

DIGITALIZACIJA I IZRADA METAPODATAKA U DRŽAVNOM ARHIVU U ZAGREBU: NACIONALNI AGREGATOR I EUROPEANA

Apstrakt: Uključivanjem u projekt *Enriching Europeana through citizen science and artificial intelligence - unlocking the 19th century* (kraće *Enrich Europeana* +) Državni arhiv u Zagrebu obvezao se dostaviti određeno gradivo na portal Europeana. Kako je to prvi projekt arhiva s Europeanom, nije postojao usklađen proces digitalizacije i izrade metapodataka. Prije početka izrade metapodataka konzultirali smo Hrvatski državni arhiv i *Europeana Publishing Framework*. Prema odrednicama projekta *Enrich Europeana* + za dostavljanje gradiva na Europeanu potrebno je koristiti nacionalne agregatore. U Hrvatskoj je nacionalni agregator razvijen u sklopu projekta *eKultura* i povezuje nekoliko različitih elemenata. U ovom radu prikazat će se razine metapodataka koje Europeana zahtjeva kod agregiranja novih sadržaja, prilagodba Državnog arhiva u Zagrebu tim zahtjevima, te proces agregiranja gradiva Državnog arhiva u Zagrebu na nacionalni agregator i posljedično na portal Europeanu.

Cljučne riječi: Agregator, Europeana, digitalizacija, metapodaci, arhiv

Napredak tehnologije i razvijanje novih usluga zatekao je neke baštinske ustanove donekle nespremne. Takav je slučaj bio i kod Državnog arhiva u Zagrebu. Globalna pandemija koja je onemogućila dotadašnji rad arhiva sa strankama, te potresi koji su oštetili spremišta i radne prostore, nagnali su Arhiv da krene s masovnijom digitalizacijom gradiva. Dodatan poticaj sustavnoj i promišljenoj digitalizaciji bilo je uključivanje Državnog arhiva u Zagrebu projektu *Enriching Europeana through citizen science and artificial intelligence - unlocking the 19th century*. U sklopu projekta Arhiv se obvezao na dostavljanje kvalitetnog digitaliziranog gradiva popraćenog s isto tako kvalitetnim i bogatim metapodacima. Do tog trena u njemu nije postojao standardizirani pristup digitalizaciji fondova, zbirki i dijelova gradiva kao ni način izrade metapodataka. Kako bi se izvršila digitalizacija nakon koje više ne bi bilo potrebno koristiti izvorno gradivo, krenulo se u određivanje najboljih postupaka i praksi za cjelovit proces digitalizacije.

U ovom radu opisat ćemo projekt zbog kojeg je Državni arhiv u Zagrebu krenuo u razvoj internog procesa digitalizacije, na koji način se pristupilo izradi metapodataka koji su bili glavni izvori za razvoj procesa digitalizacije i izrade metapodataka. Također, prikazat će se glavne funkcije aplikacije *Goobi* kao servisa za podršku tijekom digitalizacije, te proces slanja gradiva na nacionalni agregator koji je razvijen u sklopu projekta *eKultura*.

¹ arhivist, sonja.galina@daz.hr

PROJEKT ENRICH EUROPEANA +

Projekt *Enriching Europeana through citizen science and artificial intelligence – unlocking the 19th century* ili kraće *Enrich Europeana +* započeo je 1. travnja 2021. godine. U projektu sudjeluje ukupno devet partnera iz nekoliko europskih zemalja:

1. AIT Austrian Institute of Technology GmbH iz Austrije;
2. Stichting Europeana iz Nizozemske;
3. Facts & Files Historisches Forschungsinstitut Berlin Drauschke – Schreiber Partnerschaftsgesellschaft iz Njemačke;
4. Instytut Chemii Bioorganicznej Polskiej Akademii Nauk iz Poljske;
5. READ-COOP SCE iz Austrije;
6. The Provost, Fellows, Foundation Scholars and the other members of Board, of the College of the Holy and Undivided Trinity of Queen Elizabeth near Dublin iz Irske;
7. Dublin City Council iz Irske;
8. Uniwersytet Wrocławski iz Poljske;
9. Državni arhiv u Zagrebu.

Projekt je procijenjen na iznos od 1.228.755 € od čega CEF-Telecom financiranje osigurava 75% (921.566 €), a projektni partneri preostalih 25% (307.189 €). Završetak projekta bio je predviđen za 30. rujna 2022. godine², ali zbog raznih poteškoća, primarno tehničke prirode, zatraženo je produljenje trajanja projekta od šest mjeseci bez osiguravanja dodatnih financijskih sredstava. U vrijeme pisanja ovog rada odluka još nije bila konačna.

Projekt *Enrich Europeana +* nastavlja se na raniji projekt *Enrich Europeana* u kojem je sudjelovalo sedam partnera, od kojih neki nastavljaju suradnju i na *Enrich Europeana +* projektu: *AIT Austrian Institute of Technology* (Austrija), *Facts & Files* (Njemačka), *Europeana Foundation* (Nizozemska), *NET7 S. r. l.* (Italija), *Biblioteca Judeteana „Octavian Goga“ Cluj* (Rumunjska), *Österreichische Nationalbibliothek* (Austrija), te *Instytut Chemii Bioorganicznej Polskiej Akademii Nauk* (Poljska). U sklopu projekta *Enrich Europeana* razvijena je crowdsourcing platforma pomoću koje su korisnici mogli transkribirati i anotirati gradivo koje se nalazi objavljeno na *Europeani*. Korisnici su sada mogli transkribirati i anotirati veći broj kolekcija unutar *Europeane*. U svrhu promocije i dobivanja većeg broja kvalitetnih transkripcija i anotacija održano je šest događaja vezanih uz transkripciju u Austriji, Rumunjskoj i Poljskoj. Od šest događaja jedan je održan online. Na događajima je sakupljeno preko 19.000 anotacija, od čega je 13.139 dostupno na mrežnim stranicama *Europeane* nakon validacije.³

Glavni zadatak projekta *Enrich Europeana +* je nastavak razvoja *Transcribathon* platforme, crowdsourcing platforme za čitanje, transkribiranje, označavanje i dodavanje

² Grant Agreement under the Connecting Europe Facility (CEF) – Telecommunications Sector : Agreement No INEA/CEF/ICT/A2020/2288325, 1-4.

³ *Enrich Europeana*: <https://pro.europeana.eu/project/enrich-europeana> (datum posjete 25. 8. 2022)

geografskih referenci objavljenom gradivu. Cilj projekta je omogućavanje dostupnosti baštinskim ustanovama *Transcribathon* platforme kako bi uz pomoć nje i njezinih korisnika poboljšali i povećali semantički i višejezični opis digitaliziranog gradiva svojih ustanova. U sklopu ovog projekta na *Transcribathon* platformu će biti učitano većinom rukopisno gradivo 19. stoljeća koje osigurava nekoliko partnera projekta.

Projekt ima tri glavna elementa:

1. agregacija više od 100.000 novih dokumenata u *Europeanu* korištenjem postojećih nacionalnih agregatora (npr. Digital Repository of Ireland iz Dublina, Uniwersytet Wrocławski iz Poljske);
2. razvoj inovativnih AI (*artificial intelligence*) rješenja za automatizaciju transkripcije rukopisnih tekstova i semantičko obogaćivanje istih;
3. organiziranje kampanji građanske znanosti nabave iz mnoštva (*citizen science crowdsourcing*) radi poticanja korištenja *Transcribathon* platforme.

147

Projekt je podijeljen na pet aktivnosti:

1. Agregacija sadržaja i infrastruktura za razmjenu sadržaja (*Content aggregation & data exchange infrastructure*);
2. Automatizirani servis za transkripciju i obogaćivanje (*Automated transcription and enrichment services*);
3. Inovacija i skalabilnost *Transcribathon* platforme (*Innovation and scalability of Transcribathon Platform*);
4. Diseminacija i kampanje građanske znanosti nabave iz mnoštva (*Dissemination and citizen science crowdsourcing campaigns*);
5. Menadžment i planiranje održivosti projekta (*Project management and sustainability planning*).

Nabrojane aktivnosti imaju vremenski okvir za izvršavanje, no zbog mogućeg produljenja trajanja projekta i one će imati nove, prilagođene datume završetka. Svaka aktivnost obuhvaća niz zadataka koji se moraju provoditi i izvršiti. Tako se prva aktivnost sastoji od agregacije sadržaja na *Europeanu*, poboljšanja kvalitete sadržaja, standardizacije razmjene sadržaja, te validacije metapodataka kod nacionalnih agregatora. Druga aktivnost obuhvaća razvoj HRT (*Handwritten Text Recognition*) softvera, automatske transkripcije i prijevoda metapodataka, prepoznavanja tema i semantičko obogaćivanje. Treća aktivnost odnosi se na razvoj i integraciju HTR dodatka za mrežni pregled (*HTR web plugin*), standardizaciju API-a *Transcribathon* platforme i poboljšanje korisničkog sučelja *Transcribathon* platforme i ugrađivanja *Transcribathon* platforme unutra *Europeana CSP*. Četvrta aktivnost radi na promoviranju *Transcribathon* platforme kod javnosti i baštinskih ustanova organiziranjem *Transcribathon* događaja, webinarima i radionica. Diseminacijom i raznim uredničkim aktivnostima, napose u vidu pisanja blogova i izradi zajedničkih galerija na *Europeani*, želi se stvoriti što veći krug novih korisnika. Peta aktivnost odnosi se na koordiniranje i nadgledanje svih elemenata projekta, upravljanje financijama, sudjelovanja na PMB (*Project Management Board*) sastancima, te izradu modela za dugoročno održavanje *Transcribathon* platforme.⁴

⁴ Grant Agreement under the Connecting Europe Facility (CEF) – Telecommunications Sector :

ULOGE PARTNERA U PROJEKTU

Svi partneri projekta sudjeluju u navedenim aktivnostima u određenom opsegu. Tako u prvoj aktivnosti Agregacija sadržaja i infrastruktura za razmjenu sadržaja sudjeluju *Uniwersytet Wroclawski, Dublin City Council, Državni arhiv u Zagrebu, Stichting Europeana, The Provost, Fellows, Foundation Scholars and the other members of Board, College of the Holy and Undivided Trinity of Queen Elizabeth near Dublin, te Instytut Chemii Bioorganicznej Polskiej Akademii Nauk*. U aktivnosti Automatizirani servis za transkripciju i obogaćivanje sudjeluju *Facts & Files Historisches Forschungsinstitut Berlin Drauschke – Schreiber Partnerschaftsgesellschaft, Stichting Europeana, AIT Austrian Institute of Technology GmbH i READ-COOP SCE*. U trećoj aktivnosti Inovacija i skalabilnost Transcribathon platforme sudjeluju *Facts & Files Historisches Forschungsinstitut Berlin Drauschke – Schreiber Partnerschaftsgesellschaft, Stichting Europeana, AIT Austrian Institute of Technology GmbH, READ-COOP SCE, Instytut Chemii Bioorganicznej Polskiej Akademii Nauk i Državni arhiv u Zagrebu*. U četvrtoj aktivnosti Diseminacija i kampanje građanske znanosti nabave iz mnoštva sudjeluju *Uniwersytet Wroclawski, Dublin City Council, Facts & Files Historisches Forschungsinstitut Berlin Drauschke – Schreiber Partnerschaftsgesellschaft, The Provost, Fellows, Foundation Scholars and the other members of Board, College of the Holy and Undivided Trinity of Queen Elizabeth near Dublin, Instytut Chemii Bioorganicznej Polskiej Akademii Nauk, Državni arhiv u Zagrebu, Stichting Europeana, AIT Austrian Institute of Technology GmbH, te READ-COOP SCE*. U petoj aktivnosti Menadžment i planiranje održivosti projekta sudjeluju svi partneri.⁵

ULOGA DRŽAVNOG ARHIVA U ZAGREBU U PROJEKTU

Iz prethodnog poglavlja vidljivo je da je Državni arhiv u Zagrebu uključen u prvu, treću, četvrtu i petu aktivnost. Jedan od bitnijih zadataka koji ima je dostava izabranih fondova, zbirki i komada na *Europeanu*. Izabrani su:

1. HR-DAZG-857 Zbirka Ulčnik Ivan, 1848./1930. – 1745 arhivskih jedinica;
2. HR-DAZG-839 Dragutin Hirc, 1844./1921. – 141 arhivska jedinica;
3. HR-DAZG-1 Poglavarstvo slobodnog kraljevskog grada Zagreba, Zapisnici sjednica Poglavarstva, 1800./1845. – 4 arhivske jedinice;
4. Izbor pisama, 1832./1903. – 8 pisama iz fondova HR-DAZG-823 Obitelj Deželić br. 1260 i br. 1695, HR-DAZG-838 Guteša Ilija br. 42, HR-DAZG-845 Zbirka Makanec Alfred br. 982, br. 2168, br. 2568, HR-DAZG-852 Radić Stjepan br. 133, br. 161;
5. HR-DAZG-24 Gradsko poglavarstvo Zagreb, Opskrbni odjel II-3, Potrošačke kartice, 120.621 komada.

Agreement No INEA/CEF/ICT/A2020/2288325, 9-17.

⁵ Connecting Europe Facility 2014-2020 : TELECOM CALLS FOR PROPOSALS 2020 : APPLICATION FORM, PART A, General information, 41-45.

Fondovi, zbirke i komadi izabrani su zbog svojeg sadržaja i značaja za grad Zagreb (npr. osobni fond Dragutina Hirca i zbirka Ivana Ulčnika), ali i radi prezentiranja fondova koji se čuvaju u Državnom arhivu u Zagrebu kao što su to razni obiteljski i osobni fondovi. Premda je fokus projekta na gradivu 19. stoljeća, arhiv dostavlja i gradivo 20. stoljeća, točnije Potrošačke kartice iz fonda HR-DAZG-24 Gradsko poglavarstvo Zagreb, te Opskrbni odjel II-3. Iako vremenski period ne odgovara onom zadanom od strane projekta Potrošačke kartice prihvaćene su za sudjelovanje zbog ranijeg projekta koji su započeli Odsjek za informacijske i komunikacijske znanosti Filozofskog fakulteta u Zagrebu, tvrtka *Bonsai Tech* iz Zagreba, te Državni arhiv u Zagrebu. U tom projektu, za koji je iskorištena velika količina već digitaliziranog gradiva iz fonda HR-DAZG-24 Gradsko poglavarstvo Zagreb, fokus je stavljen na transkripciju strukturiranih tekstova. Sav rukom ispisan tekst u obrascima Potrošačkih kartica zahtijevao je transkripciju koja je trebala biti osigurana na nekoj platformi za nabavu iz mnoštva. Paralelno s traženjem platforme za nabavu iz mnoštva pokrenute su pripreme za projekt *Enrich Europeana* + čiji su voditelji prihvatili Potrošačke kartice za uključivanje u novi projekt. U trećoj aktivnosti Državni arhiv u Zagrebu je u ulozi koordinatora između Odsjeka za informacijske i komunikacijske znanosti Filozofskog fakulteta u Zagrebu, *Bonsai Techa* i samog projekta. Upravo ranije spomenuti projekt, te metode i tehnike koje su koristili za obradu Potrošačkih kartica, potrebno je prilagoditi za projekt *Enrich Europeana* +. U sklopu četvrte aktivnosti Državni arhiv u Zagrebu organizira *Transcribathon* događaj kao kampanju građanske znanosti nabave iz mnoštva zajedno s Filozofskim fakultetom u Zagrebu. Predstavljanje projekta i raznih aspekata istog na savjetovanjima, skupovima i konferencijama također ulazi u četvrtu aktivnost. Peta aktivnost uključuje redovito sudjelovanje predstavnika Arhiva sjednicama odbora za menadžment projekta, te sastancima vezanim uz aktivnosti u koje je Arhiv uključen.

PROCES DIGITALIZACIJE U DRŽAVNOM ARHIVU U ZAGREBU

Digitalizacija u Državnom arhivu u Zagrebu dugo je vremena bila sporadična i nesistematizirana. Digitaliziralo se gradivo koje je bilo potrebno za organiziranje izložbi, objavljivanje u monografijama, zbornicima, katalozima i slično. Nakon što se digitalizirano gradivo iskoristilo za namjenu za koju je stvoreno, često mu se izgubio trag zbog lošeg načina imenovanja datoteka, te nepostojanja točno određenog mjesta trajnog spremanja. Nastale datoteke često nisu bile u formatima koji odgovaraju dugotrajnom očuvanju. Uključivanje u projekt *Enrich Europeana* +, ali i želja da se gradivo prilikom prve digitalizacije digitalizira kako treba, primoralo je Državni arhiv u Zagrebu da promijeni i prilagodi svoje dotadašnje prakse.

UPUTE ZA DIGITALIZACIJU DOKUMENTARNOG I ARHIVSKOG GRADIVA

Za proces digitalizacije primarno su korištene *Upute za digitalizaciju dokumentarnog i arhivskog gradiva* koje je objavio Hrvatski državni arhiv u Zagrebu na početku 2021. godine. *Upute* su nastale radi pridržavanja odredbi Zakona o arhivskom gradivu i arhivima (NN 61/2018, 98/2019) koje se odnose na očuvanje vjerodostojnosti digitaliziranog gradiva i mogućnosti korištenja digitaliziranog gradiva kao izvornog. Zakonom se

propisuje stvaranje pravila za pretvorbu gradiva iz analognog u digitalni oblik koja bi bila jasno definirana, s preciznim opisom postupka pretvorbe i mogućnosti kasnije provjere ispravnosti i kvalitete. U *Uputama* se mogu pronaći detaljna pravila i savjeti za pretvorbu gradiva koja se prvenstveno izvršava u Hrvatskom državnom arhivu u Zagrebu. Ipak, *Upute* služe kao preporuka za sve posjednike arhivskog i dokumentarnog gradiva, te osobama koje sudjeluju u procesu digitalizacije.

Upute sadrže nekoliko cjelina. Prva cjelina predstavlja uvod u *Upute*. Druga cjelina, *Postupak digitalizacije*, donosi opise uobičajenih postupaka digitalizacije gradiva. Uz postupak digitalizacije, ova cjelina obrađuje i načine na koje treba dokumentirati pojedine postupke u procesu digitalizacije, ali i na koje načine je moguće postupak digitalizacije pojednostaviti i prilagoditi potrebama institucije i gradiva. Naglasak je stavljen na postupak naknadne masovne digitalizacije zbog višeg rizika za gradivo koje se digitalizira. U trećoj cjelini, *Analiza rizika u postupku digitalizacije*, definirana su osnovna svojstva gradiva koja se trebaju očuvati procesom digitalizacije kako bi digitalizirano gradivo bilo jednakovrijedno izvorniku. To su integritet, autentičnost, raspoloživost, povjerljivost, kvaliteta i prenosivost. Ova cjelina donosi najčešće rizike zbog kojih se može narušiti neko od osnovnih svojstva gradiva. Za izbjegavanje takvog scenarija preporuča se izrada analize rizika za pojedinačne informacijske sustave ili projekte digitalizacije kako bi se definirale mjere i postupci koji su potrebni da ne bi došlo do narušavanja osnovnih svojstava gradiva. Četvrta cjelina, *Mjere za očuvanje osnovnih svojstava gradiva*, donosi opis mjera i postupaka koji su potrebni radi očuvanja osnovnih svojstava gradiva. Mjere i postupci se određuju prema utvrđenoj razini rizika te se dijele na one koji se odnose na organizaciju digitalizacije i informacijski sustav koji se koristi za postupak digitalizacije gradiva i one koje se koriste tijekom digitalizacije. U prilogima *Uputa* nalaze se identifikacijski i opisni metapodaci za gradivo koje se digitalizira, preporučeni i prihvatljivi formati datoteka, te kvalitete snimaka, popis tehničkih i administrativnih metapodataka za digitalizirano gradivo.

Rezultat postupka digitalizacije koji je predviđen ovim *Uputama* je master preslika. Master preslika je vjerna kopija originala. Iz nje se po potrebi mogu izrađivati dodatne kopije koje zadovoljavaju pojedinačne zahtjeve za specifičnu primjenu. *Upute* su prvenstveno predodređene za digitalizaciju tekstualnog, slikovnog i grafičkog materijala, te trodimenzionalnih objekata koji se digitaliziraju skenerom ili digitalnom kamerom.⁶

EUROPEANA PUBLISHING FRAMEWORK

Kod provođenja najnovijih procesa digitalizacije u obzir je uzet i projekt *Enrich Europeana +*. Za ovaj projekt predviđena je određena kvaliteta gradiva i metapodataka koja je propisana *Europeana Publishing Frameworkom*. Ona se validira i ocjenjuje u sklopu aktivnosti broj 1 (Agregacija sadržaja i infrastruktura za razmjenu sadržaja). *Europeana Publishing Framework* daje smjernice o kvaliteti sadržaja i metapodataka koji su potrebni za objavu na *Europeani*. Kvaliteta sadržaja sastoji se od četiri razine, tj. *Content Tier 1, 2, 3 i 4*. Kvaliteta metapodataka sastoji se od tri razine, tj. *Metadata Tier A, B i C*. Razine koje su postignute u procesu digitalizacije imaju izravan utjecaj na dostupnost gradiva široj publici.

⁶ *Upute za digitalizaciju dokumentarnog i arhivskog gradiva*, Hrvatski državni arhiv, 2021, 3-4.

Smjernice za kvalitetu sadržaja sastoje se od ranije spomenute četiri razine. One se razlikuju po kvaliteti sadržaja te licenciranju istog. Razina *Content Tier 1* propisuje obvezu dostavljanja direktne poveznice na objekt minimalne veličine od 0.1 megapiksela, te bilo koju od licenci iz *Europeana Licensig Frameworka* (okvir za licenciranje). *Europeana API prikaz* za ovu razinu mora sadržavati metapodatke i izravnu poveznicu na objekt. Razina *Content Tier 2* propisuje obavezu dostavljanja izravne poveznice na objekt minimalne veličine od 0.42 megapiksela, te bilo koju licencu iz ranije spomenutog okvira za licenciranje. Za ovu razinu *Europeana API prikaz* mora biti jednak prvoj razini. Za razinu *Content Tier 3* propisana je obveza dostavljanja izravne poveznice na objekt minimalne veličine od 0.95 megapiksela, te dodjeljivanje bilo koje licence iz okvira za licenciranje koja dozvoljava određenu razinu ponovnog korištenja digitaliziranog objekta. *Europeana API prikaz* također mora sadržavati metapodatke i izravnu poveznicu na objekt uz mogućnost filtriranja onih objekata koji se ponovno mogu koristiti. Četvrta razina propisuje dostavljanje direktnog linka na objekt minimalne veličine 0.95 megapiksela, te licence iz okvira za licenciranje koje omogućavaju besplatnu ponovnu upotrebu digitaliziranog objekta. *Europeana API prikaz* mora sadržavati iste podatke kao i kod razine 3.

Ovisno o razini ispunjavanja zahtjeva za kvalitetu sadržaja i *Europeana* i prikazivanje sadržaja na toj platformi dobivaju različit karakter. Ako sadržaj ispunjava uvjete za *Tier 1*, *Europeana* služi kao pretraživač (*search engine*) pomoću kojeg korisnici mogu pretraživati i pregledavati kolekcije. Prednost ove razine je povećanje pronalaska kolekcija pojedine baštinske ustanove, indeksiranje od strane drugih pretraživača, te povećanje prometa na internetskim stranicama baštinskih ustanova. *Tier 2* stavlja *Europeanu* u ulogu izloga, tj. omogućava se korištenje pojedinih dokumenata mnogobrojnih baštinskih ustanova u različitim tematskim kolekcijama uz osiguravanje konteksta i relacije prema drugim zbirkama unutar *Europeane*. Na taj način se zbirke jedne baštinske ustanove prezentiraju korisnicima mnogobrojnih drugih institucija koje sudjeluju u zajedničkoj kolekciji. Za *Tier 3* *Europeana* služi kao distribucijska platforma za ponovnu nekomercijalnu upotrebu. Osiguravanjem ove razine, sadržaj se može koristiti i na drugim platformama, u sklopu raznovrsnih projekata unutar *Europeane* s postojećim partnerima, a također može biti korišten u aplikacijama i servisima trećih strana. *Tier 4* *Europeani* daje karakter besplatne platforme za ponovnu upotrebu sadržaja čime sadržaj može biti korišten na platformama otvorenog tipa kao što je Wikimedia ili biti korišten u komercijalne svrhe i upotrebljavan od strane kreativne industrije.⁷

Projekt *Enrich Europeana* + određuje da u sklopu projekta razina kvalitete sadržaja mora odgovarati *Content Tier 2* za postojeće dokumente koji su učitani u *Europeanu*, te *Content Tier 4* za dokumente koji se digitaliziraju za potrebe projekta. Svi dokumenti koji se koriste za projekt prolaze validaciju. U slučaju da se za određene dokumente ispostavi da ne odgovaraju propisanim razinama, partneri koji osiguravaju sadržaj obavezni su dokumente poboljšati i povećati razinu sadržaja.

DIGITALIZACIJA FONDOVA I ZBIRKI

U ovom radu fokusirat ćemo se na proces digitalizacije vezan uz gradivo koje je bilo određeno za projekt *Enrich Europeana* +. Gradivo se digitaliziralo i ranije, najviše u

⁷ Europeana Publishing Framework, <https://pro.europeana.eu/post/publishing-framework> (datum posjete 26. 7. 2022)

svrhu korištenja za izložbe, objavljivanje u monografijama i zbornicima radova, te za potrebe stranaka. Tijekom digitalizacije gradiva za projekt započet je rad na standardizaciji samog procesa digitalizacije u Državnom arhivu u Zagrebu kako bi digitalizirano gradivo bilo pogodno i za dugotrajno čuvanje.

Prilikom traženja najboljih formata za digitaliziranje i dugotrajno očuvanje gradiva naišli smo na već spomenute *Upute za digitalizaciju dokumentarnog i arhivskog gradiva*. Na temelju navedenih *Uputa* razradili smo postupak digitalizacije koji trenutno odgovara našim potrebama i mogućnostima. Razvijeni postupak djelomično prati opći model postupka digitalizacije koji se može pronaći u *Uputama*.⁸

Odluku o digitalizaciji gradiva zamijenila je prijava ranije spomenutih arhivskih fondova i zbirki za uključivanje u projekt *Enrich Europeana* +. Izvršena je dodatna provjera navedenog gradiva kako bi se utvrdile eventualne potrebe za dodatnom intervencijom u gradivu. Pošto je celokupno gradivo arhivistički sređeno i obrađeno, nije bilo potrebno ništa mijenjati. Osim pripreme i provjere fizičkog gradiva, izvršena je provjera stanja na serveru za slučaj da je neki od odabranih fondova i zbirki već digitaliziran. To je bio slučaj kod zbirke Ivana Ulčnika i kod Potrošačkih kartica. Dodatnom provjerom kvalitete već digitaliziranog gradiva zaključeno je da zbirka Ivana Ulčnika ne zadovoljava preporučene parametre prema *Uputama* i *Europeana Publishing Frameworkom*, te je zbirka poslana na ponovnu digitalizaciju. Potrošačke kartice bile su zadovoljavajućih parametara. Kod svakog fonda i zbirke koja se digitalizirala za projekt provjeren je redoslijed gradiva unutar kutija i druge zaštitne opreme u kojoj se gradivo čuva. Uspoređene su oznake na gradivu s inventarima kako bi se dodatno osigurali točnost i cjelovitost gradiva nakon digitalizacije. Za svaki fond i zbirku vodila se briga o fizičkoj pripremi za snimanje, odnosno vodila se briga o tome da određeno gradivo može biti snimljeno na raspoloživim uređajima za digitalizaciju. Tako je kod fonda Dragutina Hirca odlučeno da se snima svaki dokument pojedinačno na ukupno dva uređaja zbog karakteristika gradiva. Za snimanje ovog fonda korišten je plošni skener Microtek ScanMaker 1000 XL Plus, te fotografski aparat Sony DSC-HX300. Fotografski aparat korišten je za one dokumente koji zbog svoje veličine nisu mogli biti digitalizirani na plošnom skeneru (maksimalni format A3). Kod zbirke Ivana Ulčnika korišten je samo plošni skener zbog malog formata gradiva. Zapisnici sjednica zaštitno su snimani i pohranjeni na mikrofilmu, ali digitalizacijom mikrofilma nismo dobili zadovoljavajuću kvalitetu snimaka. Zbog toga se zapisnici sjednica takođe digitaliziraju na plošnom skeneru. Ostatak gradiva koje se digitalizira u sklopu projekta također je digitaliziran na navedenom plošnom skeneru. Na samoj digitalizaciji ukupno je radilo troje djelatnika čime se osigurala vrlo visoka razina kontrole nad cijelom procesom, te samostalno uočavanje eventualnih pogrešaka. Zaduženi djelatnici dobili su detaljne upute o postavkama uređaja za digitalizaciju, imenovanju datoteka i načinu spremanja istih. Svaki od izabranih fondova i zbirki dobiva svoj direktorij s punim nazivom fonda odnosno zbirke. Unutar glavnog direktorija struktura prati arhivske ili tehničke jedinice unutar fondova i zbirki. Primjerice, kod fonda Dragutina Hirca imamo ukupno 142 direktorija koji odgovaraju broju arhivskih jedinica unutar fonda. Kod Ivana Ulčnika prate se tehničke jedinice. Serija Fotografije unutar zbirke Ivana Ulčnika smještena je u ukupno 15 kutija, a upravo je to i broj direktorija u koje je

⁸ *Upute za digitalizaciju dokumentarnog i arhivskog gradiva*, Hrvatski državni arhiv, 2021, 6.

smještena digitalizirana verzija. Kod same digitalizacije istovremeno je validirana svaka pojedinačna snimka te obrezana. Po završetku digitaliziranja gradiva izvršena je provjera cjelovitosti snimanja i kvalitete snimaka. Cjelovitost snimanja i kvaliteta provjeravala se uspoređivanjem originala s digitaliziranom inačicom. U slučaju da neke snimke nisu vjerno prikazivale original, kao što je to znao biti slučaj s fotografijama kod zbirke Ivana Ulčnika, ponovno se digitalizirao sporni dokument, točno imenovao i umetao u direktorij kojem pripada. Ako je došlo do propusta kod cjelovitosti, dokumenti koji su nedostajali također su se odmah digitalizirali, imenovali, obrezivali, provjeravali i umetali u direktorije kojima pripadaju.

Izabrani fondovi i zbirke većinski se sastoje od tekstualnih dokumenata iako kod nekih, kao što je to zbirka Ivana Ulčnika, prevladavaju slikovni i grafički dokumenti s dodatkom teksta. Prema vrstama dokumenata birani su preporučeni formati koje propisuju *Upute za digitalizaciju*. Prema njima, preporučeni formati za tekstualne dokumente i knjige su JPEG 2000, TIFF i PDF/A formati, a prihvatljivi su JPEG, PNG i PDF. Preporučena rezolucija je 300 ppi (400 ppi), 24-bitna boja, te *eciRGBv2*, Adobe RGB (1998), sRGB, Pro Photo i Gray Gamma 2.2 prostor boje. *Upute* donose i dodatne napomene koje se odnose na snimanje u rezoluciji većoj od 300 ppi. Veću rezoluciju može zahtijevati određeno svojstvo izvornika, snimanje u sivoj skali kada boja nije bitna ili kod predložaka koji nisu u boji. Za tekstualne dokumente i knjige koji sadrže i fotografije, crteže ili neki sličan sadržaj preporuča se digitaliziranje kao da se radi o fotografiji, crtežu ili grafici. Kod digitalizacije fotografija, crteža i grafika preporučeni su TIFF i JPEG 2000 formati, a prihvatljivi su JPEG, PNG, DNG, te izvorni formati fotografskog aparata. Preporučena je rezolucija od 400 ppi i 24-bitna boja, ali je poželjna rezolucija od 600 ppi i 48-bitna boja. Za prostor boje preporučeni su formati *eciRGBv2*, Adobe RGB (1998), Pro Photo i sRGB. *Upute* pod Dodatna napomena stavljaju snimanje u većoj rezoluciji od 400 ppi i većoj dubini boje zbog svojstva izvornika.⁹

Kod digitalizacije izabranih fondova i zbirki slijedile su se preporučene upute uz potrebne manje izmjene kod nekih dokumenata. Većina gradiva digitalizirana je u TIFF formatu, dok je manji dio digitaliziran u JPG formatu kao izvornom formatu fotografskog aparata (dio fonda Dragutina Hirca). Većina pripremljenih fondova i zbirki digitalizirana je u rezoluciji od 800 dpi, sRGB prostorom boje, te 24-bitnom dubinom. Kod JPG formata kao izvornog formata fotografskog aparata rezolucija iznosi 350 dpi, što je maksimalan doseg uređaja. Kod zbirke Ivana Ulčnika ovi parametri su prilagođeni vrsti materijala koji se digitaliziraju. Radi se većinom o fotografijama, crtežima i drugim slikovnim materijalima za koje smo koristili rezoluciju od 600 dpi, 32-bitnu boju, te sRGB prostor boje. Potrošačke kartice već su bile digitalizirane u rezoluciji od 300 dpi. Zapisnici sjednica PGZ-a trenutno se digitaliziraju zbog ranije spomenutih problema, no za njih predviđamo rezoluciju od 600 dpi, 24-bitnu boju, te sRGB prostor boje.

Do sada ukupna veličina digitaliziranog gradiva za potrebe projekta iznosi 69,2 GB (5663 dokumenta u 179 direktorija). Potrošačke kartice koje su već ranije digitalizirane i izabrane za projekt spremljene su u ukupno 661 direktorij u kojima se nalazi nešto više od 120.000 datoteka ukupne veličine 10,6 GB.

⁹ *Upute za digitalizaciju dokumentarnog i arhivskog gradiva*, Hrvatski državni arhiv, 2021, 56-57.

METAPODACI U ARHIVU

Po završetku digitalizacije izabranih fondova i zbirki započela je izrada metapodataka. Njihova izrada u Državnom arhivu u Zagrebu do ovog projekta nije bila uobičajena jer Arhiv nema informacijski sustav u kojem bi čuvao i prezentirao digitalizirano gradivo koje se čuva na lokalnom serveru i prijenosnim tvrdim diskovima. Za svaki od izabranih fondova i zbirki postoje inventari (sumarni i analitički), ali oni nisu bili dostatni za računalnu obradu. Sve podatke koji se mogu pronaći u inventarima trebalo je oblikovati kako bi ih računalo moglo prepoznati, sistematizirati i prikazivati na točno određenom mjestu. Kod traženja rješenja za izradu metapodataka za gradivo koje je određeno za projekt trebalo je uzeti u obzir i odredbe o razini metapodataka koje projekt propisuje. Za razinu metapodataka slijede se razine prema *Europeana Publishing Frameworku*. Ovaj okvir za objavu sadržaja propisuje tri razine metapodataka: *Metadata Tier A, B i C*. Svaka od ove tri razine metapodataka ima obavezna i izborna polja koja omogućavaju povećanje pretraživosti sadržaja. Tako *Metadata Tier A* propisuje da minimalno 25% dostavljenih relevantnih EDM (*Europeana Data Model*) polja imaju barem jednu vrijednost za jezik. Također, propisuje barem jedan element za omogućavanje (*Enabling Element*) uzet iz grupe za otkrivanje (*Discovery scenario*). Kontekstualni razred nije potreban. *Metadata Tier B* propisuje da minimalno 50% dostavljenih relevantnih EDM polja metapodataka ima barem jednu vrijednost za jezik (propisana su minimalno tri različita elementa za omogućavanje od kojih dva trebaju biti iz grupe za otkrivanje). Propisuje minimalno jedan kontekstualni razred sa svim predviđenim elementima ili poveznicu na LOD (*Linked Open Data*) rječnik. *Metadata Tier C* propisuje minimalno 75% dostavljenih relevantnih EDM polja metapodataka s barem jednom vrijednosti za jezik, barem četiri različita elementa iz dvije različite grupe za otkrivanje. Nadalje propisuje minimalno dva različita kontekstualna razreda sa svim predviđenim elementima ili poveznicu na LOD rječnik. S postignućem razine A *Europeana* služi kao osnovna platforma za pretraživanje. Prednosti ove razine su povećanje posjećenosti internetskih stranica institucije s koje je neko gradivo stiglo, povećanje pronalaska gradiva koje se nalazi u kolekcijama *Europeane*, te indeksiranje od strane raznih pretraživača. S postignućem B razine metapodataka *Europeana* se koristi kao istraživačka platforma gdje se gradivo jedne institucije može koristiti u tematskim kolekcijama s drugim partnerima, pri čemu se međusobno nadopunjuju, te povećavaju vidljivost unutar *Europeane*. Razina C omogućuje razvoj *Europeane* kao platforme znanja. Gradivo koje ima C razinu metapodataka moguće je koristiti u postojećim projektima *Europeane*, od strane kreativne industrije, kao i za aplikacije i usluge.¹⁰

Prema odredbama projekta metapodaci za sadržaj koji se agregira na *Europeanu* u sklopu projekta trebaju imati razinu B (*Metadata Tier B*). Zbog toga je izrada metapodataka, odnosno upisivanje potrebnih podataka u predviđena polja, prilagođena naputcima *Europeana Publishing Frameworka*.

Za izradu metapodataka korištena je excel tablica koju je pripremio Hrvatski državni arhiv. Ova tablica izrađena je u sklopu projekta *e-Kultura - digitalizacija kulturne baštine* u kojoj Hrvatski državni arhiv sudjeluje kao jedan od glavnih partnera zajedno se

¹⁰ Europeana Publishing Framework, <https://pro.europeana.eu/post/publishing-framework> (datum posjete 26. 7. 2022)

Muzejom za umjetnost i obrt, Nacionalnom i sveučilišnom knjižnicom u Zagrebu, Hrvatskom radiotelevizijom i Ministarstvom kulture i medija koje je ujedno i nositelj projekta. U sklopu projekta *e-Kultura* partneri zajedno s tvrtkom *Ericsson Nikola Tesla* razvijaju nacionalni agregator i portal koji će služiti za prikazivanje hrvatske digitalizirane kulturne baštine. U sklopu projekta razvijeno je ili prilagođeno nekoliko platformi koje se međusobno nadopunjuju: *Goobi*, Središnja administracija, Središnji sustav i portal *eKultura*.¹¹ Aplikacija *Goobi* koristi se kao servis za podršku tijeku digitalizacije. Ovoj aplikaciji prilagođena je excel tablica koju su razradili djelatnici Hrvatskog državnog arhiva.

Excel tablica sastoji je od dva lista: Podaci i Datoteke. U listu Podaci upisuje se većina metapodataka. Oni su podijeljeni na obavezne i dodatne. Pod obavezne spadaju Identifikator, Naslov, Razina, Visa ID, Redoslijed, Vrijeme OD, Vrijeme DO, Vrsta zapisa, Vrsta sadržaja, Institucija, Zbirka, Status zapisa, Za objavu, te Uvjeti korištenja¹². Kod identifikatora stavljamo oznaku fonda ili zbirke npr. HR-DAZG-857 ili HR-DAZG-839. Upisivanjem novih elemenata, točnije serija, podserija, predmeta ili komada/jedinica, glavnom identifikatoru dodajemo bročane oznake kako bismo postigli razlikovanje viših i nižih struktura. Kod naslova upisujemo nazive prema inventarima fondova i zbirki. Nazivi se povezuju s identifikatorom na način da prvi identifikator nosi naslov fonda ili zbirke (npr. HR-DAZG-857 nosi naslov Zbirka Ulčnik Ivan). Niži identifikatori nose naslove serija, podserija, predmeta ili komada/jedinica (npr. HR-DAZG-857-1 Fotografije i tiskovine (serija), HR-DAZG-857-1-1 Accurti Antun (komad/jedinica)). Kod polja Razina za svaki pojedinačni upis definira se njegova razina unutar fonda, točnije da li je to opis za razinu fonda, serije, podserije, komada/jedinice. Izrađen je padajući izbornik s ponuđenim razinama. Polje Visa ID odnosi se na povezivanje struktura unutar jednog fonda ili zbirke. Najviša razina (fond ili zbirka) nema svoj Visa ID, a sve niže razine povezuju se na svoje odgovarajuće više razine. Tako se serija s identifikatorom HR-DAZG-857-1 Fotografije i tiskovine povezuje na svoju višu razinu HR-DAZG-857 i upravo se to upisuje u polje Visa ID. Isto tako komad/jedinica s identifikatorom HR-DAZG-857-1-1 povezuje se na svoju višu razinu, a to je serija Fotografije i tiskovine upisivanjem identifikatora, te serije HR-DAZG-857-1 u polje Visa ID. S Visa ID poljem povezano je polje Redoslijed gdje se stvara budući prikaz fonda ili zbirke na platformama. Raspored se slaže uzimajući u obzir polje Visa ID i mjesto na kojem želimo prikazati određenu nižu razinu. Za ilustraciju, serije unutar Zbirke Ulčnik Ivan složene su prema redoslijedu koji se nalazi u inventaru zbirke. Na prvom mjestu nalazi se serija HR-DAZG-857-1 Fotografije i tiskovine, na drugom serija HR-DAZG-857-2 Bilješke i novinski izresci, a na trećem serija HR-DAZG-857-3 Razno. U polja Vrijeme OD i Vrijeme DO upisujemo podatke o vremenu za pojedine razine, tj. godine, datume, periode i sl. Kod polja Vrsta zapisa definiramo da li su neki fond, zbirka, predmet ili komad/jedinica tekstualni, slikovni, zvučni materijal, 3D ili video. Za ovo polje također je napravljen padajući izbornik s ponuđenim kategorijama (TEXT, VIDEO, SOUND, IMAGE, 3D). Polje Vrsta sadržaja također ima ponuđeni padajući izbornik unutar kojeg su definirani najčešće vrste (tekstualni dokumenti, knjige/tisak, fotografije, crteži, grafike, novine, časopisi, rukopisi, karte i planovi, tehnički nacrti, zvučni zapisi, notni zapisi, razglednice, datoteke podataka, softver, mrežne stranice, film/

¹¹ Bilić, J. Projekt e-Kultura : digitalizacija kulturne baštine. *Muzeologija* (Zagreb), 56, (2019), 27-28.

¹² Nazivi metapodataka izravno preuzeti iz tablice.

video, arheološka građa, etnografska građa, građa primijenjenih umjetnosti, graditeljska baština, nematerijalna baština, prirodoslovna građa, umjetnička građa, tehnička građa). U polje Institucija upisuje se oznaka institucije. U slučaju Državnog arhiva u Zagrebu to je oznaka HR-DAZG. Oznaka se definira prilikom registriranja institucije u sustavu *e-Kultura*. U polje Zbirka također se upisuje oznaka institucije koja je povezana sa smještajem agregiranog gradiva u točan direktorij u Središnjem sustavu. Tako Državni arhiv u Zagrebu ima oznaku zbirke dazg-hr. Status zapisa prikazuje informaciju da li se radi o potpuno novom setu (NEW) ili o nadogradnji već agregiranih podataka (UPDATE). Izborom u padajućem izborniku Za objavu određujemo hoće li neki agregirani set podataka biti objavljen na portalu *e-Kultura* (Objavljen) ili će biti samo trajno pohranjen (Neobjavljen). Kod razine fonda ili zbirke nužno je odabrati DatasetRecord kako bi sustav pravilno povukao sve informacije iz tablice. U Uvjetima korištenja (sic!) biramo između osamnaest licenci za objavu sadržaja (<https://rightsstatements.org/en/> i <https://creativecommons.org/>). Dosad agregirano gradivo Državnog arhiva u Zagrebu dobilo je oznaku javnog dobra (Public Domain Mark 1.0).

Dodatna polja obuhvaćaju Sadržaj, Mjesto, MjestoVrsta, Autor, AutorVrsta, FizickaSvojstva, Dostupnost, Jezik, Pismo, NosacZapisa, KolicinaNosacaZapisa, VrstaGrade, Predmet i ImateljNaziv (sic!). Od dodatnih polja kod svih fondova i zbirki ispunjavamo polje za sadržaj, jezik, pismo i količinu nosača zapisa. Za ispunjavanje polja za jezik koriste se troslovcane oznake jezika prema ISO 639-2:1998 *Codes for the representation of names of languages – Part 2: Alpha-3 code*. Tako se za hrvatski jezik koristi oznaka hrv, za njemački ger itd. Za pismo koristimo cijeli izraz (npr. latinica). Kod pojedinih fondova stavljamo i druge metapodatke koje možemo definirati kao što je Mjesto i MjestoVrsta, Autor i AutorVrsta. Prednost ove tablice je mogućnost njezinog proširenja kao što su povećanje broja polja za mjesto, autore, jezik i pismo. Nedostatak je u tome što neka greška prilikom ispunjavanja može biti vidljiva tek prilikom učitavanja excel tablice na aplikaciju *Goobi*. Također, potrebno je voditi računa o formatu ćelija u tablici jer se promjenom samo jedne ćelije iz formata za tekst može narušiti cijela struktura fonda ili zbirke.

Prvi list *Podaci* važan je za definiranje hijerarhijske strukture fondova i zbirki. Drugi list, *Datoteke*, koristi se za povezivanje metapodataka s lista *Datoteke* i digitaliziranog gradiva koje je opisano tim metapodacima. List *Datoteke* sastoji se od polja Identifikator, Slike, Redni broj, SHA-256, Velicine i Putanje. Polje Identifikator donosi istovjetne podatke koji se nalaze u polju Identifikator na listu *Podaci*. U polje Slike unose se nazivi datoteka. U polje Redni broj unose se brojevi na temelju kojih želimo prikazati datoteke koje su vezane na određeni identifikator. Tako identifikator HR-DAZG-857-1-1 s naslovom Accurti Antun veže na sebe dvije datoteke – 0857_00001.tif i 0857_00002.tif. Odlučili smo da će se datoteka 0857_00001.tif prikazivati prva te dobiva broj 1, a datoteka 0857_00002.tif će se prikazivati druga te dobiva broj 2. SHA-256 vrijednost je koju dobivamo korištenjem aplikacije *Droid* koju je razvio Britanski nacionalni arhiv za automatsku identifikaciju većih količina dokumenata. U polje Velicina upisuje se veličina datoteka koje će se agregirati. Polje Putanja ispunjava se podacima o lokaciji svake pojedinačne datoteke. Digitalizirano gradivo koje se agregira u sklopu ovog projekta spremjeno je lokalno na serveru u Državnom arhivu u Zagrebu zbog čega putanja ima ovakvu konstrukciju: \\Server\europaana\857\1\0857_00001.tif.

Metapodaci za ranije nabrojane fondove i zbirke uključene u ovaj projekt izrađivani su po završetku digitalizacije. Metapodatke je izrađivala jedna osoba u arhivu. U trenutku pisanja ovog rada započeta je izrada metapodataka za Potrošačke kartice za koje je predloženo da se organiziranju u manje setove podataka zbog ograničenja sustava prilikom učitavanja velike količine datoteka i podataka. Maksimalni setovi biti će s oko 3.000 datoteka.

WEB APLIKACIJA GOOBI

Po završetku ispunjavanja excel tablica za metapodatke za pojedine fondove i zbirke krećemo s pokretanjem procesa na aplikaciji *Goobi*. Kao što je to već ranije spomenuto, web aplikacija *Goobi* služi kao servisa za podršku tijekom digitalizacije. Aplikacija je razvijena tako da je prilagođena svim potrebama koje imaju baštinske ustanove. To je procesno orijentirana aplikacija kod koje svi korisnici imaju dodijeljene uloge. Korisničke uloge su lokalni administrator, stručnjak za digitalizaciju, kontrolor procesa digitalizacije i objavitelj sadržaja. Jedan korisnik može imati samo jednu ulogu ili više njih.¹³ Procesi koji se pokreću na ovoj aplikaciji sastoje se od nekoliko zadataka koji se odvijaju prema unaprijed određenom redu. Za arhive je preporučeno da se koristi predložak pod nazivom ARHSlika gdje su već unaprijed definirani svi koraci unutar procesa. Taj proces sastoji se od 17 koraka: pokretanje procesa, učitavanje datoteka, provjera uređaja snimanja, provjera kvalitete snimke, validacija slike, kreiranje prezentacijskih datoteka, uređivanje metapodataka, spajanje datoteka sa snimkama, obrada datoteka, provjera cjelovitosti snimanja, exportDMSF, export priprema, priprema paketa za izvoz, te izvoz podataka. Određeni koraci vrše se automatski, a dio koraka rješavaju korisnici koji imaju za to određenu ulogu. Zadaci koje rješavaju korisnici su pokretanje procesa, učitavanje datoteka, provjera kvalitete snimke, uređivanje metapodataka, obrada datoteka, provjera cjelovitosti snimanja, priprema paketa za izvoz, te izvoz podataka. Nije moguće početi novi korak prije nego se prethodni završi. Cijeli proces započinje učitavanjem excel tablice s metapodacima za određeni fond ili zbirku. Već prilikom učitavanja excel tablice vrši se provjera ispravnosti cijele tablice i unutarnje strukture, tako da u slučaju greške aplikacija odmah javlja gdje je nastao problem. Nakon uspješnog učitavanja excel tablice slijedi učitavanje datoteka. Manje količine datoteka vrlo brzo se učitaju. Kod većeg broja datoteka (npr. više od dvije tisuće) može doći do problema zbog ograničenja sustava vezanog uz trajanje pojedinog procesa. Zbog toga je preporučljivo učitavati manje količine datoteka kako ne bi došlo do prekida koraka i neuspješnog učitavanja istih. Nakon provjere uređaja snimanja, korisnik vrši provjeru kvalitete snimke gdje može provjeriti svaku pojedinačnu datoteku i po potrebi je ispraviti ili pravilno okrenuti. Validacija slike i kreiranje prezentacijskih datoteka automatski se izvode prije uređivanja metapodataka. Korisnik u ovom koraku može provjeriti strukturu metapodataka, ispraviti pogreške u opisima, dodati nove metapodatke itd. Nakon spajanja datoteka sa snimkama slijedi obrada datoteka u kojoj vršimo ponovnu provjeru datoteka i eventualne popravke. Kod provjere cjelovitosti snimanja najbitniji je korak provjera točne paginacije za svaku pojedinačnu datoteku. Prelaskom miša preko

¹³ Upute za korisnike – Servis za podršku tijekom digitalizacije. Ugovor o javnoj nabavi Usluge razvoja i održavanja IKT sustava za projekt e-Kultura – „Digitalizacija kulturne baštine“, ugovor br. 11/2020, 12-13.

naziva pojedine jedinice prikazuje nam se broj slike ili slika koje su vezane na nju. Iduća dva koraka, exportDMSF i export priprema, vrše se automatski. Ovi koraci služe za pravičan ispis METS dokumenta u kojem su sadržane sve bitne informacije o cijelom procesu. Korak priprema paketa za izvoz vrši korisnik pri čemu se stvara SIP paket u kojem se nalaze metapodaci i reprezentacije. Posljednji korak u cijelom procesu je izvoz podataka. Preduvjet za taj korak je da je SIP paket uspješno kreiran i da je dostupan na sustavu. Podaci se izvoze prvo u Središnji sustav na definiranu lokaciju, a potom se pokreće proces prilikom čega se SIP paketi šalju u katalog ustanove.

Središnji sustav ima funkciju sigurne trajne pohrane digitalnog sadržaja u skladu s OAIS, METS, PREMIS, MODS, DC i LIDO standardima. Svaka institucija ima svoj identifikator, naziv i vrstu pod koju je registrirana, te u katalogu može vidjeti status poslanih paketa.¹⁴ Nakon uspješnog zaprimanja paketa na Središnjem sustavu, paketi s metapodacima i reprezentacijama mogu se prikazati na portalu *e-Kultura* i/ili staviti na OAI-PMH server koji omogućava pobiranje podataka od strane *Europeane*. Upravo je postojanje nacionalnog agregatora s kojeg se učitava izabrano gradivo jedan od uvjeta projekta. Također, mnogi ciljevi projekta povezani su s implementacijom novih tehnologija i usluga koje se razvijaju u sklopu projekta u nacionalne agregatore. Uzimajući u obzir rane faze razvoja hrvatskog nacionalnog agregatora, očekujemo implementacije po završetku projekta jer je trenutni glavni cilj vezan uz ovaj agregator uspješna agregacija materijala na *Europeanu* unutar trajanja projekta *Enrich Europeana* +.

ZAKLJUČAK

Projekt *Enrich Europeana* + potaknuo je Državni arhiv u Zagrebu na brzu prilagodbu novim zahtjevima koji se tiču procesa digitalizacije u ovoj ustanovi. Prilikom izbora uređaja, određivanja postavki istog i koraka tijekom samog postupka digitalizacije, vođeno je računa o preporukama iz *Uputa za digitalizaciju dokumentarnog i arhivskog gradiva*, te uvjetima koje propisuje projekt na temelju *Europeana Publishing Frameworka*. Razine kvalitete sadržaja i metapodataka osiguravaju veću pretraživost i bolji odaziv prilikom pregledavanja kolekcija na *Europeani*. Samim time i Državni arhiv u Zagrebu stječe veću prepoznatljivost među korisnicima diljem Europe.

Proces digitalizacije u Državnom arhivu u Zagrebu detaljno je razrađen i nastaviti će se primjenjivati kod budućih projekata u koje će se Arhiv uključiti kao i kod svojih vlastitih. Iako neki koraci (kao što su način imenovanja digitaliziranog gradiva, te njihovo spremanje) izgledaju banalno, njihovom točnom definicijom smanjili smo mogućnost nastanka pogrešaka tijekom cijelog procesa. Izrada metapodataka i agregacija sadržaja na *Europeanu* povezala je Državni arhiv u Zagrebu s projektom *e-Kultura* prilikom kojeg su u proces digitalizacije uvedene nove aplikacije kao što je *Goobi*. Ova web aplikacija kao podrška procesu digitalizacije svojom organizacijom procesa olakšava cijeli postupak za arhiv i njegove djelatnike. Nakon agregacije i objave sadržaja na portalu *e-Kultura* i *Europeani*, na Državnom arhivu u Zagrebu je da nastavi s ovakvom praksom, kvalitetno

¹⁴ Upute za institucijske administratore – Središnji sustav. Ugovor o javnoj nabavi Usluge razvoja i održavanja IKT sustava za projekt *e-Kultura* – „Digitalizacija kulturne baštine“, ugovor br. 11/2020, 17.

digitalizira još veći broj svojih fondova i zbirki, te objavljivanjem na portalu *e-Kultura* i *Europeana* omogućiti pristup gradivo što široj publici.

IZVORI I LITERATURA

- Bilić, Jelena. *Projekt e-Kultura, digitalizacija kulturne baštine*, Muzeologija 56, Zagreb 2019, 27-28.
- *Connecting Europe Facility 2014-2020 : TELECOM CALLS FOR PROPOSALS 2020: APPLICATION FORM, PART A, General information*
- Enrich Europeana, <https://pro.europeana.eu/project/enrich-europeana> (25. avgust 2022)
- Europeana Publishing Framework, <https://pro.europeana.eu/post/publishing-framework> (26. jul 2022)
- *Grant Agreement under the Connecting Europe Facility (CEF) – Telecommunications Sector: Agreement No INEA/CEF/ICT/A2020/2288325*
- *Upute za digitalizaciju dokumentarnog i arhivskog gradiva*, Zagreb 2021.
- *Upute za institucijske administratore – Središnji sustav*. Ugovor o javnoj nabavi Usluge razvoja i održavanja IKT sustava za projekt e-Kultura – „Digitalizacija kulturne baštine“, ugovor br. 11/2020.
- *Upute za korisnike – Servis za podršku tijekom digitalizacije*. Ugovor o javnoj nabavi Usluge razvoja i održavanja IKT sustava za projekt e-Kultura – „Digitalizacija kulturne baštine“, ugovor br. 11/2020.

Sonja GALINA

DIGITISATION AND CREATION OF METADATA IN THE STATE ARCHIVES IN ZAGREB: NATIONAL AGGREGATOR AND EUROPEANA

Summary

160

The Enrich Europeana + project encouraged the State Archives in Zagreb to quickly adapt to the new requirements regarding the digitisation process in this institution. When choosing the device, determining its settings, and the steps during the digitisation process itself, the recommendations from the Instructions for the digitisation of documentary and archival material and the conditions prescribed by the project based on the Europeana Publishing Framework were taken into account. Content and metadata quality levels ensure greater searchability and better responsiveness when browsing collections on Europeana. As a result, the State Archives in Zagreb is gaining greater recognition among users throughout Europe.

The digitization process in the State Archives in Zagreb has been elaborated in detail and will continue to be applied in future projects in which the archive will be involved, as well as in its own projects. Although some steps such as the way to name the digitized material and save them seem banal, by their exact definition we have reduced the possibility of errors during the entire process. The creation of metadata and the aggregation of content on Europeana connected the State Archives in Zagreb with the *e-Kultura* project, during which new applications such as *Goobi* were introduced into the digitization process. This web application, as a support for the digitization process, facilitates the entire process for the archive and its employees with its process organization. After the aggregation and publication of the content on the *eKultura* and Europeana portals, it is up to the State Archives in Zagreb to continue with this practice, digitize an even greater number of its holdings and collections, and by publishing them on the *eKultura* and Europeana portals, provide access to the material to the widest possible audience.