

članek

NABOR DIMENZIJ ZA OPREDELITEV KAKOVOSTI PODATKOV V BIBLIOGRAFSKIH IN NORMATIVNIH ZAPISIH

A SET OF DIMENSIONS FOR DEFINING DATA QUALITY IN BIBLIOGRAPHIC AND AUTHORITY RECORDS

Branka BadovinacInstitut informacijskih znanosti,
MariborKontaktни naslov:
branka.badovinac@izum.si**Izveček**

Za potrebe izvajanja analiz kakovosti podatkov v bibliografskih in normativnih zapisih smo oblikovali nabor dimenzij, s katerimi kvalitativno opišemo, kakšne podatke želimo v zapisih. Nabor smo oblikovali z metodo pregleda literature in tipologije napak, izdelane z analizo priporočil v okviru spremljanja kakovosti bibliografskih in normativnih zapisov dnevne produkcije v COBIB.SI v letu 2017. Končni nabor obsega 11 različnih dimenzij in zajema različne podatkovne zahteve v okviru naslednjih dveh osnovnih kategorij dimenzij: kakovost vrednosti podatka in kakovost reprezentacije (oblike) podatka. Na podlagi teh smo podali opredelitev kakovosti podatka v bibliografskih in normativnih zapisih. Izdelava nabora dimenzij je pokazala nekatere terminološke ovire strokovnih izrazov. Ustreznost nabora dimenzij je treba preveriti z analizami, kjer se bodo oblikovale tudi mere in metode merjenja ter izhodišča vrednotenja kakovosti podatkov.

Ključne besede

kakovost podatkov, bibliografski zapisi, normativni zapisi, dimenzije

Abstract

A set of dimensions was developed for the purposes of quality analysis of data of bibliographic and authority records. The dimensions are qualitative descriptions of the types of data required in records. The set was developed by literature review and typology of errors, prepared by analyzing the recommendations given to librarians within quality control of daily bibliographic records production in COBIB.SI in 2017. The final set consists of 11 dimensions; it includes different data requirements within the following two basic categories of dimensions: quality of data value and quality of data representation. Based on the dimensions, data quality in bibliographic and authority records was defined. When developing the dimensions, we had to face some terminological issues. The relevance of dimensions should be checked by actual analysis of data, when metrics and methods, and the basis for quality assessment will be formed.

Keywords

data quality, bibliographic records, authority records, dimensions

UVOD

Pojem kakovost podatkov/zapisov je v katalogizaciji pogosto izpostavljen, posebej z mislijo na končne uporabnike. Poskusi opredelitve so razpeti med pragmatičnostjo (najpogosteje glede kakovosti kataloga) in univerzalnostjo (s težnjo zaobjeti celoten življenjski krog podatka). A kot je razbrati iz mnogih študij, je pri obeh pristopih težko doseči strokovno soglasje (prim. Snow, 2017). Dojemanje in vrednotenje kakovosti podatkov pogojujejo med seboj prepleteni dejavniki, kot so končni uporabnik, enotna obdelava, organizacija/racionalizacija delovnega procesa, katalogizator in tehnologija/programska oprema. Zato se zavzemamo za pristop, da je kakovost podatkov v

katalogizaciji treba razumeti kontekstualno (Badovinac, 2017).

V nadaljevanju bomo h kakovosti podatkov pristopili pragmatično in se posredno navezali na kontekst enotne obdelave, vendar se ne bomo ukvarjali s kakovostjo standardov ali shem oziroma posledično s tem, kako ti vplivajo na razumevanje kakovosti podatkov, temveč samo z dejanskimi podatki, ki temeljijo na teh standardih (katalogizacijski pravilniki, formati). Za potrebe aktivnosti spremljanja kakovosti bibliografskih in normativnih zapisov dnevne produkcije v COBIB.SI (Dornik, et al., 2017) želimo opredeliti kakovost podatkov v bibliografskih in normativnih zapisih, ki nam bo omogočala preprostejšo in

preglednejše razumevanje analiz vsebine priporočil, ki smo jih zapisali pri pregledu zapisov.

DIMENZIJE

Kakor navaja strokovna literatura, za opredelitev kakovosti podatkov najprej potrebujemo nabor **dimenzij** (prim. Badovinac, 2017). Z dimenzijo namreč kvalitativno opišemo, kakšne podatke si želimo. Dimenzije dajejo kvalitativno razumevanje kontrole kakovosti, ki je del splošnega okvira zagotavljanja kakovosti (slika 1). Na podlagi posamezne dimenzije kasneje določimo mere in metode merjenja ter vrednotenje kakovosti podatkov.



Slika 1: Splošni okvir zagotavljanja kakovosti podatkov v bibliografskih in normativnih zapisih

Dimenzije niso neodvisne, med njimi obstajajo korelacije, ki so določene s specifikko proučevanega področja. V strokovnih virih obstaja mnogo različnih opredelitev, poimenovanj in kategorizacij dimenzij. S pragmatičnega vidika morajo biti dimenzije osnovane smiselno, da jim lahko določimo merljive attribute (prim. Batini in Scannapiecco, 2016). Z večjim številom dimenzij lahko natančneje opišemo kakovost podatka.

Dimenzije lahko oblikujemo po določenih značilnostih, tako npr. Redman, Fox in Levitin (2009) ločijo naslednje tri skupine: kakovost konceptualnega modela, kakovost vrednosti in kakovost reprezentacije. Wang in Strong (1996) ločita naslednje štiri skupine dimenzij: inherentne, kontekstualne in predstavitevne dimenzije ter dimenzije, ki omogočajo dostopnost podatkov. Inherentne so tiste dimenzije, ki zajamejo kakovost podatka takšnega, kot je; kontekstualne upoštevajo kontekst uporabe podatkov; predstavitevne so osredotočene na obliko podatkov; dostopnost pa je tisti segment, ki določa varnost in dostopnost podatka. Avtorja sta opozorila, da odločitev, katere dimenzije so pomembne za določeno opredelitev kakovosti, lahko izhaja iz teoretskega, intuitivnega ali raziskovalnega pristopa. Izbiramo pa lahko tudi med različnimi metodami določanja dimenzij, npr. intervjuji, analiza literature, študije uporabnikov ipd.

METODOLOŠKA ZASNOVA

Za potrebe analiz, ki nastajajo na podlagi aktivnosti spremljanja kakovosti bibliografskih in normativnih zapisov dnevne produkcije, želimo nabor dimenzij, ki se bo osredotočil le na podatke v bibliografskih in

normativnih zapisih. V skladu s tem se ne ukvarjamo s kakovostjo konceptualnih modelov in njihovih podatkovnih domen, ki jih v nadaljevanju poimenujemo referenčni vir (katalogizijski pravilniki, formati itn.). Naš namen v prvi vrsti torej ni ovrednotenje referenčnega vira v COBIB.SI, temveč zaznava težav, na podlagi katerih se oblikujejo nadaljnje aktivnosti za izboljšanje kakovosti podatkov, kar je naslednji nivo v splošnem okviru zagotavljanja kakovosti podatkov (npr. uvedba programskih kontrol, dopolnjevanje dokumentacije) (slika 1).

Podatek je v našem primeru vsebina ali natančneje vrednost podpolja in vrednost indikatorja v okviru posameznega podatkovnega elementa, kot to definira format COMARC (slika 2). Podatkovni element je v formatih MARC najmanjša informacijska enota, ki jo je mogoče nedvoumno prepoznati (Bibliotekarski terminološki slovar, 2009). Zapis v formatu COMARC je sestavljen iz 1) označevalcev vsebine (oznake polj, podpolj, indikatorjev), ki so namenjeni prepoznavanju podatkovnih elementov in omogočajo nadaljnje upravljanje z zapisom, in 2) vsebine podatkovnih elementov, običajno definirane s standardi, ki niso del formata, npr. s katalogizacijskimi pravili. Posamezno polje lahko vsebuje enega ali več podatkovnih elementov, ki lahko dopolnjujejo pomen indikatorja. Indikator je numerični znak v polju, ki prinaša dodatne informacije o vsebini polja/podpolja, o relacijah med polji v zapisu ali o aktivnostih za rabo podatkov (prim. Brešar, 2004).

V skladu s podatkovnimi zahtevami referenčnih virov nas specifično zanimata dve skupini dimenzij, ki se nanašajo na kakovost vrednosti (angl. *data value*) in kakovost reprezentacije (oblike) podatkov (angl. *data format*, *data representation*) (prim. Redman, et al., 2009; Badovinac, 2017). Slednje se nanaša le na tiste podatkovne elemente, v katere se podatki vnašajo v pol- ali nestrukturirani obliki, npr. blok 2XX in 3XX, in tista kodirana podpolja, v katerih oblika/struktura podatkov ni programsko kontrolirana (npr. podpolje 127a). Vrednost indikatorja pa je vedno le kodiran in strukturiran podatek. Vsi ti podatki so besedilni ali numerično določeni z naborom znakov COBISS.



Slika 2: Primer dveh podatkov v zapisu formata COMARC

Glede na opredelitev našega podatka se torej oziramo po trendu granulacije in dekonstrukcije zapisa, saj je kontrola kakovosti usmerjena na posamezne podatkovne elemente zapisa, čeprav morajo biti struktura, zastopanost in vsebina

podatkovnih elementov smiselne glede na celoten zapis opisanega vira in celotne baze zapisov (prim. Badovinac, 2017; Krstulović, 2006).

Ne nazadnje smo se odločili, da bo oblikovanje nabora dimenzij temeljilo na kombinaciji teoretskega in raziskovalnega pristopa: pri izbiri poimenovanj in pomenov dimenzij smo se poskušali približati strokovni literaturi; na osnovi analize priporočil, ki so nastala ob pregledu dnevne produkcije v letu 2017, pa smo izdelali tipologijo napak v podatkih, ki jih lahko zaznamo v okviru posamezne dimenzije.

Pregled literature

V strokovnih virih najdemo različne izraze za pojem *dimenzije*, npr. smernice, indikatorji, parametri, razsežnosti, kazalci, pogoji, zahteve, atributi, komponente ipd. Pregled literature s področja katalogizacije je pokazal manjšo uporabnost dosedanjih izsledkov. V preteklosti ni bilo večjega zanimanja za dimenzije, prevladovali sta le dve: točnost in popolnost, in sicer kot minimalna pogoja za kakovost zapisov kot celote, pogosto pa sta bili opisani ohlapno (prim. Badovinac, 2017; Snow, 2017). Novejši poskusi razširitve nabora dimenzij so za naše potrebe po eni strani zastavljeni preširoko, saj zajamejo tudi značilnosti, npr. sheme, družbeni kontekst idr. (prim. Bruce in Hilman, 2004; Stvilia in sodelavci, 2007). Po drugi strani sta izbor in namen dimenzij usmerjena na ozko področje, npr. dimenzija časovnosti pri normativnih zapisih (Moulaision, 2015) ali dimenzije, povezane z zagotavljanjem kakovosti posameznega informacijskega sistema in servisov, kot je ISSN (Oury, 2017). Pregledali smo tudi nekaj virov s področja kakovosti podatkov nasploh, npr. Batini in Scannapiecco (2016) ter Redman, Fox in Levitin (2009), ki so prav tako le posredno vplivali na oblikovanje in poimenovanje dimenzij.

Posredno smo uporabili *Kriterije za ocenjevanja bibliografskih zapisov v COBISS.SI* (2009), ki so bili zasnovani za spremljanje kakovosti 100 naključno izbranih zapisov na leto in ki so namenjeni vrednotenju zapisa kot celote (prim. Kavčič, 2012). Ti kriteriji kažejo stopnjo obveznosti, pomembnosti ali drugo značilnost podatkovnega elementa.

Tipologija napak na osnovi priporočil v okviru spremljanja kakovosti bibliografskih in normativnih zapisov dnevne produkcije v COBIB.SI

Priporočila k bibliografskim in normativnim zapisom nastajajo na podlagi metodologije zajema, vzorčenja in pregleda v okviru aktivnosti spremljanja zapisov dnevne produkcije COBIB.SI. V skladu s tem je treba

upoštevati, da gre za majhen vzorec (letno je pregledanih približno 2.500 bibliografskih zapisov s pripadajočimi normativnimi zapisi), da gre za pregled brez primarnega gradiva in da gre za pregled z metodo ekspertnega mnenja (prim. Hider in Tan, 2008). Raven zanesljivosti priporočil zvišujemo s strokovnimi posveti sodelavcev in odzivi katalogizatorjev, poleg tega pa ob zaključku aktivnosti na letni ravni preverimo odzivnost katalogizatorjev in ustreznost popravljenih zapisov ter hkrati razumljivost in ustreznost priporočil. Popis priporočil temelji na določilih referenčnega vira, kreator zapisa napake odpravi ročno (prim. Dornik, et al., 2017). Značilnost teh podatkov je, da so v času kontrole kakovosti razmeroma stabilni, zapisi so pregledani v 6 dneh po datumu kreiranja in v tem času jih redko redigirajo drugi katalogizatorji.

V popisu pregleda uporabljamo dve obliki priporočil za opis pomanjkljivosti. Prva oblika omogoča razmeroma preprosto identifikacijo tipa napake, saj je struktura naslednja: *Oznaka polja/bloka: [Vrsta napake] v polju/podpolju/bloku. Pojasnilo k napaki. Referenčni vir pojasnila*. Ta oblika ima še dve podvariaciji. V primeru manj zanesljivega priporočila uporabimo naslednjo strukturo: *Oznaka polja/bloka: Preverite vir, [vrsta napake] v polju/podpolju/bloku (?). Pojasnilo k napaki. Referenčni vir pojasnila*. Če gre za tip pomanjkljivosti, ki je po referenčnih virih neobvezen podatek, uporabimo strukturo: *Oznaka polja: Priporočamo [npr. vnos] podatka v polju/podpolju (?). Referenčni vir priporočila*. Kadar pa zaznamo pomanjkljivosti na nivoju zapisa, kot so napake pri izbiri predloge zapisa (maska), odnosu do drugih zapisov (npr. duplikati) ali odnosu do drugih baz podatkov (npr. predrugačenje prenesenih zapisov), navedemo splošno, manj strukturirano opombo.

V letu 2017 smo izvedli 40 zajemov (skupaj 22.051 bibliografskih zapisov), v vzorec je bilo vključenih in pregledanih 2.196 bibliografskih zapisov s pripadajočimi normativnimi zapisi. Te zapise je kreiralo 378 različnih katalogizatorjev iz 185 različnih ustanov. K bibliografskim in pripadajočim normativnim zapisom smo napisali 2.279 priporočil glede na posamezne podatkovne elemente. Ker nas ni zanimala pogostost posameznih priporočil, temveč raznovrstnost, smo iz vzorca izločili dvojnike priporočil v posameznem podatkovnem elementu. Na podlagi 416 enoličnih priporočil smo identificirali 15 vrst napak (tabela 1); le-te izhajajo iz podatkovnih zahtev v obliki kontrolnih vprašanj, ki jih uporabljamo pri pregledu zapisa. V tabeli 1 so z zvezdico označene vrste napak na nivoju zapisa. V tabelo smo dodali tudi dve napaki, in sicer *dvojnik zapisa* in *predelan zapis* (iz drugih baz podatkov), ki ju v vzorcu iz leta 2017 nismo zabeležili. Za boljše razumevanje vrste napake so navedeni tudi načini odprave napake.

Tabela 1: Tipologija napak glede na priporočila v okviru spremljanja kakovosti bibliografskih in normativnih zapisov dnevne produkcije v COBIB.SI v letu 2017

Kontrolna vprašanja	Vrste napake	Način odprave napake
Ali je zapis edinstven?	<i>dvojniki zapisa*</i>	zbrisati zapis
Ali zapis ustreza opisovanemu viru izvornega zapisa?	<i>predelan zapis*</i>	zbrisati zapis
Ali se vzorec (maska) ali struktura zapisa ujema z značilnostmi oz. vrsto opisovanega vira?	napačna maska*/ nejasna struktura podatkov*	spremeniti masko/ strukturo zapisa
Ali je obvezen podatek naveden?	manjka obvezen podatek	dodati podatek
Ali je podatek jasen oz. razumljiv?	nejasen podatek	preoblikovati ali zbrisati podatek
Ali je podatek naveden v pravem polju/podpolju?	podatek naveden v napačnem polju/ podpolju	prenesti podatek v drugo polje/podpolje
Ali je podatek v pravilnem vrstnem redu glede na polja/podpolja?	nepravilen vrstni red polj/podpolj	spremeniti vrstni red polj/podpolj
Ali je podatek pomensko točen?	napačen podatek	zamenjati/preoblikovati podatek
Ali je podatek točno prepisan (črkovan)?	zatiptkan podatek	prečrkovati podatek
Ali je podatek celoten?	pomanjkljiv podatek	dopolniti podatek
Ali se podatek ujema z drugimi podatki v zapisu?	protisloven podatek oz. neujemanje podatkov v dveh ali več podatkovnih elementih	uskладiti podatek v enem ali v več podatkovnih elementih
Ali je podatek predviden?	odvečen podatek	zbrisati podatek
Ali je oblika oz. podoba podatka ustrezna?	napačno oblikovan podatek	preoblikovati podatek
Ali je podatek ažuren/zapis ažuriran?	neažuriran podatek/ neredigiran zapis*	posodobiti podatek/ redigirati zapis
Ali je priporočljiv podatek naveden?	manjka priporočljiv podatek	dodati podatek

Pri izdelavi tipologije napak in, kot se kasneje pokaže, tudi pri naboru dimenzij, smo imeli največ težav z izbiro ustreznih terminov. Izraz "prepis", ki se uporablja v referenčnih virih, je nerazlikovalen, saj se lahko nanaša na vsebino podatka (kaj prepisati), mesto umestitve podatka (v katero polje, podpolje prepisati) in obliko/strukturo podatka, ki je lahko pomensko značilen (npr. invertirana oblika imena) ali nepomensko značilen (npr. pravopisna,

slovnicična, stilna pravilnost), ter tudi na dejavnost vnosa podatka, tj. pretipkavanje podatka.

Tako smo imeli težavo pri razlikovanju napake, nastale zaradi napačnega oblikovanja in strukturiranja podatka, ki vpliva na njegov pomen (in poizvedbo), ter napake, ki se nanaša na obliko oziroma podobo zapisanega podatka, tj. napake, povezane s pravopisom jezika opisovanega vira in "katalogizacijskim pravopisom", kot ga določa katalogizacijski pravilnik (npr. raba velikih malih začetnic po ISBD, raba oglatih oklepajev), ki praviloma ne vpliva na iskanje podatkov. Vse napake iz prve skupine smo obravnavali kot napačne podatke, medtem ko druga skupina pomanjkljivosti zajema obvezujoče in neobvezujoče oblikovne značilnosti zapisanega podatka.

Posebej smo izpostavili tudi napako zaradi zatiptkanja, kajti ta je po *Kriterijih za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI* (2009) in številnih študijah opredeljena kot posebej velika pomanjkljivost, saj vpliva na iskanje zapisov ter verodostojnost in ugled katalogizacijske ustanove ali servisa. V to skupino napak ne zajamemo le napak pri ročnem vnosu, tj. tipkanju podatka, temveč tudi napake, ki so posledica kopiranja in lepljenja celote ali dela podatkov oziroma drugega načina vnosa ali prenosa podatkov. Prav tako zajamemo tudi napake, ki so posledica rabe napačnih transliteracijskih tabel za vnos nelatiničnih virov podatkov.

OPIS DIMENZIJ

Vrstni red nabora dimenzij, ki sledi, ne kaže njihove pomembnosti; posamezno dimenzijo lahko uporabljamo neodvisno od drugih, vendar je pri nekaterih dimenzijah določeno, da mora podatek v okviru posamezne dimenzije zadostiti pogojem druge dimenzije kakovosti. Pri poimenovanju smo dodali tudi skrajšano ime oziroma oznako dimenzije.

Vsaka dimenzija ima podan kratek opis (definicijo), v opisu uporabljamo izraz *vir podatkov* za opisovani vir oziroma predlogo, *referenčni vir* pa je tisti vir, ki vsebuje deklarativne omejitve (izjave) glede ravnanja s podatki iz opisovanega vira. Z drugimi besedami, referenčni vir opiše podatkovne zahteve (npr. obveznost podatka, oblika podatka ipd.). V našem primeru je referenčni vir objavljena dokumentacija za vzajemno katalogizacijo v sistemu COBISS.SI (npr. priročniki, pravilniki, navodila, sklepi komisij, dokumenti za potrebe vodenja bibliografij, priročniki za programsko opremo ipd.).

Zaradi lažjega razumevanja smo pri posamezni dimenziji dodali še primere vrste napak, ki izhajajo iz primerov priporočil, in posamezne primere priporočil, ki smo jih oblikovali za spremljanje kakovosti zapisov dnevne

produkcije v letu 2017. Primeri so podani v celoti, saj je za posamezno dimenzijo potrebno kontekstualno razumevanje pomanjkljivosti, ki vpliva na posamezno dimenzijo, torej glede na podatkovni element, vrsto napake, kontekst napake in referenčni vir. V naboru smo za ilustracijo dodali še oznake napak po *Kriterijih za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI* (2009), v predstavitvi dimenzij so nekatere napake iz teh kriterijev navedene večkrat.

Semantična točnost (SEMTOČ)

Opis: Podatek je semantično točen, kadar je pomensko pravilen glede na vir podatkov in referenčni vir.

Pogoj: Podatek je strukturalno skladen.

Opomba: Primerjaj z dimenzijo Vsebinska popolnost. Angleški izraz je *semantic accuracy*.

Primeri vrste napak: napačen podatek, napačna koda, napačen indikator (kadar ta določa pomen podatka v polju/podpolju), napačna povezava med bibliografskim in normativnim zapisom (vnesen napačen avtor), napačno oblikovan/strukturiran podatek (vpliva na pomen/iskanje, npr. napačno oblikovana točka dostopa).

Primeri priporočil iz pregleda dnevne produkcije (2017):

Polje	Priporočilo
210	Napačen podatek v podpolju 210c. V podpolje 210c navedemo ime založnika ali distributerja (gl. COMARC/B, polje 210).
101	Preverite vir, napačen 1. indikator v polju 101. Kadar je enota v izvornem jeziku, je vrednost 1. indikatorja "0" (gl. COMARC/B, polje 101).
675	Napačen podatek v podpolju 675v. Skladno z uporabo 3. elektronske izdaje Univerzalne decimalne klasifikacije UDCMRF 2011 (2014) v podpolju 675v vpisujemo kodo UDCMRF 2011 (gl. COMARC/B, polje 675; UDCMRF 2011, http://udcmrf2011.nuk.uni-lj.si/Predstavitev.aspx).
100	Napačen podatek v podpolju 100b. Če v publikaciji leto izida ni navedeno, določimo pa lahko verjetno leto izida, vnesemo v podpolje 100b kodo d – publikacija, zaključena ob izidu ali v enem koledarskem letu, v podpolje 100c pa verjetno leto izida. V podpolju 210d zapišemo podatek o verjetnem letu izida v oglasnem oklepaju in dodamo vprašaj (gl. Navodila za vnos podatkov o letu izida (2009), http://home.izum.si/cobiss/e-forumi/katalogizacija/docs/Navodila_za_vnos_podatkov_o_letu_izida.pdf).

Kategorija napak po *Kriterijih za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI* (2009): B-09, B-18, B-20, B-21, B-24, B-26, B-27, B-35, B-37, B-38, B-40, B-44, N-1, N-2, N-4, N-5, N-7.

Točnost prepisa (TOČPRE)

Opis: Podatek je prepisan točno, kadar se ujema z nizom znakov glede na vir podatkov ali referenčni vir.

Pogoj: Podatek je strukturalno skladen in semantično točen.

Opomba: Angleški izraz je *syntactic accuracy*.

Primeri vrste napak: zatiptkan podatek v katerem koli polju/podpolju, napačen znak NSB/NSE, napačen prepis zaradi napačne transliteracijske tabele pri nelatiničnih virih podatkov.

Primeri priporočil iz pregleda dnevne produkcije (2017):

Polje	Priporočilo
500	Zatiptkan podatek v podpolju 500a (Diary).
200, 710, 970	Zatiptkan podatek v podpoljih 200f, 710a in 970a (International).

Kategorija napak po *Kriterijih za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI* (2009): B-12, B-13, B-21, B-27, B-33b, N-1, N-4.

Strukturalna popolnost (STRUPOP)

Opis: Strukturalna popolnost zahteva, da je podatek obvezno prisoten glede na zahteve referenčnega vira.

Pogoj: /

Opomba: Primerjaj z dimenzijo Dodana vrednost. Angleški izraz je *structural completeness*.

Primeri vrste napak: manjka koda/podatek v polju/podpolju (obvezen podatek), manjka koda/podatek v polju/podpolju glede na drugo polje/podpolje ali izvorni zapis (npr. preneseni zapisi iz ISSN, WorldCat ipd.).

Primeri priporočil iz pregleda dnevne produkcije (2017):

Polje	Priporočilo
011	Manjka podatek v podpolju 0111. Od leta 2014 se iz baze ISSN v zapisih prenaša tudi povezovalni ISSN, in sicer v podpolje 0111 ISSN-L. To podpolje povezuje izdaje na drugih medijih in ga ne brišemo (gl. COMARC/B, polje 011).
300	Manjka obvezna opomba o viru stv. nasl. v podpolju 300a. Npr. Nasl. z nasl. zaslona (gl. Prekat, pogl. 7.1.2.1).
200	Manjka podatek o odgovornosti v podpolju 200f (gl. COMARC/B, polje 200; Prekat, pogl. 1.5).
400, 810	Manjkata variantna normativna točka dostopa (polje 400) in njen vir (polje 810) za obliko imena, ki je navedena na viru (gl. COMARC/A, polji 400, 810; Značka, pogl. 2.2.1). [normativni zapis]

Kategorija napak po *Kriterijih za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI* (2009): B-10, B-17, B-25, B-29, B-30, B-33a, B-36, N3, N-6, N-7.

Vsebinska popolnost (VSEBPOP)

Opis: Podatek je vsebinsko popoln, kadar je celoten oziroma izčrpen glede na vir podatkov in zahteve referenčnega vira.

Pogoj: Podatek je strukturalno skladen, semantično točen in točno prepisan.

Opomba: Primerjaj z dimenzijo Semantična točnost. Stopnja obveznosti podatka ni pomembna. Angleški izraz je *value completeness*.

Primeri vrste napak: pomanjkljiv podatek, manjkajo znaki NSB/NSE.

Primeri priporočil iz pregleda dnevne produkcije (2017):

Polje	Priporočilo
200	Manjkajo znaki za izključitev razvrščanja in iskanja NSB/NSE (\neq – prečrtani enačaj) v podpolju 200a. V bibliografskem opisu vse začetne določne in nedoločne člene iskalnih elementov (naslov(e), ime založbe, naslov zbirke itn.) zaradi pravilnega razvrščanja po abecedi opremimo z znaki za izključitev razvrščanja NSB/NSE (\neq – prečrtani enačaj) (gl. COMARC/B, polje 200; Kavčič, I. (2011): Najpogostejša vprašanja in dileme pri kreiranju bibliografskih zapisov, http://www.nuk.uni-lj.si/infobib/images/stories/Dokumenti/Najpogostejsa_vprasanja_20120613.pdf).
215	Pomanjkljiv podatek v podpolju 215d. Navedemo ustrezen izraz za posebno oznako gradiva, ki mu enota pripada, npr. cm (gl. Prekat, pogl. 5.1).
314	Pomanjkljiv podatek v podpolju 314a. Priporočamo dopolnitev z izrazom "Mentor" (prim. Napotki za katalogizacijo doktorskih disertacij, magistrskih in diplomskih del (2006), http://home.izum.si/cobiss/obvestila_novosti/dokumenti/Napotki_za_katalogizacijo_doktorskih_disertacij_magistrskih_in_diplomskih_del_2007_01_05.pdf).

Kategorija napak po *Kriterijih za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI* (2009): B-01, B-07, B-16, B-20, B-33.

Strukturalna skladnost (STRUSKLAD)

Opis: 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.

Pogoj: Podatek je semantično točen in točno prepisan.

Opomba: Angleški izraz je *structural consistency*.

Primeri vrste napak: podatek je naveden v napačnem polju/podpolju, nepravilen vrstni red polj/podpolj, podatki so navedeni v napačni maski (predlogi), iz zapisa ni jasno, za kakšno vrsto publikacije gre, napačna predelava izvornega zapisa (npr. WorldCat, .Net, LC Names ipd.).

Primeri priporočil iz pregleda dnevne produkcije (2017):

Polje	Priporočilo
200	Podatek je naveden v napačnem podpolju (200b). Dodatek k nasl. navajamo v podpolju 200e (gl. COMARC/B, polje 200; Prekat, pogl. 1.4, dodatek A; Kavčič, I. (2011): Najpogostejša vprašanja in dileme pri kreiranju bibliografskih zapisov, http://www.nuk.uni-lj.si/infobib/images/stories/Dokumenti/Najpogostejsa_vprasanja_20120613.pdf).
225	Napačen vrstni podpolj v polju 225. Podatke vnesemo v polje v takšni obliki in zaporedju, kot predpisuje ISBD (225axv) (gl. COMARC/B, polje 225; Prekat, pogl. 6.0.1).
710	Napačen vrstni red podpolj v polju 710. Vrstni red podpolj kvalifikatorjev je dfe (gl. COMARC/B, polje 710; Prekat, pogl. 14.5.1).

Kategorija napak po *Kriterijih za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI* (2009): B-04, B-08, B-19a, B-19b, B-28, B-31, B-34, B-41, B-43, B-45, B-46, N-1, N-4, N-8.

Vsebinska skladnost (VSEBSKLAD)

Opis: Podatki so vsebinsko skladni, kadar se med seboj strukturalno in/ali pomensko ujemajo, kot to določa referenčni vir.

Pogoj: Podatki so točno prepisani in strukturalno skladni.

Opomba: Angleški izraz je *semantic coherency*. Dimenzija se običajno uporablja pri analizi bibliografskih zapisov, ki so bili pregledani brez predloge.

Primeri vrste napak: protisloven podatek oziroma neujemanje podatkov v poljih/podpoljih, nejasen podatek.

Primeri priporočil iz pregleda dnevne produkcije (2017):

Polje	Priporočilo
100, 210	Neujemanje podatkov v podpoljih 100b, 100c in 100d in 210d (gl. Navodila za vnos podatkov o letu izida (2009), http://home.izum.si/cobiss/e-forumi/katalogizacija/docs/Navodila_za_vnos_podatkov_o_letu_izida.pdf).
200, 700	Neujemanje podatkov v podpolju 200f in polju 700 (gl. COMARC/B, polji 200, 700).

Kategorija napak po *Kriterijih za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI* (2009): B-23, B-42.

Oblikovna doslednost (OBLIDOS)

Opis: Podatek je oblikovno dosleden, kadar je prepis oblike v skladu z zahtevami referenčnega vira.

Pogoj: Podatek je strukturalno skladen, točno prepisan, semantično točen, vsebinsko popoln ter vsebinsko skladen.

Opomba: Primerjaj z dimenzijama Semantična točnost in Vsebinska popolnost. Angleški izraz je *format consistency*.

Primeri vrste napak: napake v obliki, ki ne vplivajo na pomen ali iskanje, kot so npr. pravopisne in slovnične napake, oblikoslovne napake (kratice, okrajšave), raba začetnic po ISBD, napačna raba oglatih oklepajev, napačna raba (odvečna) NSB/NSE.

Primeri priporočil iz pregleda dnevne produkcije (2017):

Polje	Priporočilo
340	Napačno oblikovana opomba v podpolju 340a. Besedilo opombe je v jeziku katalogizacije (gl. Uporaba programske opreme COBISS3/Katalogizacija – ažuriranje CONOR, http://home.izum.si/IZUM/program_izobrazevanja/Prosrojnice/T63.pdf).
200, 710	Pravopisna napaka v podpoljih 200e in 710a. V angleških imenih korporacij (sem spadajo tudi imena založnikov) in sestankov pišemo vse besede, razen predlogov, členov in veznikov, z veliko začetnico (ne glede na to, kje v bibliografskem opisu se ime pojavlja), npr. Oxford University Press ipd. (gl. Kavčič, I. (2011): Najpogostejša vprašanja in dileme pri kreiranju bibliografskih zapisov, http://www.nuk.uni-lj.si/infobib/images/stories/Dokumenti/Najpogostejša_vprasanja_20120613.pdf).
301	Napačna raba velikih/malih začetnic po ISBD-ju/pravopisna napaka v podpolju 301a. Podatek navedemo z veliko začetnico (Prekat, pogl. 0.6.2; COMARC/B, polje 301).

Kategorija napak po *Kriterijih za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI* (2009): B-02, B-03, B-05, B-06, B-07, B-14, B-15, B-21, B-50, B-51, N-7.

Edinstvenost (EDIN)

Opis: Zapis je edinstven, kadar nima dvojnikov na nivoju baze zapisov glede na zahteve referenčnega vira.

Pogoj: /

Opomba: Angleški izraz je *uniqueness*.

Primeri vrste napak: zapis je dvojniki (duplikat).

Primeri priporočil iz pregleda dnevne produkcije (2017): /

Kategorija napak po *Kriterijih za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI* (2009): B-39.

Odvečnost (ODVEČ)

Opis: Podatek je odvečen, kadar ni predviden glede na referenčni vir.

Pogoj: /

Opomba: Primerjaj z dimenzijama Strukturalna skladnost in Semantična točnost. Angleški izraz je *redundancy*.

Primeri vrste napak: odvečni podatek v katerem koli polju ali podpolju.

Primeri priporočil iz pregleda dnevne produkcije (2017):

Polje	Priporočilo
230	Odvečen podatek v podpolju 230a (2.). Podatek je naveden že v prvem polju 230 (gl. COMARC/B, polje 230).
540	Odvečen podatek v podpolju 540a. Podatek je iskalen že v podpolju 200e; v polje 540 navedemo dodatne nasl., ki jih na predlogi ni (gl. COMARC/B, polje 540).
210	Polje 210 je odvečno, zapis za dogodek kreiramo, kadar za določeno intelektualno ali umetniško delo ni na voljo njegova pojavna oblika na fizičnem mediju ali spletu, zato ne uporabljamo polj in podpolj, namenjenih podatkom o založbi, distribuciji, izdaji ali izidu gradiva, ter polj in podpolj, namenjenih opisu fizičnih lastnosti dela (prim. COMARC/B, dodatek F.9).

Kategorija napak po *Kriterijih za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI* (2009): B-16, B-48, B-50.

Dodana vrednost (DODV)

Opis: Podatek je dodana vrednost, kadar je predviden, vendar po referenčnem viru ni obvezen.

Pogoj: /

Opomba: Primerjaj z dimenzijo Strukturalna popolnost. Angleški izraz je *value added*.

Primeri vrste napak: manjka priporočljiv podatek.

Primeri priporočil iz pregleda dnevne produkcije (2017):

Polje	Priporočilo
100	Priporočamo vnos kode v podpolje 100b (gl. COMARC/B, polje 100).
600-610	Priporočamo vnos predmetnih oznak v polja 600–610 (gl. COMARC/B, blok 6XX; SSG, pogl. Navodila za predmetno označevanje, http://www.nuk.uni-lj.si/ssg/navodila.html).
830	Priporočamo vnos podatkov o avtorjevi identiteti v podpolje 830a (npr. področje delovanja ipd.) (gl. COMARC/A, polje 830). [normativni zapis]

Kategorija napak po *Kriterijih za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI* (2009): B-01, B-10, B-11, B-47, B-49, B-53, N-9.

Aktualnost (AKT)

Opis: Podatek je aktualen, kadar kaže trenutnost ali pa je posodobljen glede na vir podatkov ali zahteve referenčnega vira.

Pogoj: /

Opomba: Uporablja se zlasti za normativne zapise, kjer sta potrebna redakcija kratkih zapisov in posodabljanje posameznih podatkov. Primerjaj tudi z dimenzijama Vsebinska popolnost in Oblikovna doslednost. Angleški izraz je *currency*.

Primeri vrste napak: neredigiran podatek/zapis, zastarel podatek.

Primeri priporočil iz pregleda dnevne produkcije (2017):

Nivo zapisa	Priporočilo
Normativni zapisi	Priporočamo, da v skladu s svojimi pooblastili in dostopnimi informacijami izpolnite/ažurirate normativne zapise, saj s tem pomagate drugim katalogizatorjem, da izberejo pravilno normativno točko dostopa, in ne kreirajo dvojnika. Glede na pooblastilo ažurirate polji 001 in 100; izpolnite tudi polja/podpolja 101, 102, 106, 120a, 190 (191), 820 (po potrebi) in 830 (opombo o avtorjevem področju delovanja) (gl. COMARC/A; COBISS3/ Katalogizacija, pogl. 9.9).

Kategorija napak po *Kriterijih za ocenjevanje bibliografskih in normativnih zapisov v COBISS.SI* (2009): B-06, B-51, N-9.

RAZPRAVA IN ZAKLJUČEK

Končni nabor obsega 11 različnih dimenzij, ki zajamejo različne podatkovne zahteve v okviru naslednjih dveh osnovnih kategorij dimenzij: kakovost vrednosti podatka in kakovost reprezentacije (oblike) podatka. Iz tabele 1 je razvidno, da nekatere dimenzije lahko vključujejo več vrst napak, saj se lahko nanašajo na isto podatkovno zahtevo, kot to določajo referenčni viri.

S povzetka nabora dimenzij ugotovimo, da smo popolnost in skladnost natančneje opredelili z vidika strukture in vsebine. Točnost podatka smo opredelili glede na njegov pomen in natančnost prepisa, oblikovni vidik podatka pa smo natančneje določili še z dimenzijo oblikovne doslednosti, ki se nanaša na pravopisne in oblikovne značilnosti podatka. S stopnjo obveznosti navedbe podatka

so povezane tri dimenzije: odvečnost, vsebinska popolnost in dodana vrednost. Dimenzija odvečnost je obratno sorazmerna z dimenzijo vsebinska popolnost. Slednja zahteva obvezno navedbo podatka, medtem ko odvečnost zahteva izbris redundantnih podatkov. V tem sklopu smo določili še dimenzijo dodana vrednost, ki omogoča navedbo predvidenih, a neobveznih podatkov z namenom povečati informativnost zapisa. Slednje smo nadgradili tudi z zahtevami dimenzije aktualnost, ki predvideva posodabljanje podatkov oziroma dopolnjevanje zapisov. Na nivoju zapisa smo z dimenzijo edinstvenost določili, da zapis ne sme imeti dvojnikov.

Tabela 2: Povzetek nabora dimenzij glede na kategorijo dimenzije, podatkovno zahtevo in tip napak

Oznaka dimenzije	Ime dimenzije	Kategorija dimenzije (in podatkovna zahteva)	Vrsta napake
SEMTOČ	Semantična točnost	kakovost vrednosti podatka (nanaša se na pomen podatka)	napačen podatek
TOČPRE	Točnost prepisa	kakovost vrednosti podatka (nanaša se na prepis niza znakov, ki vpliva na pomen podatka)	zatiptan podatek
STRUPOP	Strukturalna popolnost	kakovost vrednosti podatka (nanaša se na obvezno prisotnost podatka v strukturi zapisa)	manjka podatek
VSEPOP	Vsebinska popolnost	kakovost vrednosti podatka (nanaša se na celovitost vsebine podatka v podatkovnem elementu)	pomanjkljiv podatek
STRUS-KLAD	Strukturalna skladnost	kakovost vrednosti podatka (nanaša se na lokacijo in strukturo podatkovnih elementov v zapisu)	nepravilen vrstni red polj/podpolj, podatek, naveden v napačnem polju/podpolju, napačna maska/nejasna struktura podatkov, predelan zapis
VSEB-SKLAD	Vsebinska skladnost	kakovost vrednosti podatka (nanaša se na pomensko ujemanje med podatkovnimi elementi v zapisu)	protisloven podatek oz. neujemanje dveh ali več podatkovnih elementov, nejasen podatek
OBLIDOS	Oblikovna doslednost	kakovost oblike podatka (nanaša se na obliko prepisa, ki ne vpliva na pomen podatka)	napačno oblikovan podatek

EDIN	Edinstvenost	kakovost vrednosti podatka (nanaša se na edinstvenost zapisa glede na celotno bazo zapisov)	dvojniki zapisa
ODVEČ	Odvečnost	kakovost vrednosti podatka (nanaša se na redundantnost podatka v strukturi zapisa)	odvečen podatek
DODV	Dodana vrednost	kakovost vrednosti podatka (nanaša se na informativnost zapisa)	priporočljiv podatek
AKT	Aktualnost	kakovost vrednosti podatka (nanaša se na aktualnost podatka/zapisa)	nežuriran podatek/ neredigiran zapis

Na podlagi oblikovanega nabora dimenzij lahko podamo naslednjo definicijo kakovosti podatkov v bibliografskih in normativnih zapisih:

Podatek v bibliografskih in normativnih zapisih je kakovosten, kadar je naveden v edinstvenem bibliografskem ali normativnem zapisu, semantično točen in točno prepisan, strukturalno popoln, vsebinsko popoln, strukturalno skladen, vsebinsko skladen, oblikovno dosleden, predviden oziroma neodvečen, aktualen, (lahko) je dodatno informativen.

Nabor dimenzij, ki smo jih v tem članku predstavili, ni nujno zaključen, dimenzije lahko združujemo ali kategoriziramo, k naboru pa lahko po potrebi dodamo še druge dimenzije. V prihodnje želimo preveriti uporabnost nabora dimenzij v konkretnih analizah, pri čemer je treba določiti mere in metode merjenja ter končno tudi možnosti vrednotenja kakovosti podatkov v bibliografskih in normativnih zapisih.

Reference

- Badovinac, B., 2017. Izhodišča za proučevanje kakovosti podatkov v bibliografskih in normativnih zapisih: kakovost podatkov v kontekstu in raziskovalne usmeritve v katalogizaciji. *Knjižnica*, [online] 61(1–2), pp. 119–154. Dostopno na: <https://knjiznica.zbds-zveza.si/knjiznica/article/view/6165/5812> [14. 6. 2018].
- Batini, C. in Scannapieco, M., 2016. *Data and information quality: dimensions, principles and techniques*. Berlin: Springer.
- Bibliotekarski terminološki slovar*, 2009. Ljubljana: Zveza bibliotekarskih društev Slovenije.
- Brešar, T., 2004. Primerjava formatov MARC21 – UNIMARC – COMARC. *Organizacija znanja*, [online] 9(3). Dostopno na: http://home.izum.si/cobiss/oz/2004_3/html/clanek_04.html [14. 6. 2018].
- Bruce, T. R. in Hillman, D. I., 2004. The continuum of metadata quality: defining, expressing, exploiting. V: Hillmann D. in Westbrooks E. ur. *Metadata in practice*. Chicago: American Library Association. pp. 238–256. Dostopno na: <http://www.ecommons.cornell.edu/handle/1813/7895> [14. 6. 2018].
- Dornik, E., Badovinac, B., Kos, J. in Farkaš, B., 2017. Sistem zagotavljanja kakovosti COBIB.SI: izbrane aktivnosti za leto 2016. *Knjižnica*, [online] 61(1–2), pp. 191–205. Dostopno na: <https://knjiznica.zbds-zveza.si/knjiznica/article/view/6167/5814> [14. 6. 2018].
- Hider, P. in Tan, K., 2008. Constructing record quality measures based on catalog use. *Cataloging and classification quarterly*, 46(4), pp. 338–361.
- Kavčič, I., 2012. Kakovost zapisov v vzajemni bibliografsko-kataložni bazi podatkov COBIB.SI. *Knjižničarske novice*, [online] 22(6), pp. 1–19. Dostopno na: <https://old.nuk.uni-lj.si/knjiznicarskenovice/v2/podrobnostClanek.aspx?id=599> [14. 6. 2018].
- Kriteriji za ocenjevanje bibliografskih in normativnih zapisov v COBISS. SI*, 2009. [pdf] Dostopno na: https://www.cobiss.si/e-forumi/katalogizacija/docs/Nakljucni_zapisi_kriteriji_objava_20090421.pdf [14. 6. 2018].
- Krstulović, Z., 2006. Katalogizacijska pravila in kakovost bibliografskih podatkov. *Organizacija znanja*, 11(4), pp. 215–218.
- Moulaison, H. L., 2015. The expansion of the personal name authority record under Resource Description and Access: current status and quality considerations. *IFLA journal*, 41(1), pp. 13–24.
- Oury, C., 2017. Assessing the ISSN Register: defining, evaluating, and improving quality of shared international bibliographic database. *Cataloging & classification quarterly*, [online] 55(7–8), pp. 588–605. Dostopno na: <https://doi.org/10.1080/01639374.2017.1354115> [14. 6. 2018].
- Redman, T. C., Fox, C. in Levitin, A., 2009. Data and data quality. V: *Encyclopedia of library and information sciences*. New York: Taylor and Francis, pp. 1420–1431.
- Snow, K., 2017. Defining, assessing, and rethinking quality cataloging. *Cataloging & classification quarterly*, [online] 55(7–8), pp. 438–455. Dostopno na: <https://doi.org/10.1080/01639374.2017.1350774> [14. 6. 2018].
- Stvilja, B., Gasser, L., Twidale, M. B. in Smith, L. C., 2007. A framework for information quality assessment. *Journal of the American Society for Information Science and Technology*, 58(12), pp. 1720–1733.
- Wang, R. Y. in Strong, D. M., 1996. Beyond accuracy: what data quality means to data consumers. *Journal of management information systems*, 12(4), pp. 5–33.