristik iz standardov ISO/IEC 15939 in ISO/IEC 25000 – natančnost, popolnost, konsistentnost, verodostojnost, aktualnost, skladnost, natančnost in razumljivost. Za učinkovito obdelavo velikih podatkovnih datotek *BIGQA* uporablja paralelno obdelavo, delovanje modela pa so prikazali s poročili o kakovosti podatkov v za to razviti aplikaciji.

Na področju odprtih podatkov iz več virov so Yan idr. (2023) presegli zgolj kvalitativno vrednotenje podatkov in predlagali kazalnike za spremljanje podatkov in metode obdelave, ki bi zagotovili, da se vrednost podatkov med obdelavo ohrani. Kakovost podatkov so ocenili na podlagi petih ključnih dimenzij: celovitost, ustreznost, razpoložljivost, pravočasnost in legitimnost.

Na podlagi preučevanja metode so Wang idr. (2020) ocenjevanje in primerjavo dveh sorodnih zbirk podatkov o gozdovih osredotočili na dimenzije odprtosti, varnosti, celovitosti, trajnosti in razpoložljivosti podatkov ter na metapodatkovne elemente in standarde.

Orodje *TAQIH* (Álvarez Sánchez idr., 2019) za ocenjevanje kakovosti tabelaričnih podatkov se osredotoča na dimenzije natančnosti, popolnosti, dostopnosti, doslednosti, odvečnosti, berljivosti, uporabnosti in zaupanja, pri čemer se prve štiri dimenzije ujemajo tudi s standardi ISO.

Za specifični primer uporabe so Kusnirakova idr. (2022) prilagodili obstoječo metodo, pri čemer so uporabili pet dimenzij kakovosti, vključno z natančnostjo, popolnostjo (sheme in podatkov) ter doslednostjo. Te dimenzije sovpadajo s

standardi ISO in hkrati s formalnimi zahtevami čeških standardov za vladne odprte podatke, ki so bili predmet preučevanja.

Za ocenjevanje odprtih vladnih podatkov v Savdski Arabiji sta Alogaiel & Alrwais (2023) predstavila devet dimenzij, na katere sta se osredotočila v pregledu podatkovnih zbirk: popolnost, granularnost, pravočasnost, strojna uporabljivost, doslednost, natančnost, razumljivost, uporaba in odvečnost. Vsako dimenzijo, vključno s poddimenzijami, sta podrobno obrazložila in podala formule za izračun ocene. Za obdelavo algoritmov in izračun ocen za določene dimenzije sta uporabila platformo RapidMiner, končno oceno podatkovnih zbirk pa sta predstavila na lestvici od 0 do 100, pri čemer so bile ocene posamezne dimenzije različno obtežene.

Manj študij se je, v nasprotju z ostalimi, ki so obravnavale bolj zaokrožen pogled na kakovost podatkov, osredotočilo na ocenjevanje zgolj ene kvalitativne dimenzije.

Avtorji Bouchelouche idr. (2022) predlagajo odstotkovno lestvico za metrike, ki je namenjena ocenjevanju dostopnosti portalov odprtih vladnih podatkov. Osredotočajo se na preučevanje možnosti dostopa, razpoložljive formate, licenciranje in pravočasnost. Študija Raca idr. (2021) se posveča dimenziji odprtosti odprtih vladnih podatkov in ocenjuje, do katere mere so podatki razpoložljivi, dostopni, jih je mogoče najti in so pravočasno na voljo. Za končno oceno predlagajo povezavo med povprečjem navedenih štirih metrik in oceno odprtosti formata na podlagi petzvezdnične lestvice Berners-Leeja. Predlagan pristop so implementirali in testirali kot spletno storitev. Podobno sta se Ferradji & Benchikha (2022) osredotočila zgolj na časovno dimenzijo podatkov, natančneje na njihovo aktualnost in volatilnost, z namenom spodbujanja učinkovitejšega povezovanja odprtih podatkov.

6.3 Kakovost portalov

Za razliko od večine pristopov vrednotenja, ki se osredotočajo na podatke ali metapodatke, sta Molodtsov & Nikiforova (2024) predlagala edinstven pristop, usmerjen v oceno portalov odprtih podatkov. Oblikovanje in funkcionalnost teh portalov igrata pomembno vlogo pri učinkoviti ponovni uporabi podatkov. Njihov predlagani okvir vključuje 72 poddimenzij, razporejenih v devet ključnih dimenzij: večjezičnost, navigacija, splošna uspešnost portala, razumljivost podatkov, kakovost

podatkov, najdljivost podatkov, sodelovanje z javnostjo, mehanizmi in kakovost storitev povratnih informacij ter trajnost portala in sodelovanje na regijski in mednarodni ravni. Za oceno večine poddimenzij predlagajo uporabo vrednosti boolean, medtem ko morajo nekatere doseči zahtevan procentualni prag, da osvojijo točko.

V širšem okviru je največji doprinos predstavljalo delo, ki na podlagi obsežnega pregleda raziskav predlaga celostno metodologijo (Krasikov & Legner, 2023). Avtorja sta se namreč osredotočila na nadaljnjo uporabo odprtih podatkov in zato proces vrednotenja razširila dopolnila z dodatnimi fazami:

1. Konceptualizacija primerov uporabe
2. Identifikacija ustreznih odprtih podatkov
3. Tristopenjsko vrednotenje in analiza vsebine podatkovne zbirke
4. Semantična dokumentacija in integracija zbirk odprtih podatkov z internimi podatki

Vrednotenje torej zajema tri nivoje podatkovne zbirke – metapodatke, shemo in samo vsebino. Izbrane kvalitativne dimenzije so utemeljene na raziskavah Neumaier idr. (2016); Vetrò idr. (2016); Zhang idr. (2003):

1. **Metapodatki** so ovrednoteni na podlagi pričakovanih metapodatkovnih elementov:
 - a. splošne informacije (npr. format, avtorizacija za dostop, storitev iskanja),
 - b. licenciranje,
 - c. podrobnosti o izdaji (npr. izdajatelj, datum, pogostost posodabljanja),
 - d. informacije o vsebini (npr. jezik, geografska pokritost, število zapisov, atributov).
2. Vrednotenje **sheme** temelji na prisotnosti zahtevanih atributov.
3. Oceno **vsebine podatkov** sestavljajo:
 - a. splošna izpolnjenost celic (% praznih celic v celotni zbirki),
 - b. unikatnost vrstic (% podvojenih vrstic),
 - c. izpolnjenost ključnih atributov (% praznih celic v posameznem stolpcu),
 - d. skladnost z metapodatki (% celic, ki ustrezajo specifikacijam),
 - e. skladnost s formatom (% celic, ki ustrezajo definiranimu formatu).

Avtorja tudi poudarjata, da je uspešnost metode močno odvisna od natančne dokumentacije podatkovnih zbirk, kar je pomembno za nadaljnjo uporabo podatkov (npr. pri povezovanju zbirk). Zaključna faza integracije podatkov je pomembna zlasti v primeru, kadar je metoda uporabljena z vnaprej definiranimi cilji ponovne uporabe.

6.4 Zgodnje raziskave

V pregledu smo ugotovili, da številne študije izhajajo iz predhodnih raziskovalnih del. Zato vključujemo pet pogosto citiranih prispevkov, povezanih z obravnavanimi temami, saj ti predstavljajo temelj za razvoj številnih bolj specifičnih metodoloških pristopov.

Debattista idr. (2016) so za potrebe vrednotenja povezanih odprtih podatkov vzpostavili ogrodje Luzzu, ki je bilo nato preizkušeno na devetih povezanih zbirkah podatkov iz podatkovnega prostora 270a.info. Luzzu temelji na štirih ciljih – razširljivosti, interoperabilnosti, nadgradljivosti in prilagodljivosti –, ki se odražajo v glavnih komponentah ogrodja:

1. uporaba tokovnega procesiranja, ki omogoča razširljivost;
2. nabor ontologij za zajem informacij o kakovosti (vključno z merili kakovosti, problemi in poročili), ki ga je mogoče ponovno uporabiti v drugih semantičnih okvirih in orodjih;
3. preprost, deklarativen, domensko-specifičen jezik za integracijo različnih meril kakovosti omogoča nadgradnjo;
4. prilagodljiv algoritem za rangiranje zbirk podatkov upošteva uporabniško določena merila.

Za abstraktni prikaz kvalitete podatkov so uporabili *Dataset Quality Vocabulary* (daQ) (Debattista idr., 2014) in na podlagi besednjaka izbrali dva in dvajset kvalitativnih meril v devetih različnih dimenzijah.

Neumaier idr. (2016) so razvili ogrodje za avtomatizirano vrednotenje kvalitete metapodatkov s pretvorbo zapisov v standardizirano shemo DCAT, ki deluje neodvisno od programskega zaledja. Na podlagi obstoječih pristopov k homogenizaciji metapodatkov so metapodatke večjih sistemov za objavljanje odprtih podatkov, kot so CKAN, Socrata in OpenDataSoft, preslikali v shemo

DCAT. Za vrednotenje metapodatkov so uporabili pet dimenzij in 18 metrik, ki so jih povezali s ključnimi besedami iz besednjaka DCAT in jim določili funkcije za izračun. V obsežnih raziskavah so nato ovrednotili prek 260 portalov odprtih podatkov ter skupno prek milijon podatkovnih zbirk. V ta namen so razvili dve ogrodji – Open Data Portal Watch (Neumaier idr., 2016) in Open Data Portal Quality (Kubler idr., 2018), ki sta služili nadzoru, vrednotenju in razvrščanju izbranih portalov. Z uporabo kvalitativnih dimenzij so ovrednotili stanje portalov, rezultate pa so naknadno še natančneje analizirali na podlagi analitičnega hierarhičnega procesa, ki omogoča prilagajanje različnim uporabniškim preferencam (Kubler idr., 2018). Čeprav se je študija osredotočila predvsem na metapodatke portalov, so avtorji poudarili, da neustrezni opisi ali klasifikacije podatkovnih nizov neposredno vplivajo na uporabnost in možnost iskanja virov.

Vetrò idr. (2016) so predstavili ogrodje za ocenjevanje kakovosti odprtih vladnih podatkov, ki so ga razvili za potrebe ocenjevanja portalov odprtih vladnih podatkov. Nizka kakovost podatkov po njihovem mnenju namreč otežuje dostop do informacij in povečuje stroške ter pogosto vodi v decentralizirana in neusklajena prizadevanja za nadaljnjo uporabo odprtih podatkov, hkrati pa se obstoječi pristopi ne posvečajo vrednotenju kakovosti podatkov na najbolj granularni ravni, torej na ravni posameznih atributov in celic. Teoretično podlago njihovega ogrodja je predstavljala metodologija *SPDQM (SQuARE-Aligned Portal Data Quality Model)* (Moraga idr., 2009). Na podlagi raziskave med strokovnjaki so iz *SPDQM* izbrali nabor kvalitativnih karakteristik, ki so najbolj neposredno naslovile izpostavljene problematike, pri definiranju metrik pa so upoštevali principe merljivosti, interpretabilnosti, združljivosti in izvedljivosti. Vsako izmed 14 metrik, razvrščenih v sedem dimenzij (sledljivost, aktualnost, pretečenost, popolnost, skladnost, razumljivost, natančnost), so označili kot objektivno (kvantitativne metrike) oz. subjektivno (kvalitativne metrike), vse pa so jasno definirane, normalizirane in omejene na intervalni lestvici.

Prispevek Zaveri idr. (2013) ni bil pogosto citiran, ker bi predlagal novo metodologijo, temveč zaradi obsežnega pregleda dimenzij kakovosti podatkov in zaradi celovite analize stanja ocenjevanja kakovosti podatkov v literaturi. Ta sistematični pregled je bil izveden v času hitre rasti povezanih odprtih podatkov (Linked Open Data), ki je bila posledica napredka tehnologij semantičnega spleta. Avtorji so različne pristope organizirali v enotno klasifikacijsko shemo, obravnavali

terminološke nedoslednosti in predstavili seznam 23 definicij dimenzij kakovosti podatkov, od katerih so mnoge še danes uporabne, čeprav nekatere zahtevajo posodobitve in dopolnitve. Poleg tega so 21 metodologij kvalitativno analizirali, primerjali so pristope ter identificirali najpogostejše uporabljene dimenzije, vrste podatkov, ki jih podpirajo, in orodja, ki so bila osnovana na teh metodologijah.

7 Dimenzije in njihove definicije

V spodnji tabeli so navedene dimenzije, ki so se najpogostejše pojavljale v obravnavani literaturi, ter definicije, ki smo jih povzeli iz navedenih virov.

Tabela 3: Pogosto uporabljene dimenzije in njihove definicije

Dimenzija	Definicija	Vir
Berljivost	Podatki so jasni in uporabniki jih zlahka razumejo brez tehničnega predznanja.	Alogaiel & Alrwais (2023)
Doslednost	Podatki so v skladu s standardi zbirke in podatkovne vrednosti se ne spreminjajo v različnih verzijah iste zbirke.	Alogaiel & Alrwais (2023)
Jedratost/redundanca	Shema ali podatki ne vsebujejo odvečnih entitet, ni podvojenih podatkov in nepotrebnih podrobnosti.	Alogaiel & Alrwais (2023); Zaveri idr. (2013)
Licenciranje	Odobritev dovoljenja potrošniku za ponovno uporabo nabora podatkov pod določenimi pogoji.	Zaveri idr. (2013)
Natančnost	Podatki so sintaktično in semantično pravilni in predstavljajo resnična dejstva.	Fadlallah idr. (2023); Zaveri idr. (2013)
Popolnost	Prisotne so vse zahtevane informacije v določenem nizu podatkov.	Alogaiel & Alrwais (2023); Álvarez Sánchez idr. (2019); Fadlallah idr. (2023); Zaveri idr. (2013)
Pravočasnost	Podatki so primerno časovno posodobljeni.	Alogaiel & Alrwais (2023); Zaveri idr. (2013)
Razpoložljivost	Obseg podatkov, v katerem so informacije dostopne in pripravljene za uporabo.	Zaveri idr. (2013)
Razumljivost	Podatki so izraženi z ustreznimi simboli in enotami ter jih je mogoče nedvoumno razumeti in uporabljati.	Fadlallah idr. (2023); Zaveri idr. (2013)
Strojna berljivost	Podatki so na voljo v formatu, ki ga uporabnik lahko strojno prebere.	Alogaiel & Alrwais (2023)
Uporabnost	Lastnost informacij, ki uporabniku prinaša konkretno prednost pri uporabi podatkov.	Alogaiel & Alrwais (2023); Álvarez Sánchez idr. (2019)

Vir: lasten

8 Problematike

Velik del raziskav, ki obravnava kvalitativno vrednotenje podatkovnih zbirk, je metodologije osredotočil na kvaliteto objavljenih metapodatkov. Ker metodologije večinoma niso bile integrirane v celovit proces priprave podatkov za njihovo ponovno uporabo v podjetjih, je njihova aplikativna vrednost omejena (Krasikov & Legner, 2023). Omenjeni celoviti proces so namreč že Zhang idr. (2003) označili za temelj analize podatkov, ki zajema tehnike zbiranja, integracije, transformacije in prečiščevanja podatkov, in čeprav so številne metodologije nezadostno ali sploh ne integrirane, imajo kljub pomanjkljivostim pozitiven vpliv na uporabniško izkušnjo pri ponovni uporabi odprtih podatkov in jih zato ne gre zanemariti pri oblikovanju novih metodologij.

Pri pregledu obstoječih metod za kvalitativno ocenjevanje na primeru savdskoarabskega portala odprtih vladnih podatkov, sta Alogaiel & Alrwais (2023) ugotovila, da je sedanji okvir neučinkovit, ker ni jasnih indikatorjev visokakovostnih podatkov. Za izboljšanje učinkovitosti sta predlagala stalno spremljanje in vrednotenje zbirk, izboljšanje iskalnih zmogljivosti portala, vključitev vizualnih predstavitev (kot so zemljevidi in grafikoni) ter večjo ažurnost in celovitost podatkov.

Prizadevanja za izboljšanje portalov z vidika načel FAIR na data.europa.eu v okviru raziskave Wentzel idr. (2023) so v enem letu obrodila le minimalen napredek. Avtorji so izpostavili potrebo po boljšem sodelovanju ponudnikov podatkov, težavo pa je predstavljal predvsem aspekt popolnosti metapodatkov. Raziskava Šlibar & Mu (2022) je pokazala, da na ameriškem portalu odprtih vladnih podatkov v objavljenih zbirkah manjka skoraj 20 % zahtevanih polj, novozelandski portal se je slabo odrezal na področju skladnosti obveznih in neobveznih polj, najbolj zamudno pa je bilo ocenjevanje skladnosti s smernicami za objavo podatkov, saj je dokumentacija na portalih težko najdljiva in ni na voljo v strojno berljivih formatih. Na področju velikih podatkov je bila izpostavljena problematika, povezana z obsegom, hitrostjo obdelave in raznolikostjo podatkov ter hkrati poudarjen problematični del obstoječih tehnik vodilnih tehnoloških ponudnikov, kot je generalizacija kontekstov in operacij pri ocenjevanju kakovosti (Fadlallah idr., 2023).

Študije so izkazale tudi večjo potrebo po izboljšanih metapodatkovnih standardih in političnih okvirih. Wang idr. (2020) so v podatkih o gozdovih v ZDA odkrili nejasne sporazume o avtorizaciji podatkov in neustrezne varnostne ukrepe za podatke. Na podlagi tega so priporočili razvoj odprtokodnih platform, izboljšanje strojne berljivosti podatkov in spodbudili skladnost z mednarodnimi metapodatkovnimi standardi. Kusnirakova idr. (2022) so v študiji sicer odkrili le zmerne pomanjkljivosti v dimenziji popolnosti podatkov in konsistentnosti tipov podatkov, medtem ko sta natančnost in popolnost sheme pokazali znatne pomanjkljivosti.

Na področju oblikovanja portalov odprtih vladnih podatkov so Raca idr. (2021) opozorili na nedosledne formate datotek in pomanjkanje enotnega standarda za objavo virov in naborov podatkov. Ti izzivi ovirajo dostopnost in interoperabilnost podatkov ter uporabnikom otežijo pridobivanje ter učinkovito uporabo podatkov.

Priznanje

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

Viri in literatura

- Albertoni, R., Browning, D., Cox, S. J. D., Gonzales Beltran, A., Perego, A., & Winstanley, P. (2024). *Data Catalog Vocabulary (DCAT)—Version 3*. <https://www.w3.org/TR/vocab-dcat-3/>
- Alogaiel, N. F., & Alrwais, O. A. (2023). An Assessment of the Quality of Open Government Data in Saudi Arabia. *IEEE Access*, 11, 61560–61599. <https://doi.org/10.1109/ACCESS.2023.3285611>
- Álvarez Sánchez, R., Beristain Iraola, A., Epelde Unanue, G., & Carlin, P. (2019). TAQIH, a tool for tabular data quality assessment and improvement in the context of health data. *Computer Methods and Programs in Biomedicine*, 181, 104824. <https://doi.org/10.1016/j.cmpb.2018.12.029>
- Berners-Lee, T. (2006). *Linked Data—Design Issues*. <https://www.w3.org/DesignIssues/LinkedData.html>
- Berners-Lee, T. (2012). *5-star Open Data*. <https://5stardata.info/en/>
- Bouchelouche, K., Ghomari, A. R., & Zemmouchi-Ghomari, L. (2022). Enhanced analysis of Open Government Data: Proposed metrics for improving data quality assessment. V *ISLA 2022—International Symposium on Informatics and its Applications*. <https://doi.org/10.1109/ISIA55826.2022.9993482>
- Corcho, O. (2024). *Open data needs for researchers and academics*. Urad za publikacije Evropske unije. <https://data.europa.eu/doi/10.2830/603563>
- Debattista, J., Auer, S., & Lange, C. (2016). Luzzu—A Methodology and Framework for Linked Data Quality Assessment. *Journal of Data and Information Quality*, 8(1), 4:1-4:32. <https://doi.org/10.1145/2992786>
- Debattista, J., Lange, C., & Auer, S. (2014). Representing dataset quality metadata using multi-dimensional views. *ACM International Conference Proceeding Series, 2014-September*(September), 92–99. <https://doi.org/10.1145/2660517.2660525>

- Direktiva 2003/98/ES Evropskega parlamenta in Sveta z dne 17. novembra 2003 o ponovni uporabi informacij javnega sektorja, Evropski parlament in Svet, 13, 032 DD (2003). <http://data.europa.eu/eli/dir/2003/98/oj/slv>
- Direktiva 2007/2/ES Evropskega parlamenta in Sveta z dne 14. marca 2007 o vzpostavitvi infrastrukture za prostorske informacije v Evropski skupnosti (INSPIRE), Evropski parlament in Svet, 108 OJ L (2007). <http://data.europa.eu/eli/dir/2007/2/oj/slv>
- Direktiva 2013/37/EU Evropskega parlamenta in Sveta z dne 26. junija 2013 o spremembi Direktive 2003/98/ES o ponovni uporabi informacij javnega sektorja, Evropski parlament in Svet, 175 OJ L (2013). <http://data.europa.eu/eli/dir/2013/37/oj/slv>
- Direktiva (EU) 2019/1024 Evropskega parlamenta in Sveta z dne 20. junija 2019 o odprtih podatkih in ponovni uporabi informacij javnega sektorja, Evropski parlament in Svet, 172 OJ L (2019). <http://data.europa.eu/eli/dir/2019/1024/oj/slv>
- Evropska komisija. (2010). *Evropska digitalna agenda*. Urad za publikacije Evropske unije. <https://eur-lex.europa.eu/legal-content/SL/TXT/?uri=CELEX:52010DC0245&qid=1744008275152>
- Evropska komisija. (2015). *Strategija za enotni digitalni trg za Evropo*. Urad za publikacije Evropske unije. <https://eur-lex.europa.eu/legal-content/SL/TXT/?uri=CELEX%3A52015DC0192&qid=1698141732808>
- Evropska komisija. (2017). *Evropski okvir interoperabilnosti – strategija za izvajanje*. Urad za publikacije Evropske unije. <https://eur-lex.europa.eu/legal-content/SL/TXT/HTML/?uri=CELEX:52017DC0134&from=FI>
- Evropska komisija. (2020). *Oblikovanje digitalne prihodnosti Evrope*. Urad za publikacije Evropske unije. <https://eur-lex.europa.eu/legal-content/SL/TXT/?uri=CELEX:52020DC0067>
- Evropska komisija. (2021). *Digitalni kompas do leta 2030: Evropska pot v digitalno desetletje*. Urad za publikacije Evropske unije. <https://eur-lex.europa.eu/legal-content/sl/TXT/?uri=CELEX:52021DC0118>
- Fadlallah, H., Kilany, R., Dhayne, H., El Haddad, R., Haque, R., Taher, Y., & Jaber, A. (2023). BIGQA: Declarative Big Data Quality Assessment. *Journal of Data and Information Quality*, 15(3). <https://doi.org/10.1145/3603706>
- Ferradi, M. A., & Benchikha, F. (2022). Enhanced metrics for temporal dimensions toward assessing Linked Data: A case study of Wikidata. *Journal of King Saud University - Computer and Information Sciences*, 34(8, Part A), 4983–4992. <https://doi.org/10.1016/j.jksuci.2021.05.010>
- Fritze, H., Kügeler, A., Jirka, S., & Fechner, T. (2024). *2024 Geospatial Trends*. Urad za publikacije Evropske unije.
- Graux, H. (2024a). *Report on granular data governance systems for open data: A primer on the impact of the data governance act on open data ecosystems*. Urad za publikacije Evropske unije. <https://data.europa.eu/doi/10.2830/796169>
- Graux, H. (2024b). *The Symbiosis between Data Protection and Open Data*. Urad za publikacije Evropske unije.
- Graux, H. (2024c). *What is data ownership, and does it still matter under EU data law?: An exploration of traditional concepts of data ownership, and of the expected impact of the Data Act*. Urad za publikacije Evropske unije. <https://data.europa.eu/doi/10.2830/052185>
- Graux, H., Gryffroy, P., Gad-Nowa, M., & Boghaert, L. (2024). *The role of artificial intelligence in processing and generating new data: An exploration of legal and policy challenges in open data ecosystems*. Urad za publikacije Evropske unije. <https://data.europa.eu/doi/10.2830/412108>
- Hafidz, I., Adzanni, G. A., & Aini Rakhmawati, N. (2023). Open Data Portal Quality (ODPQ) Framework Based Metric for Assessing the Quality of Open Data Portals in Indonesian Local Governments. 2023 *International Conference on Smart-Green Technology in Electrical and Information Systems*, 127–132. <https://doi.org/10.1109/ICSGTEIS60500.2023.10424389>
- Huyer, E., & Knippenberg, L. V. (2020). *The economic impact of open data: Opportunities for value creation in Europe*. Urad za publikacije Evropske unije. <https://doi.org/10.2830/63132>
- Krasikov, P., & Legner, C. (2023). A Method to Screen, Assess, and Prepare Open Data for Use: *Journal of Data and Information Quality*, 15(4), 1–25. <https://doi.org/10.1145/3603708>

- Kubler, S., Robert, J., Neumaier, S., Umbrich, J., & Le Traon, Y. (2018). Comparison of metadata quality in open data portals using the Analytic Hierarchy Process. *Government Information Quarterly*, 35(1), 13–29. <https://doi.org/10.1016/j.giq.2017.11.003>
- Kusnirakova, D., Ge, M., Walletzky, L., & Buhnova, B. (2022). Interoperability-oriented Quality Assessment for Czech Open Data. V A. Cuzzocrea, O. Gusikhin, W. VanDerAalst, & S. Hammoudi (Ur.), *Proceedings of the 11th International Conference on Data Science, Technology and Applications* (str. 446–453). INSTICC. <https://doi.org/10.5220/0011291900003269>
- Lämmel, P., Dittwald, B., Bruns, L., Tcholtchev, N., Glikman, Y., Cuno, S., Flügge, M., & Schieferdecker, I. (2020). Metadata Harvesting and Quality Assurance within Open Urban Platforms. *Journal of Data and Information Quality*, 12(4). <https://doi.org/10.1145/3409795>
- Mednarodna organizacija za standardizacijo. (2008). *Software engineering—Software product Quality Requirements and Evaluation (SQuaRE)—Data quality model (ISO/IEC 25012:2008)*. <https://www.iso.org/standard/35736.html>
- Molodtsov, F., & Nikiforova, A. (2024). An Integrated Usability Framework for Evaluating Open Government Data Portals: Comparative Analysis of EU and GCC Countries. *Proceedings of the 25th Annual International Conference on Digital Government Research*, 899–908. <https://doi.org/10.1145/3657054.3657159>
- Moraga, C., Moraga, M. Á., Calero, C., & Caro, A. (2009). SQuaRE-aligned data quality model for web portals. *Proceedings - International Conference on Quality Software*, 117–122. <https://doi.org/10.1109/QSIC.2009.23>
- Moro, M., & Page, M. (2024). *2023 open data best practices in Europe*. Urad za publikacije Evropske unije. <https://data.europa.eu/doi/10.2830/68229>
- Neumaier, S., Umbrich, J., & Polleres, A. (2016). Automated Quality Assessment of Metadata across Open Data Portals. *Journal of Data and Information Quality*, 8(1), 2:1-2:29. <https://doi.org/10.1145/2964909>
- Nogueras-Iso, J., Lacasta, J., Urena-Camara, M. A., & Ariza-Lopez, F. J. (2021). Quality of Metadata in Open Data Portals. *IEEE Access*, 9, 60364–60382. <https://doi.org/10.1109/ACCESS.2021.3073455>
- Open Knowledge Foundation. (2015). *The Open Definition*. <https://opendefinition.org/>
- Page, M., Behrooz, A., & Moro, M. (2024). *Open data maturity report 2024*. Urad za publikacije Evropske unije. <https://data.europa.eu/doi/10.2830/8656811>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372. <https://doi.org/10.1136/bmj.n71>
- Pizzamiglio, A. (2024). *Exploring business models for public open data resources*. Urad za publikacije Evropske unije. <https://data.europa.eu/doi/10.2830/6425248>
- Raca, V., Velinov, G., Cico, B., & Kon-Popovska, M. (2021). Measuring the Government Openness using an Assessment Tool: Case study of Six Western Balkan Countries. *2021 10th Mediterranean Conference on Embedded Computing*, 1–5. <https://doi.org/10.1109/MECO52532.2021.9460163>
- RDF Working Group. (2014). *RDF - Semantic Web Standards*. <https://www.w3.org/RDF/>
- Šlibar, B., & Mu, E. (2022). OGD metadata country portal publishing guidelines compliance: A multi-case study search for completeness and consistency. *Government Information Quarterly*, 39(4), 101756. <https://doi.org/10.1016/j.giq.2022.101756>
- Uredba (EU) 2021/694 Evropskega parlamenta in Sveta z dne 29. aprila 2021 o vzpostavitvi programa Digitalna Evropa in razveljavitvi Sklepa (EU) 2015/2240 (Besedilo velja za EGP), 166 OJ L (2021). <http://data.europa.eu/eli/reg/2021/694/oj/slv>
- Uredba o posredovanju in ponovni uporabi informacij javnega značaja. (2016). Uradni list RS. <https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2016-01-0994?sop=2016-01-0994>

- Uredba o spremembah in dopolnitvah Uredbe o posredovanju in ponovni uporabi informacij javnega značaja.* (2022). Uradni list RS. <https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2022-01-3683?sop=2022-01-3683>
- Van Nuffelen, B. (2024). *DCAT-AP 3.0*. <https://semiceu.github.io/DCAT-AP/releases/3.0.0/>
- Vetrò, A., Canova, L., Torchiano, M., Minotas, C., 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>
- Wang, B., Wen, J., & Zheng, J. (2020). Research on Assessment and Comparison of the Forestry Open Government Data Quality Between China and the United States. V *Communications in Computer and Information Science: Let. 1179 CCIS* (str. 370–385). https://doi.org/10.1007/978-981-15-2810-1_36
- Wentzel, B., Kirstein, F., Jastrow, T., Sturm, R., Peters, M., & Schimmler, S. (2023). An Extensive Methodology and Framework for Quality Assessment of DCAT-AP Datasets. V *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Let. 14130 LNCS* (str. 262–278). https://doi.org/10.1007/978-3-031-41138-0_17
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Yan, T., You, Z., Zhang, Y., & Hua, R. (2023). Multi-source Open Data Quality Evaluation Model in the Web 3.0 Era. 2023 6th International Conference on Data Science and Information Technology, 203–207. <https://doi.org/10.1109/DSIT60026.2023.00038>
- Zakon o spremembah in dopolnitvah Zakona o dostopu do informacij javnega značaja (ZDIJZ-E).* (2015). Uradni list RS. <https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2015-01-4086/zakon-o-spremembah-in-dopolnitvah-zakona-o-dostopu-do-informacij-javnega-znacaja-zdijz-eh=ZDIJZ>
- Zakon o spremembah in dopolnitvah Zakona o dostopu do informacij javnega značaja (ZDIJZ-G).* (2022). Uradni list RS. <https://www.uradni-list.si/glasilo-uradni-list-rs/vsebina/2022-01-3466/zakon-o-spremembah-in-dopolnitvah-zakona-o-dostopu-do-informacij-javnega-znacaja-zdijz-g?h=ZDIJZ>
- Zaveri, A., Rula, A., Maurino, A., c, R., Lehmann, J., & Auer, S. (2013). Quality Assessment Methodologies for Linked Open Data. *Semantic Web Journal*.
- Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. *Applied Artificial Intelligence*, 17(5–6), 375–381. <https://doi.org/10.1080/713827180>
- Zuiderwijk, A., Janssen, M., Poulis, K., & Kaa, G. V. D. (2015). Open data for competitive advantage: Insights from open data use by companies. *ACM International Conference Proceeding Series*, 27-30-May-2015, 79–88. <https://doi.org/10.1145/2757401.2757411>

O avtorjih

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. **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 Artolog.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 Artolog.si, Tocen.si, and Vodostaji.si.*

