

METODOLOGIJA ZA KVALITATIVNO VREDNOTENJE ODPRTIH PODATKOV

MATEVŽ PESEK, JURE JUVAN, KLARA ŽNIDERŠIČ,
MATIJA MAROLT

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

Glede na hitro naraščajočo količino odprtih podatkov in dejstvo, da so ti široko dostopni, je pomembno, da so podatki ne le dostopni, temveč tudi dosegajo primerno kakovost za nadaljnjo uporabo. Ta je namreč ključna za zagotavljanje pravih, uporabnih in zanesljivih rezultatov, zato je pred ponovno uporabo smiselno oceniti, v kakšnem stanju so podatki. V ta namen smo razvili metodologijo za kvalitativno vrednotenje odprtih podatkov, pri čemer smo elemente metodologije izbrali na podlagi pregleda obstoječih raziskav in pristopov k vrednotenju kakovosti podatkov. Pristop, ki ga predlagamo, ni specifično prilagojen posameznim strokovnim področjem, temveč vključuje splošne kazalnike kakovosti, ki omogočajo široko uporabnost metodologije. V okviru ciljnega raziskovalnega projekta smo metodologijo prilagodili uporabi na portalu Odprtih podatkov Slovenije (OPSI), s čimer želimo spodbuditi uporabo slovenskih odprtih podatkov v različnih raziskovalnih in poslovnih področjih.

DOI
[https://doi.org/
10.18690/um.fov.2.2025.56](https://doi.org/10.18690/um.fov.2.2025.56)

ISBN
978-961-286-963-2

Ključne besede:
odprti podatki,
kakovost podatkov,
podatkovne zbirke,
metodologija,
kvalitativno vrednotenje

DOI
[https://doi.org/
10.18690/um.fov.2.2025.56](https://doi.org/10.18690/um.fov.2.2025.56)

ISBN
978-961-286-963-2

Keywords:

open data,
data quality,
datasets,
methodology,
qualitative assessment

METHODOLOGY FOR QUALITATIVE EVALUATION OF OPEN DATA

MATEVŽ PESEK, JURE JUVAN, KLARA ŽNIDERŠIČ,

MATIJA MAROLT

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

The increasing availability of open data and open data platforms emphasizes the need for reliable evaluation methods to ensure data quality. Effective assessment is critical as the promotion of open data reuse gains momentum, while the quality of this data, often published as a by-product of government processes, can be inconsistent. This paper explores how existing research addresses the challenge of open data assessment and highlights the lack of methodologies applicable to Slovenian open data portals. Existing frameworks are usually limited to assessing either metadata or specific datasets. To fill this gap, we propose a novel methodology based on a comprehensive review of current approaches. This framework is designed to assess multiple dimensions of open data quality and serves as a basis for the development of automated tools for assessing datasets and data portals.



1 Uvod

Pojem odprti podatki predstavlja vse podatke, do katerih lahko vsakdo dostopa, jih uporablja, spreminja ali deli za katerikoli namen. Zaradi razpršenosti podatkov se pojavljajo težave pri njihovem kvalitativnem vrednotenju in ponovni uporabljivosti (Vetrò idr., 2016). Večina raziskav se osredotoča na metapodatke, saj je to področje standardizirano z uporabo RDF slovarjev in različnih lestvic kriterijev, po drugi strani pa so raziskave na polju podatkov samih bistveno manj raziskane zaradi raznolikosti v njihovi vsebini in vsebini podatkovnih zbirk. Problematike nastajajo zaradi pomanjkljivih, nenatančnih, nepopolnih, zastarelih in neveljavnih podatkov, nejasnih vrednosti, prevelike količine informacij za procesiranje in manjka bistvenih podatkov, nekatere izmed njih pa so bolj tehnične narave, kot so neustrezno opredeljeni ali težko dostopni formati, odsotnost standardov in standardne programske opreme za obdelavo odprtih podatkov ter razdrobljenost programske opreme in aplikacij. Ker je večina metodologij iz pregleda literature usmerjena zgolj v formate, je vsebina prepuščena uporabniku, hkrati pa je med seboj težko primerljiva zaradi raznolikosti v modalnosti in uporabljenih tipih.

Ročni pregled podatkovnih zbirk na portalu OPSI je razkril podobne težave, zlasti sintaktične napake (manjkajoče črke, šumniki, posebni znaki), raznolikost formatov ter vprašljivo povezljivost podatkov in ocenjevanje njihovih podatkovnih zbirk. Najpogostejši format PCAXIS ima denimo omejena, zastarela orodja in deluje zgolj na operacijskem sistemu Windows. Glavni cilj je ustvariti koherentno in celovito metodo za učinkovito vrednotenje kakovosti podatkovnih zbirk in njihovih odprtih podatkov za uporabo tako na ravni državnega portala odprtih podatkov OPSI kot potencialno tudi na ravni drugih (evropskih) podatkovnih portalov.

2 Pregled področja

V pregledu literature smo se osredotočili na pregled temeljnih standardov za zagotavljanje kvalitete podatkov, metapodatkovne in podatkovne študije, evropske uredbe na področju odprtih podatkov ter ostale evropske dokumente, slovensko zakonodajo ter analitična poročila na področju odprtih podatkov s portala *data.europa.eu* (Evropski podatkovni portal).

Metapodatkovne študije so področje kvalitativnega ocenjevanja podatkov obravnavale z orodji, kot je *LMZZM* (Debattista idr., 2016), ki je bilo vzpostavljeno za potrebe vrednotenja povezanih odprtih podatkov in temelji na štirih ciljih – razširljivost, interoperabilnost, nadgradljivost in prilagodljivost – ter na metodologiji Dataset Quality Ontology (daQ) (Debattista idr., 2014), ki je namenjena abstraktnemu prikazu podatkov. Med novejša raziskava spada tudi Data Quality Vocabulary (DQV) (Albertoni idr., 2020), ki ga je razvil Konzorcij svetovnega spleta (W3C) in v glavnem temelji na slovarju DCAT.

Podatkovne študije se osredotočajo na problematiko kvalitete podatkov, pri čemer se je zaradi svoje široke zasnove in aplikativnosti kot najrelevantnejša izkazala študija *Open data quality measurement framework: Definition and application to Open Government Data* (Vetrò idr., 2016), od katere smo prevzeli metodološki okvir in 8 metrik. Izpostaviti velja tudi metodologijo *LANG* (Zhang idr., 2019), ki predstavlja dvostopenjsko zasnovan model za vrednotenje sintaktičnih in semantičnih vidikov kvalitete in je bila avtomatizirana. Študija *A Method to Screen, Assess, and Prepare Open Data for Use* (Krasikov idr., 2023) se osredotoča na podjetniški vidik uporabe odprtih podatkov in vključuje sistematični pregled, oceno ter pripravo odprtih podatkov ne zgolj s tehničnega stališča, temveč tudi s stališča koristnosti odprtih podatkov v specifičnih kontekstih.

Najaktualnejša evropska uredba na področju odprtih podatkov, *Direktiva o odprtih podatkih in ponovni uporabi informacij javnega sektorja* (Direktiva (EU) 2019/1024, 2019), podaja predloge za prenovu in dopolnitev področja, pri čemer obravnava razpoložljive formate, zagotavljanje dostopa do dinamičnih podatkov v realnem času z ustreznimi tehničnimi sredstvi, medsebojno povezovanje vzpostavljenih zbirk na več nivojih EU ter povečanje ponudbe dragocenih javnih podatkov za ponovno uporabo, hkrati pa tudi podatke javnih podjetij, raziskovalnih organizacij in organizacij, ki financirajo raziskave.

V slovenski zakonodaji se na naslovno tematiko odprtih podatkov navezuje *Zakon o dostopu do informacij javnega značaja* (ZDIJZ-E, 2015; ZDIJZ-G, 2022) ter *Uredba o posredovanju in ponovni uporabi informacij javnega značaja* (Uredba o posredovanju, 2016; Uredba o spremembah uredbe o posredovanju, 2022), ki v pravni red Republike Slovenije med drugim prenašata zgoraj opisano evropsko *Direktivo (EU) 2019/1024*.

2. Iskanje literature na področju kvalitativnega vrednotenja odprtih podatkov. Uporabili smo ključne besede: »odprti podatki«, »metodologija«, »kvaliteta odprtih podatkov« in »ocenjevanje odprtih podatkov«.
3. Nabor in pregled člankov, ki ustrezajo raziskovalni nameri, pri čemer smo jih natančno predelali ter izpisali metodološke principe, problematike ter predloge za rešitve na danem področju.
4. Sinteza vseh izsledkov v obliki novega metodološkega orodja, ki naslavlja obstoječe problematike in izpostavlja potencialne pomanjkljivosti.
5. Jasna opredelitev dimenzij in metrik, preko izračuna katerih dobimo vrednosti, ki predstavljajo rešitve za aktualne problematike.

Tabela 1: Definicije metrik, ustrezne za portal OPSI

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 %]
Unikatnost	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, vrp: število naključnih vrednosti (črke, številke, simboli itn.)	$m_5 = \left(1 - \frac{vrp}{vr}\right) \cdot 100$	[0 %, 100 %]
	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]
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 %]

Dimenzija	Metrika (<i>m</i>)	Spremenljivke	Formula	Merilo
Uparljivost	Odstotek navedenih podatkovnih tipov	pt: število podatkovnih tipov, ptnm: število navedenih podatkovnih tipov	$m_{10} = \left(1 - \frac{ptnm}{pt}\right) \cdot 100$	[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 (Vetrò idr., 2016)	cn: število celic z napakami (šumniki, posebni znaki, številke itn.), c: število celic	$m_{16} = \left(1 - \frac{cn}{c}\right) \cdot 100$	[0 %, 100 %]
Razumljivost	Jasnost poimenovanja	p: podatki, pt: podatkovni tipi, spsp: 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 %]
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]
	Petvezdični 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 ter URI za identifikacijo, kp: kontekstualizacija podatkov ter 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

3.1 Definicije dimenzij in metrik

Zaradi široke in splošno uporabne zasnove metodologije SPDQM (Vetrò idr., 2016; Moraga idr., 2009) smo iz nje prevzeli nekatere dimenzije, metrike, spremenljivke in formule, pri čemer smo nekatere določili na podlagi ročnega pogleda, druge pa smo pridobili iz ostalih okvirjev, označenih v Tabela 1. Ker predlagana metodologija temelji tako na metodologijah, ki so usmerjene v splošno obravnavo, kot tudi tistih, ki so prilagojene specifičnim kontekstom, lahko pričakujemo ugodne rezultate znotraj konteksta portala OPSI.

3.1.1 Popolnost

Dimenzija, ki določa, do katere stopnje podatki vsebujejo vrednosti za vse pričakovane značilnosti:

- Odstotek popolnih celic.
- Odstotek popolnih vrstic.
- Odstotek popolnih stolpcev.
- Odstotek popolnosti metapodatkov v podatkovnih zbirkah: metrika je izračunana na podlagi števila metapodatkov in števila manjkajočih metapodatkov.
- Odstotek predvidenih vrednosti v celicah: označuje odsotnost naključnih vrednosti v celicah znotraj podatkovnih zbirk. Spremenljivki za izračun metrike sta število vrednosti in število naključnih vrednosti, 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:

- Odstotek nepodvojenih vrstic.
- Odstotek nepodvojenih vrednosti v celicah: izračunan na podlagi podvojenih vrednosti v celicah znotraj podatkovne zbirke.

- Pearsonov korelacijski koeficient: opisuje, do katere stopnje sta v podatkovnih zbirkah povezana dva različna stolpca. Spremenljivki za izračun metrike sta izbrani stolpec in stolpec za primerjavo. Da bi izračunali Pearsonov korelacijski koeficient, kovarianco stolpcev delimo s standardnima odklonoma stolpcev.

3.1.3 Doslednost

Dimenzija, ki določa, do katere stopnje se podatki pojavljajo v enaki obliki:

- **Odstotek pravilno poimenovanih stolpcev:** predstavlja stolpce, ki so poimenovani brez slovničnih napak in s šifrantom oz. besedilom in je izračunan na podlagi števila nepričakovano poimenovanih stolpcev.
- **Odstotek doslednosti podatkovnih tipov:** 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.

3.1.4 Uparljivost

Dimenzija, ki določa, do katere stopnje se podatki med seboj povezujejo:

- **Odstotek navedenih podatkovnih tipov:** v procentih označuje prisotnost navedenih podatkovnih tipov v podatkovnih zbirkah. Število navedenih podatkovnih tipov, na podlagi katerega izračunamo metriko, označuje podatkovne tipe (zaporedna številka, datum, ime občine itd.), ki ne obstajajo znotraj podatkovne zbirke.
- **Odstotek veljavnih podatkovnih tipov:** v procentih označuje prisotnost veljavnih podatkovnih tipov v podatkovnih zbirkah. Število neveljavnih podatkovnih tipov, na podlagi katerega izračunamo metriko, označuje podatkovne tipe (zaporedna številka, datum, ime občine itd.), ki vsebujejo napake v zapisu znotraj podatkovne zbirke.
- **Odstotek strojno berljivih podatkovnih zbirk.**

3.1.5 Ponovna uporabljivost

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

- Odstotek podatkovnih zbirk z licenco.
- Odstotek podatkovnih zbirk s kontaktno točko: 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.
- Odstotek podatkovnih zbirk z informacijo o izdajatelju.

3.1.6 Natančnost

Dimenzija, ki določa, do katere stopnje se podatki zanesljivo pojavljajo v pravilni obliki in ne dopuščajo napak:

- **Odstotek sintaktično natančnih celic:** spremenljivki za izračun metrike sta število celic z napakami pri kodiranju (pretvorba šumnikov in posebnih znakov) in skupno število celic.

3.1.7 Razumljivost

Dimenzija, ki določa, do katere stopnje so podatki razumljivi in enostavni za uporabo:

- **Jasnost poimenovanja:** ocenjuje, do katere subjektivne stopnje je v podatkovnih zbirkah prisotna jasnost poimenovanja. Spremenljivke, ki predstavljajo predpogoj oz. so upoštrevane 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.

- **Odstotek stolpcev v razumljivem podatkovnem tipu:** v procentih označuje prisotnost stolpcev v razumljivem podatkovnem tipu v podatkovnih zbirkah. Število nerazumljivih podatkovnih tipov, ki je poleg števila stolpcev spremenljivka v formuli za izračun metrike, predstavlja podatkovne tipe, pri katerih na podlagi danega imena ni razumljivo označeno, katere podatke vključujejo.

3.1.8 Aktualnost

Dimenzija, ki določa, do katere stopnje so podatki objavljeni pravočasno:

- **Zamuda v objavi:** opisuje, od kdaj do kdaj so bili določeni podatki dostopni.

3.1.9 Skladnost

Dimenzija, ki določa, do katere stopnje so podatki skladni z določenimi standardi:

- **Skladnost z eGMS:** opisuje, do katere mere so podatkovne zbirke skladne z e-Government Metadata Standard.
- **Petzvezdični odprti podatki:** ocenjuje skladnost podatkovnih zbirk s pogoji petzvezdičnega sistema za ocenjevanje, pri čemer je vsaka podatkovna zbirka vrednotena s toliko zvezdicami, kot izpolnjuje pogojev.

3.1.10 Sledljivost

Dimenzija, ki določa, do katere stopnje je mogoče spremljati sledi podatkovnih zbirk:

- **Sled izdelave:** spremlja število sledi o izdelavi podatkovne zbirke.
- **Sled posodobitev:** spremlja število sledi o posodobitvah podatkovne zbirke.

4 Zaključek

Jasno je, da je kvalitativno ocenjevanje odprtih podatkov problematično. Idealu se najbolj približajo Vetrò idr. (2016), ki poskušajo zajeti rešitve, ki bi delovale v splošnem, na ključnih portalih, hkrati pa tudi na posamičnih portalih. Ker smo pri gradnji metodologije upoštevali mnoge predloge iz pregleda literature, lahko trdimo, da metodologija upošteva širok in raznolik spekter problematičnih aspektov. Poleg izdelave nove metodologije je smiselno tudi preveriti njeno učinkovitost z avtomatiziranim orodjem, ki bi tako kot pri metodologiji LANG (Zhang idr., 2019) dalo skoraj identične rezultate ročnemu pregledu, pri čemer bi z manjšim časovnim naporom hkrati zagotovili višjo kvaliteto in zanesljivost rezultatov. Pod to uvrščamo tudi predlog MQA, ki poziva k nujnosti rednega preverjanja podatkov in po zadnjem pregledu še ponovnega preverjanja dostopnosti, s čimer zagotovimo aktualnost in ažurnost podatkovnega portala.

Z ubranim pristopom smo izčrpno in široko zajeli predloge in rešitve tujih metodologij na osnovi različnih podatkovnih portalov. Predlagali smo rešitve za problematike, ki so prisotne na specifičnem portalu OPSI, ter odprli pot za nadaljnje raziskovanje na področju kvalitativnega vrednotenja odprtih podatkov, hkrati pa z metodologijo vzpostavljamo podlago za izdelavo avtomatiziranega orodja za preverjanje podatkovnih zbirk.

Priznanje

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

Literatura

- Albertoni, R., Isaac, A. (2020). Introducing the data quality vocabulary (DQV). *Semantic Web*, 12(1), 81-97. <https://doi.org/10.3233/SW-200382>
- Berners-Lee, T. (2006). Linked data-design issues. <http://www.w3.org/DesignIssues/LinkedData.html>
- Chowdhary, K., Chowdhary, K. R. (2020). Natural language processing. *Fundamentals of artificial intelligence*, 603-649. https://doi.org/10.1007/978-81-322-3972-7_19
- Debatista, J., Auer, S., Lange, C. (2016). Luzzu—a methodology and framework for linked data quality assessment. *Journal of Data and Information Quality (JDIQ)*, 8(1), 1-32. <https://doi.org/10.1145/2992786>

- Debattista, J., Lange, C., & Auer, S. (2014, September). Representing dataset quality metadata using multi-dimensional views. In Proceedings of the 10th International Conference on Semantic Systems (pp. 92-99). <https://doi.org/10.1145/2660517.2660525>
- Evropski parlament and Svet Evropske unije. (2019). Direktiva (EU) 2019/1024 evropskega parlamenta in sveta z dne 20. junija 2019 o odprtih podatkih in ponovni uporabi informacij javnega sektorja (Direktiva (EU) 2019/1024).
- Krasikov, P., Legner, C. (2023). A Method to Screen, Assess, and Prepare Open Data for Use: A Method to Screen, Assess, and Prepare Open Data for Use. *ACM Journal of Data and Information Quality*, 15(4), 1-25. <https://doi.org/10.1145/3603708>
- Moraga, C., Moraga, M. A., Caro, A., & Calero, C. (2009, August). SPDQM: SQuaRE-aligned portal data quality model. In Ninth International Conference on Quality Software, QSIC (pp. 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>
- The Official Portal for European Data – data.europa.eu.
- Uredba o posredovanju in ponovni uporabi informacij javnega značaja (Uredba o posredovanju). (2016).
- Uredba o spremembah in dopolnitvah uredbe o posredovanju in ponovni uporabi informacij javnega značaja (Uredba o spremembah uredbe o posredovanju). (2022).
- 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).
- Zakon o spremembah in dopolnitvah zakona o dostopu do informacij javnega značaja (ZDIJZ-G). (2022).
- Zhang, R., Indulska, M., Sadiq, S. (2019). Discovering data quality problems: the case of repurposed data. *Business & Information Systems Engineering*, 61, 575-593. <https://doi.org/10.1007/s12599-019-00608-0>

