



**PAM** Pokrajinski  
arhiv  
Maribor

*Moderna*  
**arhivistika**

Časopis arhivske teorije in prakse  
Journal of Archival Theory and Practice

ISSN 2591-0884

<https://doi.org/10.54356/MA>

Letnik 6 (2023), št. 2 / Year 6 (2023), No. 2

Maribor, 2023

Prejeto / Received: 19. 05. 2023

1.01 Izvirni znanstveni članek

1.01 Scientific article

<https://doi.org/10.54356/MA/2023/MJUO4040>

## **STANJE IN PERSPEKTIVE VZAJEMNEGA METAPODATKOVNEGA KORPUSA SLOVENSKE JAVNE ARHIVSKE SLUŽBE**

**Dr. Miroslav NOVAK**

arhivski svetnik, Pokrajinski arhiv Maribor, Slovenija

[miroslav.novak@pokarh-mb.si](mailto:miroslav.novak@pokarh-mb.si)

### **Izvleček:**

*Namen: Vzajemni arhivski metapodatkovni korpus Slovenske javne arhivske službe obsega več milijonov podatkovnih entitet, ki so bile zajete v obdobju 2009/10 do 2022. Osnovni namen raziskave je, da na podlagi kvantitativnih metod opredelimo stopnjo kakovosti zajetih arhivskih metapodatkov. Na tej osnovi je v nadaljevanju opravljena validacija njegove potencialne uporabnosti tako z vidika posrednih ali neposrednih uporabnikov kot tudi sistemov, ki temeljijo na umetni inteligenci.*

*Metoda/pristop: Za potrebe tega prispevka je bilo uporabljenih več metod. Osnovo predstavljajo statistične metode, ki so bile uporabljene za obdelavo podatkov iz letnih statističnih poročil o stanju in prirastu podatkov v vzajemni arhivski podatkovni bazi. Ob tem je bila uporabljena primerjalna metoda pridobljenih rezultatov iz začetka leta 2023 s podatki primerljivih analiz iz leta 2013 in nato 2016. Za predstavitev kompleksnosti obvladovanja celotnega problema pa so bile uporabljene še naslednje metode: metoda analize, opisna metoda, metoda povzemanja in izkustvena metoda. Na podlagi pridobljenih statističnih podatkov in s pomočjo analize SWOT so ob koncu prikazane prednosti, slabosti in priložnosti ter nevarnosti obstoječe arhivske strokovne prakse generiranja slovenskega arhivskega metapodatkovnega korpusa. Podrobne analize obdelav rezultatov po posameznih arhivih ali nižjih popisnih enot niso predmet te raziskave.*

*Rezultati: Na podlagi opravljenih analiz statističnih podatkov avtor ugotavlja, da število zapisov nominalno raste iz leta v leto. Njihova rast pa je opredeljena s povprečno pol milijona zapisi in njihovimi povezavami letno. Hkrati z nominalno rastjo zapisov pa podrobna analiza trendov kaže, da so ti različni skozi daljše obdobje po posameznih modulih. Rezultati tudi kažejo, da bo treba izvesti nekatera uravnoteženja zajemov med moduli, saj se bo le na tej osnovi celoten korpus razvijal enakomerno, hkrati pa bodo uporabnikom in drugim sistemom zagotovljene celovite informacije o ohranjenih arhivskih entitetah v slovenskih javnih arhivih.*

*Sklepi/ugotovitve: Rezultati statističnih obdelav podatkov o vzajemnem podatkovnem korpusu SJAS kažejo na relativno visoko stopnjo profesionalnosti zaposlenih v SJAS. Hkrati s tem nakazujejo, da bo treba posvetiti ustrezno pozornost standardizaciji opisovanja kontekstov, doslednejšemu povezovanju zapisov o enotah popisa z drugimi zapisi v sami podatkovni zbirki in drugimi zapisi v drugih relevantnih podatkovnih zbirkah. Prav tako bo treba dosledneje uporabljati statistične metode na nižjih ravneh popisnih enot, saj bo le tako mogoče sproti ugotavljati morebitne pomanjkljivosti v zapisih in izvajati potrebne popravke. Hkrati pa bo treba nekatere rešitve prilagoditi zahtevam standarda ISO 24083:2021. Razumevanje gibanja statističnih podatkov v podatkovnem korpusu SJAS ni pomembno le za upravjalce podatkovne zbirke SJAS, ampak tudi za različne uporabnike arhivskega gradiva in drugih storitev v javnih arhivih pa tudi za morebitne nadaljnje implementacije v sistemih, ki jih s skupnim imenom označujemo »umetna inteligenca«.*

**Ključne besede:**

*vzajemna podatkovna zbirka, arhivski metapodatki, statistične metode, standardizacija, slovenska javna arhivska služba.*

**Abstract:**

**Status and perspectives of the mutual metadata corpus of the Slovenian Public Archival Service**

*Purpose: The mutual archival metadata corpus of the Slovenian Public Archives Service (hereinafter: SJAS) comprises several million data entities that were captured in the period 2009/10 to 2022. The basic purpose of the research is to define the quality level of captured archival metadata based on quantitative methods. On this basis, the validation of its potential usefulness is carried out both from the point of view of indirect or direct users, as well as systems based on artificial intelligence.*

*Method / approach: Several methods were used for the purposes of this paper. The basis is represented by statistical methods that were used to process data from annual statistical reports on the state and growth of data in the mutual archival database. At the same time, the comparative method of the obtained results from the beginning of 2023 with the data of comparable analyzes from 2013 and then 2016 was used. To present the complexity of managing the entire problem, the following methods were also used: analysis method, descriptive method, summary method and experiential method. On the basis of the obtained statistical data and with the help of a SWOT analysis, the advantages, disadvantages, opportunities and threats of the existing archival professional practice of generating the Slovenian archival meta data corpus are shown at the end. Detailed analyzes of processing results by individual archives or lower units of description are not the subject of this research.*

*Results: Based on the analysis of statistical data, the author notes that the number of records nominally grows from year to year. Their growth was defined by an average of half a million records and their links annually. At the same time as the nominal growth of records, a detailed analysis of trends shows that these are different over a longer period by individual modules. The results also show that some catch-balancing between modules will need to be done. After all, only on this basis will the entire corpus develop evenly, and at the same time users and other systems will be provided with comprehensive information about preserved archival entities in Slovenian public archives.*

*Conclusions / findings: The results of the statistical processing of data about the mutual data corpus of SJAS show a relatively high level of professionalism among the employees in SJAS. At the same time, they indicate that the necessary attention will have to be paid to the standardization of describing contexts, to a more consistent connection of records about units of description with other records in the database itself and records in other modules of the Slovenian archival metadata corpus. It will also be necessary to implement selected statistical methods more consistently at the lower levels of units of description, as this is the only way it will be possible to identify potential deficiencies in the records and implement the necessary corrections. At the same time, some solutions will have to be adapted to the requirements of the ISO 24083:2021 standard. Understanding the movement of statistical data in the SJAS data corpus is not only important for the managers of the SJAS database, but also for various users of archival material and other services in public archives, as well as for possible further implementations in systems known as "artificial intelligence".*

**Keywords:**

*Mutual database, archival metadata, statistical methods, standardization, Slovenian public archival service.*

## 1 Uvod

Razlogov za uvajanje intenzivne digitalizacije arhivske dejavnosti je veliko. Omenimo naj le naraščanje količine podatkov, ki jo v arhivih generirajo, obdelujejo, ohranjajo in uporabljajo v digitalnem okolju; zahteve po splošni dostopnosti do arhivskega gradiva s pomočjo tehnološke infrastrukture itd. Strateška usmeritev glede digitalizacije v arhivih je med drugim tudi skladna s splošnimi trendi razvoja slovenske družbe v digitalnem okolju (prim. Digitalna Slovenija 2030).

Arhivski strokovni izzivi v tej zvezi pa niso samo v segmentu implementacije sodobnih tehnik in poizvedovalnih metod, ampak so veliko globlji. Ti se kažejo v zmožnosti posameznikov in sistemov, da natančno razpoznavajo entitete v množici in na tej osnovi oblikujejo odgovore na zastavljena vprašanja. Pri tem pa ni več toliko pomembno, ali je odgovore oblikoval človek ali stroj, ampak je pomembna stopnja relevance in nepristranost odgovorov na zastavljena vprašanja.

Arhivisti se z izzivi razpoznavanja arhivskih entitet srečujejo že od samih začetkov arhivske dejavnosti. Iz izkušenj je znano, da je izziv razpoznavanja vedno premo sorazmeren s pomanjkanjem kontekstnih in drugih metapodatkov za razumevanje vsebin, ki so v fokusu raziskovanja (Novak, 2021). Iz tega lahko izpeljemo spoznanje, da s kvalitativnimi pristopi in posledično s kvantitativnim povečevanjem količine podatkov o arhivskem gradivu in/ali iz arhivskega gradiva omogočamo lažje razpoznavanje arhivskih entitet.

V nadaljevanju te raziskave se bomo omejili le na kvantitativne vidike upravljanja podatkovnih struktur v Vzajemnem metapodatkovnem korpusu Slovenske javne arhivske službe (dalje: VMK SJAS). Na kvalitativno stopnjo doseženih rezultatov popisovanja v slovenskih javnih arhivih pa lahko le sklepamo glede na dosežene kvantitativne rezultate. Poglobljena kvalitativna analiza VMK SJAS ni predmet raziskave, prav tako niso predmet kvantitativnih raziskav podrejene celote popisnih enot na nižjih ravneh.

VMK SJAS sestavljajo obsežne in kompleksne podatkovne strukture, ki odgovarjajo na mnoga vprašanja uporabnikov, a hkrati s tem pogosto generirajo nove arhivske dileme in nove zahteve, ki se manifestirajo tudi kot nezmožnost hitrega (pre)oblikovanja in implementiranja novih rešitev v obstoječo arhivsko teorijo in prakso. To nezmožnost v slovenski arhivski teoriji in praksi kompenzirajo z vedno višjo stopnjo standardizacije na področju arhivskega popisovanja. Izkušnje kažejo, da le-ta prispeva k izboljšanju kvalitete informacijske ponudbe. Hkrati pa standardizacija na nivoju zapisov omogoča uporabo kompleksnejših statističnih obdelav podatkov in s tem bolj jasen uvid v realno stanje VMK SJAS.

V Sloveniji je prišlo do prvih poskusov izgradnje podatkovnih baz arhivskega gradiva s pojavom prvih računalnikov drugi polovici osemdesetih let prejšnjega stoletja. Pomembno prelomnico razvoja na tem področju predstavlja poskus vzpostavitve sistema AERISS (Novak, Drobnjak, 2002), ki je v nadaljnjem razvoju predstavljal dobro osnovo leta 2008 vzpostavljene podatkovne zbirke scopeArchiv v Arhivu Republike Slovenije in leto pozneje tudi vzajemne podatkovne zbirke v slovenskih regionalnih arhivih SIRAnet (Cvelfar, 2010). Leta 2020 je bil na podlagi obeh podatkovnih zbirk vzpostavljen VMK SJAS. Celoten proces vzpostavitve tega korpusa je bil izveden v okviru projekta e-Arh.si.

VMK SJAS opredeljujemo kot kompleksen soodvisen sistem, ki ga sestavlja 18 aktivnih vertikalnih modulov<sup>1</sup> in s tem povezanih vzajemnih zbirk metapodatkov. Zato predstavlja osrednje digitalno metapodatkovno okolje arhivskega strokovnega dela zaposlenih v SJAS. Z organizacijsko-vsebinskega stališča pa sistem sestavlja 7 navideznih lokalnih zbirk, ki jih po horizontali vzajemno gradijo vsi slovenski javni arhivi.

V praksi nekateri pogosto VMK SJAS neustrezno omejujejo le na modul Popisovanje. Res je, da v tem modulu lahko delujejo zapisi o popisnih enotah posameznega arhiva kot zaključena celota, vendar je treba upoštevati, da so ti v VMK SJAS praviloma povezani z drugimi zapisi, npr. o deskriptorjih, o prevzemih, o tehničnih enotah ali z zapisi o popisnih enotah iz drugih arhivov. Celoten korpus je torej veliko več kot le vsota vseh v sistem vključenih zbirk podatkov. S stališča analize podatkov je zato treba upoštevati ne samo število zajetih zapisov o posameznih tipih entitet (PE, TE, lokacije, prevzemi itd.) oz. njihovih kardinalnih števil, ampak tudi število vzpostavljenih povezav med zapisi.

Vzpostavljanje povezav med zapisi predstavlja pomemben kazalnik uspešnosti izgradnje VMK SJAS. Hkrati pa s tem lahko določamo stopnjo medsebojne vpetosti zapisov o arhivskih entitetah. V tehničnem in vsebinskem smislu povezave postavljajo relacijska razmerja, ki omogočajo kompleksnejšo praktično vsakodnevno uporabo podatkovne zbirke, predvsem pa zagotavljajo natančno arhivsko strokovno obravnavanje tako povezanih vsebin.

Z večanjem števila zapisov in vzpostavljanjem njihovih povezav se premo sorazmerno povečuje tudi kompleksnost razmerij med entitetami obdelav. Na ta način se teoretično približujemo intencam upravljanja zapisov v kontekstu oz. obdelavi arhivskih vsebin v kontekstih (RiC). Prav zato morajo biti arhivski strokovni delavci posebej pozorni na oblikovanje zahtev popisovanja in zajemanja podatkov ter s tem povezanih strateških usmeritev razvoja VMK SJAS kot celote.

Uspešnost zajemanja podatkov v VMK SJAS lahko sledimo skozi primerjave rezultatov statističnih obdelav podatkov. Podobne trende opažamo tudi v drugih segmentih arhivske dejavnosti, kjer so intenzivno uporabljene različne statistične metode. V tem kontekstu naj izpostavimo primer javno objavljene analize iz leta 2009, ki jo je pripravil Jure Maček z naslovom Pregled dela v slovenskih državnih regionalnih arhivih (Maček, 2009). Dober vpogled v dejansko uporabo statističnih metod v arhivih dobimo tudi na podlagi:

- izdelav poročil o delu in doseženih rezultatih arhivskega strokovnega dela, poslovanja arhivov itd.,
- oblikovanja podatkov za potrebe različnih vodnikov in drugih arhivskih informativnih pomagal,
- pridobljenih rezultatov upravljanja skladiščnih kapacitet in mikroklima v skladiščih,
- spremljanja uporabe arhivskega gradiva in izvajanju drugih storitev v arhivih,
- popisovanja različnih arhivskih entitet itd.

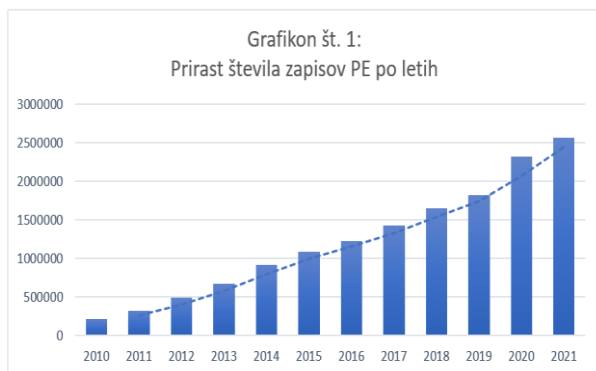
<sup>1</sup> Korpus podatkovne zbirke SJAS sestavljajo zbirke podatkov v naslednji moduli: modul Popisne enote, modul Deskriptorji, modul Zajem, modul Akcesija, modul Lokacije, modul Tehnične enote, modul Izposoja, modul Reprodukcijske, modul Partnerji, modul Dosjeji, modul Dokumenti, modul Dogodki, modul Aktivnosti, modul Klasifikacije, modul Obrazci, modul Podatkovni elementi, modul Nabori, Modul Poročila in modul Sistemski uporabniki (Scope).

S tovrstnimi aktivnostmi sledimo tudi namenu standarda ISO 24083:2021 z naslovom Informatika in dokumentacija – Mednarodna statistika arhivov<sup>2</sup>, ki opredeljuje vsebine in načine zbiranja podatkov za statistične namene s področja arhivske dejavnosti (Novak, 2022).

VMK SJAS predstavlja osnovo za izvajanje arhivskega strokovnega dela in uporabe arhivskega gradiva v slovenski javni arhivski službi. Več kot desetletje sistematičnega zbiranja podatkov o razvoju in obsegu vzajemne podatkovne zbirke ter dve opravljeni poglobljeni raziskavi iz leta 2013 (Novak, Semlič Rajh, 2023) in 2016 (Semlič Rajh, 2016) omogočajo osnovni vpogled v razvoj in obseg vzajemne podatkovne zbirke. Zadnja, neobjavljena analiza s tega področja pa je bila izvedena v postopkih fuzije podatkovnih baz ARS in SIRAnet v skupno VMK SJAS. Ta baza je v začetku leta 2023 vsebovala 7.000.000 zapisov o arhivskih entitetah in njihovih medsebojnih povezavah.

## 2 KVANTITATIVNA ANALIZA ZAJETIH PODATKOV V VMK SJAS

### 2.1 Značilnosti zajemanja popisnih enot



Slovenski arhivski strokovni delavci zajemajo podatke o arhivskem gradivu v modul Popisne enote sistema scopeArchiv (Scope). Zajemanje je izvedeno na različne načine. Do sedaj običajna načina sta s tipkovnico ali na podlagi paketnih zajemov s pomočjo namenske programske opreme, ki je del programskega paketa scopeArchiv, iz drugih, aktivnih, opuščeni ali začasno vzpostavljenih arhivskih podatkovnih okolij<sup>3</sup>.

<sup>2</sup> V povzetku standarda lahko preberemo, da ta določa smernice za arhivsko skupnost o zbiranju in poročanju statističnih podatkov za naslednje namene:

- strateško načrtovanje in upravljanje arhivov;
- za potrebe združevanja in primerjave statističnih podatkov o delovanju in uspešnosti na regionalni, nacionalni in mednarodni ravni;
- za poročanje o rezultatih zainteresirani javnosti, kot so različne finančne ustanove, politika v najširšem pomenu besede, raziskovalci itd.;
- za spodbujanje vloge in vrednosti arhivov v družbi, še posebej za napredek učenja in raziskovanja, izobraževanja in kulture ter družbenega in gospodarskega življenja; za izboljšanje upravljanja procesov, krepitev preglednosti in podpiranje dobrega upravljanja arhivskih ustanov.

<sup>3</sup> Paketni zajemi podatkov se izvajajo tako, da se spremenljivke v podatkovno zbirko prenesejo s pomočjo dveh zunanjih orodij – Finding Aids Assistant in Transfer Assistant – ali s pomočjo posebnega vtičnika Data Import.

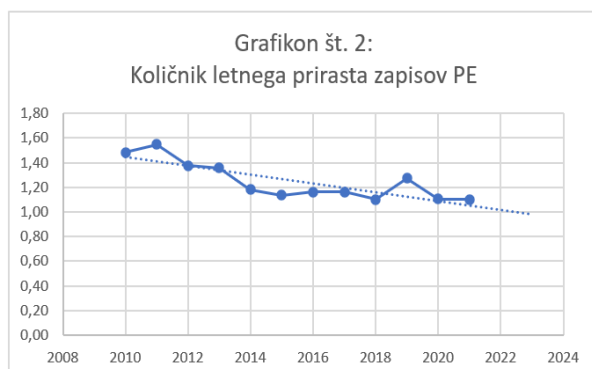
Osnovna analiza števila zajetih zapisov o popisnih enotah (PE) kaže, da je bilo teh konec leta 2010 nekaj več kot 214.000. To število se je do konca leta 2022 povzpelo na več kot 2.820.000 (grafikon št. 1). Krivulja prirasta je relativno linearna, izjemno leto je bilo 2020, ko je število zapisov poskočilo zaradi združevanja podatkovne baze slovenskih regionalnih arhivov s podatkovno bazo Arhiva Republike Slovenije v VMK SJAS.

Količniki letnega prirasta zapisov popisnih enot glede na predhodno leto kažejo intenziteto zajemanja podatkov v tekočem letu glede na predhodno. Pričakovana vrednost količnika je vsaj 1, kar pomeni, da je bilo v predhodnem letu zajeto enako število zapisov kot v tekočem letu.

Analiza vrednosti količnika kaže, da je bila intenziteta zajemanja zapisov od leta 2014 dalje praviloma z negativnim trendom, odklon najdemo le leta 2017 (tabela št. 1). Sicer pa je razpon vrednosti v celotnem obravnavanem obdobju med 1,10 do 1,55. Pri tem je pomembno, da vrednost količnika ves ta čas nikoli ni padla pod 1. Relativno visoki količniki na začetku prejšnjega desetletja so lahko odraz splošnega absolutno nizkega števila zajetih zapisov in množičnega uvažanja podatkov iz lokalnih podatkovnih okolij (grafikon št. 2), trend dolgoročno ni ugoden, zato kaže temu problemu posvetiti večjo arhivsko strokovno in tudi organizacijsko pozornost.

leto	količnik
2010	↑ 1,48
2011	↑ 1,55
2012	→ 1,37
2013	→ 1,36
2014	↓ 1,18
2015	↓ 1,14
2016	↓ 1,16
2017	→ 1,16
2018	↓ 1,10
2019	→ 1,27
2020	↓ 1,10
2021	↓ 1,10

**Tabela 1: količnik prirasta zapisov popisnih enot glede na predhodno leto**



Rešitve, ki se uveljavljajo pri sodobnem prevzemanju arhivskega gradiva, da skladno z veljavno zakonodajo ustvarjalec ali predajnik arhivskega gradiva zagotovi podroben popis arhivskega gradiva, so v tem kontekstu zelo zaželeni. To velja še posebej za primer, ko pozneje v pristojni arhivski ustanovi ni potreben večji vložek v normalizacijo tako izdelanih zapisov. Vedno bolj uveljavljene rešitve popisovanja arhivskega gradiva s

pomočjo »outsourcinga« lahko spremenijo trend prirasta zapisov, vendar le ob intenzivnem popisovanju še nepopisanega arhivskega gradiva v pristojnih arhivskih ustanovah. V zvezi s popisovanjem gradiva, ki je že v arhivskih skladiščih, zaznavamo problem, da so obstoječi aktivni uporabniki na sistemu zelo obremenjeni še z drugimi deli in nalogami, tako da popisovanje arhivskega gradiva v tem kontekstu ni vedno prioriteta, zato bo treba razmisliti tudi o alternativnih rešitvah, kot so npr.: uvajanje dodatne delovne sile v obliki »crowdsourcinga«, razvijanje drugih, bolj racionalnih metod oblikovanja zapisov, uvajanje izboljšav v procese digitizacije v arhivih, iskanje rešitev v smeri intenzivne uporabe generativne umetne inteligence itd.

## 2.2 O uporabnikih



Podatki o uporabnikih sistema so zajeti v modulu Uporabniki. Slednje delimo v »aktivne« in »pasivne«. Aktivni uporabniki so tisti, ki zajemajo, dopolnjujejo, popravljajo ali uporabljajo zapise v dostopnih modulih in jih v posameznih okoljih implementirajo pri vsakdanjem arhivskem strokovnem delu.

Druga skupina, pasivni uporabniki, pa so tisti, ki so v preteklosti prispevali zapise v vzajemno podatkovno zbirko, a so po določenem času prenehali z delom na sistemu, ker so odšli v pokoj ali so zamenjali službo in podobno.

Leto	Število vnosov
2011	1519,38
2012	1973,34
2013	1996,65
2014	2517,24
2015	1777,18
2016	1563,51
2017	1995,47
2018	2158,05
2019	1592,19
2020	3598,51
2021	1620,91
2022	1758,82

Tabela 2: Količnik med uporabniki in zapisi popisnih enot po letih

Zapisi o pasivnih uporabnikih zaradi sledenja avtorstva vsebin popisnih enot ostajajo v podatkovni bazi, vendar za njih niso potrebne licence. Enako velja za različne sistemske uporabnike, ki upravljajo s sistemom ali zagotavljajo dostope do podatkov v drugih sistemih.

Za potrebe te raziskave so zanimivi predvsem aktivni uporabniki (grafikon št. 3), saj prispevajo zapise v sistem. Podatki kažejo, da absolutno število aktivnih uporabnikov skozi obravnavano obdobje raste, vendar je v preteklosti tudi nekajkrat zanihalo. Sicer je rast števila aktivnih uporabnikov pozitiven pokazatelj, saj kaže, da lahko vedno več arhivskih strokovnih delavcev, zaposlenih v javnih arhivih, neposredno prispeva podatke v sistem.

Če pa postavimo število uporabnikov v relacijo do števila novo zajetih zapisov v posameznem letu, lahko ugotovimo, da prihaja do navideznega neskladja, saj večje število uporabnikov od leta 2014 dalje generira manjši letni prirast, kot je to uspelo manjšemu številu uporabnikov v začetnih letih delovanja sistema (grafikon št. 2).

Sicer pa, kot kaže statistika, vsak aktivni uporabnik izdelava povprečno 2005 novih zapisov popisnih enot na leto. Odklon od povprečja predstavlja leto 2020, ko je prišlo do vzpostavitve VMK SJAS in se je to letno povprečje dvignilo na 3598 zapisov (tabela št. 3).

V kontekstu te analize pa je pomembno tudi ugotavljanje povprečnega absolutnega števila zajetih zapisov o popisnih enotah glede na število aktivnih uporabnikov na sistemu po letih. Če je v letu 2011 bilo v poprečju dostopnih 4.670 zapisov PE po uporabniku, se je to število dvignilo v letu 2022 na 19.332 zapisov (tabela št. 2).

Leto	Količnik
2011	4670,24
2012	5582,16
2013	7336,11
2014	9547,68
2015	11528,00
2016	13091,51
2017	14425,80
2018	15631,20
2019	17223,39
2020	16732,89
2021	17336,26
2022	19332,55

Tabela 3: Količnik dostopnih zapisov popisnih enot po uporabniku in po letih

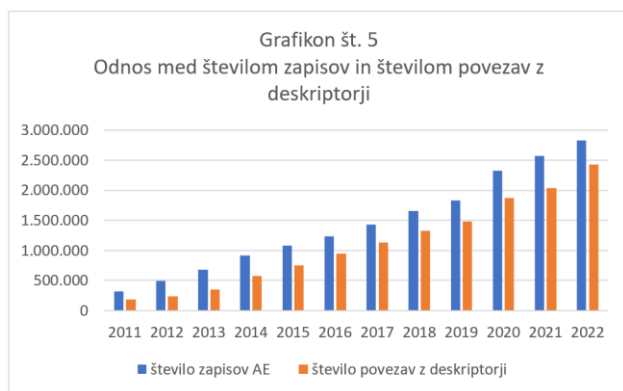




Grafična predstavitev teh podatkov pa pokaže rahlo odstopanje od linearne napovedi v letu 2020, ko je prišlo do fuzije podatkovne zbirke regionalnih arhivov z zbirko Arhiva Republike Slovenije. Kot kaže trend, se je krivulja v naslednjih dveh letih uskladila s trendom iz predhodnih let in je v naraščanju (grafikon št. 4). Trend je logična posledica splošnega povečanja števila zapisov v modulu Popisne enote v obravnavanem obdobju.

### 2.3 O povezavah med zapisi popisnih enot in normativnimi vsebinami (deskriptorji)

Pomemben sestavni del VMK SJAS predstavljajo normativne vsebine. Te so potrebne za natančno opredeljevanje in razločevanje različnih entitet arhivske obdelave. Hkrati pa omogočajo natančnejše poizvedovanje.



Popisne enote zajemamo v modulu Popisne enote, normativne vsebine pa v modulu Deskriptorji. Med zapisi v modulu Popisne enote in zapisi v modulu Deskriptorji so lahko vzpostavljene eno- ali dvosmerne relacije.

Celota normativnih zapisov je organizirana v sedem tezavrov, in sicer: tezaver fizičnih oseb, tezaver družin, tezaver pravnih oseb/korporativnih teles, tezaver zemljepisnih imen, tezaver stvarnih gesel, tezaver funkcij/aktivnosti in tezaver klasifikacije fondov in zbirk.

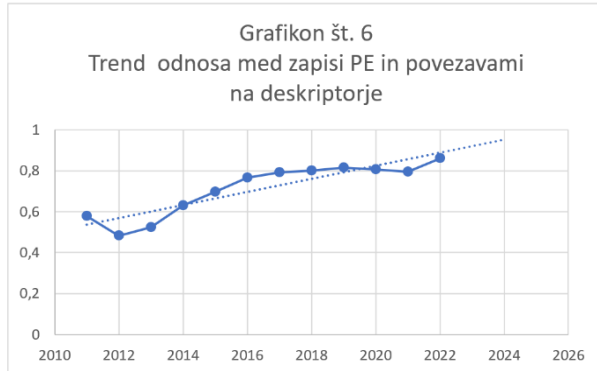
Zapisi o popisnih enotah, ki spadajo k natančno določeni normativni vsebini, so povezani s posebnimi obojestranskimi povezavami, kar pomeni, da se tudi na zapisu normativne vsebine nizajo povezave na pripadajoče popisne enote.

Praksa popisovanja arhivskega gradiva kaže, da bi bilo treba vsak zapis opredeliti z vsaj dvema deskriptorjema, in sicer z deskriptorjem, ki opredeljuje ustvarjalca arhivskega gradiva v fokusu popisovanja, in deskriptor, ki opredeljuje kraj ustvarjalca, oz. deskriptor zemljepisne entitete, ki je predmet popisa popisne enote. Na nivoju fondov in podfondov je treba upoštevati vsaj še povezavo na tezaver klasifikacije fondov in zbirk. To pomeni, da bi vsak zapis popisne enote na nivojih serije in nižje moral imeti vsaj dve povezavi na normativne vsebine. Vsak zapis popisne enote na nivojih, višje od serije, pa bi moral imeti vsaj tri povezave na normativne vsebine.

Leto	Količnik
2011	↓ 0,578693
2012	↓ 0,4827169
2013	↓ 0,5238931
2014	⇒ 0,6301478
2015	⇒ 0,6972367
2016	↑ 0,766112
2017	↑ 0,7909231
2018	↑ 0,8007426
2019	↑ 0,8134721
2020	↑ 0,8056424
2021	↑ 0,7938939
2022	↑ 0,8596143

Tabela 4: Količnik med zapis popisnih enot in zapisi o deskriptorjih po letih

Če pogledamo statistične podatke predmetnega področja, opazimo, da število povezav med zapisi deskriptorjev in zapisi popisnih enot iz leta v leto narašča. Če je bilo leta 2011 takšnih povezav nekaj več kot 183.000, jih je bilo v letu 2022 več kot 2.400.000 (grafikon št. 5). Tak trend je tudi skladen s splošnim razvojem števila zapisov v vzajemni podatkovni zbirki SJAS in rastočega zavedanja arhivskih strokovnih delavcev o dejanskih potrebah po tovrstnih povezavah. Analiza trenda vzpostavljanja povezav kaže na naraščanje, vendar z rahlimi odkloni v letu 2012 in v letu 2020, ko je prišlo do vzpostavitve VMK SJAS (grafikon št. 6).



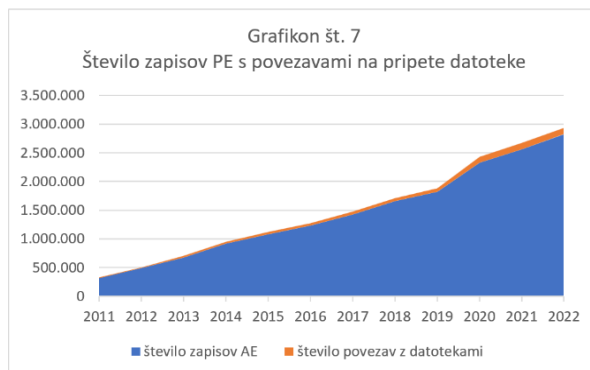
S strateškega stališča razvoja in uporabe VMK SJAS pa je zelo poveden količnik med številom zapisov popisnih enot in številom povezav do zapisov o deskriptorjih. Vrednosti po letih se

gibljejo od 0,48 do 0,85 povezave (tabela št. 4). Pričakovana vrednost je 2 za zapise od nivoja serije in nižje, ter vrednost 3 za nivoje, višje od nivoja serije. Čeprav je trend povezovanja zapisov PE z deskriptorji v smeri naraščanja, se v povprečju še ne približujemo vrednosti 1, kar bi pomenilo, da je vsak zapis teoretično povezan z vsaj enim deskriptorjem. V tem segmentu ima slovenska javna arhivska služba še veliko prostora, ki ga bo treba sistematično zapolniti. Na tej točki pa lahko zaznamo tudi nevarnost, da bi orodja, ki delujejo na podlagi generativne umetne inteligence, zaradi pomanjkanja podatkov v tem segmentu ob generiranju odgovorov močno halucinirala. To lahko zelo poveča stopnjo entropije celotnega sistema.

## 2.4 O povezavah med zapisi popisnih enot in pripetimi datotekami

V VMK SJAS je mogoče vnesti tudi vsebine, ki so organizirane v obliki datotek. Na razpolago sta dva načina:

- neposredni zajem v podatkovno zbirko v obliki podatkovnih objektov (»blob«) – to so praviloma grafični formati datotek ali
- kot povezave med zapisi v podatkovni zbirki in datotekami, ki se nahajajo na medijih z zahtevanimi uporabniškimi pravicami, do katerih je mogoče dostopati preko vmesnika scopeArchiv; v tem primeru so datoteke lahko v različnih formatih, običajno v formatu za dolgoročno hrambo.



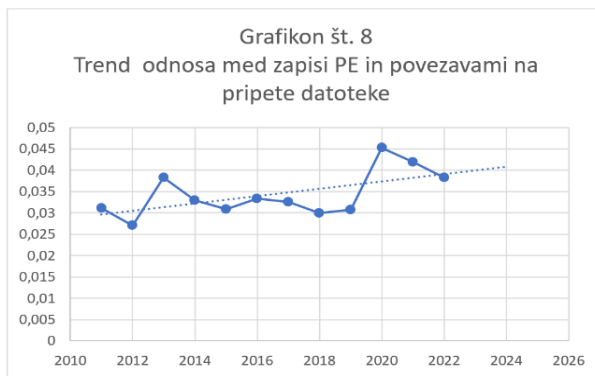
Pri statističnih obdelavah pripetega števila datotek pa je treba upoštevati dejstvo, da te lahko predstavljajo bodisi le en digitizat, bodisi eno stran arhivskega gradiva, ali pa več tisoč digitizats, orga-niziranih v eni datoteki. Zato podatkov o datotekah

ne moremo enačiti s številom dostopnih digitiziranih zapisov preko vmesnika scopeArhiv ali preko portala Virtualna arhivska čitalnica<sup>4</sup>.

V VMK SJAS tako ločimo vsebine datotek, ki se pojavljajo v posebnem elementu – galeriji, v posebnih elementih znotraj obrazca za popis popisnih enot ali kot prednastavljena rešitev kot sestavni del zapisa v posebnem zavihku »Datoteke«. Ob tem velja omeniti, da za pripete datoteke velja različna politika njihovega objavljavanja. Če so vsebine pripetih datotek sestavni del podatkovne zbirke (kot »blob«), potem je njihova objava preko portala odvisna od statusa zapisa. Če pa so datoteke povezane z zapisom v podatkovni zbirki, je za njihovo dostopnost preko portala potrebno zaključiti zapis, datoteko pa postaviti v okolje, do katerega imajo dostop tudi zunanji uporabniki. Enako velja tudi za datoteke, katerih povezave med zapisi in datotekami so izvedene preko zavihka »Datoteke«. V tem primeru je treba datoteke, ki jih želimo objaviti na spletu, dodatno opredeliti s posebnim statusom.

Leto	Količnik
2011	↓ 0,03112011
2012	↓ 0,027070822
2013	⇒ 0,038324429
2014	↓ 0,032999955
2015	↓ 0,030838883
2016	⇒ 0,033360095
2017	↓ 0,032588222
2018	↓ 0,029902101
2019	↓ 0,030694881
2020	↑ 0,045287531
2021	↑ 0,041976938
2022	⇒ 0,038229574

**Tabela 5: Količnik med zapisi popisnih enot in povezavami na pripete datoteke po letih**



#### Prirast števila

datotek v VMK SJAS znaša od malo manj kot 10.000 datotek v letu 2011 do skoraj 108.000 datotek v letu 2022. Povprečno je bilo na leto zajetih 8.992 datotek (grafikon št. 7).

Vrednost količnikov med številom zapisov PE in številom pripadajočih datotek po letih se giblje od 0,027 do 0,045 (tabela št. 5). Pri tem ugo-

tavljamo, da je trend povezovanja zapisov popisnih enot s pripadajočimi datotekami v porastu, se pa intenziteta iz leta v leto opazno spreminja. Spreminjanje vrednosti količnika lahko utemeljujemo z različnimi dejavniki. Izreden skok leta 2020 je posledica vzpostavitve vzajemnega podatkovnega korpusa SJAS. Predhodna nihanja količnika lahko razlagamo predvsem z intenziteto digitizacije na letni ravni (grafikon št. 8) in s tem povezano politiko prikazovanja digitiziranih zapisov s pomočjo Virtualne arhivske čitalnice oz. njenega predhodnika – arhivskega portala Query.

## 2.5 O povezavah med zapisi popisnih enot in zapisi o tehničnih enotah

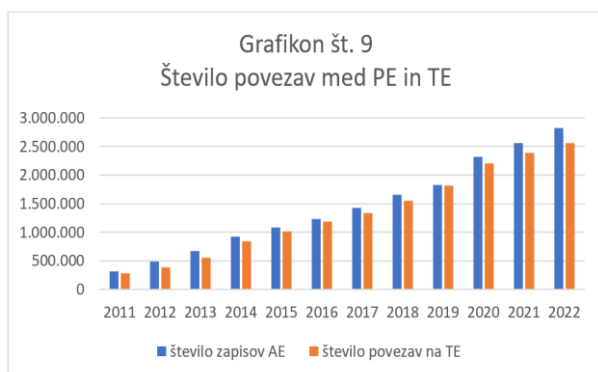
Teoretično je vsaka arhivska vsebina, shranjena na fizičnem mediju in popisana, tudi odložena v tehnično enoto. To pomeni, da mora biti vsak zapis o popisni enoti fizičnega arhivskega gradiva povezan z natančno enim zapisom o tehnični enoti. Iz tega sledi predpostavka, da mora biti vrednost količnika med številom zapisov popisnih enot in številom zapisov tehničnih enot natančno 1.

Zbrani statistični podatki kažejo, da v obravnavanem obdobju število povezav raste premo sorazmerno s številom zajetih zapisov popisnih enot. V letu 2011 je bilo tako izdelanih 277.569 povezav med zapisi popisnih enot in zapisi tehničnih enot. Število teh

<sup>4</sup> Virtualna arhivska čitalnica ali VAC je portal, preko katerega lahko dostopajo zunanji uporabniki do podatkov o arhivskem gradivu, ki ga hrani SJAS v podatkovnem korpusu SJAS. Podrobneje o tem glejte v prispevku Koncilija, Jenuš, Hajtnik, ki je bil objavljen leta 2021.

povezav se je leta 2022 povečalo na 2.564.439 (grafikon št. 9). To pa tudi pomeni, da se v okviru SJAS ta funkcionalnost precej uporablja, saj podatke o tehničnih enotah potrebujejo uporabniki v arhivskih čitalnicah in tudi arhivski strokovni delavci pri vsakdanjem arhivskem strokovnem delu.

Če analiziramo rezultate na letnem nivoju, sicer opazimo nihanja količnika predvsem v letih 2012, 2018 in 2022 (grafikon št. 10). Kljub temu je splošni trend v smeri naraščanja in je v tem smislu ugoden, saj kaže le na manjše nedoslednosti pri zajemanju podatkov.



Podrobna analiza količnika skozi čas pa pokaže, da je bila najnižja vrednost 0,786, in sicer v letu 2012, ter najvišja v letu 2019, ko se vrednost količnika dvigne na 0,99 (tabela št. 6). Povprečni količnik je 0,912, kar pomeni, da v poprečju nekaj več kot 9 % zapisov popisnih enot v VMK SJAS ni povezanih z zapisi o tehničnih enotah.

Če količnik prevedemo v absolutne vrednosti, pomeni to v letu 2023 skoraj četrt milijona osirotelih zapisov. Ta podatek pa zahteva splošni razmislek o tem, kako izdelati manjkajoče povezave za nazaj in kako z metodološkega stališča omejiti tovrstne pojave v prihodnje.



Problem je z metodološkega stališča popisovanja arhivskega gradiva še nekoliko širši, saj regionalni arhivi načeloma zapisujejo te podatke tudi v zapis o popisnih enotah in na osnovi tega nato izdelajo aktivne povezave v zavihku

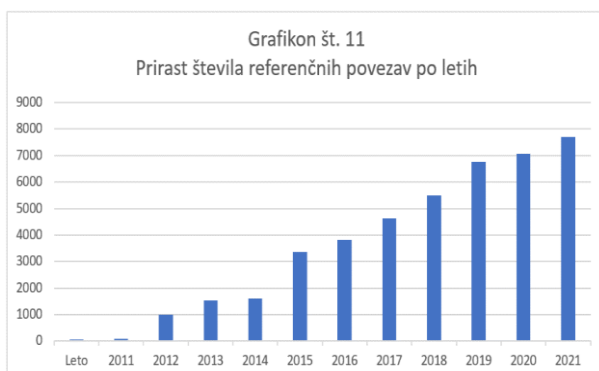
Leto	Količnik
2011	⇒ 0,874023856
2012	⇩ 0,786261018
2013	⇩ 0,814658583
2014	⇒ 0,920498769
2015	⇩ 0,933679515
2016	⇩ 0,964278459
2017	⇩ 0,937024299
2018	⇩ 0,936867308
2019	⇩ 0,993176237
2020	⇩ 0,949859236
2021	⇩ 0,930632412
2022	⇒ 0,908553001

**Tabela 6: Količnik med številom zapisov popisnih enot in zapisi o tehničnih enotah po letih**

Tehnične enote.

V Arhivu Republike Slovenije so razvili nekoliko drugačno metodo. Pri popisovanju arhivskega gradiva vzpostavljajo le aktivne povezave v zavihku, kar pomeni, da bo treba v primeru pomanjkljivo vzpostavljenih relacij med tehničnimi in popisnimi enotami te ponovno pridobiti v fizičnem okolju. Enako velja tudi v primeru regionalnih arhivov, če v opisu popisne enote ni zajetega podatka o tehnični enoti.

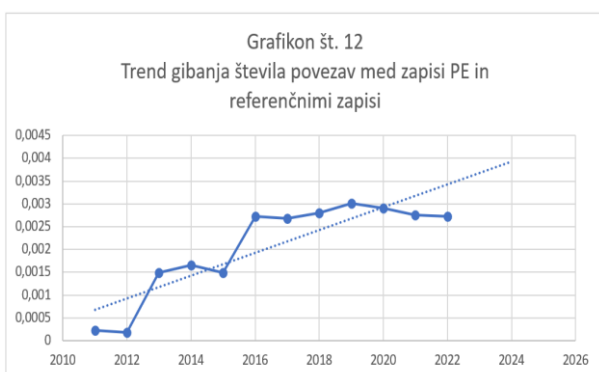
## 2.6 O povezavah med zapisi popisnih enot in referenčnimi zapisi



povezanih zapisov popisnih enot arhivskega gradiva in s tem neposredno dvignejo uporabnost takega arhivskega informativnega pomagala. Relacije, ki se pri tem vzpostavljajo, so obojestranske.

Leta 2011 je bilo v podatkovni bazi slovenskih regionalnih arhivov vzpostavljenih le 70 referenčnih povezav, do leta 2022 se je njihovo število povzpelo na več kot 7.700. Letno povprečje znaša 642 izdelanih povezav (grafikon št. 11).

V celotnem korpusu zapisov je moč množice vzpostavljenih relacij zelo šibka. To dejstvo nakazuje na to, da slovenska arhivska stroka še ni zadovoljivo opredelila pomena vzpostavljanja vsebinskih relacij ali pa teh relacij ne vzpostavlja zaradi različnih vzrokov. Omenimo naj samo nizko stopnjo izraženosti tovrstnih relacij v arhivskem gradivu, racionalizacijo potrebnega časa za zajemanje podatkov itd. Ne glede na te razloge pa je dobro, da trend gibanja števila teh povezav raste (grafikon št. 12).



Velikih sprememb na predmetnem področju kratkoročno ni pričakovati. Tovrstne povezave se standardno vzpostavljajo sukcesivno, ne pa tudi paketno. Za boljše rezultate in bolj racionalno vzpostavljanje tovrstnih relacij bi bilo treba razviti in implementirati posebno programsko rešitev znotraj odjemalca scopeArchiv.

Zapise o popisnih enotah lahko arhivski strokovni delavci povezujejo z medsebojnimi referencami. Razlogi za tovrstne povezave so praviloma vsebinski – ko se en zapis popisne enote neposredno nanaša na enega ali več drugih zapisov o popisnih enotah. S vzpostavljenimi referenčnimi povezavami lahko arhivski strokovni delavci izgradijo mrežo vsebinsko

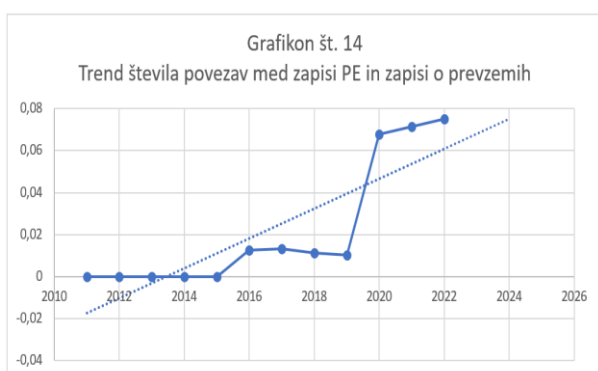
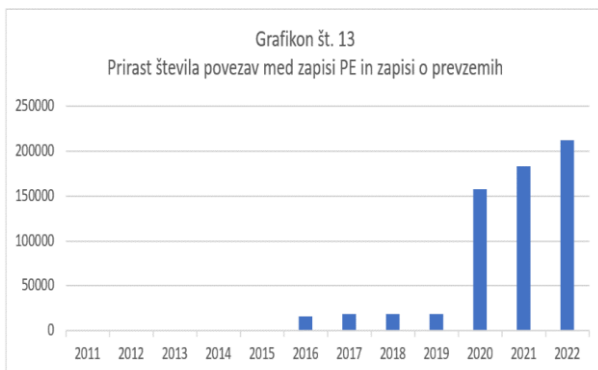
Leto	Količnik
2011	↓ 0,00022
2012	↓ 0,000175
2013	⇒ 0,001491
2014	⇒ 0,001664
2015	⇒ 0,001495
2016	↑ 0,002729
2017	↑ 0,002674
2018	↑ 0,002797
2019	↑ 0,003008
2020	↑ 0,002907
2021	↑ 0,002756
2022	↑ 0,002731

**Tabela 7: Količnik med zapisi popisnih enot in referenčnimi povezavami po letih**

Podrobnejša analiza tega področja pa izkazuje, da je bilo največ aktivnosti na njem leta 2019, saj takrat vrednost koeficienta naraste na 0,003. Sicer pa je bil dosežen leta 2012 najnižji količnik, tj. 0,000175 (tabela št. 7). Povprečni letni količnik za celotno obravnavano obdobje znaša 0,00205.

Velikih sprememb na predmetnem področju kratkoročno ni pričakovati. Tovrstne povezave se standardno vzpostavljajo sukcesivno, ne pa tudi

## 2.7O povezavah med zapisi popisnih enot in zapisi o prevzemih



nadzora materialnega varstva fizičnega arhivskega gradiva.

Statistični rezultati kažejo, da se število povezav med zapisi PE in zapisi o prevzemih dvigne po letu 2020 (grafikon št. 13), tj. po izvedeni združitvi obeh podatkovnih baz. Razlog za porast je, da regionalni arhivi tem relacijam v preteklosti niso posvečali večje pozornosti, Arhiv republike Slovenije pa jih je vzpostavljala sistematično že v obdobju pred vzpostavitvijo vzajemnega metapodatkovnega korpusa.

Količnik med številom prevzemov in številom zapisov o popisnih enotah je v zadnjih treh letih v porastu (grafikon št. 14), vendar je še daleč pod pričakovano vrednostjo, saj znaša v letu 2022 le 0,075 (tabela št. 8).

Izziv v tem segmentu je obsežen in zahteva veliko znanstveno-raziskovalnega dela, saj predstavlja z vidika digitalizacije arhivskih strokovnih postopkov v javnih arhivskih ustanovah dosežen rezultat le malo več kot 7 % pričakovanih povezav zapisov o popisnih enotah z zapisi o prevzemih. Zato v tem segmentu ostaja še veliko možnosti za izboljšave in s tem natančno vzpostavljanje relacij med obema vrstama zapisov.

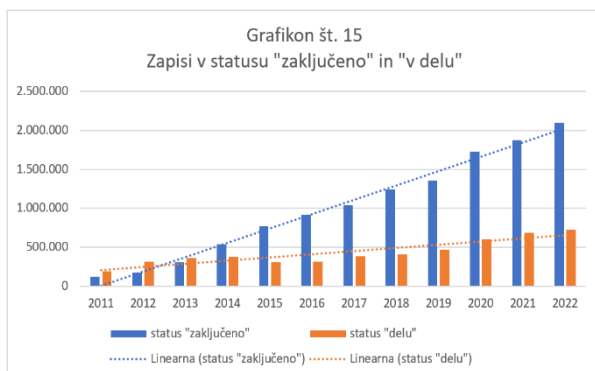
Med pomembnejše povezave, ki jih arhivski strokovni delavci vzpostavljajo v postopkih strokovnega dela znotraj metapodatkovnega korpusa SJAS, spadajo relacije med zapisi o popisnih enotah in zapisi o prevzemih.

Pri tem velja pravilo, da lahko en prevzem predstavlja enega ali več fondov oz. zbirk ali pa, da je arhivski fond ali zbirka sestavljen/-a iz enega ali več prevzemov. Zaradi teh spremenljivk je treba povezati vsak zapis popisne enote z vsaj enim zapisom o prevzemu. Iz tega lahko izpeljemo pravilo, da je količnik med številom zapisov o prevzemih in številom zapisov o popisnih enotah vsaj 1. Dosledna izvedba teh relacij bi omogočala natančen pregled prevzetega gradiva do nivoja popisnih enot. Med drugim bi s tem dosegli boljše razumevanje (historičnih) kontekstov, v nekaterih primerih pa omogočili tudi višjo raven

Leto	Količnik
2011	0
2012	0
2013	0
2014	0
2015	0
2016	0,012782
2017	0,013132
2018	0,01132
2019	0,010308
2020	0,067809
2021	0,07139
2022	0,075004

**Tabela 8: Količnik med zapisi popisnih enot in zapisi o prevzemih po letih**

## 2.8 O statusu zapisov »v delu« in »zaključeno«



Zapisi o popisnih enotah imajo lahko status »v delu« ali status »zaključeno«. Če je zapis v statusu »zaključeno«, to pomeni, da ga ni mogoče neposredno spreminjati, hkrati pa so zapisi s tem statusom javno dostopni preko Arhivske virtualne čitalnice, pod pogojem, da niso dodatno zaščiteni, npr. zaradi varstva osebnih podatkov. S stališča splošne uporabnosti je priporočljivo, da so zapisi zaključeni v najkrajšem možnem času in s tem tudi dostopni javnosti.

Konec leta 2022 je bilo 2.098.632 zapisov v statusu »zaključeno« in 723.921 zapisov v statusu »v delu«. To pomeni da je približno 1/4 zapisov popisnih enot v statusu »v delu« (grafikon št. 15). Podatek ni vzpodbuden niti za uporabnike niti za arhivske strokovne delavce. Znano je, da morajo biti zapisi, ki niso kakorkoli varovani, ali zapisi, ki opredeljujejo varovane arhivske vsebine, skladno z veljavno zakonodajo dostopni javnosti.

Zapisi s statusom »v delu«, ki so sicer z arhivskega strokovnega stališča zaključeni, vendar pa njihov status ni bil spremenjen iz »v delu« v »zaključeno«, predstavljajo v sistemu različna tveganja in s tem višanja stopnje entropije. Take zapise mogoče zelo hitro, namerno ali nenamerno, manipulirati (npr. prestavljati, spreminjati itd.), celo brisati vsebine popisnih elementov ali celotnih zapisov, če niso referenčno ali kako drugače povezani z drugimi zapisi.

Količniki med številom zapisov o popisnih enotah in številom zapisov, ki imajo status »v delu« oz. »zaključeno«, kažejo trend naraščanja zapisov, ki so v statusu »zaključeno«. Ta trend je ugoden in kaže na zavedanje arhivskih strokovnih delavcev, da je treba ustrezno urediti status zapisov v sistemu, ko je zaključeno popisovanje na obravnavani celoti arhivskega gradiva.

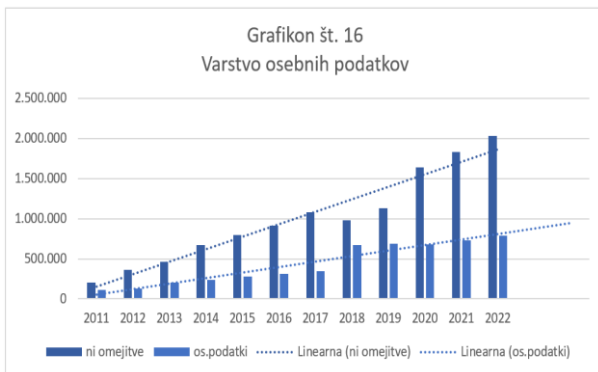
S stališča nadaljnje izgradnje sistema bo treba potrebno pozornost posvetiti zapisom s statusom »v delu«. Statistični podatki kažejo, da so količniki za zapise s statusom »v delu« sicer v negativnem trendu, vendar nominalno naraščajo (tabela št. 9). Razlogov za take rezultate je več in jih je treba iskati v posameznih lokalnih okoljih.

Ne glede na ugodne trende bo v tem segmentu še treba veliko postoriti, predvsem na področju organizacije zajemanja, redaktiranja in zaključevanja zapisov popisnih enot. Pri tem je dobro, da je spreminjanje statusa zapisov dobro strojno podprto, kar pomeni, da je mogoče zelo velikemu številu zapisov spremeniti status v kratkem času. Pri tem je treba poudariti, da zaključevanje ne vpliva le na status zapisa popisne enote, ampak neposredno tudi na nekatere povezane entitete, kot so priponke, deskriptorji itd., ki so skladno s spremembo statusa zapisov javno dostopni ali pa tudi ne.

Leto	"Zaključeno"	"V delu"
2011	↓ 0,385483	↑ 1,59380003
2012	↓ 0,353191	↑ 1,83132947
2013	↓ 0,460723	⇒ 1,1705022
2014	⇒ 0,591314	↓ 0,69114828
2015	↑ 0,713664	↓ 0,40044404
2016	↑ 0,745427	↓ 0,34151292
2017	↑ 0,730723	↓ 0,3685076
2018	↑ 0,751498	↓ 0,33067478
2019	↑ 0,743944	↓ 0,34418714
2020	↑ 0,742566	↓ 0,34668123
2021	↑ 0,731485	↓ 0,36708228

Tabela 9: Količnik med zapisi popisnih enot in njihovim statusom, "v delu" / "zaključeno"

## 2.9 O statusu »osebnih podatkov«



Zajeti zapisi popisnih enot imajo lahko različne omejitve glede dostopa do njih samih in s tem tudi do gradiva, na podlagi katerega so nastali. V nadaljevanju posvečamo posebno pozornost tistim zapisom, katerih arhivsko gradivo ima omejitve zaradi zakonodaje glede varstva osebnih podatkov (GDPR, 2016; ZVOP2, 2022), tajnih podatkov po Zakonu o tajnih podatkih (ZTP, 2023), poslovne skrivnosti (ZDavP-2, 2022), zahtev

zasebnega izročitelja in podobno.

Statistični podatki kažejo, da je v podatkovnem korpusu SJAS daleč največ omejitev zaradi varstva osebnih podatkov. Tako konec leta 2022 nima nobene tovrstne omejitve 2.029.394 zapisov, 788.547 zapisov pa ima status »osebni podatki« (grafikon št. 16). Le 4.613 zapisov ima eno od ostalih omejitev. Slednje so statistično komaj zaznavne in jih v raziskavi vodimo pod skupnim imenom »ostalo«. Podatki kažejo, da je nominalno število statusov zapisov v trendu naraščanja, hkrati pa se dolgoročno v ozadju sistema spreminja njihova dostopnost, ko preteče predvideni rok omejitve dostopa.

Pomemben vpliv na omejitve dostopa do opisov popisnih enot je imela uredba GDPR v letu 2018. Takratne uskladitve statusov zapisov so vplivale na to, da se je njihovo število s statusom »osebni podatki« povečalo za nekaj 100.000 in se posledično za toliko zmanjšalo število zapisov s statusom »ni omejitve« (tabela št. 10).

Znano je, da zapisi s statusom »osebni podatki« postanejo skladno z zakonom dostopni brez omejitev po 75 letih od nastanka dokumenta, iz katerega je uporabljen osebni podatek. Določanje oz. spreminjanje statusa zapisom glede njihovega posebnega režima varovanja je sistemsko zelo dobro urejeno in za arhivske strokovne delavce ne predstavlja večjega izziva, razen dosledne uporabe te funkcionalnosti.

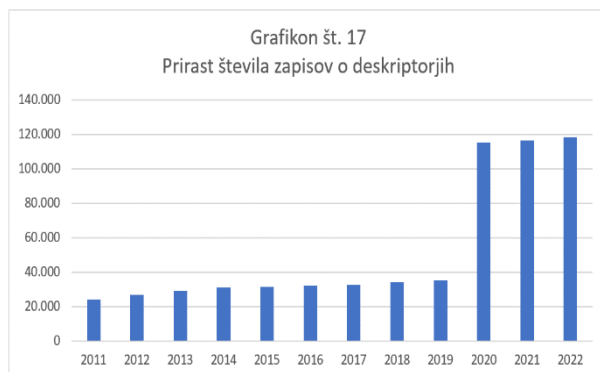
Leto	"Ni omejitve"	"Osebni podatki"	"Ostalo"
2011	⇒ 0,653686047	⇒ 0,529487704	↓ 0,000145562
2012	↑ 0,731451662	↓ 0,366551539	↓ 0,001617238
2013	⇒ 0,691500647	⇒ 0,440287203	↑ 0,013270977
2014	↑ 0,736923357	↓ 0,351502042	↑ 0,015622038
2015	↑ 0,738566229	↓ 0,348508248	↑ 0,013534201
2016	↑ 0,743245176	↓ 0,341295309	↑ 0,01217637
2017	↑ 0,75763258	↓ 0,31638534	↑ 0,011144087
2018	↓ 0,592869123	↑ 0,68285778	⇒ 0,005615758
2019	↓ 0,619012433	↑ 0,611484818	⇒ 0,006526301
2020	↑ 0,704081738	↓ 0,41750814	⇒ 0,006662152
2021	↑ 0,71373734	↓ 0,398587766	⇒ 0,00624172
2022	↑ 0,718992345	↓ 0,388562793	⇒ 0,00585

Tabela 10: Količnik razmerja med zapisi popisnih enot in njihovim statusom glede varovanja podatkov po letih

Analiza količnikov (tabela št. 10) kaže na stopnjo doslednosti implementacije statusov zapisov popisnih enot ter na količino posebej varovanih arhivskih vsebin, ki jih strokovno obravnavajo arhivski delavci skladno z letnimi programi dela. Kljub temu pa je spremljanje absolutnega števila zapisov, ki vsebujejo varovane podatke, ali njihovih količnikov pomembno tako zaradi arhivskih strokovnih kot tudi pravno reguliranih zahtev.

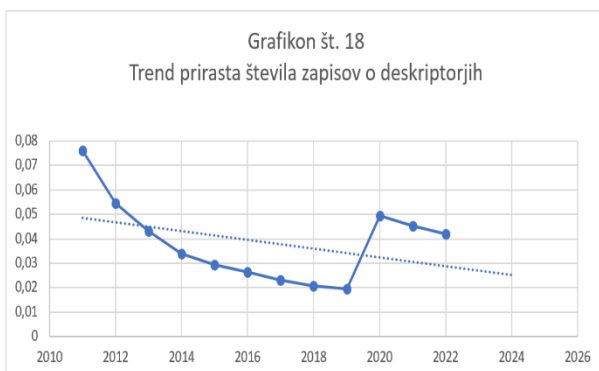


## 2.10 O dekriptorjih v podatkovni bazi SJAS



2020 se je nominalno število zapisov v modulu zelo povečalo (grafikon št. 17). S tem pa so se mnoge vrednosti dekriptorjev podvojile ali celo večkratno pomnožile. Dvojnike in multiplikate zapisov arhivski strokovni delavci združujejo s pomočjo posebnega vtičnika tako, da prenašajo vsebine opisov dekriptorjev in njihove povezave na izbrane unikatne zapise o dekriptorjih. Do konca leta 2022 so bili preverjeni in odstranjeni vsi duplikati oz. multiplikati zapisov o dekriptorjih za vse tezavre, razen za tezaver zemljepisnih imen. Ta je še vedno razdeljen v dva dela – prvi del obsega skoraj 14.000 zapisov entitet, drugi del pa skoraj 12.000 zapisov. Oba dela bo treba v perspektivi združiti v en tezaver zemljepisnih entitet.

Analiza zajema opisov dekriptorjev po letih kaže, da so bili na začetku izvedeni paketni uvozi predvsem zemljepisnih imen in pravnih oseb/korporacij. V nadaljevanju nato ni bilo vloženega pričakovanega intelektualnega vložka za intenzivno izgradnjo in uporabo tega kontekstnega okolja tako v okviru same izgradnje tezavrov kot tudi njihove predstavitve skozi Virtualno arhivsko čitalnico. Prav zato ne preseneča rahlo negativni trend prirasta števila zapisov o dekriptorjih (grafikon št. 18).



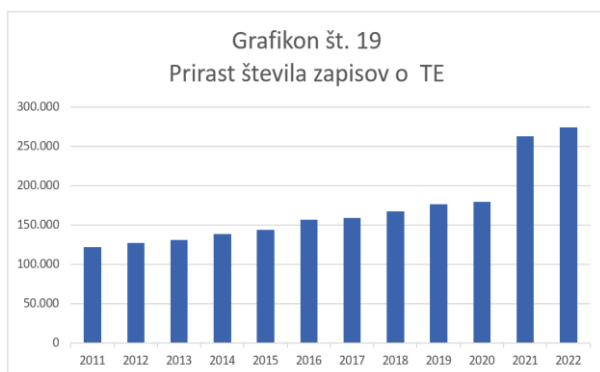
Moč množice zapisov o dekriptorjih je v upadanju glede na moč množice zapisov o popisnih enotah. Pri tem moramo ločiti zmanjševanje števila dekriptorjev zaradi združevanja zapisov o geografskih entitetah s pomočjo posebnega vtičnika od splošnega trenda upadanja količnika razmerja med zapisi o popisnih enotah in zapisih o dekriptorjih (tabela št. 11).

Ko bo zaključen postopek združevanja zemljepisnih imen, bo potrebna ponovna analiza, na podlagi katere bo mogoče ugotoviti, ali so pridobljene informacije v procesu urejanja in popisovanja ustrezno izražene tudi skozi izdelane dekriptorje. Z njihovo dosledno izdelavo se bo posledično zvišala stopnja razpoznavanja in razumevanja vsebin popisnih enot in s tem splošna uporabnost celotnega metapodatkovnega korpusa.

Leto	Količnik
2011	0,0760574
2012	0,0545569
2013	0,0432406
2014	0,0337899
2015	0,029259
2016	0,0263205
2017	0,0229765
2018	0,0206958
2019	0,0193632
2020	0,0495307
2021	0,0453455
2022	0,0419199

**Tabela 11: Količnik razmerja med zapisi popisnih enot in dekriptorji po letih**

## 2.11 O tehničnih enotah



Zapisi o tehničnih enotah (TE) so zajeti in sistemizirani v modulu Tehnične enote in predstavljajo logično povezavo z zapisi o lokacijah tehničnih enot v arhivskem skladišču, hkrati pa so povezani z zapisi o arhivskih vsebinah, ki so vložene v tehnične enote.

Število zapisov o tehničnih enotah se v praksi le počasi povečuje, še posebej potem, ko je bila po javnih arhivih sistemsko zamenjana tehnična oprema arhivskega gradiva v arhivskih skladiščih. Prirast števila tehničnih enot pa je v obravnavanem obdobju omejen predvsem na nove prevzeme.

Dinamika zajemanja zapisov o tehničnih enotah je neposredno odvisna od politike uporabe sistema scopeArchiv po posameznih arhivih.

Absolutno število zapisov o tehničnih enotah narašča. Leta 2011 je bilo v tem modulu paketno zajetih 122.391 zapisov. To število je v letu 2022 naraslo na 274.625 (grafikon št. 19). Rezultati so skladni z dinamiko novih prevzemov arhivskega gradiva in učinkom vzpostavitve vzajemnega podatkovnega korpusa SJAS. Če pa gledamo trend količnika med zapisi arhivskih popisnih enot in tehničnih enot, opazimo, da ta pada od leta 2011 dalje (tabela št 12). To pomeni, da je bilo intenzivno zajemanje zapisov tehničnih enot izvedeno takoj na začetku vzpostavitve vzajemne podatkovne zbirke regionalnih arhivov in nato leta 2021 po vzpostavitvi VMK SJAS.

Metoda vnaprejšnje oz. sprotne priprave zapisov o tehničnih enotah se je izkazala v času od leta 2011 dalje za pravilno, saj te zajemajo, kontrolirajo in posodablajo zaposleni, ki so zadolženi za arhivska skladišča, in so zato ti zapisi po svojem statusu tudi normativni.

Trend količnika relacij med številom zapisov popisnih enot in številom zapisov tehničnih enot dokazuje tudi, da arhivski strokovni delavci v okviru korpusa SJAS arhivskega gradiva ne popisujemo več le po tehničnih enotah, ampak po vsebinah (grafikon št. 20). Kot izziv v tem segmentu opredeljujemo dejstvo, da obstoječe število zapisov predstavlja le kakšno tretjino tehničnih enot, ki obstajajo v arhivskih skladiščih SJAS.

Leto	Količnik
2011	0,3853912
2012	0,2592513
2013	0,1940876
2014	0,1515072
2015	0,1329972
2016	0,1278976
2017	0,1115041
2018	0,1010859
2019	0,0967191
2020	0,0773852
2021	0,1025487

**Tabela 12: Količnik razmerja med zapisi popisnih enot in tehničnih enot po letih**

## 2.12 O lokacijah v skladiščih



Podobne trende kot pri zapisih o tehničnih enotah lahko zasledimo tudi pri zapisih o lokacijah. Te so bile v veliki meri paketno zajete do leta 2011, ko je njihovo število znašalo 35.093. Nato so bile izvedene v glavnem le dopolnitve oz. spremembe zapisov lokacij. Leta 2022 pa je število zapisov naraslo na 104.710 (grafikon št. 21).

Tudi v tem primeru je viden opazen prirast zapisov leta 2020, ko je bila izvedena vzpostavitev VMK SJAS. V tem letu se namreč poveča število zapisov s 45.564 na 95.932. Pri tem število zapisov o lokacijah hkrati predstavlja tudi število polic v arhivskih skladiščnih regalih.

Analiza količnika med zapisi popisnih enot in zapisi lokacij pa kaže na veliko večjo intenziteto zajemanja podatkov o arhivskih popisnih enotah v primerjavi z zapisi o lokacijah. Ta se giblje od 0,02 v letu 2018 in 2019 do 0,11 v letu 2011. Količniki kažejo, da področje skladiščnega poslovanja skozi VMK SJAS ni v fokusu arhivskih strokovnih delavcev. Razloge je iskati predvsem zmožnostih obvladovanja lokacij manjših skladiščnih prostorov tudi brez informacijske podpore.

S stališča upravljanja skladišč pa je teoretično zanimiv tudi odnos med zapisi o tehničnih enotah in lokacijah. S 104.710 zajetimi opredelitvami lokacij v arhivskih skladiščih je povezanih 274.625 opisov tehničnih enot.

Podatki iz VMK SJAS v tem segmentu še niso praktično uporabni v celoti, saj bi po njih sklepali, da imajo slovenski javni arhivi še zelo veliko prostih skladiščnih kapacitet, skoraj 2/3, kar pa ne ustreza dejanskemu stanju. Prav zato bi bilo treba izdelati zapise o lokacijah v skladiščih skladno z dejanskim stanjem, da bodo tudi ti podatki natančni in uporabni za nadaljnje upravljanje skladiščnih kapacitet v okviru SJAS.

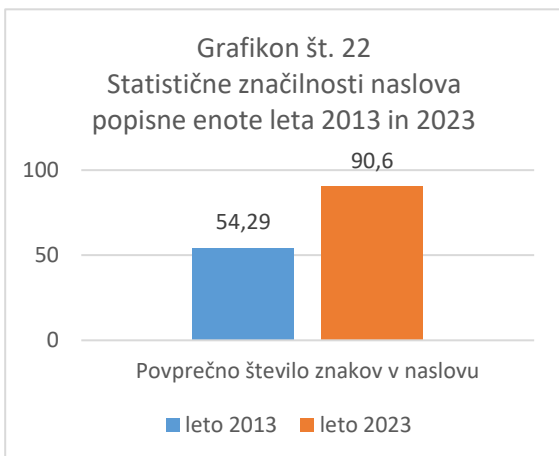
Leto	Količnik
2011	0,1105027
2012	0,0790363
2013	0,0682346
2014	0,0502446
2015	0,0430349
2016	0,0365732
2017	0,0316612
2018	0,02729
2019	0,0249573
2020	0,0412456
2021	0,040736
2022	0,0370976

**Tabela 13: Količnik razmerja med zapisi popisnih enot in lokacij po letih**

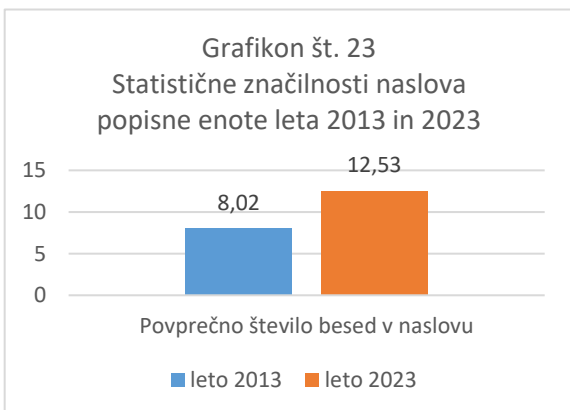
## 2.13 O oblikovanju naslovov popisne enote

S stališča vsestranske uporabe podatkov iz vzajemnega podatkovnega korpusa SJAS spada med najpomembnejše elemente Naslov PE. Ta je standardiziran s standardom ISADg2 in služi razpoznavanju vsebine PE. Na višjih popisnih nivojih je razpoznavanju arhivskih entitet ob elementu Naslov PE namenjenih še več elementov, npr. element Vsebina. Na nižjih popisnih nivojih pa je element Naslov PE pogosto edini vir za razpoznavanje vsebine popisne enote.

Za upravljanje vzajemnega podatkovnega korpusa SJAS lahko s kvantitativnega stališča pri tem elementu sledimo predvsem moči množice zapisov v različnih kontekstih. Na tej osnovi lahko ugotavljamo intenziteto zajemanja in upravljanja z vsebinami v tem elementu. Moč množice besed ali moč množice črk v naslovu popisne enote pa ob kvantitativnih rezultatih nakazujejo tudi na nekatere kvalitativne vidike oblikovanja vsebin tega elementa, ki jih bi bilo treba v nadaljevanju celovite analize VMK SJAS podrobneje raziskati.

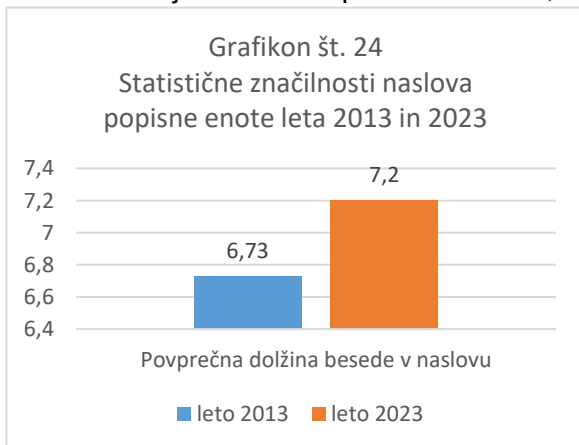


Prva kvantitativna raziskava s področja oblikovanja naslova je bila opravljena leta 2013. Pridobljeni rezultati služijo kot referenca za raziskavo iz leta 2023. V obeh navedenih raziskavah je bila uporabljena enaka raziskovalna metoda. Najprej je bilo izvedeno poizvedovanje za naslovi popisnih enot, ki so bili ustvarjeni v podatkovni zbirki v določenem obdobju. Vsi zadetki poizvedbe so bili preneseni v urejevalnik besedil, ki omogoča štetje besed, odstavkov in črk s presledki. Pridobljeni numerični podatki so bili nato obdelani s statističnimi metodami.



V letu 2013 je bilo obravnavanih skupaj 94.125 zapisov, ki so vsebovali 6.626.536 besed. Te so bile sestavljene iz 44.329.234 znakov s presledki. Leta 2023 je bil vzorec nekoliko manjši, saj je raziskovalni korpus sestavljalo 67.060 zapisov, ki so vsebovali 840.037 besed. Te so imele skupaj 6.069.120 znakov s presledki. Leta 2013 so bili v raziskavo zajeti vsi zapisi v bazi, leta 2023 pa le tisti, ki so nastali od 1. 1. 2023 do 31. 3. 2023.

Leta 2013 je bilo povprečno število uporabljenih znakov v naslovu zapisa popisne enote 52,29. To število se je leta 2023 povečalo na 90,50 (grafikon št. 22). Leta 2013 je bilo v enem naslovu zapisa popisne enote uporabljene povprečno 8,02 besede. Ta vrednost se je leta 2023 povečala na 12,53 (grafikon št. 23). Leta 2013 je povprečna beseda v naslovu popisne enote vsebovala 6,73 znaka. Leta 2023 pa se je ta vrednost povečala na 7,2 (grafikon št. 24).



Na podlagi pridobljenih rezultatov lahko ugotovimo, da so se značilnosti oblikovanja naslovov popisnih enot spremenile v smeri povečanja vseh treh dejavnikov – števila uporabljenih znakov s presledki za oblikovanje vsebine naslova, števila besed, s katerimi arhivski strokovni delavci opredeljujejo vsebino naslova popisne enote, kakor tudi dolžina

poprečne uporabljene besede v naslovu popisne enote. Pridobljeni kvantitativni podatki sicer ne kažejo na dejansko kvaliteto zajetih vsebin v naslovih popisnih enot, ampak nakazujejo na povečanje podatkovnih potencialov, ki so zajeti v vzajemnem metapodatkovnem korpusu SJAS. To pa je dobro izhodišče za uvajanje naprednejših oblik poizvedovanj – ob predvidevanju, da bodo v nadaljevanju izvedene še zelo natančne kvalitativne raziskave v tem segmentu.

### 3 Razprava

Periodične analize vzajemnega metapodatkovnega korpusa SJAS, pa četudi le kvantitativne, so potrebne iz več razlogov. Prvi se nanaša na razumevanje in kvantifikacijo dinamike povečevanja oz. nazadovanja posameznih entitet znotraj korpusa kot celote in s tem tudi njegovih sestavnih delov. Drugi pomemben razlog pa je, da je mogoče na tej osnovi izdelati njegovo grobo splošno oceno stabilnosti in robustnosti.

Na podlagi zgoraj izvedenih analiz in pridobljenih rezultatov so bile zaznane naslednje prednosti arhivskega metapodatkovnega korpusa SJAS:

- Na podlagi konstantne rasti števila zapisov lahko sklepamo, da je v okviru sistema dosežena relativno visoka stopnja strokovnega soglasja glede oblikovanja zapisov, njihovega upravljanja in uporabe.
- Na podlagi gibanja števila uporabnikov sistema v relaciji do števila zajetih zapisov, npr. popisnih enot in števila deskriptorjev, in med njimi izdelanih relacij lahko sklepamo, da je vzpostavljen dober sistem vzajemnosti pri zajemanju podatkov tako na nivoju posameznih arhivskih ustanov kot tudi na nivoju celotnega sistema.
- Sistem v njegovi osnovi gradi relativno malo število arhivskih ustanov (7), zato je stopnja doseganja dogovorov relativno visoka, kar se odraža v visoki fleksibilnosti celotnega sistema.
- Vsakdanja uporaba sistema kaže, da ta s stališča končnih uporabnikov temelji na solidni infrastrukturi, komunikacijska infrastruktura obeh ponudnikov, HKOM-a in Arnesa, pa je zelo stabilna in zagotavlja primerne hitrosti delovanja.
- Isti podatki kažejo tudi na to, da celoten sistem, ki temelji na podatkovni bazi ORACLE in Microsoftovem sistemskem okolju, zagotavlja robustnost podatkovnega okolja in relativno dobro podporo končnemu uporabniku.
- Pomembna značilnost podatkovnega korpusa SJAS je, da število zapisov konstantno raste – s povprečno vrednostjo več kot 200.000 zapisov letno.
- Uporabniški vmesnik sistema je v slovenskem jeziku, kar je pomembno za končnega aktivnega uporabnika.
- Število povezav med zapisi popisnih enot in z zapisi o deskriptorjih, tehničnih enotah, datotekah itd. narašča praviloma premo sorazmerno, kar pomeni relativno dobro kontekstno povezane arhivske vsebine.

V procesih izgradnje podatkovnega korpusa SJAS pa na podlagi zgornje analize zaznavamo tudi naslednje slabosti:

- Nedosledno uveljavljen sistem vsebinske validacije zapisov, kar lahko zaznamo pri kvantitativni analizi naslova popisnih enot.
- Na podlagi istih podatkov lahko sklepamo tudi na nedoslednosti pri oblikovanju zapisov predvsem iz zgodnjih obdobj vzpostavitve vzajemnega korpusa SJAS.

- V nekaterih segmentih, npr. pri količniku letnega prirasta zapisov popisnih enot ali pri trendu prirasta zapisov o deskriptorjih, zaznavamo v zadnjih letih negativne trende rasti.
- Kljub relativno visokemu številu povezav med zapisi obstaja v sistemu še veliko takih, ki lebdijo v bazi in niso povezani z drugimi zapisi.
- Na nivoju podatkovnega korpusa SJAS je zaznana približno četrtnina zapisov popisnih enot, ki so v statusu »v delu«. Mnogi so ostali v tem statusu, čeprav je bilo arhivsko strokovno delo na pripadajočem arhivskem gradivu že zaključeno.
- Podatki kažejo, da je na letnem nivoju zaznati neuravnoteženo intenziteto zajemanja podatkov, npr. pri povezavah med zapisi popisnih enot in referenčnimi zapisi, podobne trende sledimo tudi v nekaterih modulih, npr. v modulu Prezemi.
- Pomanjkanje sistemskega odpravljanja pravopisnih in drugih napak, do katerih prihaja pri zajemanju podatkov.

Na podlagi zgoraj predstavljenih statističnih rezultatov in možnosti, ki jih ponuja sodobna informacijska tehnologija, so bile zaznane naslednje priložnosti arhivskega metapodatkovnega korpusa SJAS:

- Vzpostaviti celovito podatkovno zbirko vsega fizičnega arhivskega gradiva, ki se hrani v slovenskih javnih arhivih, in pri tem uporabiti najsodobnejše tehnološke rešitve, vključno s sistemi za množične pretvorbe v digitalne ekvivalente ter generativno razpoznavanje arhivskih vsebin.
- Na podlagi intenzivnega arhivskega strokovnega sodelovanja ustvarjalcev in pristojnih arhivov izgraditi sistem prevzemanja tako, da bodo pri ustvarjalcih oblikovane metapodatkovne vsebine popolnoma zadoščale potrebam po kvalitetnih arhivskih opisih prevzetega gradiva v kontekstih.
- Obstoječa zasnova in sama izgradnja arhivskega metapodatkovnega korpusa do določene stopnje že omogoča izvedbo popisovanja arhivskega gradiva v kontekstu. Na osnovi pridobljenih izkušenj bi bilo mogoče v prihodnosti implementirati celoten multidimenzionalni model popisovanja arhivskega gradiva – še posebej tistega v elektronski obliki.
- Dosežena stopnja standardizacije na nivoju celotnega metapodatkovnega korpusa SJAS predstavlja dobre osnove za neposredno implementacijo novih tehnoloških rešitev, kot so generativna umetna inteligenca (UI), strojno učenje, strojno razpoznavanje vsebin itd.

Zgoraj pridobljeni podatki pa kažejo možnosti pojavljanja različnih groženj celovitosti in neposredni uporabnosti arhivskega metapodatkovnega korpusa SJAS:

- Nekateri kazalci kažejo, da vzajemni metapodatkovni korpus SJAS ni imun na dejavnike, ki lahko vzpodbudijo ali povzročajo entropijo sistema in s tem izgubo zaupanja arhivske stroke in javnosti v sistem. V tem kontekstu gre predvsem za trende upadanja števila izdelanih povezav popisnih enot z deskriptorji ali za veliko število zapisov, ki so v statusu »v delu«.
- Z večanjem števila zapisov v metapodatkovnem korpusu SJAS in z večanjem fluktuacije aktivnih uporabnikov na sistemu se zaradi različnih dejavnikov lahko zmanjša stopnja zajemanja kontekstnih podatkov, kar lahko vodi v nižjo stopnjo razpoznavanja arhivskih vsebin in s tem pristranost in haluciniranje sistemov, ki temeljijo na UI.

- Relativno slabo izvedeno sledenje spremembam v zapisih in poznejše sledenje dostopom do arhivskih vsebin lahko omogoči nekontrolirano odtujevanje zaščitenih podatkov.
- Relativno veliko grožnjo celotnemu sistemu zaznavamo v uporabi metod popisovanja, ki niso skladne z večnivojskim pristopom in mednarodnimi standardi za popisovanje arhivskega gradiva.

#### 4 Zaključek

Na podlagi pridobljenih statističnih podatkov lahko ugotovimo, da število zapisov in medsebojnih povezav v celotnem arhivskem metapodatkovnem korpusu SJAS naraste v povprečju za pol milijona entitet na leto.

Pričakovan prirast v tem segmentu pa bo ob upoštevanju tehnoloških in arhivskih organizacijskih trendov v prihodnje še veliko večji. Hkrati s tem je pričakovana tudi višja stopnja kompleksnosti arhivskih metapodatkovnih struktur, ki bodo opredeljevale predvsem elektronske oblike arhivskega gradiva.

Za natančno obvladovanje obstoječih in novih metapodatkovnih struktur bodo potrebne nove metode arhivskega strokovnega dela, pa tudi nove rešitve informacijske podpore arhivskim strokovnim postopkom. Vse to bo treba razviti in validirati, vendar implementacije ne bo mogoče izvesti čez noč. Arhivski strokovni delavci morajo na tem področju delovati sistematično in proaktivno z jasno izdelanimi arhivskimi strokovnimi cilji. V nasprotnem primeru lahko pričakujemo entropijo celotnega metapodatkovnega korpusa.

Do sedaj znani statistični podatki o značilnostih slovenskega metapodatkovnega korpusa so pomembni predvsem z arhivskega strokovnega in managerskega stališča. Na tej osnovi lahko arhivski strokovni delavci začenjajo podrobnejše raziskave posameznih problemov. Rezultati raziskav pa morajo voditi do potrebnih optimizacij v delovnih postopkih zajemanja metapodatkov, v sami podatkovni bazi ali na ravni uporabniškega vmesnika. Za pravilno nadaljnje spremljanje gibanja rasti celotnega korpusa in za ustrezno mednarodno primerljivo interpretacijo zbranih podatkov o entitetah arhivske strokovne obdelave v okviru SJAS pa bo treba dosledneje implementirati standard ISO 24083:2021.

Kljub dobrim rezultatom pa je treba opozoriti še na nekatere odprte probleme. Izpostavimo naj samo t. i. »plavajoče« zapise, ki niso kontekstno povezani z drugimi zapisi. K temu je treba prišteti še arhivske strokovne aktivnosti pri izgradnji vzajemnega korpusa z negativnimi trendi ali rešitvami. Te ustvarjajo dvome v pravilnost tako na nivoju vsebinskih kot tudi kontekstnih metapodatkov itd. Vsekakor pa ti pojavi generirajo splošno informacijsko pristranost in omejujejo možnost njihove poznejše uporabe ter implementacije sodobnih informacijskih rešitev, npr. s področja generativne umetne inteligence.

Pridobljeni statistični podatki, predvsem tisti, ki se nanašajo na oblikovanje naslovov popisnih enot, generirajo pomembna nova arhivska metodološka vprašanja, ki so povezana predvsem z definicijo in razumevanjem kvalitete zajetih podatkov. V povezavi s tem se pojavljajo še druga vprašanja. Izpostavimo naj samo dve izmed njih. Kako zagotoviti nepristranost odgovorov na podlagi vsebin, ki sicer obstajajo, a imajo zajeto nezadovoljivo strojno berljivo količino podatkov o arhivski entiteti? Kako v metodološkem smislu opredeliti stopnjo halucinacije in/ali pristranosti različnih sistemov zaradi napak ali pomanjkljivo izdelanih opisov arhivskih entitet v vzajemnem metapodatkovnem korpusu SJAS?

## 5 Viri in literatura

- Cvelfar, B. (2010).** Vzajemno delovanje slovenskih arhivov v projektu Sira\_net. *Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja : zbornik referatov dopolnilnega izobraževanja s področij arhivistike, dokumentalistike in informatike*, št. 9, str. 395–412. Maribor : Pokrajinski arhiv Maribor. Sneto 20. 11. 2023 s spletne strani <http://www.pokarh-mb.si/uploaded/datoteke/039cvelfar2010.pdf>.
- Digitalna Slovenija 2030 - Krovna strategija digitalne preobrazbe Slovenije do leta 2030. (2023).** Ministrstvo za digitalno preobrazbo, Ljubljana. Sneto 31. 10. 2023 s spletne strani [https://www.gov.si/assets/ministrstva/MDP/Dokumenti/DSI2030-potrijena-na-Vladi-RS\\_marec-2023.pdf](https://www.gov.si/assets/ministrstva/MDP/Dokumenti/DSI2030-potrijena-na-Vladi-RS_marec-2023.pdf).
- GDPR. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).** Pridobljeno 27. 7. 2023 s spletne strani: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.
- ISADg2. General International Standard Archival Description - Second edition. (2000).** Pridobljeno 20. 7. 2023 s spletne strani: <http://www.ica.org/10207/standards/isadg-general-international-standard-archival-description-second-edition.html>.
- ISO 24083:2021 Information and documentation — International archives statistics.** Pridobljeno 22. 7. 2023 s spletne strani: <https://www.iso.org/standard/77765.html>.
- Koncilija, Ž., Jenuš, G., Hajtnik, T. (2021).** Virtualna arhivska čitalnica (VAČ) in izzivi digitalizacije arhivskih čitalniških postopkov. *Moderna arhivistika, IV, 2021, št. 2*, str. 129–148. Maribor : Pokrajinski arhiv Maribor.  
DOI: <https://doi.org/10.54356/MA/2021/AXJY3787>
- Maček, J. (2009).** Pregled dela v slovenskih državnih regionalnih arhivih. *Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja : zbornik referatov dopolnilnega izobraževanja s področij arhivistike, dokumentalistike in informatike*, št. 8, str. 39–54. Maribor : Pokrajinski arhiv. Pridobljeno 11. 7. 2023 s spletne strani: [http://www.pokarh-mb.si/uploaded/datoteke/Radenci/radenci2009/04\\_macek\\_2009.pdf](http://www.pokarh-mb.si/uploaded/datoteke/Radenci/radenci2009/04_macek_2009.pdf).
- Novak, M. (2021).** O validaciji metod arhivskega strokovnega in znanstvenega dela. *Arhivi v službi človeka - človek v službi arhivov: digitalna transformacija v arhivistiki 9. znanstvena konferenca z mednarodno udeležbo Za človeka gre: digitalna transformacija v znanosti, izobraževanju in umetnosti*, str. 31–45. Maribor : AMEU - ECM, Alma Mater Press.
- Novak, M. (2022).** Statistika i statističke metode u savremenoj arhivskoj teoriji i praksi. *Arhivski pogledi*, št. 3. str. 163–180. Tuzla : Društvo arhivskih zaposlenika Tuzlanskog kantona. Pridobljeno 11. 11. 2023 s spletne strani: <http://www.daztk.net/arhivskipogledi3.pdf>.
- Novak, M., Drobnjak, V. (2002).** Zasnova in strateške usmeritve slovenskega vzajemnega arhivskega informacijskega sistema. *Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja : zbornik referatov dopolnilnega izobraževanja s področij arhivistike, dokumentalistike in informatike*, str. 322–333. Maribor : Pokrajinski arhiv Maribor.
- Novak, M., Semlič Rajh, Z. (2013).** Merjenje rezultatov uspešnosti izgradnje arhivskih podatkovnih zbirk s kvantitativno-kvalitativno metodo na primeru podatkovne zbirke SIRAnet. V: *Standardizacija (p)opisov arhivskega gradiva in uskladitev strokovnih praks v slovenskih javnih in cerkvenih arhivih : zbornik referatov*, str. 19–39. Ljubljana : Arhivsko društvo Slovenije.
- Poročilo o zajetih podatkih v sistemu SIRAnet za obdobje 1. 1. 2012 do 31. 12. 2012. Interno gradivo. Pokrajinski arhiv Maribor.
- Poročilo o zajetih podatkih v sistemu SIRAnet za obdobje 1. 1. 2013 do 31. 12. 2013. Interno gradivo. Pokrajinski arhiv Maribor.



Poročilo o zajetih podatkih v sistemu SIRAnet za obdobje 1. 1. 2014 do 31. 12. 2014. Interno gradivo. Pokrajinski arhiv Maribor.

Poročilo o zajetih podatkih v sistemu SIRAnet za obdobje 1. 1. 2015 do 31. 12. 2015. Interno gradivo. Pokrajinski arhiv Maribor.

Poročilo o zajetih podatkih v sistemu SIRAnet za obdobje 1. 1. 2016 do 31. 12. 2016. Interno gradivo. Pokrajinski arhiv Maribor.

Poročilo o zajetih podatkih v sistemu SIRAnet za obdobje 1. 1. 2017 do 31. 12. 2017. Interno gradivo. Pokrajinski arhiv Maribor.

Poročilo o zajetih podatkih v sistemu SIRAnet za obdobje 1. 1. 2018 do 31. 12. 2018. Interno gradivo. Pokrajinski arhiv Maribor.

Poročilo o zajetih podatkih v sistemu SIRAnet za obdobje 1. 1. 2019 do 31. 12. 2019. Interno gradivo. Pokrajinski arhiv Maribor.

Poročilo o zajetih podatkih v sistemu SIRAnet/SJAS za obdobje 1. 1. 2020 do 31. 12. 2020. Interno gradivo. Pokrajinski arhiv Maribor.

Poročilo o zajetih podatkih v sistemu SJAS za obdobje 1. 1. 2021 do 31. 12. 2021. Interno gradivo. Pokrajinski arhiv Maribor.

Poročilo o zajetih podatkih v sistemu SJAS za obdobje 1. 1. 2022 do 31. 12. 2022. Interno gradivo. Pokrajinski arhiv Maribor.

**Product overview – scope archive software.** Pridobljeno 10. 10. 2023 s spletne strani: <https://www.scope.ch/en/product-overview/>.

**RiC. (2021).** Records in Contexts, Conceptual Model, Consultation Draft v0.2. Pridobljeno 10. 3. 2023 s spletne strani: [https://www.ica.org/sites/default/files/ric-cm-02\\_july2021\\_0.pdf](https://www.ica.org/sites/default/files/ric-cm-02_july2021_0.pdf).

**Scope. Product overview – scope archive software.** Pridobljeno 20. 11. 2023 s spletne strani: <https://www.scope.ch/en/product-overview/>.

**Semlič Rajh, Z. (2016).** Kdor zna iskati, ta najde : uporabniki in analiza uporabe vzajemne arhivske podatkovne zbirke SIRAnet. V: *Gostenčnik, N. (ur.) Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja. Popisovanje arhivskega gradiva [Elektronski vir] : zbornik mednarodne konference, str. 91–112.* Maribor : Pokrajinski arhiv. Sneto 11. 11. 2023 s spletne strani: [http://www.pokarh-mb.si/uploaded/datoteke/Radenci/radenci\\_2016/091-112\\_semlic\\_2016.pdf](http://www.pokarh-mb.si/uploaded/datoteke/Radenci/radenci_2016/091-112_semlic_2016.pdf).

**ZDavP-2. Zakon o davčnem postopku** (Uradni list RS, št. 13/11 – uradno prečiščeno besedilo, 32/12, 94/12, 101/13 – ZDavNepr, 111/13, 22/14 – odl. US, 25/14 – ZFU, 40/14 – ZIN-B, 90/14, 91/15, 63/16, 69/17, 13/18 – ZJF-H, 36/19, 66/19, 145/20 – odl. US, 203/20 – ZIUPOPDVE, 39/22 – ZFU-A, 52/22 – odl. US, 87/22 – odl. US in 163/22). Pridobljeno 27. 10. 2023 s spletne strani: <http://pisrs.si/Pis.web/pregledPredpisa?id=ZAKO4703#>.

**ZTP. Zakon o tajnih podatkih** (Uradni list RS, št. 50/06 – uradno prečiščeno besedilo, 9/10, 60/11, 8/20 in 18/23 – ZDU-10). Pridobljeno 27. 10. 2023 s spletne strani: <http://www.pisrs.si/Pis.web/pregledPredpisa?id=ZAKO2133>.

**ZVOP2. Zakon o varstvu osebnih podatkov** (Uradni list RS, št. 163/22). Pridobljeno 27. 10. 2023 s spletne strani: <http://www.pisrs.si/Pis.web/pregledPredpisa?id=ZAKO7959>.

## SUMMARY

### STATUS AND PERSPECTIVES OF THE MUTUAL METADATA CORPUS OF THE SLOVENIAN PUBLIC ARCHIVAL SERVICE

**Miroslav NOVAK, Ph. D.**

Archival councillor, Regional Archives Maribor, Slovenia

[miroslav.novak@pokarh-mb.si](mailto:miroslav.novak@pokarh-mb.si)

Based on the obtained statistical data, we can conclude that the number of records and interconnections in the entire archival metadata corpus of the Slovenian Public Archival Service (SJAS) grows by half a million entities per year on average.

Considering technological and archival organizational trends, the expected increase in this segment will be much greater in the future. At the same time, a higher level of complexity of archival metadata structures is also expected, which will primarily define electronic forms of archival material.

New methods of archival professional work, as well as new information support solutions for archival professional processes, will be needed for the accurate management of existing and new metadata structures. All of these will need to be developed and validated, but their implementation will not be done overnight. Archival professionals must work systematically and proactively in this area with clearly defined archival professional goals. Otherwise, I can expect the entropy of the entire metadata corpus.

The statistical data known so far about the characteristics of the Slovenian metadata corpus are important above all from an archival professional and managerial point of view. On this basis, archival professionals can start more detailed research into individual problems. The results of the research must lead to the necessary optimizations in the work procedures of metadata capture, in the database itself or at the level of the user interface. For proper further monitoring of the growth trend of the entire corpus and an adequate internationally comparable interpretation of the collected data on the entities of archival professional processing within the framework of SJAS, it will be necessary to implement ISO standard 24083:2021 more consistently.

Despite the good overall results in the development of the SJAS mutual metadata corpus, some open problems should be pointed out. Let us highlight only examples of so-called "floating" records that are not contextually connected to other records. These examples should be added to those archival professional activities in building a mutual corpus, which is defined by negative trends or solutions. All these phenomena create doubts about the correctness of content, context, and other metadata. In any case, these phenomena generate a general information bias and limit the possibility of their later use and the implementation of modern information solutions, e.g., from the field of generative artificial intelligence.

The obtained statistical data, especially those relating to the title of the units of descriptions, generate important new archival methodological questions, which are mainly related to the definition and understanding of the quality of the data covered. In connection with this, other questions arise. Let's highlight just two of them. How to ensure the impartiality of answers with generative artificial intelligence based on contents that do exist, but have an unsatisfactory amount of machine-accessible data about the

archival entity? How to methodologically define the degree of hallucination and/or bias of different systems due to errors or poorly developed descriptions of archival entities in archival metadata corpus?