

MEDIATEKA – DIGITALIZACIJA IN TRAJNO HRANJENJE ZVOKOVNIH VSEBIN RTV SLOVENIJA

Bojan Kosi
Martin Žvelc

Oddano: 26. 8. 2010 – Sprejeto: 28. 1. 2011

Strokovni članek
UDK 621.396:004:005.921.1"746"(497.4)

Izveleček

RTV Slovenija hrani velike količine zvokovnih in slikovnih gradiv, različnih dokumentov in notnega gradiva, večinoma še v analogni obliki zapisa. Vsesplošna digitalizacija je močno spremenila postopke in načine ustvarjanja vsebin, ki se v digitalni obliki zapisa zapisujejo na različne medije. Takšni zapisi zahtevajo tudi nove načine hranjenja. V prispevku sta predstavljena razvoj in struktura Mediateke, oddelka RTV Slovenije in modela hranjenja zvokovnih vsebin, saj je bilo to področje zaradi hitrih tehnoloških sprememb v najbolj kritičnem stanju. Intenzivno delo v Mediateki je steklo leta 2008 in po dveh letih uspešnega dela se Radio Slovenija lahko pohvali s sodobnim sistemom za trajno hranjenje zvokovnih vsebin. Digitalni arhiv Radia Slovenija ustreza vsem standardom in predpisom, ki veljajo za sodobne arhivske sisteme. V članku je opisano tudi upravljanje z zvokovnimi vsebinami, ki poteka s pomočjo programske opreme Mediarc, in ga prav tako lahko uporabimo pri realizaciji trajnega digitalnega video arhiva TV Slovenija.

Ključne besede: RTV Slovenija, Radio Slovenija, Mediateka, digitalizacija, digitalni arhivi, zvokovne vsebine

KOSI, Bojan; MARTIN ŽVELC. Mediatheque – digitization and preservation of audio content in RTV Slovenia. Knjižnica, 55(2011)1, pp. 115–128

Abstract

RTV Slovenia's archives contain large amounts of audio and video materials, various documents and music scores, and most of them are still in the analogue format. Widespread digitization has revolutionized the processes and ways of creating content in the digital format, recorded on different media. Such records also require new ways of preservation. In the article the development and structure of the Mediateque department at RTV Slovenia is presented. Also an overview to the preservation model of audio content is given. Due to rapid technological changes the audio content was the most critical and the first to be digitized. The intensive work in Mediateque began in 2008 and after two years Radio Slovenia has developed modern system of permanent storage of audio content. Radio Slovenia's Digital Archive meets all the standards and regulations applicable to modern archival systems. In the article the application of Mediarc software is also presented, which as it could be used for digitizing and permanent storage of TV Slovenia's video archives.

Keywords: RTV Slovenia, Radio Slovenia, Mediateque, digitization, digital archives, audio content

1 Uvod

RTV Slovenija hrani velike količine zvokovnih in slikovnih gradiv, različnih dokumentov in notnega gradiva, večinoma še v analogni obliki zapisa. Vseobšerna digitalizacija je močno spremenila postopke in načine ustvarjanja vsebin, ki se v digitalni obliki zapisa zapisujejo na različne medije. Takšni zapisi zahtevajo tudi nove načine hranjenja. V skladu s predpisi smo se odločili za izgradnjo sistema za dolgoročno e-hrambo gradiva pri ustvarjalcu. Najprej smo se lotili zvokovnih vsebin, saj je bilo to področje zaradi hitrega tehnološkega razvoja v najbolj kritičnem stanju.

Že sredi 80. let prejšnjega stoletja, ko se je kot medij za zapis zvoka v digitalni obliki pojavil CD medij, smo na Radiu Slovenija pričeli z uvajanjem digitalnih tehnoloških novitet, najprej na področju snemalne tehnike. Pred nekaj leti je zamrla proizvodnja magnetofonskih trakov, enako se je zgodilo tudi z magnetofoni, ki so bili desetletja tehnološka osnova vseh RTV ustanov. To pa je bil tudi povod za intenzivno iskanje in načrtovanje sistema, ki bi zagotovil trajno hranjenje digitalno ustvarjenega gradiva in omogočil, da se tudi vsebine na analognih medijih digitalizirajo in trajno shranijo. Projektna naloga je bila izdelana leta 2004, po postavitvi in testiranju tehnološke opreme je bil leta 2007 ustanovljen nov oddelek, ki smo ga imenovali Mediateka.

Intenzivno delo v Mediateki je steklo leta 2008, s tem se je pričelo sistematično digitaliziranje in arhiviranje zvokovnih vsebin, ki so na analognih nosilcih v arhivih Radia Slovenija in Regionalnih centrih Koper in Maribor. Prav tako se zajemajo in arhivirajo vsebine, ki so na digitalnih nosilcih zvoka (CD, CDR, DAT), in nove vsebine, ki vsakodnevno nastajajo v produkciji RTV Slovenija. Tako nastaja centralni digitalni arhiv zvokovnih vsebin, ki omogoča trajno hranjenje in enostavno uporabo v vseh programih, ki jih v eter pošilja nacionalna RTV. Obstoječe vsebine so zaščitene pred nadaljnjim propadanjem, nove pa se kvalitetno arhivirajo in skupaj predstavljajo neprecenljivo vrednost za našo medijsko hišo, posredno pa tudi za ves slovenski narod.

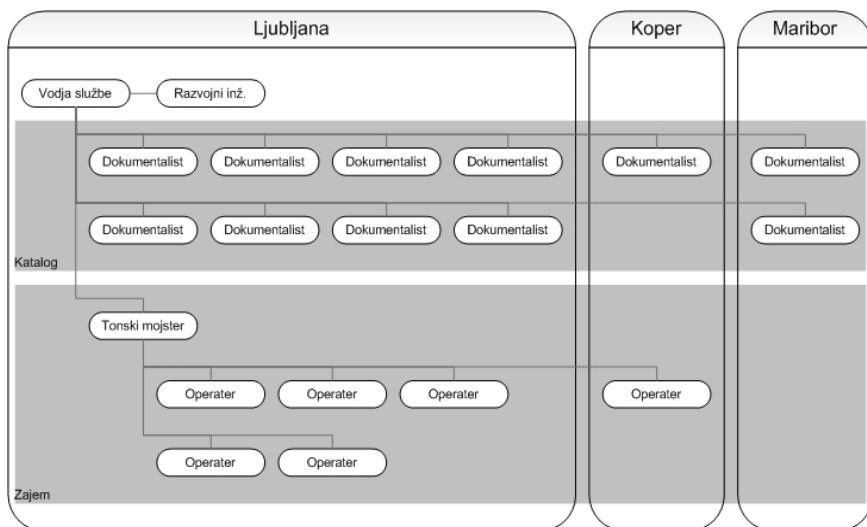
2 Mediateka

Mediateka – služba za digitalizacijo in zajem vsebin ter razvoj in upravljanje centralnega digitalnega arhiva zvokovnih vsebin RTV Slovenija – je bila ustanovljena zaradi zahtevnosti postopka digitalizacije zvokovnih arhivov s kataložko obdelavo podatkov. Proces digitalizacije in kataložke obdelave zvokovnih vsebin poteka hkrati na treh lokacijah (Radio Slovenija, RC Maribor in RC Koper), zato so tudi posamezna delovna mesta temu primerno lokacijsko ločena (Slika 1).

Zaposleni delavci v Mediateki na različnih delovnih mestih opravljajo dela, ki so potrebna za nemoten potek procesov digitalizacije in kataložke obdelave podatkov o digitaliziranih zvokovnih vsebinah. Zaposleni v Mediateki so razvrščeni v tri skupine. Prva skupina skrbi za digitalizacijo in zajem, druga za preverjanje tehnične kakovosti vsebin in metapodatkov, naloga tretje skupine je kataložka obdelava podatkov (Slika 1).

V prvi polovici leta 2010 je bilo v Mediateki zasedenih 16 delovnih mest, preostala štiri delovna mesta pa bodo aktivirana predvidoma do konca leta 2010. Zasnovana je bila sledeča kadrovska struktura:

- *vodja službe* organizira in vodi delo celotne službe in skrbi za njen nadaljnji razvoj,
- *razvojni inženir* skrbi za administracijo in razvoj računalniškega sistema,
- *dokumentalisti* skrbijo za kataložko obdelavo zajetih vsebin,
- *tonski mojster* skrbi za nemoten potek zajema in digitalizacije zvokovnih vsebin ter bdi nad kakovostjo digitaliziranih posnetkov,
- *operaterji* izvajajo zajem zvokovnih vsebin po navodilih tonskega mojstra.



Slika 1: Shema delovnih mest Mediateke RTV Slovenija

Delo je organizirano tako, da so posamezna opravila znotraj različnih delovnih procesov med seboj čim bolj usklajena. Procesni dela so v veliki meri avtomatizirani in centralno nadzorovani. Vsi glavni procesi dela so vodeni in nadzorovani prek računalniškega sistema, ki skrbi, da so posamezna opravila in naloge pravočasno dodeljene posameznim delavcem ali avtomatiziranim modulom računalniškega sistema. Digitalne vsebine se v visoki kakovosti in kakovosti za poslušanje shranjujejo v primarno pomnilniško enoto in identično sekundarno enoto na drugi lokaciji, za dodatno varovanje je poskrbljeno s kopiranjem na tračno knjižnico, kasete pa se hranijo v Regionalnem centru Koper.

3 Tehnična osnova modela hranjenja zvokovnih vsebin

Na osnovi projektne dokumentacije je bila že pred ustanovitvijo Mediateke izbrana in nato tudi nameščena strojna in programska oprema, ki je omogočila začetek delovanja digitalnega zvokovnega arhiva. Pri načrtovanju sistema je bilo najpomembnejše vodilo kakovost pretvorbe zvokovnih zapisov iz analogne v digitalno obliko in kakovost zajema vsebin z digitalnih nosilcev zvoka (CDR, CD, DAT) (Watkinson, 2001). Poleg priporočil organizacije EBU so bili upoštevani tudi izsledki mednarodnih institucij, ki se ukvarjajo z zaščito pred propadanjem avdiovizualnih vsebin (Bradley, 2009; International, 2010). V pomoč so bile

tudi izkušnje drugih RTV ustanov v evropskem prostoru, ki so se lotile takšnega projekta pred RTV Slovenija. Ne bomo naštevali vseh zahtevanih parametrov in tehničnih lastnosti, ki so navedene v projektni dokumentaciji, omenimo naj le, da je format zapisa zvoka v datotečni obliki BWF (ang. broadcast wave file) z možnimi vzorčnimi frekvencami 192/96/48/44,1 kHz, procesiranje je 32-bitno, v pomnilniško enoto se shranjujejo datoteke s 24- in 16-bitno resolucijo. Pregledovanju zvokovnih vsebin so namenjene MP3 datoteke. V Mediateki ves čas skrbimo tudi za ustrezno nadgrajevanje računalniške opreme in rast pomnilniških kapacitet za trajno hranjenje ter 100 % varovanje digitalnih zapisov.

Računalniški sistem Mediateke je zgrajen na osnovi sistema, ki ga je za arhiviranje zvokovnih vsebin razvilo avstrijsko-nemško podjetje NOA – *Audio solutions* (2010) in je trenutno konfiguriran tako, da omogoča:

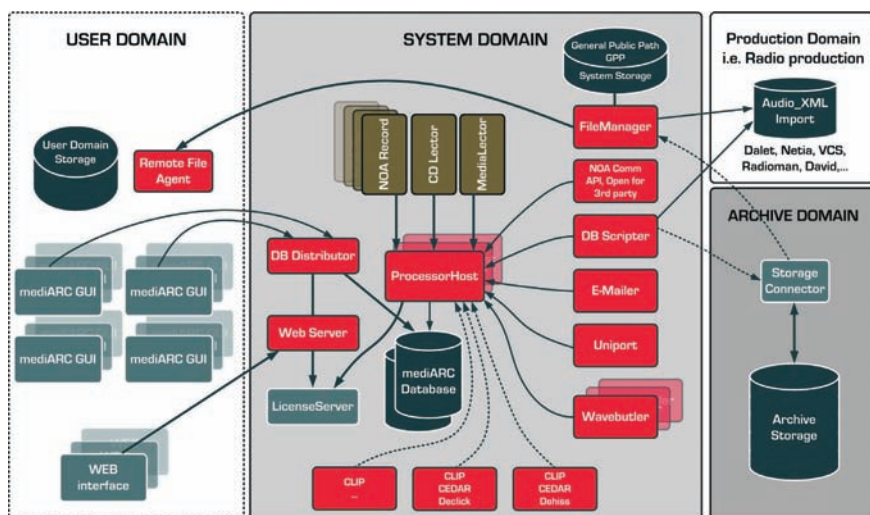
- zajem analognih zvokovnih zapisov z magnetofonskih trakov in pretvorbo v digitalno obliko z visoko kakovostjo,
- zajem digitalnih zvokovnih vsebin, ki se hranijo na CD, CDR in DAT nosilcih zvoka,
- zajem novih, vsakodnevno ustvarjenih digitalnih zvokovnih vsebin (povezava z vsemi produkcijskimi okolji na RTV Slovenija),
- kataloško obdelavo arhiviranih zvokovnih vsebin,
- iskanje, naročanje in distribucijo arhiviranih digitalnih vsebin za produkcijske potrebe vseh organizacijskih enot RTV Slovenija.

Jedro programske opreme predstavlja NOA MediARC – *sistem za upravljanje medijskih vsebin*, ki ustreza modelu OAIS (Reference, 2002). Sistem v osnovi lahko nadzoruje različne vrste medijskih vsebin (avdio, video, slikovno gradivo, dokumente), še posebej pa je specializiran za upravljanje zvokovnih vsebin, v bližnji prihodnosti tudi videovsebin.

Pri upravljanju medijskih vsebin so seveda izjemno pomembni metapodatki, ki skupaj z avdio ali/in videodatotekami tvorijo medijske vsebine. V sistem je vgrajen katalog vsebin, njegova struktura se lahko spreminja in razširja glede na sedanje in morebitne prihodnje potrebe. Modularna zgradba in možnosti prilaganja celotnega sistema NOA MediARC pa omogočajo njegovo učinkovito uporabo danes in v prihodnosti (Slika 2).

Za boljše razumevanje sistema mediARC si oglejmo še njegovo modularno zgradbo (ohranili smo izvirne angleške izraze):

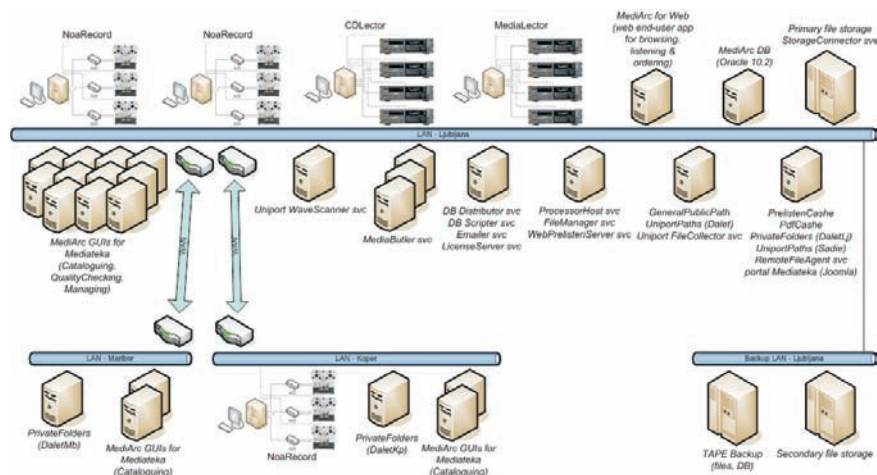
- *mediARC Database* predstavlja NOA shemo podatkovne baze, razvite s tehnologijo Oracle 10g.
- *DB Distributor* je vmesnik do podatkovne baze in omogoča povezavo aplikacije mediARC GUI na bazo.



Slika 2: Shema mediARC (primer radijskega arhiva)

- *ProcessorHost* je vmesnik do podatkovne baze in omogoča povezavo vseh avtomatiziranih modulov na podatkovno bazo.
- *FileManager* dostavlja datoteke med različnimi domenami z uporabo različnih metod. Predstavlja most med uporabniško, sistemsko in arhivsko domeno.
- *Storage Connector* skrbi za zapis arhivskih datotek v arhivski strežnik (NAS) in ciklično nadzira stanje arhivskih kapacitet.
- *RemoteFile Agent* je v uporabniški domeni namenjen prevzemu naročenih datotek iz sistemske domene in prenosu do končnega uporabnika.
- *DB Scripter* izvaja skripte, ki se prožijo kot posamezna *opravila* v *procesih* ali pa se samodejno prožijo v določenem ciklusu.
- *WaveButler* oz. *MediaButler* omogoča hkratno konverzijo več avdio ali video-datotek. Za to uporablja odprtokodno knjižnico orodij »FFMPEG«.
- *UniPort Processor* je orodje za avtomatizirano zajemanje digitalnih zvokovnih vsebin (datotek) in njihovih metapodatkov.
- *CLIP* je orodje za povezavo sistema z zunanji orodji, ki jih lahko vključimo v procese, ki so nadzorovani prek mediARC-a. Npr. z restavracijskimi orodji Cedar (DeHiss, DeClick) lahko v proces naročanja vsebine vključimo samodejno odstranjevanje šumov in pokov.
- *E-Mailer* pošlje elektronsko sporočilo na poljuben elektronski naslov in tako obvesti uporabnika o končanem izvozu naročenih vsebin ali pa administratorju sporoči napako, ki se je pripetila med izvajanjem določenega *procesa*.

Spodnja shema (Slika 3) prikazuje računalniške strežnike, delovne postaje in povezave, ki tvorijo računalniški sistem Mediateke RTV Slovenija.



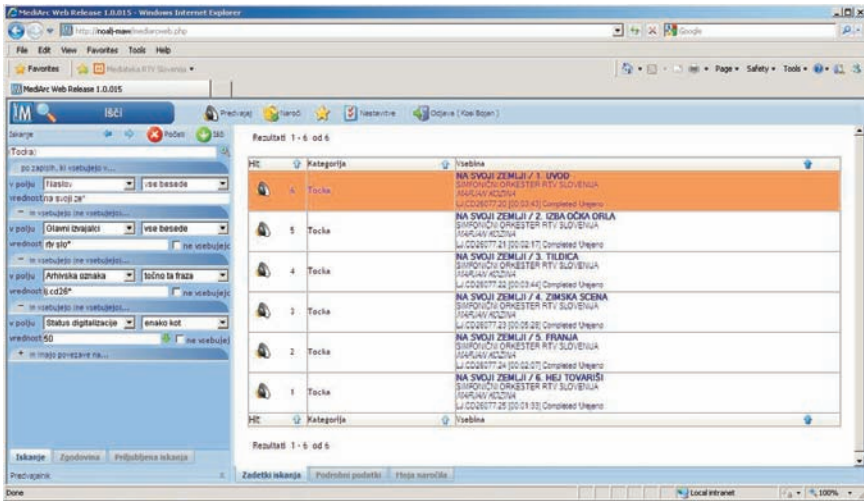
Slika 3: Mrežna shema računalniškega sistema Mediateke RTV Slovenija

Večina vseh delovnih postaj za zajem, aplikacijskih strežnikov, podatkovna baza in primarni arhivski datotečni strežnik so na primarni lokaciji v Ljubljani, kjer je na sekundarni lokaciji tudi sekundarni arhivski strežnik in LTO tračna knjižnica za varnostno kopiranje arhivskih datotek in podatkovne baze. Na oddaljenih lokacijah v Kopru in Mariboru je postavljena še dodatna infrastruktura za zajem gradiva in kataloško obdelavo vsebin, predvideno pa je tudi varno hranjenje kaset LTO z arhivskimi datotekami in varnostno kopijo podatkovne baze.

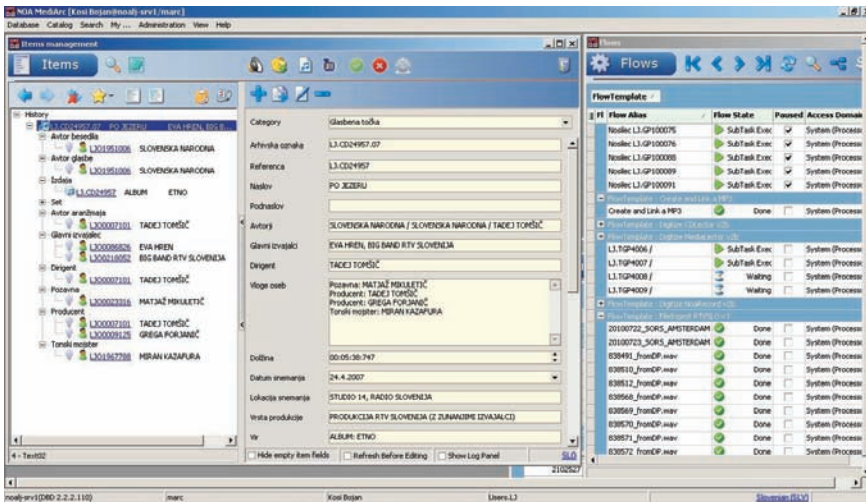
Končnim uporabnikom arhiva je namenjena spletna aplikacija »MediArc WEB«, ki teče v spletnem strežniku Apache in omogoča iskanje, poslušanje in naročanje zelenih vsebin (Slika 4). Prijava v sistem je mogoča le z uporabniškim imenom in geslom, kar omogoča tudi nadzor nad dostopom do arhiva, vodenje evidence naročil in sledljivost naročenega gradiva.

Zaposlenim v Mediateki je namenjena aplikacija »MediArc GUI«, ki poleg iskanja in poslušanja vsebin služi za urejanje kataloga in vso ostalo administracijo sistema (Slika 5).

Uporabnik lahko poslušar arhivske posnetke v nizki ločljivosti (MP3, 128 kbps), pri naročilu posnetka pa se iz arhivske datoteke v visoki ločljivosti (BWF) ustvari datoteka v zelenem formatu zapisa. Ko uporabnik sproži naročilo posnetka, se zažene *proces*, ki poskrbi, da se naročilo izvede po točno določenem postopku. Vsak *proces* je sestavljen iz več *opravil*, ki imajo vsako zase točno določeno funk-



Slika 4: MediARC WEB



Slika 5: MediARC GUI

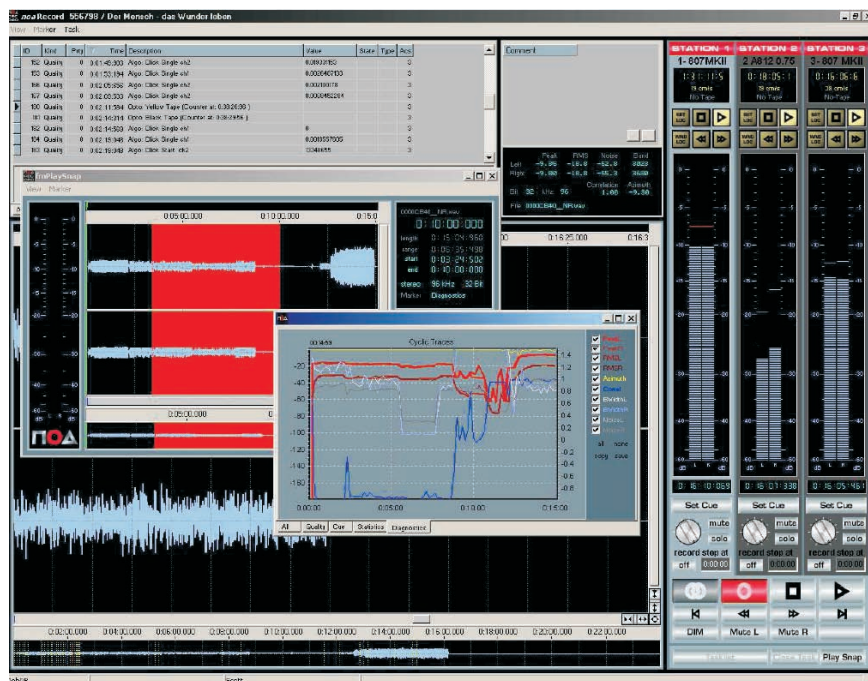
cijo. Npr. neko *opravilo* lahko iz arhivske datoteke ustvari datoteko v drugem formatu zapisa, neko drugo *opravilo* pa zapiše kataloške podatke naročenega posnetka v XML datoteko.

V sistemu obstaja tako imenovana *predloga procesa* za naročilo posnetka, iz katere se za vsako naročilo posebej ustvari samostojen *proces*, ki nato poskrbi, da se

naročilo izvrši z vnaprej določenimi *parametri procesa*. Tudi sam zajem in kataloško urejanje vsebin poteka prek vnaprej določenih *procesov* oz. njihovih *predlog*. Administrator sistema lahko obstoječe *predloge* popravlja, spreminja njihove *parametre* ali pa ustvari nove *predloge procesov*, po katerih bodo potekali posamezni procesi dela v Mediateki. Vgrajena administratorsko-razvojna orodja spodbujajo kreativnost pri načrtovanju *procesov* in omogočajo realizacijo zelo kompleksnih rešitev.

Za zajem zvokovnega gradiva uporabljamo tri različne vrste aplikacij, ki so povezane s centralnim sistemom NOA mediARC, ki prek *procesov* skrbi za nemoteno in tekoče delo operaterjev:

- (ang.) *CDLector* omogoča hkratni prepis štirih CD/CDR-jev,
- (ang.) *MediaLector* omogoča hkratni prepis štirih DAT-ov,
- (ang.) *NoaRecord* pa omogoča hkratni prepis treh magnetofonskih trakov ali vinilnih plošč. Aplikacija *NoaRecord* skrbi za celovit nadzor nad zajemom in digitalizacijo vsebin z magnetofonskih trakov (Slika 6).

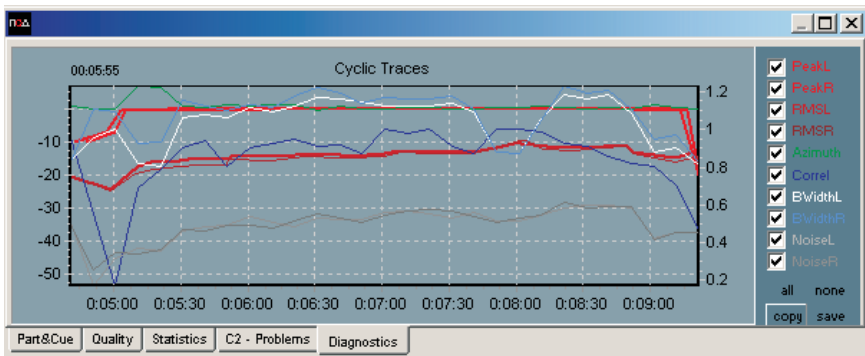


Slika 6: NoaRecord – aplikacija za zajem in digitalizacijo vsebin z magnetofonskih trakov

Pri zajemu vsebin je pomembna funkcija sistema tudi avtomatsko generiranje metapodatkov o kakovosti zajetega avdiosignala. Sistem med zajemanjem in digitaliziranjem analizira avdiosignal in hkrati zapisuje podatke o različnih parametrih:

- kakovost (fizične poškodbe trakov, prekoračitev maksimalne analogne in digitalne ravni signala),
- statistika (maks. jakost, povpr. jakost, razmerje signal/šum, ...),
- diagnostika (azimut, amplituda, panorama, raven šuma).

Te informacije se zapisujejo v posebno datoteko v obliki XML zapisa, ki jo imenujemo (ang.) *Marker lista*. Določene podatke lahko prikažemo tudi v obliki grafa, kar nam je v pomoč pri preverjanju kakovosti zajetega materiala in kasnejši uporabi vsebin (Slika 7).

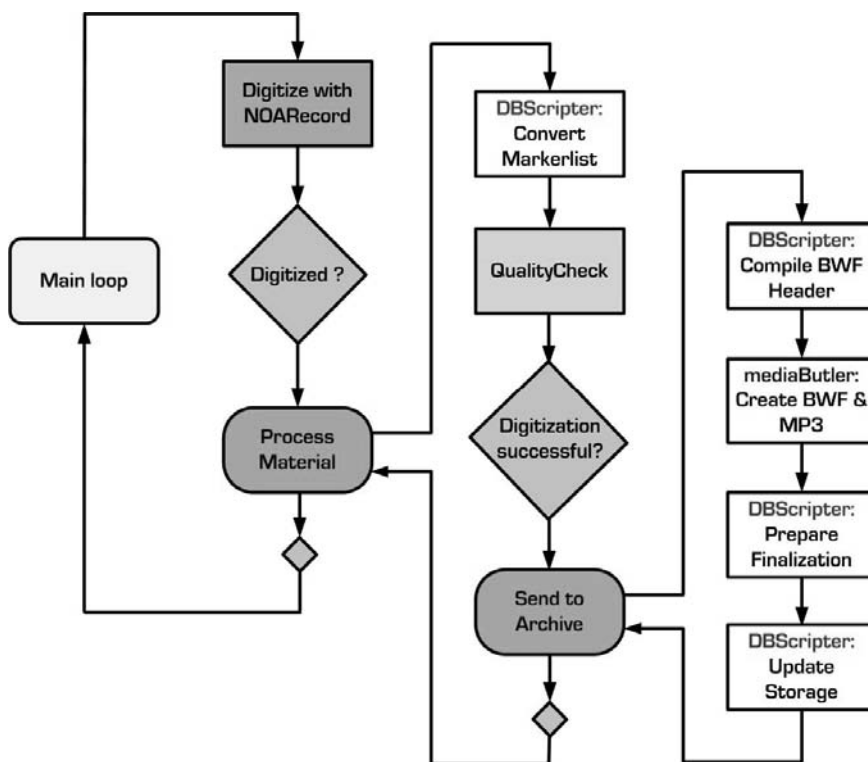


Slika 7: Diagnostika zajetih vsebin v aplikaciji NoaRecord

V nadaljevanju (Slika 8) je prikazan in opisan proces zajema in arhiviranja zvočnih vsebin z magnetofonskih trakov.

Celoten proces se začne z zajemom vsebine v aplikaciji *NoaRecord* – opravilo *Digitize with NoaRecord*, kjer operater na seznamu opravil za zajem izbere ustrezno oznako nosilca zvoka, ki ga nato zajame – digitalizira. Če je vsebina z nosilca zvoka – traku uspešno prepisana, sledi skupek več opravil – *Process material*. Avtomatiziran servis *DBScripter* najprej obdela seznam *markerjev*, ki so nastali ob zajemu vsebine, nato pa sledi kontrola kakovosti zajete vsebine (raw audio 96kHz, 32bit floating point), ki jo izvede tonski mojster. Pri tem so mu v pomoč različni metapodatki, ki so bili v obliki *markerjev* avtomatsko generirani med zajemom gradiva (poročilo o napakah, statistika, diagnostika, ...).

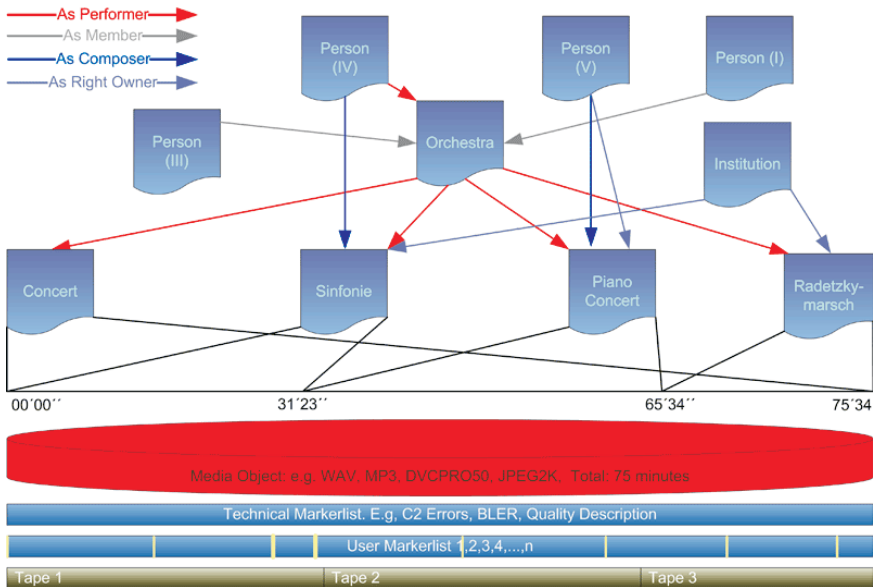
Po kontroli kakovosti sledi naslednja skupina opravil, ki so potrebna za uspešno arhiviranje zajete vsebine. *DBScripter* najprej uredi podatke v glavi BWF arhivske



Slika 8: Proces zajema in arhiviranja vsebine z magnetofonskih trakov

datoteke, sledi mu *MediaButler*, avtomatiziran servis za pretvorbo medijskih datotek, ki iz zajete surove avdiovsebine najprej ustvari BWF arhivsko datoteko (48 ali 96 kHz, 24bit) in nato še MP3 datoteko (128kbps), ki je namenjena poslušanju vsebin pri brskanju in iskanju v katalogu. Na vrsti je ponovno *DBScripter*, ki najprej poskrbi za zapis vseh zajetih metapodatkov v podatkovno bazo, nato pa se poveže z avtomatiziranim servisom *StorageConnector*, ki shrani BWF in MP3 datoteko v primarni arhivski datotečni strežnik (On-line NAS).

Za konec pogledajmo še osnovni princip kataloške obdelave zajetih vsebin (Slika 9). V procesu digitalizacije nosilcev vsebin se na osnovi enega ali več nosilcev ustvari tako imenovani *media objekt*, v njem so združene računalniške datoteke z zajetimi vsebinami in metapodatki, ki so nastali pri zajemu vsebin. Objekt je osnova za kataloško obdelavo, kjer s postavitvijo markerjev na samem objektu – posnetku, le-tega navidezno razdelimo na posamezne segmente, ki predstavljajo posamezne kataloške zapise (glasbene skladbe, radijske oddaje, zvočni efekti, ...).



Slika 9: Princip relacij med kataloškiimi zapisi in arhiviranimi datotekami

Ustvarjene kataloške zapise, povezane s prej določenimi segmenti, nato podrobno popišemo in jih povežemo tudi med seboj oz. z drugimi kataloškiimi zapisi. Pri povezovanju posameznih kataloških zapisov pa določimo tudi njihova medsebojna razmerja. Npr. *Oseba1* je kot *Član* povezana na *Orkester*, *Orkester* pa je kot *Glavni izvajalec* povezan s *Simfonijo* ... Pri ustvarjanju in povezovanju različnih tipov zapisov kataloški sistem ni omejen, vendar je pri tem pomembno, da ostaja struktura kataloga dovolj pregledna in učinkovita za uporabo in nadzor arhiviranih vsebin.

4 Zaključek

Po dveh letih je arhiv zvokovnih vsebin postal nepogrešljiv vsebinski vir pri ustvarjanju programov RTV Slovenija. Tudi za uporabnike v naših regionalnih centrih, ki so povezani v centralni sistem prek mikrovalovnih zvez. Uredniki, glasbeni redaktorji, novinarji in drugi ustvarjalci lahko s svojega namiznega računalnika prek internih spletnih povezav iščejo po razkošni knjižnici, poslušajo in z enim samim klikom naročijo prepis zelenega posnetka na ustrezno lokacijo. Sistem pravic omogoča selektivno dostopanje do arhiviranih vsebin in nadzorovano uporabo. Zajem in digitaliziranje vseh vsebin, ustvarjenih v preteklosti,

bo končan v naslednjih štirih letih. Kataloška obdelava pa bo trajala dlje, saj je neodvisna od procesa zajema. Pomembna je verodostojnost podatkov o vsebinah, zato jih dokumentalisti skrbno preverjajo, z naknadnim vpisovanjem v tem trenutku morda neznanih podatkov se bo bogatila vrednost arhiva tudi v prihodnosti. Skupaj z vsakodnevnim prirastom novih vsebin postaja digitalni arhiv RTV Slovenija velika kulturna dediščina tudi prihodnjih rodov.

Navedeni viri

1. Bradley, K. (Ur.). (2009). *Guidelines on the production and preservation of digital audio objects* (2nd ed.). Aarhus: International Association of Sound and Audiovisual Archives.
2. *International Association of Sound and Audiovisual Archives (IASA)*. (2010). Pridobljeno 23. 8. 2010 s spletne strani: <http://www.iasa-web.org/>
3. *NOA Audio solutions*. (2010). Pridobljeno 23. 8. 2010 s spletne strani: <http://www.noa-audio.com/>
4. *Reference model for an open archival information system (OAIS)*. (2002). Washington: Consultative Committee for Space Data Systems. Pridobljeno 23. 8. 2010 s spletne strani: <http://public.ccsds.org/publications/archive/650x0b1.pdf>
5. Watkinson, J. (2001). *Art of digital audio* (3rd ed.). Oxford: Focal Press.

Bojan Kosi je zaposlen v Mediateki, RTV Slovenija.

Naslov: Kolodvorska 2, 1550 Ljubljana

Naslov elektronske pošte: bojan.kosi@rtvslo.si

Martin Žvelc je zaposlen v Mediateki, RTV Slovenija.

Naslov: Kolodvorska 2, 1550 Ljubljana

Naslov elektronske pošte: martin.zvelc@rtvslo.si