

Prejeto / Received: 30. 07. 2021
<https://doi.org/10.54356/MA/2021/KMXI9438>

1.02 Pregledni znanstveni članek

1.02 Review Article

POSTOPEK PREVZEMA IN PROBLEMATIKA OBLIKOVANJA SPREJEMNIH INFORMACIJSKIH PAKETOV (SIP) NA PRIMERU ZVOČNIH ZAPISOV DRŽAVNEGA ZBORA REPUBLIKE SLOVENIJA

Jože Glavič

Zgodovinski arhiv Ljubljana, Slovenija
joze.glavic1@guest.arnes.si

Vesna Gotovina

Arhiv Republike Slovenije, Slovenija
vesna.gotovina@gov.si

Klavdija Krivec

Arhiv Republike Slovenije, Slovenija
klavdija.krivec@gov.si

Dr. Žiga Koncilija

Arhiv Republike Slovenije, Slovenija
ziga.koncilija@gov.si

Dr. Tatjana Hajtnik

Arhiv Republike Slovenije, Slovenija
tatjana.hajtnik@gov.si

Izvleček:

Vzpostavitev postopka izročitve oz. prevzemanja elektronskega arhivskega gradiva predstavlja eno izmed ključnih nalog v okviru projekta vzpostavitve slovenskega elektronskega arhiva oz. projekta e-ARH.si (2016–2021). Skladno s tem je bilo izvedenih več pilotnih prevzemov. Razvita je bila ustrezna metodologija, dokumentacija in orodja, namenjena učinkovitemu in varnemu sodelovanju z ustvarjalcem oz. v podporo izročitvi in prevzemanju elektronskega arhivskega gradiva v obliki sprejemnih informacijskih paketov (SIP) v repozitorij. Prispevek se osredotoča na ustvarjalca in specifično zvočnih zapisov kot arhivskega gradiva ter zlasti na samo izvedbo prevzema zvočnih zapisov Državnega zbora Republike Slovenije, nastalih januarja 2017 – od začetka sodelovanja z ustvarjalcem do oblikovanja ustreznih navodil oz. dokumentacije, priprave SIP ter do izročitve elektronskega arhivskega gradiva pristojnemu arhivu oz. v slovenski elektronski arhiv.

Ključne besede:

projekt e-ARH.si, nestrukturirani zvočni zapisi, SIP, Urejevalnik SIP, elektronski arhiv

Abstract:

Ingest Procedure and the Challenges of Creating Submission Information Packages (SIP) on the Case of Audio Records of the National Assembly of the Republic of Slovenia

The establishment of the procedure for the acquisition of electronic archival records is one of the key tasks within the project of establishing a Slovenian electronic archive or project e-ARH.si (2016-2021). Accordingly, several pilot acquisitions have been carried out. Appropriate methodology, documentation and tools have been developed for effective and safe cooperation with the creator or in support of the acquisition of electronic records in the form of SIP to the repository. The article focuses on the creator and the specifics of audio records as archival material and especially on the actual acquisition process of audio records of the National Assembly of the Republic of Slovenia, created in January 2017 – from the beginning of cooperation with the creator to the creation of appropriate instructions or. documentation,

preparation of SIP packages and to the delivery of electronic archival records to the competent archive or in the Slovenian electronic archive.

Key words:

project e-ARH.si, unstructured audio records, SIP, SIP creator, electronic archive

1. Uvod

Dolgoletna praksa prevzemanja fizičnega arhivskega gradiva je pokazala, da je sodelovanje med pristojnim arhivom in ustvarjalcem krožni proces, pri katerem je izmenjava informacij in znanj nujna. Ustvarjalec predstavi način ustvarjanja in specifiko lastnega dokumentarnega in arhivskega gradiva, predstavniki pristojnega arhiva pa na podlagi teh informacij ustrezno svetujejo in izdajajo navodila za uspešen in po standardih arhivske stroke izveden prevzem. Vzorec se ponavlja, dokler niso vsa pomembna vprašanja razrešena, končni rezultat pa je prevzeto arhivsko gradivo v arhivskem depozitu.

S pojavom novih medijev, na katerih je zapisano gradivo, oz. s pojavom elektronskega arhivskega gradiva (dalje e-AG) pa sam proces sodelovanja med ustvarjalcem in pristojnim arhivom pridobiva povsem nove razsežnosti. Tako z vidika zahtevanih znanj, novih uporabljenih orodij in metod, kakor tudi z vidika pojavnosti novih izzivov, ki jih je potrebno ob tem ves čas razreševati. Na eni strani smo soočeni z neobstoječo prakso, ki tako sočasno nastaja. Po drugi strani pa s tem dobivamo priložnost, da nastale izzive razrešujemo na način, da je izpolnjen primarni cilj prevzemanja fizičnega in elektronskega gradiva – ohranitev arhivskega gradiva po ustreznih standardih in zakonodaji za naše zanamce.

V nadaljevanju tega članka bomo poskusili predstaviti način, izzive in metode, ki so se pokazali na primeru pilotnega prevzema zvočnih (avdio) posnetkov oz. zapisov, ki nastajajo pri delu Državnega zbora Republike Slovenija (dalje Državni zbor RS). Če lahko iz samega procesa izluščimo le en pomemben zaključek, je to ta, da je za uspešno izveden proces prevzema e-AG sedaj nujen interdisciplinarni pristop – poleg samih izkušenj in znanja s področja arhivistike je sedaj potrebno vključiti tudi znanja s področja informatike.

2. O ustvarjalcu in arhivskem gradivu

V Arhivu Republike Slovenije (dalje Arhiv RS) smo za nekdanjo Skupščino Slovenije (dalje Skupščina) in Državni zbor RS pripravili do sedaj že pet pisnih strokovnih navodil za odbiranje arhivskega gradiva iz dokumentarnega gradiva. V zadnjih dveh različicah je bila za sejno gradivo narejena temeljita analiza dokumentov, ki nastajajo znotraj sejnega dosjeja (sejne mape). Narejen je seznam vseh možnih vrst dokumentov, ki lahko nastanejo na sejah.

Vrste gradiva, ki nastajajo znotraj sejne mape, so se z leti spreminjale, ni pa vedno nujno, da se hranijo prav vsi naštetih dokumenti znotraj posamezne seje. V obdobju, za katerega je bil izveden testni primer, so nastajali sejni dosjeji rednih, izrednih in slavnostnih oz. žalnih sej, redne in nujne seje odborov in komisij ter redne in nujne seje preiskovalnih komisij. V sejni mapi redne seje lahko npr. nastane 24 različnih vrst dokumentov, med temi so tudi zvočni zapisi, ki so predmet testne izročitve in prevzema.

V poslovnikih nekdanje Skupščine in Državnega zbora RS imajo magnetofonski oz. zvočni zapisi¹ posebno nalogo iz čisto operativnega pomena, saj se »...o delu na seji državnega zbora vodijo dobesedni zapisi, magnetofonski zapis/zvočni zapis in magnetogram seje.«² Govornik ima najkasneje v treh dneh po svojem nastopu na seji državnega zbora pravico do redakcijskih popravkov v magnetogramu svojega govora. Popravki ne smejo spremeniti smisla in bistva njegovega izvajanja. V dvomu odloči o dopustnosti popravka predsednik državnega zbora. Popravke magnetograma govornik potrdi s podpisom.«³ Nato se iz magnetogramov sej v državnem zboru pripravijo še zapisi sej.

V Državnem zboru RS snemajo večino sej, razen tistih, ki so zaprte in se odvijajo v t. i. tihi sobi (npr. nekatere seje delovnega telesa, vezane na nadzor nad delom varnostnih in obveščevalnih služb ipd.). Zvočni zapisi, ki so s pisnimi strokovnimi navodili postali tudi kulturni spomenik, so nastajali na različnih medijih. V obdobju tik pred osamosvojitvijo Slovenije in vse do leta 2018 so bili zapisani na magnetne trakove (v analogni obliki). Zaradi posebnosti uporabe teh trakov jih v Arhivu RS nismo prevzemali, saj so imeli opremo za snemanje in predvajanje le v Skupščini oz. v Državnem zboru RS. Skupaj z Državnim zborom RS smo se o tem, da bi bilo smiselno magnetne trakove prepisati v digitalno obliko, začeli pogovarjati takoj po letu 2000, ko smo začeli s postopnim prevzemanjem arhivskega gradiva Skupščine zadnjih let pred osamosvojitvijo Slovenije. Zaradi razvoja tehnologije v zadnjih letih in izjemno drage opreme ter trakov je postalo snemanje na magnetne trakove za državni zbor preveliko finančno breme, zato so se po zgledu drugih parlamentov v Evropski skupnosti odločili za digitalno obliko zvočnih zapisov sej. Od leta 2018 dalje so tako zvočni zapisi le še v digitalni obliki. Zaradi vsega omenjenega hranijo zvočne zapise v dveh različnih oblikah in z dvema različnima načinoma evidentiranja oz. (od uvedbe digitalne oblike) zajema gradiva. Do leta 2008 so uporabljali za evidentiranje kartotečna kazala, od državnozborskih volitev 2008 dalje pa uporabljajo informacijski sistem za upravljanje dokumentarnega gradiva, t. i. UDIS. Dokler so snemali gradivo na magnetne trakove, ga niso evidentirali v informacijski sistem UDIS, od uvedbe digitalnih zvočnih zapisov pa ga evidentirajo in uvozijo v sistem. To so začeli izvajati tudi za nazaj, torej za obdobje pred letom 2018.

3. Zakaj so zvočni zapisi tako izjemno dragocen arhivski vir?

Zvočni zapisi omogočajo globlje spoznavanje določenega časa, stanja duha, načina komunikacije posameznih delegatov v času skupščine in nato poslancev v določenem mandatnem obdobju, možnost psiholoških raziskav itn. Zvočni zapisi lahko dopolnijo sliko določenih pomembnih osebnosti, za katere hranimo veliko arhivskega gradiva v pisni obliki, tako njihovega zasebnega gradiva kakor tudi tistega, ki je nastajalo pri različnih javnopravnih osebah. Spoznavamo lahko tudi manj znane delegate ali poslance. Spremljamo lahko razvoj slovenskega jezika, ki se je v zadnjih tridesetih letih močno spremenil. Zvočni zapisi so dragocen vir za popestritev arhivskih razstav, saj je za večino obiskovalcev spisovno gradivo nezanimivo in ne vzbudi velike pozornosti. Včasih se morda kdo vpraša, zakaj se ne bi glede avdio/video zapisov zadovoljili samo s tem, kar hranijo v izjemno dragocenem arhivu javne Radiotelevizije Slovenija. Odgovor

¹ V poslovnikih skupščine in državnega zbora se izraz magnetofonski zapisi uporablja do leta 2017, od tedaj dalje pa uporabljajo izraz zvočni zapisi.

² Magnetogram je „dobesedni zapis na magnetofonski trak posnetega pogovora, govora“. (Vir: Fran SSKJ).

³ Citat povzet iz poslovnikov državnega zbora.

je preprost. Zvočni zapisi, ki so nastali pri poslovanju v Državnem zboru RS, so avtentični zapisi celotne seje.

Kako zanimivo je lahko zvočno gradivo, si lahko ogledamo na spletnih straneh državnega arhiva⁴ v Nemčiji za obdobje združitve obeh Nemčij. Želja vsakega arhivista je, da bi kaj podobnega imeli tudi v Sloveniji za obdobje osamosvojitve.

4. Specifike zvočnih zapisov

Predhodno poglavje nam predstavi pomembnost ustvarjalca in gradiva, ki nastaja pri njegovem delovanju. Za razumevanje postopka priprave na prevzem arhivskega gradiva in posledično priprav dodatnih strokovno-tehničnih navodil pa je potrebno poznavanje okoliščin nastanka gradiva, zato je to poglavje namenjeno predstavitvi procesa njegovega nastajanja, njegovi specifičnosti in izzivom, ki nastajajo pri samem prevzemu gradiva v arhiv.

V Državnem zboru RS se plenarne seje in seje delovnih teles snemajo. Snemanje poteka v več dvoranah, kjer so nameščeni konferenčni mikrofonski sistemi kot vir avdio signala. Vsi ti avdio signali iz dvoran so povezani v tako imenovani avdio studio ali centralno tehnološko sobo, kjer jih snemajo z za to posebej nameščeno opremo. Kot že omenjeno, sta se do junija 2018 v okviru celotne izdelave sejnih zapisnikov v pisani obliki zapisovali tudi dve vrsti avdio oz. zvočnih zapisov. To so analogni arhivski zvočni zapisi na magnetih trakovih za hrambo in digitalni zapisi za izdelavo prepisov. Po juniju 2018 so zapisovanje na magnetne trakove za hrambo nadomestili z zapisovanjem na digitalne zvočne zapise, ki se v digitalni obliki tudi arhivirajo (Navodila za delo).

Digitalni zapisi za izdelavo prepisov so namenjeni temu, da se uporabijo kot podlaga za izdelavo tipkanih zapisov sej, za dobesedni zapis, imenovan tudi magnetogram, v fizično obliko. Snemajo se v formatu .mp3 v dolžinah po pet minut z medsebojnim prekrivanjem petih sekund. Digitalni snemalni sistem zapise zapisuje v datotečni sistem, kjer se sinhronizirajo v posebni aplikaciji za izdelavo prepisov. Oblika in dolžina zapisa sta prilagojeni za lažjo obdelavo prepisovalcev. Izkušeni prepisovalci tak zvočni zapis prepišejo in uredijo v približno 15–20 minutah. Metapodatki, potrebni za izdelavo prepisa, so zagotovljeni iz zalednih aplikacij in so: tip in zaporedna številka seje, prostor, govorec ter dolžina govora. Za izdelavo prepisov se uporabljata urejevalnik besedila in namenski predvajalnik zvoka. Urednik prepisov nato združi posamezne dele prepisa v integralno besedilo in ga objavi v aplikaciji. Istočasno se to pojavi tudi na spletu. Na prepisih sej se nato opravi še redakcija za potrebe tiskanih publikacij sejnih zapisov (Bojan Verbič, 2017).

Za izdelavo analognih arhivskih zvočnih zapisov za hrambo (torej zapisov na magnetofonske trakove) je Državni zbor RS uporabljal dvostezni magnetofon proizvajalca Studer/Revox – Revox B77. Hitrost snemanja na magnetofonske trakove je 4,75 cm/s, zapis pa je monofonski. Trak se ob koncu snemanja obrne tako, da se lahko posnameta obe strani, s čimer je zagotovljena racionalnost uporabe. Na posamezen trak je lahko zapisanih več različnih sej, zato se ob tem vodi evidenca, ki omogoča njihovo identifikacijo.

Po zgoraj opisanem postopku nastajajo magnetni trakovi, ki vsebujejo posnete govore sej plenarne seje in seje delovnih teles, ter zapisani tipkani prepisi teh sej. Le-ti se hranijo v arhivu Državnega zbora RS. Ocenjeno skupno število magnetnih trakov v letu 2017 se giblje med 7.000 in 7.500, kar skupaj predstavlja od 40.000 do 45.000 ur

⁴ Nem. Das Bundesarchiv. Dosegljivo na:
<https://www.bundesarchiv.de/DE/Navigation/Finden/Toene/toene.html>.

avdio vsebin (Bojan Verbič, 2017). Kot je že omenjeno, je potrebno vse to gradivo po Zakonu o varstvu dokumentarnega in arhivskega gradiva ter arhivih (dalje ZVDAGA) ter po Uredbi o varstvu arhivskega in dokumentarnega gradiva (dalje UVDAG) prevzeti in dolgoročno hraniti v Arhivu RS.

5. O razlogih za digitalizacijo zvočnih zapisov

Državni zbor RS nima ustreznih zmožnosti za hrambo magnetogramskih trakov pod pogoji, kot jih določa UVDAG. Sem sodi tako zagotavljanje zadostnih prostorskih, kadrovskih in tehničnih kot tudi s tem povezanih finančnih sredstev. Zaradi slabše dostopnosti in povišane cene magnetnih trakov, ki jih je potrebno redno prepisovati, je postal strošek vzdrževanja obstoječega načina snemanja nesmotrn in težko upravičljiv. Državni zbor RS je tako sprejel načrt opustitve snemanja na magnetne trakove in prehod na novejšie informacijske sisteme, ki bodo omogočali hrambo avdio zapisov zgolj v digitalni obliki. Poleg tega Državni zbor RS načrtuje tudi digitalizacijo trenutno že hranjenih magnetnih trakov. Te spremembe so sedaj mogoče tako zaradi novih tehnoloških možnosti kakor tudi zaradi pobud Evropske unije, ki članice (tudi finančno) vzpodbuja k izvajanju različnih projektov na področjih digitalizacije (ReNPK 14-17 2013).

Z vidika arhivske stroke je v zvezi z digitalizacijo zvočnih zapisov ključno, da je ciljna digitalna oblika ustrezna za dolgoročno hrambo. Zato je pomembna izbira ustreznega formata in pripadajočih metapodatkov. Pri digitalizaciji zapisov se je potrebno prav tako posvetiti ustrezni kakovosti pretvorbe in zagotavljanju revizijskih sledi, vse z namenom zagotavljanja temeljnih načel dolgoročne hrambe DUCAT.⁵ Kako in na kakšen način to dosegamo, mora biti opredeljeno v notranjih pravilih organizacije ob upoštevanju ustrezne zakonodaje (ZVDAGA, UVDAG) in Pravilnika o enotnih tehnoloških zahtevah za zajem in hrambo gradiva v digitalni obliki.

Vprašanje koristi in smotrnosti digitalizacije zvočnih zapisov v primeru Državnega zbora RS je večplastno. Najbolj očiten in enostaven razlog zanjo je pragmatičnost, ki jo s pretvorbo v digitalno obliko pridobimo. Uslužbenci Državnega zbora RS, ki so odgovorni za snemanje in zapis na magnetne trakove, so ocenili, da se uporabljeni trakovi za snemalnik Studer/Revox – Revox B77 pridobijo na tržišču čedalje težje. Skladno s tem naraščajo tudi njihove cene, s čimer nakup magnetnih trakov predstavlja vedno večji strošek. Poleg cene medijev zapisa se pojavlja tudi težava s samo fizično količino gradiva, ki ga je treba hraniti v posebej za to namenjenih prostorih. Prostor hrambe mora imeti po smernicah, zapisanih v Pogojih za izvajanje materialnega varstva arhivskega gradiva v depojih in na razstavah (Pogoji za izvajanje materialnega varstva), stalno temperaturo pod 15°C in 30- do 40-odstotno relativno zračno vlažnost v stalno zatemnjenem prostoru (ZVDAGA). Zagotovitev takšnega prostora za hrambo med 7.000 in 7.500 magnetofonskih trakov predstavlja ne samo velik finančni vložek, temveč tudi velik arhitekturni zalogaj za načrtovanje prostorov Državnega zbora RS. Ne smemo pozabiti, da so zvočni zapisi na magnetogramskih trakovih opredeljeni kot arhivsko gradivo, kar pomeni, da mora biti poskrbljeno mdr. tudi za njihovo dolgoročno hrambo. V to je zajeto tako fizično zagotavljanje prostora kakor tudi hramba trakov v smislu stalnega nadzora, saj so magnetni trakovi po svoji naravi podvrženi naravni razgradnji. Skladno s tem je potrebno skrbeti za redno pregledovanje in presnemavanje trakov, kar zahteva svoj stroškovni in kadrovski davek.

⁵ Dostopnost, uporabnost, celovitost, avtentičnost in trajnost.

Kljub temu da predstavlja digitalizacija poslovanja oz. projekt prehoda na izključno digitalno snemanje in hrambo visok začetni strošek nakupa nove opreme, je ta rešitev dolgoročno finančno, kadrovsko in prostorsko smotrnejša.

Pri pragmatičnosti moramo omeniti tudi dostopnost in uporabo arhivskega gradiva. Trenutno moramo v primeru, da želi uporabnik pregledati določen posnetek, za tega zaprositi. Nato je treba v arhivu fizično kopijo poiskati in jo dati uporabniku v uporabo v prostorih arhiva oz. institucije, na za to namenjeni tehnični opremi. V primeru digitaliziranih objektov pa se lahko hranijo posnetki na ustrezno varovanih strežnikih, kjer se jih z ustreznimi informacijskimi oz. spletnimi rešitvami poišče v bazi podatkov (preko ustreznih metapodatkov) in nato tudi lažje pregleda. Pri magnetnih trakovih je treba posnetek previjati, da poiščemo določen segment zapisa seje, pri digitalni različici pa le klikamo po ustrezni kontrolni plošči. Prav tako bo lahko isti posnetek pregledovalo več oseb hkrati.

Dolgoročna e-hramba s tem odpravi številne pomanjkljivosti klasičnega arhiviranja, nudi pa tudi mnogo dodatnih prednosti, ki jih lepo povzame Hajtnik, ko pravi, da e-hramba omogoča »...preprostejše iskanje zelenih vsebin po raznovrstnih atributih dokumentov, prostorsko so neomejeni, dostopni časi do vsebin pa so v primerjavi s klasičnimi arhivi zanemarljivi. In kar je še pomembno, omogočene so učinkovitejše kontrole nad uporabo gradiva (beleži se lahko vsak vpogled, sprememba ali izpis dokumenta, kar zagotavlja revizijsko sledljivost skozi ves čas hrambe)«. Omeni pa še druge prednosti: »Z uvedbo e-hrambe boste dosegli večjo dostopnost gradiva (omogočen hkraten dostop do istega gradiva in to z oddaljenih mest), večjo preglednost gradiva (učinkovitejše iskanje dokumentov) in zagotovili njegovo večjo varnost (več kopij gradiva na geografsko oddaljenih lokacijah, večjo varnost in preglednost dostopov).« (Hajtnik, 2011, str. 493–494).

Da bi prejšnje trditve v kontekst postavili, v nadaljevanju citiramo okrožnico Oddelka za parlamentarne službe Avstralije, ki deluje v okviru avstralskega parlamenta. Citat se nanaša na izdajo njihovih prepisov debat v parlamentu in se glasi:

»In this age of instant information, stakeholders expect faster access to the Hansard record. Given the availability of the audio-visual record of parliamentary proceedings, the Department of Parliamentary Services considers it timely to review the use, relevance and appropriate format for the Hansard record.« (Kenny, 2010).

V citatu je izpostavljeno, da je treba zaradi duha časa, ko uporabniki zahtevajo hitrejši dostop do posnetkov, dobro premisliti o pomembnosti in primernih formatih izdaje njihovega gradiva. Nadaljnja vsebina okrožnice nakazuje na vedno večjo pomembnost zvočnih posnetkov v primerjavi s preteklostjo, ko so imeli glavno vlogo tipkani prepisi.

Naslednji vidik smotrnosti digitalizacije je kulturna vrednost gradiva. Merila, ki določajo, ali je gradivo smiselno označiti kot arhivsko, so navedena v 41. členu ZVDAGA. Zakaj so zvočni zapisi kot arhivski vir pomembni, je opredeljeno že uvodoma. V tem primeru lahko vidimo, da magnetni trakovi, poleg kulturne vrednosti za zgodovino in kulturo, ustrezajo tudi drugim kriterijem. To so predvsem izvirnost podatkov in informacij, pomen gradiva z vidika kulturne raznolikosti, pomen avtorja, specifičnost kraja, pomen javnopravne osebe in ne nazadnje notranje in zunanje značilnosti gradiva.

Posebej pa bi tudi izpostavili sam pomen zvoka. Zvočni zapisi, poleg same vsebine pogovorov, slikajo tudi način izražanja in ostale značilnosti sporočanja, ki se v tiskani besedi izgubijo. Za primer lahko vzamemo govor *»I have a Dream«* aktivista in borca za človekove pravice Martina Luthra (MLA 1963). Nekaj je brati govor, drugo je poslušati zvočne zapise, s katerimi se prenesejo ne samo sporočilo, ampak tudi dramatičnost in teža govora, pa tudi ostale karakteristike kot so barva glasu, hitrost govora itn. To je

dodaten razlog, da se poleg tipkanih oblik vsebin pogovorov ohranijo tudi njihovi zvočni posnetki.

Celota in prepletenost vseh naštetih razlogov nam jasno kaže, da je digitalizacija poslovanja – oz. ožje: zvočnih zapisov – več kot upravičen proces. Pilotni prevzem takšnih digitiziranih zvočnih zapisov pa pomeni bogat prispevek k ohranjanju in dopolnitvi kulturne dediščine Slovenije in slovenskega naroda ter odličen primer poskusa ohranjanja »duha časa« iz obdobja s konca 20. stoletja.

6. Izvedba pilotnega prevzem zvočnih zapisov

Delež gradiva pri ustvarjalcih, ki nastaja v digitalni obliki, in s tem tudi delež e-AG, se eksponentno povečuje, zato je sposobnost arhivov za prevzemanje digitalnih vsebin izjemno pomembna. Arhivsko gradivo, ki nastaja pri poslovanju organizacije, je lahko že v svoji izvorni obliki v digitalni obliki, lahko pa gradivo po svojem izvoru nastane v fizični ali analogni obliki in se za potrebe e-hrambe digitalizira, kot to velja za naš pilotni testni primer zvočnih zapisov sej Državnega zbora RS. Kadar je arhivsko gradivo v digitalni obliki, arhiv v skladu z 21. členom UVDAG, kot del pisnega strokovnega navodila, ob začetku odbiranja izda še strokovno-tehnično navodilo (dalje STN).

Za pilotni prevzem zvočnih zapisov sej Državnega zbora RS, ki so digitalizirane, je bilo potrebno pripraviti pisno strokovno navodilo, ki vsebujejo tudi STN ob izročitvi.⁶ Temeljni namen STN ob izročitvi gradiva v arhiv je opredeliti sprejemni informacijski paket (SIP). Pristojni arhiv izda STN ustvarjalcu za vsako posamezno vrsto e-AG,⁷ te se med seboj razlikujejo po strukturi zapisov, produkcijskem okolju in zahtevnosti. STN določajo vsebino in strukturo SIP s tehničnega vidika, za pomoč pri analizi gradiva v digitalni obliki poslovanja smo uporabili odgovore, ki jih je ustvarjalec podal v predpripravljenem vprašalniku (R_KC1.1.3_pSTN_Vprašalnik_zvočni zapisi).

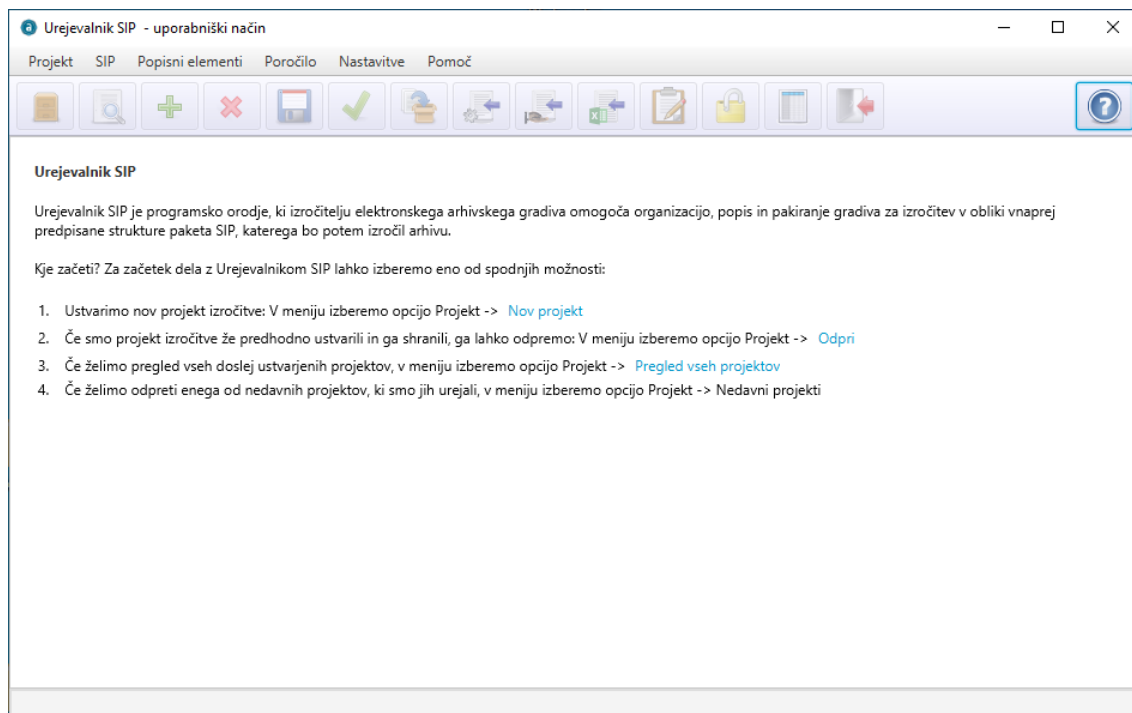
V preteklosti je bilo prevzemanje e-AG zahtevno in kompleksno ter je zahtevalo veliko koordinacije, usklajevanj in preverjanj. V sklopu projekta Slovenskega elektronskega arhiva e-ARH.si za obdobje 2016–2020 (dalje e-ARH.si) smo postopek poenostavili in standardizirali z uporabo v ta namen razvitega programskega orodja, ki smo ga nadgradili v verziji 2.0. Z Urejevalnikom sprejemnega informacijskega paketa (dalje Urejevalnik SIP), smo delno digitalizirali postopek izročitve arhivskih vsebin v digitalni obliki. Orodje je na voljo brezplačno⁸ in s polno podporo pristojnega arhiva. V prvi vrsti je namenjeno ustvarjalcem e-AG in je enostavno za uporabo, ko se z njim nekoliko spoznamo. Ustvarjalcem e-AG omogoča zapakirati gradivo, da ga lahko predajo v pristojni arhiv, kjer se izvede uvoz v elektronski arhiv.

⁶ *Obstaja tudi Strokovno-tehnično navodilo za varstvo gradiva do izročitve v pristojni arhiv, s katerim se določi vsebinska in tehnična urejenost gradiva, metapodatki, dovoljene oblike zapisov (formati), spremljajoča dokumentacija, pogostost oblikovanja sprejemnih informacijskih paketov in njihove izročitve pristojnemu arhivu, uporaba certificirane opreme, obvladovanje gradiva z notranjimi pravili in s tem skladnost s Pravilnikom o enotnih tehnoloških zahtevah za zajem in hrambo gradiva v digitalni obliki.*

⁷ *Vrste e-AG glede na strukturo zapisa in okolje, v katerem gradivo nastane oz. se vzdržuje, so: nestrukturirano gradivo (datotečni sistemi; sem sodijo še elektronska pošta, spletne strani in družbena omrežja), podatkovne zbirke, dokumentni sistemi.*

⁸ *V kolikor ste ustvarjalec arhivskega gradiva po ZVDAGA in bi potrebovali programsko orodje za pripravo SIP za izročitev v pristojni arhiv, se obrnite na vašega pristojnega arhivista ali posredujte zahtevo na elektronski naslov glavne pisarne Arhiva Republike Slovenije. Pridobili boste ustrezna navodila, kako do orodja, kaj morate pred tem urediti vi in kaj bo pred uporabo orodja za vas uredil pristojni arhiv (Novičnik projekta e-ARH.SI 2019).*

Urejevalnik SIP omogoča gradnjo SIP v dveh načinih delovanja – naprednem in uporabniškem. Napredni način je namenjen arhivistom, ki določijo strukturo paketa SIP s kreiranjem sheme profila. Shema definira dovoljene formate gradiva, datotečno strukturo, omeji število datotek v mapi ipd. Na drugi strani uporabniški način ustvarjalcem omogoča popisovanje gradiva, polnjenje sheme profila z dejanskim gradivom in ustvari SIP za prenos gradiva v elektronski arhiv. V Urejevalnik SIP je vgrajena tudi validacija sprejemnega informacijskega paketa. Poleg parametrov, ki jih v shemi določi arhivist, orodje preverja tudi ustreznost formatov in z algoritmi izračuna bitne vsote, ki služijo dokazovanju celovitosti in avtentičnosti gradiva.



Slika 1: Vstopno okno aplikacije Urejevalnik SIP v uporabniškem načinu.

7. Opredelitev SIP in njegove strukture ter izziv velikih datotek zvočnih zapisov

Prevzem e-AG je ena izmed pomembnejših prednostnih nalog v okviru projekta e-ARH.si. Na podlagi izkušenj iz preteklih prevzemov e-AG in povratnih informacij s strani ustvarjalcev javnega arhivskega gradiva smo ob koncu leta 2019 uspešno zaključili z uradnim prevzemom aplikacije Urejevalnik SIP, ki omogoča pripravo SIP za izročanje e-AG. Vpeljane izboljšave in novosti, ki zagotavljajo enostavno izročanje digitalni arhivskih vsebin, smo preizkusili s piloti na vseh vrstah arhivskega gradiva.

V pilotnih prevzemih smo izpeljali celoten proces prenosa e-AG od ustvarjalca v arhiv. Začetnemu uvodnemu sestanku z ustvarjalcem je na strani ustvarjalca sledilo odbiranje arhivskega gradiva na podlagi veljavnega pisnega strokovnega navodila, definiranje SIP s pomočjo odgovorov ustvarjalca iz vprašalnika, priprava in izdaja STN ter Urejevalnika SIP skupaj s shemo profila s strani arhiva, priprava gradiva v skladu s STN pri ustvarjalcu, gradnja in validacija SIP, poskusni prenos v arhiv, testiranje in

validacija SIP, izročitev gradiva, prenos v elektronski arhiv ter predaja Zapisnika o izročitvi in prevzemu e-AG.

V okviru pilotnega prevzema smo se osredotočili na digitalizirane zvočne zapise sej, ki so izvoženi iz informacijskega sistema UDIS, in sicer za obdobje januar 2017. Digitalizirano gradivo je v datotekah v formatu .wav. Posamezno sejo sestavlja več datotek .wav, njihovo število pa je odvisno od prekinitev sej. Datoteke posamezne seje imajo enako klasifikacijsko številko. Glede na definicije, ki jih uporabljamo v okviru arhiva, sodijo datoteke sej glede v kategorijo nestrukturiranega gradiva,⁹ po vsebini¹⁰ pa med zvočne zapise.

Ob pregledu digitaliziranih zvočnih zapisov smo ugotovili, da se skupna velikost datotek posamezne seje giblje med 100 MB in 3,5 GB, vendar so pogoste datoteke, katerih velikost je bližje zgornji velikostni meji. Pri prenosu gradiva v repozitorij je izjemnega pomena velikost SIP. V Sloveniji za upravljanje zapisov in e-hrambo uporabljamo aplikacijo scopeArchiv¹¹, ki jo je razvilo švicarsko podjetje ScopeSolutions. Po specifikaciji švicarskega zveznega arhiva SFA, različica 4.0 (2015), je največja priporočena velikost za SIP 8 GB, vendar je zaradi hitrejšega prenosa podatkov preko modula scopeIngest¹² v repozitorij priporočena velikost SIP 2 GB. Podobne tehnične omejitve veljajo tudi za dostopanje in ogledovanje gradiva preko Virtualne arhivske čitalnice¹³ (VAČ). Podobne rezultate smo dobili tudi pri našem testiranju raznolikih zmogljivosti (SIP z veliko datoteko, SIP z velikim številom datotek, ipd.). Kadar so velikosti SIP večje ali gre za več SIP, velja usmeritev, da se je potrebno o prenosu SIP posvetovati in uskladiti z Arhivom RS.

V Državnem zboru RS potekajo seje dva- do šestkrat mesečno, to so v povprečju 4 seje na mesec, letno povprečje pa lahko zaokrožimo na 50 sej. V primeru prevzema celotnega gradiva, kjer je povprečna seja zvočnih zapisov velika npr. 2 GB, je to lahko SIP, ki obsega 100 GB (50 sej velikosti 2GB). Zaradi velikostnih izzivov pri prenosu gradiva, uporabi modula scopeIngest ter ogledovanju gradiva v VAČ smo se v pilotu odločili, da bo posamezen SIP hranil posamezno sejo.

Za gradnjo sheme profila mora biti izpolnjen predpogoj poznavanja hierarhične strukture SIP, map in datotek. V shemi opredeli arhivist dovoljene formate gradiva, datotečno strukturo, število datotek v mapah ipd., na strani ustvarjalca pa shema med drugim omogoča popis gradiva (avtomatski in ročni), vnos gradiva ter kreiranje SIP za prenos v repozitorij. Glede na prejeto gradivo ter metapodatke na nivoju sej smo se odločili, da bomo shemo profila definirali na sledeči način:

- 1. nivo: SIP, datoteka z metapodatki .md-xml,
- 2. nivo: seja z metapodatki, spremljevalna dokumentacija,

⁹ Nestrukturirano gradivo je gradivo, ki se pri prenosu v arhiv nahaja v datotečnem sistemu (mape, podmape, datoteke). Laično razumevanje pojma, da gre za gradivo, ki ni strukturirano, je zavajajoče.

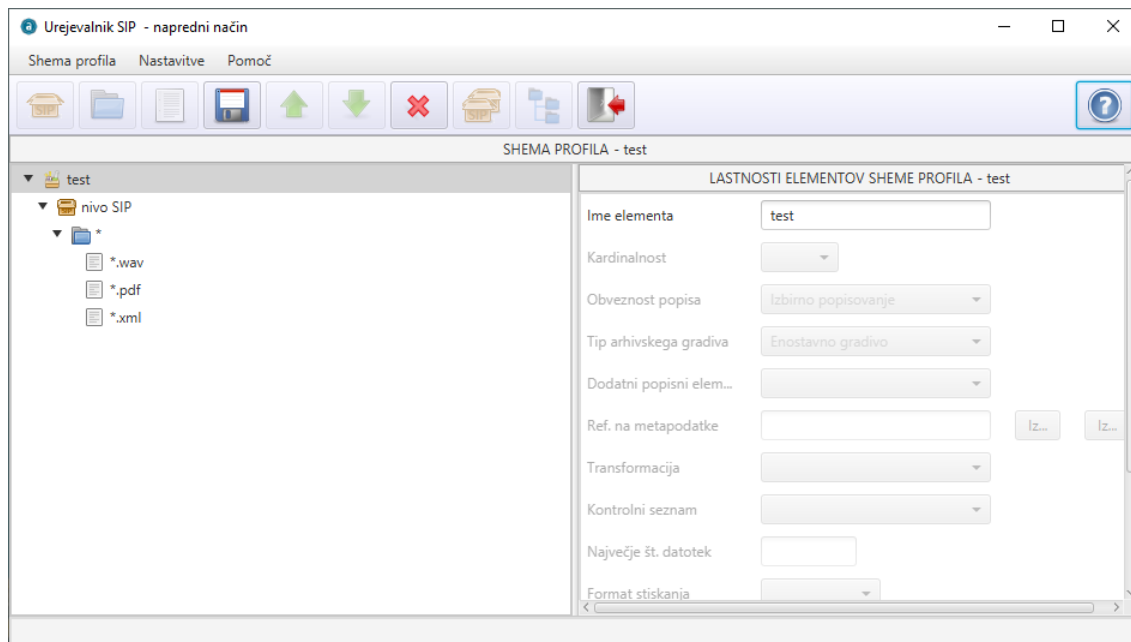
¹⁰ Vrste e-AG po vsebini delimo na besedilne in mešane zapise, slikovne zapise, zvočne zapise, avdiovizualne zapise, prostorske podatke in tehnične risbe.

¹¹ ScopeArchiv je aplikacija, ki omogoča upravljanje z zapisi in digitalno hrambo ter zajema postopke arhivskega dela od vnosov do uporabe za javne in zasebne arhive. Aplikacija omogoča prenos digitalnih podatkov v arhiv skladno z OAIS in ISO 20652. Programska oprema sledi mednarodnim standardom ISAD (G), ISAAR (CPF), ISDIAH, ISDF in standardom metapodatkov EAD, Dublin Core in PREMIS.

¹² Modul scopeIngest je vmesnik med popisi zapisov in repozitorijem Fedora Commons.

¹³ Virtualna arhivska čitalnica (VAČ), razvita v okviru projekta e-ARH.si, je spletna storitev, ki uporabnikom omogoča oddaljen dostop do arhivskega gradiva in zajema tudi digitalizacijo arhivskih čitalniških postopkov.

- 3. nivo: vse datoteke posamezne seje.



Slika 2: Shema profila v naprednem načinu Urejevalnika SIP prikazuje definiranje nivojev.

8. Priprava metapodatkov za avtomatski uvoz v SIP

Ustvarjalec je pripravil metapodatkovno datoteko za vsako sejo v .xml formatu. Urejevalnik SIP omogoča več načinov dodajanja metapodatkov določenim mapam, datotekam znotraj SIP. Prav tako omogoča dodajanja metapodatkov na nivoju SIP. Najbolj zamudno je ročno vpisovanje metapodatkov skozi grafični vmesnik Urejevalnika SIP. Za avtomatiziran proces uvoza pripravljenih metapodatkov lahko uporabimo opcijo, kjer pravilno uvozimo izvirne datoteke .xml s pomočjo translacijske datoteke .xslt¹⁴. Za ta način je potrebno znanje XSLT, tudi ustrezne označbe polj, ki jih zahteva Urejevalnik SIP. Ta so definirana v katalogu metapodatkov.

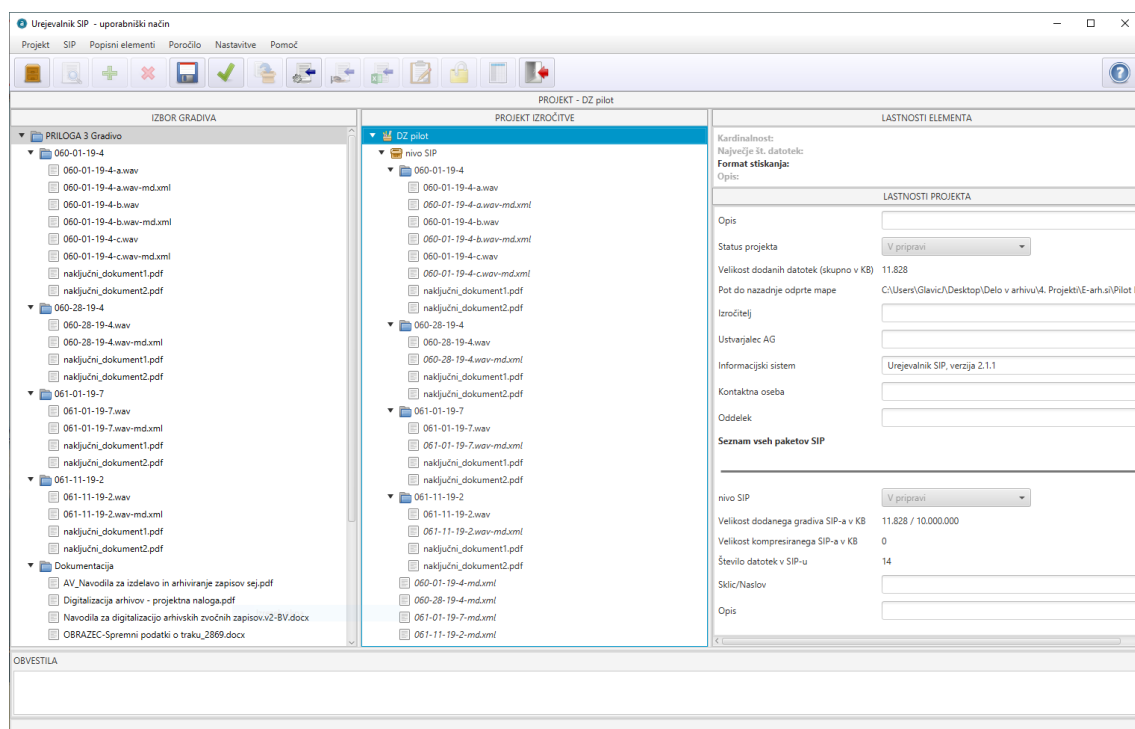
Na voljo pa je tudi opcija, kjer s pomočjo programskega jezika Python pretvorimo izvirno .xml v tekstovno obliko, nato pa le-to pretvorimo v ustrezno obliko .md-xml, ki jo zahteva Urejevalnik SIP. V našem primeru je omenjeno pretvorbo uredil informatik Arhiva RS mag. Boris Domajnko. Postopek je potekal tako, da smo mu pristojni arhivisti pripravili ustrezno konkordančno tabelo, z originalnimi metapodatki v enem stolpcu in informacijo, v katero polje se preslikajo metapodatki v grafičnem vmesniku Urejevalnika SIP. S pomočjo pripravljenega programa Python je potekala izdelava ustreznih datotek md-xml, poimenovanih z imenom map sej. Te so se nato v Urejevalniku SIP dodale na nivo SIP. Zaradi ohranjanja avtentičnosti smo originalne nespremenjene podatke vključili v SIP za posamezno sejo.

¹⁴ Ang. Extensible Stylesheet Language Transformations, tj. jezik za preoblikovanje dokumentov .xml v obliko, ki jo želimo.

9. Funkcionalnost multiSIP kot rešitev za večje število strukturno enakih SIP

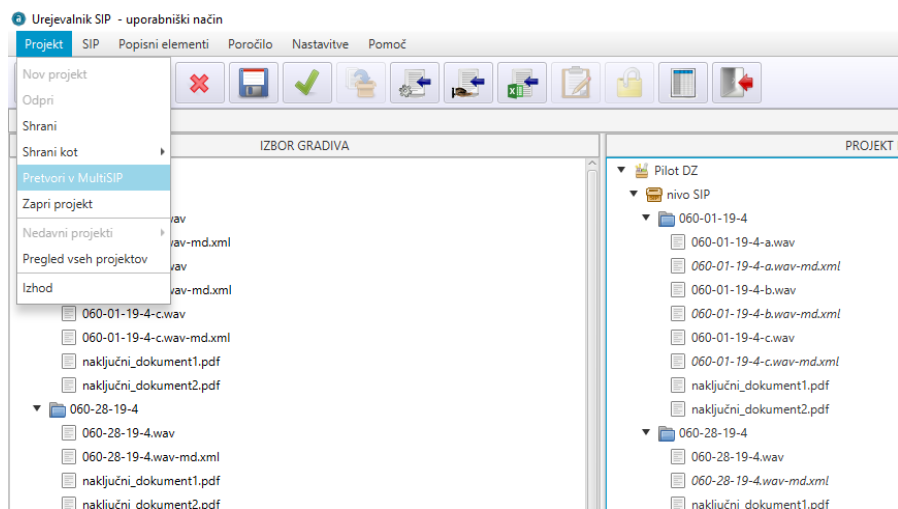
V času izvajanja pilotnih prevzemov različnih vrst e-AG v okviru projekta e-ARH.si smo osvojili mnogo novih spoznanj, ki jih v specifikaciji za razvoj Urejevalnika SIP 2.0 ni bilo mogoče predvideti. Arhivsko gradivo je pri ustvarjalcih v različnih strukturah in nivojih urejenosti. Kot omenjeno, smo na pilotu prevzema zvočnih zapisov sej Državnega zbora RS ugotovili, da je najsmotrneje prevzeti posamezno sejo v posameznem sprejemnem informacijskem paketu. Letno povprečje števila sej je približno 50, kar pomeni, da je potrebno prevzeti 500 sej za obdobje desetih let oziroma da gre za prevzem 500 SIP. Za definiranje vsakega SIP je potrebno kreirati shemo profila. V našem primeru smo želeli orodje nadgraditi s funkcionalnostjo, ki omogoča kreiranje poljubnega števila shem profilov za prevzem poljubnega števila sej, kar postane precej zamudno, če je potrebno celoten proces kreiranja sheme profila izvesti večkrat. Rešitev je v razbitju gradiva na več enakih paketov z enako strukturo. Funkcionalnost, poimenovana multiSIP, omogoča v primeru večjega števila map arhivskega gradiva razbitje enega SIP, ki ga lahko kreiramo v orodju, v več SIP z enakimi nastavitvami glede na število map na krovnem nivoju pri ustvarjalcu. Na podlagi te in nekaterih drugih željenih funkcionalnosti je bila izdelana specifikacija potrebnih prilagoditev in nadgradnje programskega orodja Urejevalnik SIP 2.0.

Za boljše razumevanje si oglejmo del procesa pretvorbe v več SIP. V podoknu Izbor gradiva (glej sliko 3) je prikazano gradivo v datotečnem sistemu pri ustvarjalcu. V podoknu Izbor gradiva je gradivo preneseno v pripravljeno shemo profila kot več map znotraj enega SIP.



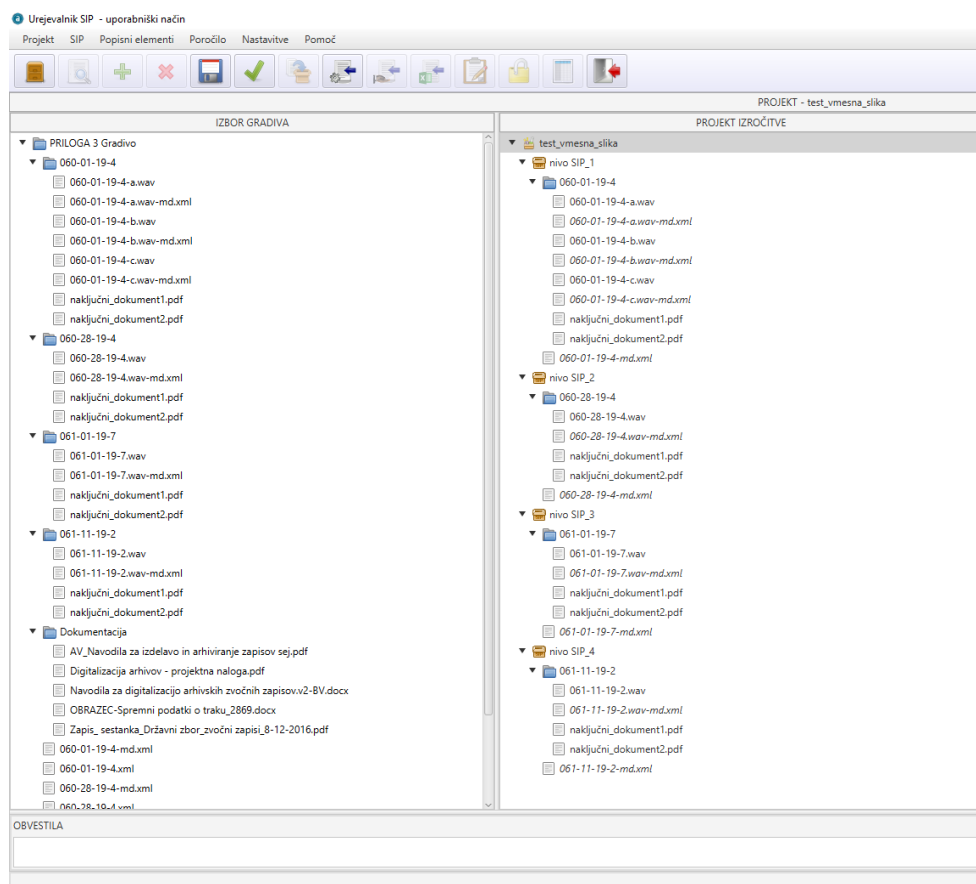
Slika 3: Gradivo na datotečnem sistemu pri ustvarjalcu (podokno IZBOR GRADIVA) in gradivo, preneseno v shemo profila (podokno PROJEKT IZROČITVE) pred pretvorbo v več SIP.

Z izbiro možnosti »*Pretvori v MultiSIP*« ustvarjalec sproži pretvorbo map znotraj enega SIP v več SIP z enako strukturo.



Slika 4: Opcija multiSIP v uporabniškem načinu Urejevalnika SIP.

Z izbiro funkcionalnosti multiSIP se vsaka mapa spremeni v SIP, ki ima enako strukturo.



Slika 5: Rezultat pretvorbe multiSIP so posamezni strukturno enaki SIP.

Po avtomatskem uvozu metapodatkov iz vključenih datotek .md-xml in validaciji pripravljenega gradiva je sledila izdelava SIP.

Končni rezultat opisanega postopka so bili izdelani paketi, s hash datoteko in avtomatsko generiranimi poročili, ki se jih nato preda v pristojni arhiv, tam pa se v primeru, da je SIP ustrezno pripravljen, izvede uvoz preko scopeIngesta v repozitorij.

10. Zaključek

Kot rezultat uspešnega pilotnega prevzema zvočnih zapisov Državnega zbora RS, torej oblikovanja SIP in prenosa SIP v repozitorij, smo vzporedno tudi dopolnjevali interne predloge dokumentov: R_KC1.1.3_pSTN_Vprašalnik_zvočni zapisi, v katerem smo skupaj z ustvarjalcem odgovorili na (tehnične, zakonodajne, arhivske strokovne itn.) okoliščine nastanjanja gradiva, ki je predmet prevzema; KC1.1.5 Načrt_prevzema, v katerem smo definirali in začrtali način poteka sodelovanja med ustvarjalcem in pristojnim arhivom; plod sodelovanja pa se najbolj odraža v dokumentu R_KC1.1.8.2.3 Dodatno strokovno-tehnično navodilo do izročitve. Tu so podrobno zapisani sam postopek prevzema, natančno definirana vsebina prevzema, definirana struktura SIP in metapodatkov.

Opisani postopek prevzema gradiva v arhiv, skupaj z izzivi in problematiko, ki se pri njem pojavijo, je le en primer izmed mnogih. Pri vsakem primeru prevzema elektronskega gradiva se soočamo z različnimi spremenljivkami, od različne vrste gradiva in različnih formatov do urejenosti gradiva, oblik zapisov metapodatkov itd. Opis primera pred vami nakaže eno izmed možnih rešitev za specifičen primer gradiva. To so nestrukturirani zvočni posnetki. Kot smo opisali, je izziv takega gradiva sama količina gradiva in velikost posamezne datoteke. S podobnimi primeri gradiva se srečujemo tudi drugod, na primer pri posnetkih sej mestnih občin. Tudi tam nastaja istovrstno gradivo s podobnimi značilnostmi. Tako naj opisan primer služi kot eden izmed načinov spopadanja z izzivi, ki jih prinese tako gradivo.

11. VIRI IN LITERATURA

- Das Bundesarchiv: Originaltöne zur Wiedervereinigung.** Pridobljeno 29. 7. 2021 s spletne strani: <http://wiedervereinigung.bundesarchiv.de/tonaufnahmen.html>.
- Digitalizacija arhivskih zvočnih zapisov sej Državnega zbora in njegovih delovnih teles ter dogodkov.** Interni dokument Državnega zbora Republike Slovenije.
- Fran SSKJ: Slovar slovenskega knjižnega jezika.** Pridobljeno 29. 7. 2021 s spletne strani: www.fran.si.
- Hajtnik, T. (2011).** Vse, kar mora vodstvo organizacije vedeti o e-hrambi. *Tehnični in vsebinski problemi klasičnega in elektronskega arhiviranja*, 10. zbornik referatov dopolnilnega izobraževanja s področja arhivistike, dokumentalistke in informatike.
- KC1.1.3_pSTN_Vprašalnik_zvočni zapisi. (2020).** Interni dokument projekta e-ARH.si.
- KC1.1.5 Načrt_prevzema. (2020).** Interni dokument projekta e-ARH.si.
- KC1.1.8.2.3 dSTN. (2021).** Interni dokument projekta e-ARH.si.
- Kenny, D. (2010).** *Hansard as a record in the 21st century*. Pridobljeno 14. 7. 2021 s spletne strani: https://www.aph.gov.au/binaries/visitors/circulars/2010/infocirc_28_10.pdf.
- King, M. Luther. (1963).** *Martin Luther King delivers the "I have a dream" speech from the podium at the March on Washington Bob Adelman*. Pridobljeno 23. 7. 2021 s spletne strani: <https://www.loc.gov/item/2013645765/>.
- MLA.** *March on Washington for Jobs and Freedom*. Pridobljeno 21. 1. 2019 s spletne strani: http://openvault.wgbh.org/catalog/A_76C3B93B557D4976A032C27C72ACED18#at_89.0_0_s.
- Navodila za delo, Digitalizacija arhivskih zvočnih zapisov sej Državnega zbora in njegovih delovnih teles ter dogodkov. (2020).** Interni dokument Oddelka za razvoj informacijskega sistema Državnega zbora Slovenije.
- Novičnik projekta e-ARH.si (2019).** Pridobljeno 15. 7. 2021 s spletne strani: <https://www.gov.si/novice/2020-03-27-resitve-za-ustvarjalce-arhivskega-gradiva/>.
- Pravilnik o enotnih tehnoloških zahtevah za zajem in hrambo gradiva v digitalni obliki. (2020).** Uradni list RS, št. 118/2020.
- ReNPK14–17 (2013):** Resolucija o nacionalnem programu za kulturo 2014–2017. Pridobljeno 29. 7. 2021 s spletne strani: <http://www.pisrs.si/Pis.web/pregledPredpisa?id=RESO96>.
- Uredba o varstvu dokumentarnega in arhivskega gradiva (2017).** Uradni list RS, št. 42/17.
- Verbič, B (2017).** *Digitalizacija avdio zapisov Državnega zbora*. Projektna naloga. Ljubljana.
- Zakon o varstvu dokumentarnega in arhivskega gradiva ter arhivih, Priloga 1. (2014).** Uradni list RS, št. 30/06, 24/14 – Odl. US in 51/14.

SUMMARY

INGEST PROCEDURE AND THE CHALLENGES OF CREATING SUBMISSION INFORMATION PACKAGES (SIP) ON THE CASE OF AUDIO RECORDS OF THE NATIONAL ASSEMBLY OF THE REPUBLIC OF SLOVENIA

Jože Glavič

Historical Archives Ljubljana, Slovenia
joze.glavic1@guest.arnes.si

Vesna Gotovina

Archives of the Republic of Slovenia,
Slovenia
vesna.gotovina@gov.si

Klavdija Krivec

Archives of the Republic of Slovenia,
Slovenia
klavdija.krivec@gov.si

Žiga Koncilija, Ph. D.

Archives of the Republic of Slovenia,
Slovenia
ziga.koncilija@gov.si

Tatjana Hajtnik, Ph. D.

Archives of the Republic of Slovenia, Slovenia
tatjana.hajtnik@gov.si

The long-standing practice of physical records acquisition has shown that cooperation between the competent archival institution and the creator of archival records is a circular process in which the exchange of information and knowledge is necessary. If we can draw only one important conclusion from the process of digital records acquisition, it is that an interdisciplinary approach is now necessary for a successful takeover process — in addition to the experience and knowledge from the field of archiving, the knowledge from the field of informatics must be included. Ingest of digital records is one of the most important tasks within the project of establishing the Slovenian Public Electronic Archives e-ARH.si (2016-2021). Based on experience from the past, the pilot ingest of digital records and feedback from the creators of public archival records, we have successfully completed the process of developing documentation, methodology and application (SIP 2.0 Editor). The application enables the preparation of the Submission Information Package (SIP) for the ingest of digital records.

Within the e-ARH.si project, several pilot ingests were carried out encompassing the entire process of transferring digital records from the creator to the electronic archives. The initial introductory meeting with the creator of archival records was followed by: the selection of archival records on the basis of valid written-professional instructions, the definition of SIP structure, preparation and issuance of professional-technical instructions, provision of SIP Editor and the profile scheme issued by the competent archival institution, preparation of archival records in accordance with professional-technical instructions at the creator of archival records, construction and validation of SIP, experimental transfer to the repository, testing and validation of the SIP, delivery of records, final transfer to the repository and handing over the Ingest Minutes.

The focus of this pilot transfer were digitized sound recordings of the Slovenian National Assembly (January 2017) sessions/meetings, which were exported from the UDIS information system. Each session consists of several .wav files and their number depends on the number of session interruptions. According to the definitions we use in

the electronic archives, session files fall into the category of unstructured records, and in terms of content into sound records. When reviewing the audio recordings, we found out that the total file size of an individual session ranges between 100 MB and 3.5 GB, more common being those whose size is closer to the upper size limit. In Slovenia, we use the scopeArchiv application for professional processing of archival records, where the maximum recommended size for an individual SIP is 8 GB, but due to optimal data transfer via the scopeIngest module to the repository, the recommended SIP size is 2 GB. The same restriction applies to accessing and viewing archival records through the Virtual Archival Reading Room. We also encountered similar technical limitations in cases where we have a SIP with a very large number of files. When the size of the SIP is larger or there are several SIPs, it is necessary to consult the competent archival institution about the transfer.

As a result of the successful pilot ingest of audio recordings of the Slovenian National Assembly, the formation of the SIP and its transfer to the repository, internal document templates were adjusted. A questionnaire in which we, together with the creator of archival records, answered the (technical, legislative, archival-professional, etc.) circumstances of how records are created. In the ingest plan, we defined and outlined the manner of cooperation between the creator of the archival records and the competent archival institution. The final product of cooperation is represented in the document Additional professional and technical instructions for sound recordings of the Slovenian National Assembly. This document details the ingest process itself, the precise definition of the ingest content, the structure of the SIP package and metadata definition.