

# Spremljanje kakovosti zapisov dnevne produkcije v COBIB.SI v obdobju od 2015 do 2020

## Quality control of daily bibliographic records production in the period from 2015 to 2020

**Branka Badovinac<sup>1</sup>**

**IZVLEČEK:** V prispevku predstavljamo rezultate analize spremljanja kakovosti podatkov v okviru aktivnosti spremljanja kakovosti zapisov dnevne produkcije v COBIB.SI v obdobju od 2015 do 2020. V vzorec je bilo vključenih 13.478 bibliografskih zapisov s pripadajočimi normativnimi zapisi. Analiza temelji na dimenzijah kakovosti podatkov, ki smo jih pripisali beleženim pomanjkljivostim pregledanih bibliografskih in normativnih zapisov. Predstavili smo posamezne pomanjkljivosti v okviru določenih podatkovnih elementov. V rezultate smo vključili tudi okvirno oceno kakovosti zapisov glede na število poslanih sporočil katalogizatorjem in uspešnost metode spremljanja glede na delež popravljenih zapisov. Ocenili smo, da je približno 70 % pregledanih zapisov dobrih ali zadosti dobrih in da je odzivnost katalogizatorjev razmeroma visoka. V diskusiji smo izpostavili težave pri nekaterih podatkovnih elementih in pomen seznanjanja z novostmi v katalogizacijski praksi.

**KLJUČNE BESEDE:** spremljanje kakovosti dnevne produkcije, kakovost podatkov, COBIB.SI, bibliografski zapisi, normativni zapisi, 2015–2020

**ABSTRACT:** In this paper, the results of an analysis of the data quality performed within quality control of daily bibliographic records production in the period from 2015 to 2020 are presented. The sample included 13,478 bibliographic records with associated authority records. The analysis is based on the data quality dimensions that were attributed to the recorded errors of the reviewed bibliographic and normative records. Also, the individual errors in the context of specific data elements are presented. Furthermore, the results include an indicative quality score of records in terms of the number of messages sent to cataloguers and the performance of the monitoring method in terms of the proportion of records corrected. According to our estimations, some 70% of the records reviewed were good or good enough and the response rate of cataloguers was relatively high. In the discussion, problems with some data elements were highlighted and it was stressed how important it is to keep up to date with changes in the Slovenian cataloguing practice.

**KEYWORDS:** quality control of daily production, data quality, COBIB.SI, bibliographic records, authority records, 2015–2020

---

## 1 Uvod

Brez dvoma so mnogi programi, osredotočeni na (ne)uporabnike, v preteklih desetletjih doprinesli k popularizaciji branja in uporabe knjižnic v Sloveniji. Še vedno pa velja, da je knjižnični katalog največje okno v osnovno poslanstvo knjižnične dejavnosti, ki zajema zbiranje, obdelavo in posredovanje virov. Sodeč po javnomnenjski raziskavi uporabnikov in neuporabnikov splošnih knjižnic je COBISS najbolj prepoznavna spletna storitev (Raziskava, 2020). Z razvojem repozitorija in razširitvijo sistema COBISS na raziskovalno in študijsko okolje

---

<sup>1</sup> Mag. Branka Badovinac, Institut informacijskih znanosti, Maribor, Slovenija, [branka.badovinac@izum.si](mailto:branka.badovinac@izum.si).

je pričakovati, da se bo uporabnost COBISS-a tudi v tej skupini uporabnikov nekoliko zvišala v primerjavi z ugotovitvami iz prejšnjih let (Vilar et al., 2012).

Tudi knjižnični katalogi so se medtem razvijali in v mnogih pogledih prevetrili. Uporabniška opravila v razmerju s knjižničnimi katalogi so okvirno že dolgo znana, a so bila premalo celostno povzeta. Posodobljeni teoretični pristopi so prinesli nov model in pojmovni aparat ter zahteve po tehnološki prenovi. Poglavitni izziv pa je, kako spoznanja prenesti na izvedbeno raven.

Znano je, da večina končnih uporabnikov na podlagi podatkov o naslovu in/ali avtorju v katalogu išče že poznani vir (angl. *known-item search*) (Wakeling, 2017), tako kot je to bilo značilno tudi za čas listkovnih katalogov, ki so se uporabljali predvsem za informacijo o dostopnosti vira (Hafter, 1979). Zato se vrstijo pozivi in razvijajo projekti, ki odpirajo knjižnične baze na svetovni splet in se povezujejo z drugimi deležniki, z namenom izboljšati brskalno funkcijo odkrivanja virov. K temu se pridružujejo še težnje po posodobitvi vsebinske obdelave in po boljši izrabi univerzalnih sistemov predmetnega označevanja ter želja po uvedbi dostopa do celotnih besedil že na vstopni točki kataloga.

Ne glede na stavnico tekme knjižnic s tehnološkimi velikani svetovnega spleta menimo, da imajo knjižnični katalogi prednost vsaj v dveh pogledih.

Prva prednost je edinstvenost podatkov in virov, ki jih ponujajo knjižnični sistemi, tudi iz okolij manj razširjenih jezikov. V teh, enako kot v slovenskem knjižničnem sistemu, je katalogizacija praviloma ročna, podprta z intelektualnim delom katalogizatorjev, saj ni veliko možnosti za prenos podatkov iz drugih partnerskih okolij (npr. iz založniškega). Edinstvenost v vzajemni bibliografski bazi COBIB.SI lahko razumemo tudi v smislu »razširjene slovenike«, saj zajema opise domoznanskega gradiva in dela raziskovalcev za potrebe slovenskega sistema SICRIS. Poleg tega že nekaj več kot tretjino zapisov v vzajemni bazi COBIB.SI predstavljajo zapisi za sestavne dele in članke (Letno, 2021, str. 10), kar kaže na večjo granulacijo opisa virov, s katero so se nekateri tuji in večji online knjižnični katalogi začeli ukvarjati šele nedavno.

Druga prednost knjižničnih katalogov je kredibilnost podatkov, ki temelji na domišljeni praksi natančnega opisovanja za identifikacijo in razlikovanje množice raznovrstnih virov. V tem pogledu je kakovost podatkov pogosto izpostavljen pojem, saj se z njo soočimo ob vsakokratnem upravljanju podatkov in posodabljanju knjižničnih katalogov. Poleg tega vemo, da vsaka napaka neposredno vpliva na zadovoljstvo uporabnika in da je odprava napak v primerjavi s kontrolo vnosa podatkov običajno stroškovno višja (Redman, Fox in Levitin, 2009).

Z namenom razumevanja pomanjkljivosti zapisov v bazi COBIB.SI bomo v prispevku podrobneje predstavili rezultate aktivnosti spremljanja kakovosti dnevne produkcije v obdobju prvih petih let od uvedbe novega načina vzorčenja zapisov iz baze COBIB.SI. V uvodnih dveh poglavjih sta predstavljena pristop h kakovosti podatkov in metoda spremljanja dnevne produkcije, nato sledita pregled analize zabeleženih pomanjkljivosti pregledanih zapisov in okvirna ocena kakovosti pregledanih zapisov v obdobju od leta 2015 do leta 2020. V diskusiji in zaključku so podani predlogi, ki bi lahko prispevali k zmanjšanju napak že ob vnosu podatkov.

## 2 Kakovost podatkov v kontekstu

Na podlagi strokovne literature lahko ocenimo, da se v katalogizaciji kakovost podatkov pojavlja v okviru petih temeljnih, med seboj povezanih kontekstov. To so končni uporabnik, katalogizator, tehnologija in programska oprema, enotna obdelava (standardizacija) ter organizacija/racionalizacija delovnega procesa (Badovinac, 2017).

*Končni uporabnik* je osrednji pojem kakovosti podatkov v katalogizaciji, vendar je zelo malo neposrednih raziskav, ki bi ločile kakovost podatkov od kakovosti funkcionalnosti in vmesnikov katalogov ipd. Sicer pa se v tem kontekstu najpogosteje poudarja obseg podatkov, ki so za končnega uporabnika (ne)potrebni.

*Katalogizator* je tip uporabnika, ki se ga najpogosteje obravnava v raziskavah, povezanih s kakovostjo podatkov v katalogizaciji. Raziskovalce predvsem zanima mnenje katalogizatorjev o kakovosti zapisov in vzrokih, ki vplivajo na kakovost podatkov. Študije povzemajo tudi, katera znanja katalogizatorji potrebujejo in/ali kako uporabljajo katalogizacijske priročnike in orodja.

*Tehnologija in programska oprema* je kontekst, ki zajema orodja in njihove funkcionalnosti, ki vplivajo na kakovost podatkov in katalogov. Izpostavlja se, da je tudi tehnologija poseben tip uporabnikov, ki potrebuje določeno kakovost podatkov za uspešno upravljanje in transformacijo podatkov.

*Enotna obdelava (standardizacija)* je kontekst, ki poudarja kakovost podatkov v okviru mednarodnih in lokalnih diskusij o oblikovanju vsebinskih standardov (pravilnikov) in shem (formatov). V tem kontekstu se izvajajo tudi različne analize kakovosti podatkov in zapisov z vidika posameznih dimenzij kakovosti.

*Organizacija/racionalizacija delovnega procesa* je izvedbeni nivo, kjer se kakovost podatkov postavi v razmerje z ekonomsko učinkovitostjo, saj predstavlja katalogizacija velik strošek za knjižnice. Na podlagi Grahamove (1990) teze, da je kakovost zapisov v obratnem sorazmerju s produktivnostjo, lahko v tem kontekstu izpostavimo dve diskurzivni poziciji. V prvi se poudarjata čim večja natančnost zapisa in pomen intelektualnega vložka dobro usposobljenega kadra, v drugi pa sta pomembni časovna učinkovitost pri določitvi obsega nabora virov in potrebnih podatkov ter avtomatizacija katalogizacijskega postopka.

Za potrebe aktivnosti spremljanja kakovosti v COBISS.SI smo se umestili v kontekst enotne obdelave (standardizacije), pri čemer se ukvarjamo s kakovostjo vrednosti oz. dejanskih podatkov v podatkovnih elementih formata COMARC. Za ta namen smo razvili pristop, kjer smo najprej oblikovali nabor dimenzij za opis značilnosti podatkov, ki jih lahko merimo in ovrednotimo glede na zastavljene standarde.

## 3 Opredelitev kakovosti podatkov v bibliografskih in normativnih zapisih

V okviru aktivnosti spremljanja kakovosti zapisov dnevne produkcije v COBIB.SI smo h kakovosti pristopili pragmatično. Na osnovi tipologije napak smo postavili enajst dimenzij (tabela 1), s katerimi definiramo in opazujemo kakovost podatkov v bibliografskih in normativnih zapisih, ne glede na spremembe v katalogizacijskih pravilih in formatu (referenčnih virih). Podatek je dejanska vrednost, navedena v podatkovnem elementu (podpolje, indikator). Po tej definiciji je podatek v bibliografskih in normativnih zapisih kakovosten, kadar je ta naveden v edinstvenem zapisu (EDIN) in ni odvečen (ODVEČ). Podatek

mora biti semantično točen (SEMTOČ), točno prepisan (TOČPRE), strukturalno popoln (STRUPOP), vsebinsko popoln (VSEBPOP), strukturalno skladen (STRUSKLAD), vsebinsko skladen (VSEBSKLAD), oblikovno dosleden (OBLIDOS), aktualen (AKT) in (lahko) dodatno informativen (DODV) (Badovinac, 2018).

Tabela 1: Dimenzije kakovosti podatka v bibliografskih in normativnih zapisih (Badovinac, 2018)

Dimenzija (oznaka)	Pomen	Pogoj	Vrsta napake
<b>Edinstvenost (EDIN)</b>	Zapis je edinstven, kadar nima dvojnikov na nivoju celotne baze glede na zahteve referenčnega vira.	/	dvojniki zapisa
<b>Odvečnost (ODVEČ)</b>	Podatek je odvečen, kadar ni predviden glede na referenčni vir.	/	odvečen podatek
<b>Semantična točnost (SEMTOČ)</b>	Podatek je semantično točen, kadar je pomensko pravilen glede na vir podatkov in referenčni vir.	STRUSKLAD	napačen podatek
<b>Točnost prepisa (TOČPRE)</b>	Podatek je prepisan točno, kadar se ujema z nizom znakov glede na vir podatkov ali referenčni vir.	STRUSKLAD SEMTOČ	zatiptan podatek
<b>Strukturalna popolnost (STRUPOP)</b>	Strukturalna popolnost zahteva, da je podatek obvezno prisoten glede na zahteve referenčnega vira.	/	podatek manjka
<b>Vsebinska popolnost (VSEBPOP)</b>	Podatek je vsebinsko popoln, kadar je celoten oziroma izčrpen glede na vir podatkov in zahteve referenčnega vira.	STRUSKLAD SEMTOČ TOČPRE	pomanjkljiv podatek
<b>Vsebinska skladnost (VSEBSKLAD)</b>	Podatki so vsebinsko skladni, kadar se med seboj strukturalno in/ali pomensko ujemajo, kot to določa referenčni vir.	TOČPRE STRUSKLAD	protisloven podatek oz. neujemanje dveh ali več podatkovnih elementov, nejasen podatek
<b>Strukturalna skladnost (STRUSKLAD)</b>	Podatek je strukturalno skladen, kadar je naveden v pravilnem polju/podpolju in so ta polja/podpolja v pravilnem vrstnem redu, kot to določa referenčni vir. Na nivoju zapisa so podatki strukturalno skladni glede na predpisani vzorec (masko) zapisa ali izvorni zapis.	SEMTOČ TOČPRE	nepavilni vrstni red polj/podpolj, podatek, naveden v napačnem polju/podpolju, napačna maska/nejasna struktura podatkov, predelan zapis
<b>Oblikovna doslednost (OBLIDOS)</b>	Podatek je oblikovno dosleden, kadar je prepis oblike v skladu z zahtevami referenčnega vira.	STRUSKLAD SEMTOČ TOČPRE VSEBPOP VSEBSKLAD	napačno oblikovan podatek
<b>Aktualnost (AKT)</b>	Podatek je aktualen, kadar kaže trenutnost ali pa je posodobljen glede na vir podatkov ali zahteve referenčnega vira.	/	nežuriran podatek/neredigiran zapis

Dimenzija (oznaka)	Pomen	Pogoj	Vrsta napake
<b>Dodana vrednost (DODV)</b>	Podatek ima dodano vrednost, kadar je predviden, vendar po referenčnem viru ni obvezen.	/	priporočljiv podatek

Na osnovi dimenzij lahko določimo tudi, koliko je možnih tipov slabih podatkov pri posameznem podatkovnem elementu. Pri podatkovnem elementu, npr. v *podpolju 101d<sup>2</sup>* v sistemu COBISS, kjer navedemo kodo jezika povzetka, določimo, da mora podatek zadostiti štirim dimenzijam: koda mora biti navzoča, kadar vir vsebuje povzetek (STRUPOP), koda mora biti semantično točna (SEMTOČ) oz. vsebinsko skladna s podatkom v podpolju 320a (VSEBSKLAD) in koda ne sme obstajati, kadar povzetka na viru ni (ODVEČ)<sup>3</sup>. Zahteve glede na veljavne referenčne vire v sistemu COBISS kažejo, da so v primeru podpolja 101d možne štiri vrste slabih podatkov. Na tem mestu lahko že uvidimo, da je v bibliografskih in normativnih zapisih možnih zelo veliko pomanjkljivosti.

#### 4 Spremljanje kakovosti dnevne produkcije v COBIB.SI

Osnovni namen aktivnosti spremljanja kakovosti zapisov dnevne produkcije v vzajemni bibliografski bazi COBIB.SI je takojšna ročna odprava pomanjkljivosti v pregledanem zapisu. S poslanim sporočilom o pomanjkljivosti katalogizatorjem posredujemo tudi morebiti spregledano novost v katalogizacijski praksi<sup>4</sup>.

Sredi leta 2015 je bila uvedena nova metoda vzorčenja in pregledovanja. Z metodo naključno vzorčimo 10 % bibliografskih zapisov, kreiranih na določen dan, skupaj s pripadajočimi normativnimi zapisi. Predviden je vsaj en tedenski zajem zapisov v bazi COBIB.SI. Iz zajema so izločeni zbirni zapisi in zapisi, označeni za brisanje, predhodni zapisi (CIP), zapisi iz konverzij in paketno uvoženi zapisi, zapisi prvega vnosa in zapisi, ki so jih kreirali ali redigirali katalogizatorji v Narodni in univerzitetni knjižnici (NUK).<sup>5</sup> Slednji ima vzpostavljeno lastno kontrolo zapisov. Vzorčenje prvih 10 % zapisov iz zajema izvedemo z Excelovo funkcijo RANDBETWEEN(1;XXX).

Pregled zapisov zaključimo vsaj v šestih dneh od dneva zajema, saj je baza COBIB.SI dinamična. Pregled zapisa se izvede ročno, brez primarnega vira (izjema so le prosto dostopni viri na spletu). Zapis pregledamo na osnovi okvirnih kontrolnih vprašanj (Badovinac, 2018), saj nimamo »popolnega zapisa«, s katerim bi podatke lahko primerjali in ugotovili odstopanja.

Opombo o pomanjkljivosti zapišemo strukturirano in podprto z javno objavljeno dokumentacijo katalogizacijske prakse (referenčnimi viri). O pomanjkljivostih obvestimo katalogizatorje prek e-pošte le, kadar ima zapis najmanj eno večjo pomanjkljivost. Slednje je

<sup>2</sup> Zaradi berljivosti in dolžine besedila bomo v sestavku navajali le oznake podatkovnega elementa (polje, podpolje, indikator). Imena in pomeni podatkovnih elementov so navedeni v priročnikih COMARC/B (2021) in COMARC/A (2021).

<sup>3</sup> V sistemu COBISS so nekatere dimenzije kakovosti pri posameznih podatkovnih elementih dosežene že s programskim segmentom COBISS3/Katalogizacija, npr. kontrola vnosa podatka s šifranti. Sicer bi lahko pri navedenem podatkovnem elementu določili, da mora podatek zadostiti tudi dimenziji točnost prepisa (TOČPRE).

<sup>4</sup> S terminom katalogizacijska praksa zajamemo vsa katalogizacijska pravila in navodila za vnos v format COMARC ter za vodenje bibliografij v sistemu COBISS.SI.

<sup>5</sup> Iskalna zahteva: dm=[datum] not (dt=c or rs=d or rs=p or cr=\*old or cr=ctk springer\* or cr=uplsi\* or cr=knt izum\_ or cr=nuk\* or re=nuk\* or rs=i).

določeno v skladu s *Kriteriji za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI* (2009), pri čemer kot večjo pomanjkljivost obravnavamo tudi napačne podatke, vsebinsko in strukturno neskladne podatke, manjkajoče obvezne podatke ali pomanjkljive podatke ter odvečne podatke v katerem koli podatkovnem elementu. Že na tem mestu je treba poudariti, da vključuje programski segment COBISS3/Katalogizacija skoraj 200 različnih programskih kontrol, ki preverjajo pomanjkljivosti že ob shranjevanju zapisa (gl. COBISS3/Katalogizacija, pogl. 7.1.10 ter dodatka B1 in B2).

Ob zaključku koledarskega leta vse zabeležene pomanjkljivosti kodiramo z dimenzijami, torej podatkovnemu elementu pripišemo dimenzijo, na katero vpliva slab podatek. Analiza na osnovi dimenzij nam omogoča pregled, kakšne so pomanjkljivosti v podatkovnem elementu. Ker ne beležimo, koliko je možnih vseh pomanjkljivosti na pregledani zapis, ne moremo izračunati deleža in povprečja pomanjkljivosti na posamezen pregledan zapis. Prav tako bi v tem primeru potrebovali otežitev glede na pomembnost podatkovnega elementa, ki bi korigirala tovrstne rezultate. Kljub temu nam analiza, predvsem kronološki pregled, daje nekaj uporabnih rezultatov za diskusijo in načrtovanje nadaljnjih aktivnosti za izboljšave kakovosti podatkov v sistemu COBISS.SI (npr. programske kontrole, identifikacija istovrstnih napak, ažuriranje dokumentacije ipd.).

Z enomesečnim zamikom po zadnjem zajemu v letu vzorčimo 10 % pregledanih zapisov, za katere smo katalogizatorjem poslali sporočila, z namenom, da preverimo, ali in kako so bile pomanjkljivosti odpravljene. S tem postopkom preverimo tudi ustreznost poslanih priporočil.

#### 4.1 Vzorec za obdobje od junija 2015 do decembra 2020

Od junija 2015 do konca leta 2020 smo zajeli 133.027 bibliografskih zapisov in od tega vzorčili 13.478 bibliografskih zapisov s pripadajočimi normativnimi zapisi (tabela 2). V tem obdobju smo pregledali vsaj en zapis vsake od 246 zajetih ustanov, pri čemer je treba izpostaviti, da imajo nekatere ustanove več oddelkov<sup>6</sup>. Pregledali smo vsaj en zapis vsakega od 574 zajetih katalogizatorjev ne glede na ustanovo oz. lokalno bazo, za katero so kreirali zapis<sup>7</sup>.

Tabela 2: Povzetek pregleda dnevne produkcije v obdobju od 2015 do 2020

Kazalec/leto	2015 (jun.– dec.)	2016	2017	2018	2019	2020	Skupaj
Štev. zajemov	24	50	40	42	41	40	237
Štev. zajetih zapisov	12.445	29.630	20.051	23.261	23.609	24.031	133.027
Štev. vzorčenih zapisov	1.251	2.956	2.202	2.323	2.347	2.398	13.478
Štev. zapisov brez napak*	672	1.461	920	934	1.052	837	5.876
Štev. zapisov z večjo pomanjkljivostjo*	388	691	509	518	771	599	3.180
Štev. zapisov z manjšo pomanjkljivostjo*	186	747	750	828	477	922	4.204
Štev. neocenjenih zapisov	5	57	23	43	49	40	217

<sup>6</sup> Značilnosti vzorca glede na število zapisov na ustanovo (brez oddelkov) po osnovni opisni statistiki so naslednji: max = 1032, min = 1,  $\bar{x}$  = 54,7, SD = 107, Me = 17,5.

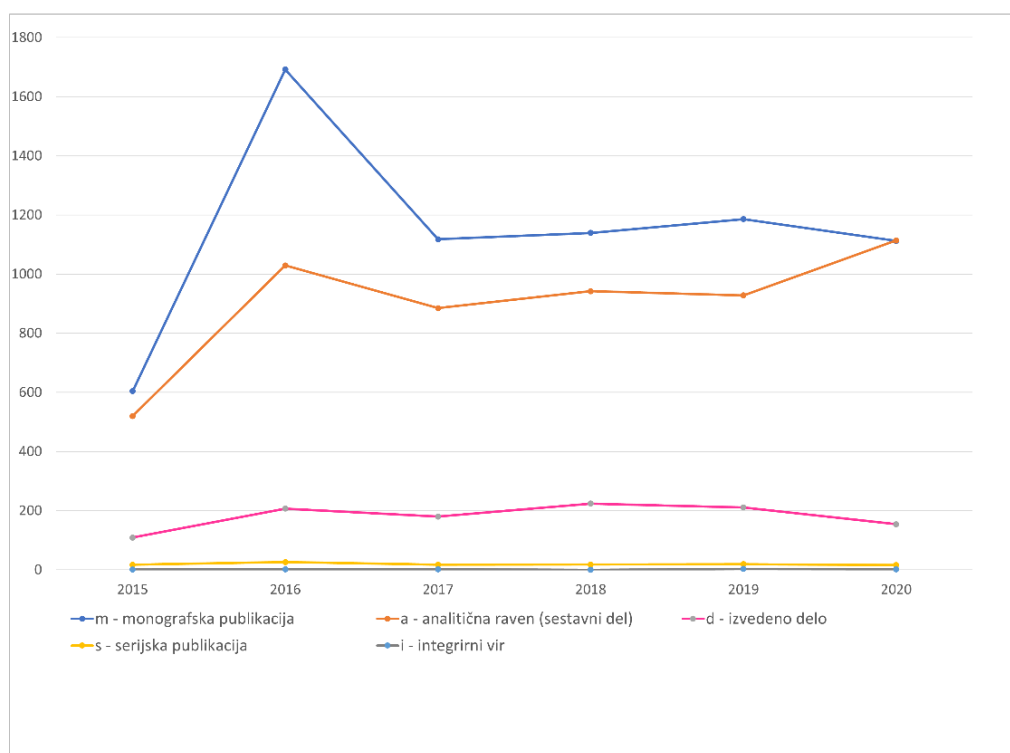
<sup>7</sup> Značilnosti vzorca glede na število zapisov na katalogizatorja po osnovni opisni statistiki so naslednji: max = 305, min = 1,  $\bar{x}$  = 23,4, SD = 31,6, Me = 12,5.

Kazalec/leto	2015 (jun.– dec.)	2016	2017	2018	2019	2020	Skupaj
Štev. poslanih sporočil	326	260	435	452	411	486	2.370
Štev. podatkovnih elementov s pomanjkljivostjo (bib. zapisi)**	873	2.389	2.012	2.131	1.917	2.112	11.434

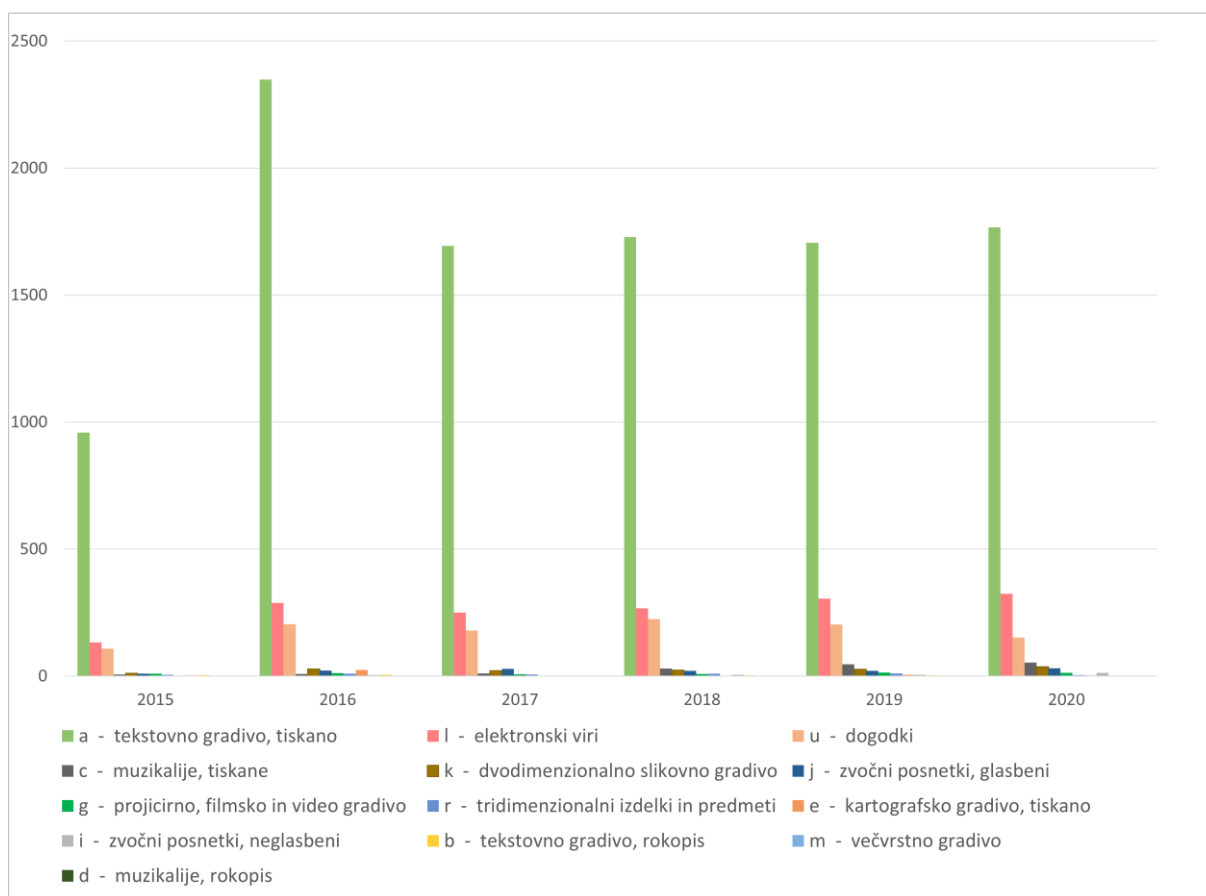
\* Vključene so tudi pomanjkljivosti pripadajočih normativnih zapisov.

\*\* Vključuje en podatkovni element ali kombinacijo več podatkovnih elementov, pa tudi nivo polja in splošno opombo.

Glede na bibliografski nivo smo povečini pregledali zapise za monografske vire in sestavne dele, glede na vrsto zapisa pa predvsem tiskano gradivo. Opaziti je, da se delež elektronskih virov z leti počasi zvišuje (sliki 1 in 2). Vzorec je razmeroma skladen z značilnostmi letnega prirasta zapisov v COBIB.SI (COBISS, Kazalci rasti, 2021).



Slika 1: Pregledani bibliografski zapisi po letih glede na bibliografskih nivo – šifrant podpolja 001c (obdobje od 2015 do 2020, n = 13.478)



Slika 2: Pregledani bibliografski zapisi po letih glede na vrsto zapisa – šifrant podpolja 001b (obdobje od 2015 do 2020,  $n = 13.478$ )

## 4.2 Okvirna ocena kakovosti zapisov glede na število poslanih sporočil

Pri pregledu zapisov lahko podamo okvirno oceno pregledanega bibliografskega zapisa s pripadajočimi normativnimi zapisi. Uporabimo štiristopenjsko lestvico: 1) zapis brez napak, 2) zapis z najmanj eno manjšo pomanjkljivostjo, 3) zapis z najmanj eno večjo pomanjkljivostjo in 4) neocenjeni zapis (nedorečena katalogizacijska praksa, izbrisan zapis, status katalogizatorja: novinec ipd.).

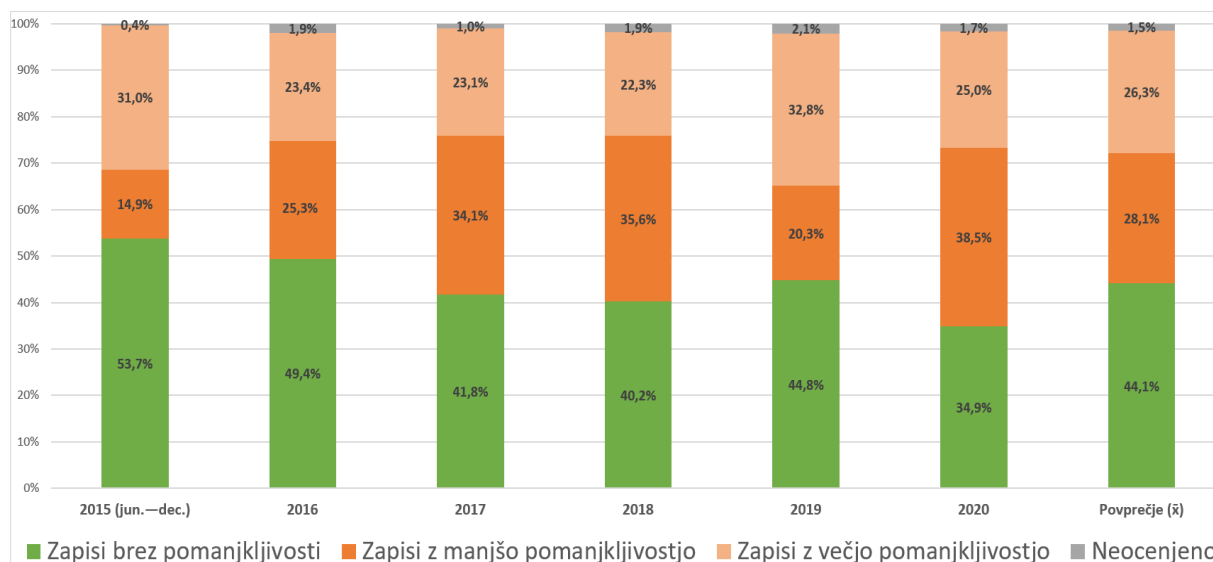
Sporočilo kreatorju/redaktorju zapisa pošljemo le, če ima zapis najmanj eno večjo pomanjkljivost. V opazovanem obdobju smo poslali 2.370 sporočil s priporočili za 3.503 pregledane bibliografske zapise s pripadajočimi normativnimi zapisi (tabela 2). Sporočilo je prejelo 453 različnih katalogizatorjev za zapise, kreirane za 215 različnih ustanov/baz<sup>8</sup>. Sicer pa je v našem vzorcu 481 različnih katalogizatorjev kreiralo zapis, ki je imel pripisano najmanj eno pomanjkljivost na zapis.

Aritmetična sredina kaže, da je v opazovanem obdobju delež zapisov brez napak predstavljal 44,1 %, zapisov (bibliografski s pripadajočimi normativnimi) z najmanj eno manjšo pomanjkljivostjo je bilo 28,1 %, zapisov z najmanj eno večjo pomanjkljivostjo pa 26,3 %.

<sup>8</sup> Pri katalogizatorjih glede na število sporočil osnovna opisna statistika kaže max = 135, min = 1, PO = 7,08, SD = 9,56, Me=5, medtem ko je za ustanove v vzorcu značilno naslednje: max = 239, min = 1, PO = 14,77, SD = 27,04, Me = 6.



Povprečni delež neocenjenih zapisov je bil 1,5-odstoten. S slike 3 je razvidno tudi, da se je delež z najmanj eno večjo pomanjkljivostjo v letu 2020 v primerjavi s prejšnjimi leti povečal, saj je bilo v tem letu v katalogizacijski praksi uvedenih več novosti.



Slika 3: Delež bibliografskih zapisov glede na oceno pomanjkljivosti (obdobje od junija 2015 do 2020,  $n = 13.478$ , v %)

### 4.3 Analiza pomanjkljivosti glede na dimenzije kakovosti podatka

#### 4.3.1 Bibliografski zapisi

Pri 6.241 različnih bibliografskih zapisih smo opazili 11.434 pomanjkljivosti, ki so vplivale na posamezno dimenzijo kakovosti podatka. Opazovano obdobje kaže manjša odstopanja pri številu podatkovnih elementih s slabim podatkom (tabela 3).

Tabela 3: Število podatkovnih elementov s pomanjkljivostjo (bibliografski zapisi, obdobje od 2015 do 2020)

Leto	Štev. podatkovnih elementov s slabim podatkom*
2015 (jun.–dec.)	873
2016	2.389
2017	2.012
2018	2.131
2019	1.917
2020	2.112
<b>Skupaj 2015–2020</b>	<b>11.434</b>

\*En podatkovni element ali kombinacija dveh ali treh podatkovnih elementov (podpolje, indikator, polje); vključena tudi splošna opomba.

Analiza slabih podatkov je pokazala tudi, da smo 9.392 pomanjkljivosti beležili pri 283 različnih podatkovnih elementih, 2.042 splošnih opomb pa se je nanašalo na raven celotnega zapisa; 55 odstotkov slabih podatkov se je pojavilo pri le 9 različnih podatkovnih elementih, na nivoju polja 210 in sklopa predmetnih oznak ter splošne opombe na nivoju zapisa (tabela 4). Na nivoju polja 210 beležimo posebej odvečnost pri izvedenih delih, katerih navodila so bila

objavljena leta 2013. Na nivoju zapisa pa smo s splošno opombo predvsem opozorili na strukturno neskladnost podatkov (npr. iz zapisa ni bilo jasno, za kakšen tip vira gre, ipd.).

*Tabela 4: Dvanajst najpogostejših podatkovnih elementov s slabimi podatki (55 % od vseh, n = 11.434)*

Podatkovni element	Štev. slabih podatkov
100e	2.551
600–610*	781
300a	679
675v	346
200a	329
200e	323
200f	285
105a	283
100b	256
215a	198
210*	149
Splošna opomba**	114
Skupaj	6.297

Legenda: \* – nivo polja, \*\* – nivo zapisa.

Glede na dimenzije se je večina priporočil nanašala na pomanjkanje priporočljivih podatkov (DODV; 38,9 %). Sledile so težave z obliko podatkov (OBLIDOS; 12,2 %), točnostjo podatkov (SEMTOČ; 12 %) in manjkajočimi podatki (STRUPOP; 11 %). Preostala tretjina slabih podatkov se je nanašala na strukturalno skladnost, odvečnost, zatipkanost, vsebinsko skladnost in vsebinsko popolnost podatka. Dvojnikov zapisov (EDIN) je bilo v tem obdobju zanemarljivo malo (4 primeri) (tabela 5).

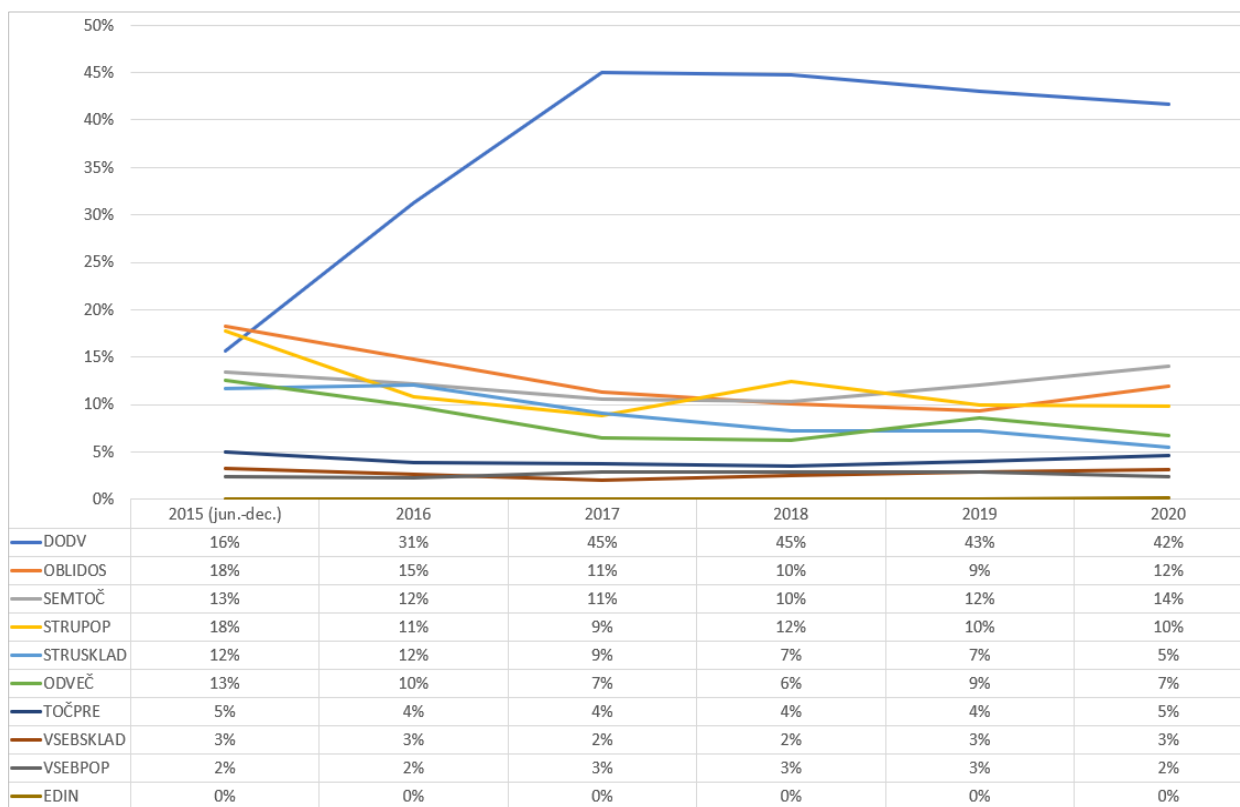
*Tabela 5: Slabi podatki glede na dimenzije kakovosti podatka v bibliografskih zapisih (obdobje od junija 2015 do 2020, n = 11.434)*

Dimenzija	Štev. slabih podatkov	Delež (%)
DODV	4.449	38,9
OBLIDOS	1.390	12,2
SEMTOČ	1.368	12,0
STRUPOP	1.254	11,0
STRUSKLAD	981	8,6
ODVEČ	915	8,0
TOČPRE	462	4,0
VSEBSKLAD	308	2,7
VSEBPOP	303	2,6
EDIN	4	0,03
Skupaj	11.434	100

Legenda: DODV – dodatna vrednost; OBLIDOS – oblikovna doslednost; SEMTOČ – semantična točnost; STRUPOP – strukturalna popolnost; STRUSKLAD – strukturalna skladnost; ODVEČ – odvečnost; TOČPRE – točnost prepisa; VSEBSKLAD – vsebinska skladnost; VSEBPOP – vsebinska popolnost; EDIN – edinstvenost

Čeprav pri večini dimenzij beležimo zmanjševanje deleža slabih podatkov, je v tem obdobju mogoče opaziti nekaj nihanj (slika 4). Od leta 2017 COBISS+ omogoča fasetno navigacijo; le-ta

vključuje tudi podatke, ki po veljavni katalogizacijski praksi v bibliografskem zapisu niso obvezni. Zaradi te funkcionalnosti COBISS+ smo pri pregledovanju začeli beležiti oz. priporočati tudi vnos teh podatkov, kar se kaže v povišanem deležu slabih podatkov pri dimenziji DODV od leta 2017 naprej. Večji delež slabih podatkov za SEMTOČ in OBLIDOS v letu 2020 pa pripisujemo težavam z usvajanjem obsežnega sklopa novosti v katalogizacijski praksi. Vzroka za manjše odstopanje pri dimenziji STRUPOP v letu 2018 ni bilo mogoče ugotoviti, se pa večinoma navezuje na manjkajoče podatke o odgovornosti v bloku 7XX pri kreiranju zapisov za vodenje bibliografij raziskovalcev.



*Legenda: DODV – dodatna vrednost; OBLIDOS – oblikovna doslednost; SEMTOČ – semantična točnost; STRUPOP – strukturalna popolnost; STRUSKLAD – strukturalna skladnost; ODVEČ – odvečnost; TOČPRE – točnost prepisa; VSEBSKLAD – vsebinska skladnost; VSEBPOP – vsebinska popolnost; EDIN – edinstvenost*

*Slika 4: Slabi podatki glede na dimenzije kakovosti podatka v bibliografskih zapisih po letih (obdobje od junija 2015 do 2020, n = 11.434, v %)*

Analiza po posameznih dimenzijah kaže na naslednje ugotovitve:

- DODV: O pomenu podatkov, ki prispevajo dodano vrednost k opisu vira, smo že pisali v povezavi z uvedbo fasetne navigacije v COBISS+ (Badovinac, 2020). Tu gre predvsem za manjkajoče priporočljive podatke v podpoljih 100e, 105f, 105g in 105z, ki so v naši analizi povezani kar z dvema tretjinama (59,3 %) vseh slabih podatkov v okviru celotnega DODV. V okviru te dimenzije se kaže tudi večji delež manjkajočih podatkov v poljih 600–610 (17,5 %), od tega nekaj manj kot tretjina manjkajočih predmetnih oznak v slovenskem jeziku. Razmeroma velik odstotek predstavlja tudi manjkajoč podatek v podpolju 105a (4,8 %) in podpolju 100b (4,4 %).

- OBLIDOS: Glede na oblikovanje podatkov smo beležili težave predvsem v podpoljih 300a, 200f, 215c, 200e, 215a, 210a, 200a, 210c, 230a in 200g. Ti skupaj predstavljajo 68,3 % vseh pomanjkljivosti glede na dimenzijo OBLIDOS. Napake v oblikovanju vplivajo na predstavitev podatka v izpisu ISBD in »katalogizacijski pravopis«.
- SEMTOČ: Napake glede na semantično točnost so bile izrazite v podpolju 675v (19,2 %), kar je posledica spremembe v oznaki v začetku leta 2020. Napačne podatke smo v celotnem obdobju večkrat zasledili še v naslednjih podatkovnih elementih: 856u (6,3 %), 856ind2 (5,5 %), 702ind1 (4,4 %), 100b (4,2 %), 710ind1 (3,1 %), 200a (3 %), 200ind1 (2,6 %) in 101i (2,5 %).
- STRUPOP: Obvezni podatki so manjkali predvsem v območju opomb (blok 3XX) (31 %), poljih opisa po ISBD v bloku 2XX (27 %) in v podpoljih kodiranih podatkov bloka 1XX (25 %). Manjkajoči podatki so izstopali v podpoljih 300a, 101d, 105a, 101d, 105a, 230a, 215d, 200b, 102a in 320a. Posebej izstopa tudi manjkajoč podatek v podpolju 215a (in morebitnem podpolju 215c), ki je posledica uvedbe obveznosti podatka za elektronske vire, dostopne na spletu, v letu 2020.
- STRUSKLAD: Težave s strukturalno skladnostjo podatkov beležimo predvsem pri polju 200 (38,8 %), kjer so bili podatki navedeni bodisi v napačnem podpolju bodisi je bil vrstni red podpolj napačen. V okviru te dimenzije predstavljajo razmeroma velik delež tudi splošne opombe na zapis v celoti (11 %), pri katerih smo poudarili, da iz zapisa ni jasno, za kakšen tip vira gre. Sledile so pomanjkljivosti, vezane na polje 215 (napačen vrstni red podpolj) in polje 300 (podatek je pogosto naveden v napačnem polju).
- ODVEČ: Največ odvečnih podatkov smo beležili v polju 210 (15,4 %) pri zapisih za izvedena dela. Po pogostosti si sledijo odvečni podatki v podpoljih 510a in 300a ter polju 712 (zapisi, preneseni iz drugih sistemov), podpoljih 001g in 001h (zapisi, prevzeti iz COBISS.net) ter podpolju 540a.
- TOČPRE: Zatičkane podatke smo zasledili pri 50 različnih podatkovnih elementih, največ v podpolju 200a, kar 15,58 %. Sledile so tipkarske napake v podpoljih 610a (10,2 %) in 675v (9,7 %), 200e (9,7 %), 200f (8,7 %), 300a (7,4 %), 215 (5,6 %).
- VSEBSKLAD: Vsebinska neskladnost podatkov je bila izrazita zlasti pri kombinaciji dveh ali več podatkovnih elementov v podpoljih 100b, 100c, 100d in 210d (skupaj 47,8 %). Razmeroma velik delež neskladnosti je bil opažen v podpoljih 102a in 210a (16,6 %) ter kombinaciji podpolj 200f, 200g in 300a ter v bloku 7XX (7,1 %).
- VSEBPOP: Pomanjkljive podatke smo zasledili v polju 200 (36 %; predvsem v podpolju 200e (16,2 %)) in v polju 215 (15,5 %; zlasti v podpoljih 215a (5,9 %) in 215h (5,3 %)) ter v polju 225 (14,2 %; zlasti v podpoljih 225v (7,6 %) in 225a (5,6 %)) ter v podpolju 300a (13,5 %).
- EDIN: Zapisov dvojnikov je bilo v opazovanem obdobju malo (samo 4 primeri), saj omogoča programski segment COBISS3/Katalogizacija preverbo že pri shranjevanju zapisa.

Glede na posamezni bibliografski nivo smo pregledali tudi vrste pomanjkljivosti, pri katerih bi izpostavili naslednje:

- *Monografske publikacije* predstavljajo največjo skupino virov, opisanih v pregledanih zapisih. Zasledili smo 6.156 pomanjkljivosti pri 224 različnih podatkovnih elementih. Pri splošnih opombah smo beležili predvsem strukturalno neskladnost, posebej pri poljih

200, 215 in 225 ter poljih za odgovornost 701, 702 in 710. Pogoste težave so bile vezane na oblikovanje podatka v skladu z ISBD-jem in »katalogizacijskim pravopisom« (polji 200 in 300). Napačni podatki so bili najdeni v podpolju 675v, poleg tega pa tudi v podatkovnih elementih 702ind1, 100b, 856u, 710ind1, 200ind1, 100i, 200a, 856ind2, 200a idr. Največkrat zatipkani so bili podatki v podpoljih 200a, 610a, 675v, 300a in 200f. Manjkale so obvezne opombe v podpolju 300a ter podatki v podpoljih 100g, 100i, 101d, 102a, 105a, 105e, 200b, 200f, 210a, 215d idr. Podatki v podpoljih 001g, 300a, 540a in 100d so bili pogosto odvečni. Priporočali smo tudi vnos različnih neobveznih podatkov – v podpolju 100e in poljih 600–610 ter tudi v bloku 7XX, kjer so navedeni podatki o odgovornosti.

- Pri *sestavnih delih* smo zasledili 4.407 pomanjkljivosti pri 117 različnih podatkovnih elementih, od tega nekaj splošnih opomb in neujemanje med podatkovnimi elementi. Največkrat smo priporočali podatek v podpolju 100e (39 %). Sledila so neujemanja podatkov (13 %). Priporočali smo tudi vnos podatka v podpolje 105a, glede na podpolje 215c. Opozorili smo na obliko podatka v podpoljih 300a, kar je bilo značilno tudi za večino podatkovnih elementov v blokih 2XX in 3XX. Večkrat smo zapisali tudi priporočilo za vnos podatkov v polje 701, ki se je nanašalo na zapise, povezane z vodenjem bibliografij raziskovalcev.
- Pri *izvedenih delih* so se pomanjkljivosti pokazale v 49 različnih podatkovnih elementih, kjer smo beležili 425 slabih podatkov. Pri bibliografskem nivoju pogosto ni bila jasna struktura podatkov, ki bi razlikovala med objavljenimi viri in dogodki. Sicer pa se je z leti zmanjšal delež odvečnih podatkov v polju 210 in drugih poljih, ki so predvideni v opisu za objavljene vire (npr. 200b, 856u idr.).
- Čeprav predstavljajo *serijske publikacije* manjši delež pregledanih zapisov v vzorcu (1 % v posameznem letu), pa je, glede na število pomanjkljivosti v podatkovnih elementih, ta skupina virov med najslabše opisanimi. Pri 80 različnih podatkovnih elementih smo beležili 423 pomanjkljivosti, najpogosteje so bile manjkajoče ali nepopolne obvezne opombe v podpoljih 300a, 311a, 326a in 210d ter odvečnost podatka v polju 712.
- *Integrirni viri* so bili v vzorcu zastopani z manj kot enim odstotkom, zato posamične analize za ta bibliografski nivo nismo izvedli.

#### 4.3.2 Normativni zapisi

Pri pripadajočih normativnih zapisih smo opazili 3.033 slabih podatkov, od tega se več kot dve tretjini nanašata na priporočilo za ažuriranje celotnega normativnega zapisa (AKT) v skladu s katalogizatorjevim pooblastilom za urejanje teh zapisov. Preostale pomanjkljivosti so se nanašale na 908 različnih normativnih zapisov, kjer so manjkali priporočljivi (DODV, 16,8 %) in drugi podatki (STRUPOP, 11,5 %). Preostalih pomanjkljivosti je bilo glede na delež slabih podatkov malo (tabela 6).

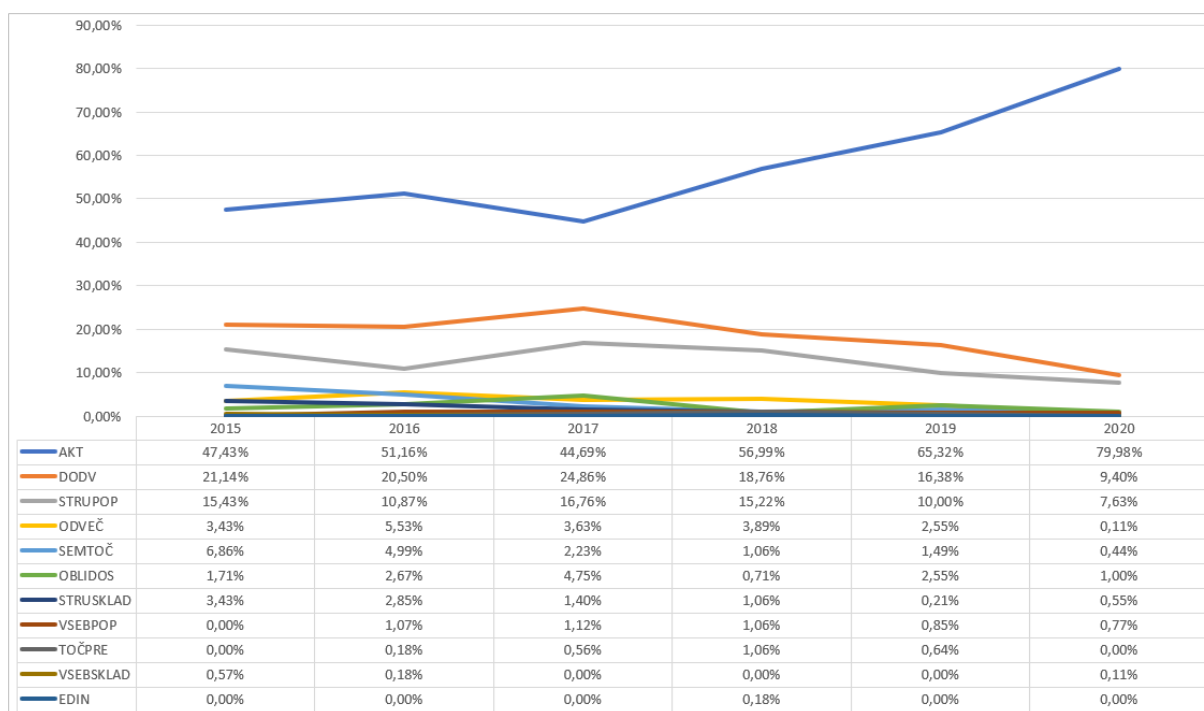
Tabela 4: Slabi podatki glede na dimenzije kakovosti podatka v pripadajočih normativnih zapisih (obdobje od junija 2015 do 2020, n = 3.033)

Dimenzija	Štev. slabih podatkov	Delež slabih podatkov (v %)
AKT	1.882	62,1
DODV	509	16,8
STRUPOP	350	11,5

Dimenzija	Štev. slabih podatkov	Delež slabih podatkov (v %)
<b>ODVEČ</b>	85	2,8
<b>SEMTOČ</b>	65	2,1
<b>OBLIDOS</b>	60	2,0
<b>STRUSKLAD</b>	39	1,3
<b>VSEBPOP</b>	27	0,9
<b>TOČPRE</b>	12	0,4
<b>VSEBSKLAD</b>	3	0,1
<b>EDIN</b>	1	0,0
<b>Skupaj</b>	<b>3.033</b>	<b>100</b>

Legenda: AKT – aktualnost; DODV – dodatna vrednost; STRUPOP – strukturalna popolnost; SEMTOČ – semantična točnost; ODVEČ – odvečnost; OBLIDOS – oblikovna doslednost; STRUSKLAD – strukturalna skladnost; VSEBPOP – vsebinska popolnost; TOČPRE – točnost prepisa; VSEBSKLAD – vsebinska skladnost; EDIN – edinstvenost

Analiza glede na opazovana leta kaže, da kljub priporočilom katalogizatorji s pooblastilom za urejanje normativnih zapisov ne ažurirajo povezanih normativnih zapisov (AKT). Opaziti pa je, da pada delež manjkajočih priporočljivih podatkov (DODV) (slika 5), zlasti podatek o viru oblike točke dostopa v polju 810 in delež manjkajočih podatkov (STRUPOP), predvsem variantna točka dostopa v polju 400 (slika 4).

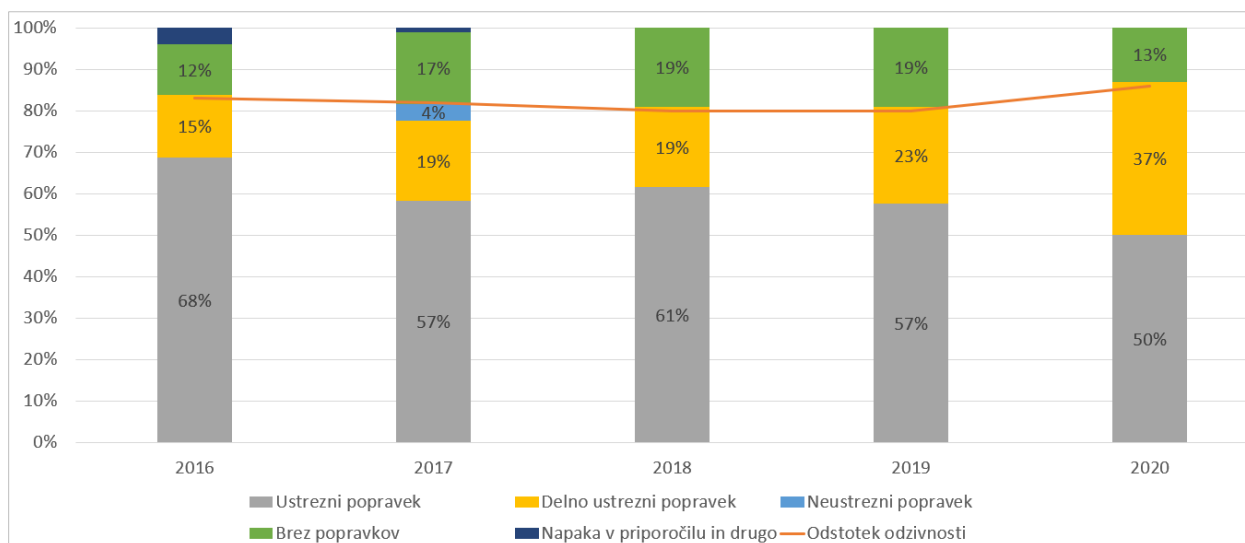


Legenda: AKT – aktualnost; DODV – dodatna vrednost; STRUPOP – strukturalna popolnost; SEMTOČ – semantična točnost; ODVEČ – odvečnost; OBLIDOS – oblikovna doslednost; STRUSKLAD – strukturalna skladnost; VSEBPOP – vsebinska popolnost; TOČPRE – točnost prepisa; VSEBSKLAD – vsebinska skladnost; EDIN – edinstvenost

Slika 5: Slabi podatki glede na dimenzije kakovosti podatka v pripadajočih normativnih zapisih po letih (obdobje od junija 2015 do 2020, n = 3.033, v %)

#### 4.4 Uspešnost spremljanja kakovosti dnevne produkcije: odprava pomanjkljivosti

Uspešnost metode spremljanja zapisov preverjamo ob zaključku aktivnosti na letni ravni tako, da naključno izberemo in pregledamo 10 % zapisov, za katere smo poslali priporočila. Slika 6 kaže, da katalogizatorji vnašajo popravke v pregledane zapise, odzivnost katalogizatorjev je približno 80-odstotna. S slike 6 je tudi razvidno, da je delež ustreznih popravkov v povprečju več kot 58-odstoten. Delno ustrezno popravljene zapise so predvsem tisti, pri katerih je število popravkov na zapis večje.



Slika 6: Odzivnost in ustreznost popravkov glede na poslana priporočila po posameznih letih (obdobje 2016–2020; n = 279)

## 5 Diskusija

Čeprav je baza COBIB.SI dinamična, spreminjajo pa se tudi število katalogizatorjev in obseg njihovih pooblastil ter število ustanov, smo z načinom vzorčenja zapisov pri spremljanju dnevne produkcije uspeli zajeti nekatere značilnosti slovenske vzajemne bibliografsko-kataložne baze COBIB v opazovanih letih. Pri rezultatih je treba upoštevati, da vzorec ne zajema zapisov iz NUK-a, ki v COBIB.SI prispevajo približno 10 % letnega prirasta. Tudi programski segment COBISS3/Katalogizacija s programskimi kontrolami preprečuje vnos nekaterih napak že ob kreiranju novih zapisov. Poleg tega so z metodo pregleda brez primarnega vira nekateri podatki težje preverljivi.

Na podlagi deleža poslanih sporočil o pomanjkljivostih lahko ocenimo, da je približno 45 % bibliografskih zapisov s pripadajočimi normativnimi zapisi brez napak. Od preostalih 55 % zapisov je polovica takšnih, ki imajo najmanj eno manjšo pomanjkljivost, medtem ko je pri drugi polovici zapisov zaradi večje pomanjkljivosti potreben en popravek ali več popravkov. Ocenimo lahko, da je glede na obstoječo katalogizacijsko prakso približno 70 % zapisov dobrih ali zadosti dobrih.

Sicer pa so nasveti med katalogizatorji dobro sprejeti, odziv je približno 80-odstoten. Težave pri popravljanju zapisov se kažejo tam, kjer je seznam priporočil za posamezen zapis obsežnejši ali pa je zaradi ene pomanjkljivosti treba vnesti popravke v več podatkovnih elementov.

Rezultati analize slabih podatkov v splošnem kažejo vpliv posodabljanja katalogizacijske prakse. Vendar bi z vidika dimenzij verjetno najbolj ustrezalo mnenje, da »katalog ni na pol poln napak, temveč na pol prazen« (Šauperl in Semlič Rajh, 2020). Najbolj izstopa delež manjkajočih priporočljivih podatkih (DODV). V sistemu COBISS.SI sicer velja, da so knjižnice zavezane k navajanju podatkov, ki so po katalogizacijskem pravilniku in formatu COMARC obvezni, o vnosu drugih podatkih pa presodijo glede na potrebe uporabnikov kataloga. Z vidika funkcionalnosti kataloga na nivoju vzajemne baze smo pri spremljanju kakovosti dnevne produkcije opozarjali zlasti na manjkajoče priporočljive podatke, povezane z uvedbo fasetne navigacije v COBISS+. Nekaj več zavedanja pri vnosu podatkov smo zasledili pri podatkovnih elementih za predmetno označevanje, saj ocenjujemo, da je zapisov brez ene predmetne oznake približno 9 %. Predvidevamo lahko, da se bo slednje v prihodnjih letih zmanjšalo tudi zaradi uvedbe Splošnega geslovnika COBISS.SI (SGC) v letu 2021.

Opaziti je, da se velik delež manjkajočih podatkov nanaša tudi na način vodenja bibliografij raziskovalcev. Poleg naših rezultatov, ki se vežejo na obstoječo katalogizacijsko prakso, Peteh in Južnič (2020) denimo pozivata tudi k vnosu predmetnih oznak v tujem jeziku, saj zaradi izmenjave podatkov z drugimi sistemi te prispevajo k večji najdljivosti in prepoznavnosti del slovenskih raziskovalcev. Podobno je tudi z obsegom navajanja odgovornosti v bloku 7XX, ki z obstoječo prakso nepopisovanja avtorjev popači izpis avtorstva pri bibliografskih virih s številnimi avtorji (prim. Dodatna, 2018).

Delež manjkajočih obveznih (STRUPOP) in pomanjkljivih podatkov (VSEBPOP) v pregledanih zapisih je razmeroma majhen (skupaj 13,6 %). Po drugi strani pa je zanimivo, da beležimo 8 % odvečnih podatkov (ODVEČ), kar izhaja iz nepoznavanja pravil in sprememb v katalogizacijski praksi. Verjetno je vzrok tudi uporaba metode kreiranja novih zapisov na podlagi prirejanja podobnega zapisa (COBISS3/Katalogizacija, 2021, pogl. 8.3.2).

Slabi podatki glede na semantično točnost (SEMTOČ) in točnost prepisa (TOČPRE) so z vidika kakovosti podatkov najbolj problematični. V prvo skupino uvrščamo tiste pomanjkljivosti, pri katerih ne prepoznamo dejanske vrednosti podatka, pri drugi skupini slabih podatkov pa podatek prepoznamo oz. je delno razumljiv (take napake so npr. odvečni presledki, ločila, zamenjane črke). V opazovanem obdobju je bilo teh napak skupaj 16 % od vseh zabeleženih pomanjkljivosti. Natančnejša analiza kaže, da je napačnih podatkov, ki bi grobo omejevali najdljivost vira (podpolja 200a in 7XX3 ter 7XXa), malo. To potrjuje tudi predhodna analiza deleža tipkarskih napak v podpolju 200a, ki so se znatno zmanjšale po uvedbi črkovalnika (Badovinac, 2019). Bolj problematične so napake v navedbi spletnih povezav (podpolje 856u), ki končnim uporabnikom omejuje dostop do elektronskih virov prek spleta. Sicer pa smo razmeroma obsežno neseznanjenost knjižničarjev o spremembi in obliki podatka v podpolju 675v v letu 2020 omilili z možnostjo pomoči pri vnosu podatka v programskem segmentu COBISS3/Katalogizacija (2021, pogl. 7.3.2.4).

Podobno težo pomembnosti, kot jo ima točnost, ima tudi strukturalna skladnost (STRUSKLAD), ki predstavlja 9 % vseh slabih podatkov. Pri tem je problematično zlasti polje 200, kjer so podatki navedeni v napačnih podatkovnih elementih (npr. dodatki k naslovu, odgovornosti).



Ker gre za semantiko podatka, je tovrstne pomanjkljivosti težko preverjati s programskimi kontrolami, katalogizator mora zato biti seznanjen s pravili uporabe posameznih podatkovnih elementov in kako njihov vrstni red vpliva na pomen podatka. Posebej pa mora biti pozoren, katero strukturo podatkovnih elementov uporabi pri opisu bibliografskega vira, ki je izhodišče za identifikacijo vira.

Vnos podatkov v format COMARC, ki temelji na strukturi reprezentacije podatkov po standardu ISBD, zahteva vnos vsebinsko istih podatkov v različne podatkovne elemente, ki so lahko razpršeni po formatu. To povzroča napako neskladnosti podatkov v posameznem zapisu. Kot kažejo rezultati analize pri dimenziji vsebinska skladnost (VSEBSKLAD), se slabi podatki navezujejo zlasti na podatkovne elemente, povezane z letnico izida.

Oblikovna doslednost (OBLIDOS) je po analizi z 12-odstotnim deležem druga največja skupina slabih podatkov, ki pa nima večjega vpliva na končne uporabnike. Zanimivo je, da se v tej skupini slabih podatkov kaže predvsem neupoštevanje osnovnih pravil prepisa podatkov v skladu s »katalogizacijskim pravopisom«, npr. uporaba velikih začetnic po ISBD in uporaba oglatih oklepajev (Prekat, pogl. 0.4 in 0.6). V letu 2014 so bili v programski segment COBISS3/Katalogizacija, v blok 3, dodani šifranti za pomoč pri vnosu, ki pomagajo ohraniti konsistentnost navajanja opomb v poljih: 300, 320, 324, 327, 328 in 337 (gl. COBISS3/Katalogizacija, 2021, pogl. 7.3.2.4).

Občasno smo pri spremljanju kakovosti opazili tudi nekaj lokalnih praks in načinov strukturiranja podatkov, s katerimi se je mogoče izogniti programskim kontrolam. Vendar pa ocenjujemo, da gre v primerjavi z ugotovitvami izpred več kot dveh desetletij, ko je Dimec (1994) pisala o problematiki odstopanja od standardov, bolj za izjeme kot pravilo. Raba lokalne prakse v zapisih je oz. utegne biti težava pri tistih nekaj ustanovah, pri katerih je število zapisov s tovrstno prakso visoko.

Pri normativnih zapisih se je pokazalo, da kljub priporočilom katalogizatorji s pooblastilom za urejanje normativnih zapisov le-teh ne ažurirajo. Vzroki so verjetno večplastni, predvidevamo, da katalogizatorji bodisi ne razumejo obsega svojih pooblastil bodisi nimajo dovolj časa za redakcijo, ki zajema usklajevanje več normativnih zapisov, preverjanje informacij in komunikacijo z avtorji. Kadar pa se v normativne zapise posega, se podatki le delno posodablajo. Takšni zapisi niso primerni za sinhronizacije z bibliografskim zapisom ali za prenos podatkov v druge sisteme (Kos, 2019; Krajnc Vobovnik, 2020; COBISS3/Katalogizacija, 2021, pogl. 9.12). Pomanjkanje informacij v normativnih zapisih povzroča povezovanje avtorjev z napačnimi bibliografskimi zapisi.

## 6 Zaključek

Rezultati naše analize kažejo na razmeroma dobre podatke v zapisih. Kljub temu menimo, da bi se nekatere pomanjkljivosti preprečile že s seznanitvijo s spremembami v katalogizacijski praksi in usvajanjem teh sprememb. Zato pozivamo katalogizatorje, da uredijo prijave na prejemanje sporočil (npr. Krasovka; IZUM – portal Izobraževanje).

Poznavanje osnovnih izhodišč katalogizacije je bistveno. Ta temeljijo na vsebinskem standardu PPIAK in ISBD-ju (deloma povzeto v priročnikih *Prekat* in *Značka*) in priročnikih formata COMARC. Poleg osnovnih dokumentov je treba poznati natančnejša navodila in rešitve

uporabe katalogizacijskih pravil. Eden izmed korakov je bila ureditev dodatnih navodil za katalogizacijo na IZUM-ovem portalu Izobraževanje (*Seznam priročnikov in drugih navodil*).

V prihodnje bo aktivnosti spremljanja kakovosti v COBISS.SI treba podpreti tudi z dodatnimi načini preverjanja semantične točnosti podatkov, ki je zdaj omejena na podatke v zapisu in na vire, ki so prosto dostopni prek spleta. Eden izmed že podanih predlogov je lažje sporočanje o napakah v zapisih prek COBISS+ na nivoju prikazanega zapisa.

## Zahvala

Zahvala gre najprej katalogizatorjem, saj s svojim delom prispevajo k razvoju sistema COBISS in z razumevanjem sprejemajo priporočila za izboljšanje kakovosti podatkov. Zahvala gre tudi sodelavcem IZUM-a in NUK-a, ki so delili svoje izkušnje in znanja ter prisluhnili mojim vprašanjem pri razreševanju katalogizacijskih dilem.

## Reference

Badovinac, B., 2017. Izhodišča za proučevanje kakovosti podatkov v bibliografskih in normativnih zapisih: kakovost podatkov v kontekstu in raziskovalne usmeritve v katalogizaciji. *Knjižnica*, 61(1/2), str. 119–154. Dostopno na: <https://knjiznica.zbds-zveza.si/knjiznica/article/view/6165> [25. 2. 2021].

Badovinac, B., 2018. Nabor dimenzij za opredelitev kakovosti podatkov v bibliografskih in normativnih zapisih. *Organizacija znanja*, 23(1/2), str. 2–10. Dostopno na: [https://www.cobiss.si/oz/HTML/OZ\\_2018\\_1\\_2\\_final/6/index.html](https://www.cobiss.si/oz/HTML/OZ_2018_1_2_final/6/index.html) [27. 5. 2021].

Badovinac, B., 2019. »Pikice in vejice« pod drobnogledom: spremljanje kakovosti zapisov v letu 2018. *Blog COBISS*, 27. 5. 2019. Dostopno na: <https://blog.cobiss.si/2019/05/27/pikice-in-vejice-pod-drobnogledom/> [25. 2. 2021].

Badovinac, B., 2020. Več podatkov zagotavlja boljšo fasetno navigacijo v COBISS+. *Blog COBISS*, 11. 6. 2020. Dostopno na: <https://blog.cobiss.si/2020/06/11/vec-podatkov-zagotavlja-boljso-fasetno-navigacijo-v-cobiss/> [27. 5. 2021].

COBISS, *Kazalci rasti*, 2021. Dostopno na: <https://www.cobiss.si/kazalci/> [24. 2. 2021].

*COBISS3/Katalogizacija*, 2021. Maribor: IZUM. Dostopno na: [https://home.izum.si/izum/e-prirocniki/1\\_COBISS3\\_Katalogizacija/Cel\\_1\\_COBISS3\\_Katalogizacija.pdf](https://home.izum.si/izum/e-prirocniki/1_COBISS3_Katalogizacija/Cel_1_COBISS3_Katalogizacija.pdf) [25. 2. 2021].

COMARC/A, 2021, Maribor: IZUM. Dostopno na: [http://home.izum.si/izum/e\\_manuals\\_html/COMARC\\_B/slv/ch1.html](http://home.izum.si/izum/e_manuals_html/COMARC_B/slv/ch1.html) [25. 2. 2021].

COMARC/B, 2021, Maribor: IZUM. Dostopno na: [http://home.izum.si/izum/e\\_manuals\\_html/COMARC\\_A/slv/ch1.html](http://home.izum.si/izum/e_manuals_html/COMARC_A/slv/ch1.html) [25. 2. 2021].

Dimec, Z., 1994. Strokovna usposobljenost za katalogizacijo v COBISSu: analiza problematike in možni ukrepi za izboljšanje. *Knjižnica*, 38(3/4), str. 75–92.

*Dodatna navodila za katalogizacijo bibliografskih enot, ki se upoštevajo pri vrednotenju raziskovalne uspešnosti*, 2018. Maribor: IZUM. Dostopno na: [https://home.izum.si/izum/e-prirocniki/druga\\_navodila/Navodila\\_za\\_bibliografije\\_2018-11-23.pdf](https://home.izum.si/izum/e-prirocniki/druga_navodila/Navodila_za_bibliografije_2018-11-23.pdf) [25. 2. 2021].

Graham, P. S., 1990. Quality in cataloguing: making distinctions. *Journal of Academic Librarianship*, 16(4), str. 213–218.

Hafter, R., 1979. Type of search by type of library. *Information Processing & Management*, 16(5), str. 261–264. Dostopno na: [https://doi.org/10.1016/0306-4573\(79\)90032-3](https://doi.org/10.1016/0306-4573(79)90032-3) [27. 5. 2021].

Jon R. Hufford., 1991. Use studies and OPACs. *Technical Services Quarterly*, 9(1), str. 57–70, Dostopno na: [https://doi.org/10.1300/J124V09N01\\_06](https://doi.org/10.1300/J124V09N01_06) [27. 5. 2021].

Kos, J., 2019. Kako sestavimo normativni zapis za osebno ime. *Blog COBISS*, 8. 10. 2019. Dostopno na: [https://blog.cobiss.si/2019/10/08/sestavimo\\_normativni\\_zapis/](https://blog.cobiss.si/2019/10/08/sestavimo_normativni_zapis/) [25. 2. 2021].

Krajnc Vobovnik, A., 2020. Ažuriranje zapisov v VIAF in priporočila za kakovostnejši prikaz slovenskih normativnih zapisov v VIAF. *Organizacija znanja*, 25(1–2), 2025005 (str. 1–18). Dostopno na: <https://doi.org/10.3359/oz2025005> [13. 8. 2021]

*Kriteriji za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI*, 2009. Dostopno na: [http://home.izum.si/IZUM/program\\_izobrazevanja/COBISS3\\_Katalogizacija/Kriteriji\\_za\\_ocenjevanje\\_bibliografskih\\_in\\_normativnih\\_zapisov.pdf](http://home.izum.si/IZUM/program_izobrazevanja/COBISS3_Katalogizacija/Kriteriji_za_ocenjevanje_bibliografskih_in_normativnih_zapisov.pdf) [14. 6. 2018].

Letno poročilo o delu IZUM za leto 2020, 2021. Maribor: IZUM. Dostopno na: [https://www.izum.si/wp-content/uploads/2021/04/Letno\\_porocilo\\_2020.pdf](https://www.izum.si/wp-content/uploads/2021/04/Letno_porocilo_2020.pdf) [27. 5. 2021].

Peteh, M. in Južnič, P., 2020. O pomenu predmetnih oznak in podatka o afiliaciji v raziskovalnih knjižnicah v COBISS.SI: praksa izpolnjevanja polj bloka 6XX in podpolja 70X8. *Organizacija znanja*, 25 (1–2), 2020, 2025002, (str. 1–12). Dostopno na: <https://doi.org/10.3359/oz2025002> [27. 5. 2021].

*Raziskava med člani, uporabniki in neuporabniki slovenskih splošnih knjižnic*, 2020. Ljubljana: Ninamedia. Dostopno na: [https://www.knjiznice.si/wp-content/uploads/2020/06/ZSK\\_javnomnenska\\_raziskava\\_Ninamedija\\_2020.pdf](https://www.knjiznice.si/wp-content/uploads/2020/06/ZSK_javnomnenska_raziskava_Ninamedija_2020.pdf) [27. 5. 2021].

Redman, T. C., Fox, C. in Levitin, A., 2009. Data and data quality. Str. 1420–1431. V: *Encyclopedia of library and information sciences*. New York: Taylor and Francis.

Šauperl, A. in Semlič Rajh, Z., 2020. Kakovost podatkov v knjižničnih katalogih in arhivskih podatkovnih zbirkah. V: Katić, T. in Tomašević, N. ur. *Mirna Willer: Festschrift*. Zadar: Sveučilište u Zadru. Str. 153–172. Dostopno na: <https://morepress.unizd.hr/books/press/catalog/view/62/61/989-1> [25. 2. 2021].

Vilar, P., Bartol, T., Pisanski, J. in Južnič P., 2012. Are librarians familiar with information seeking behaviour of teachers and researchers in their respective institutions? V: *Libraries in the Digital Age (LIDA): Zadar, Croatia, 18 - 22 June 2012: proceedings*. Zadar: University of Zadar. Dostopno na: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.958.471&rep=rep1&type=pdf> [12. 4. 2021].

Wakeling, S. et al., 2017. Users and uses of a global union catalog: a mixed-methods study of WorldCat.org. *Journal of the Association for Information Science and Technology*, 68(9), str. 2166–2181. Dostopno na: <https://doi.org/10.1002/asi.23708> [27. 5. 2021].