

1.09 Objavljeni strokovni prispevek na konferenci
1.09 Published Professional Conference Contribution

Bojan Kosi*, Martin Žvelc**

MEDIATEKA - SLUŽBA ZA DIGITALIZACIJO IN TRAJNO HRAMBO AUDIO VSEBIN

Izvleček:

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. Najprej smo se lotili zvokovnih vsebin, saj je bilo to področje zaradi hitrega tehnološkega razvoja najbolj kritično. Intenzivno delo v Mediateki je steklo leta 2008 in po štirih letih uspešnega dela se RTV Slovenija lahko pohvali s sodobnim sistemom za trajno hrambo zvokovnih vsebin. Digitalni arhiv ustreza vsem standardom in predpisom, ki veljajo za sodobne arhivske sisteme. Upravljanje z zvokovnimi vsebinami poteka s pomočjo programske opreme NOA-Mediarc, ki jo uporabljajo tudi mnoge druge RTV-ustanove po svetu. Zaradi izredne odprtosti in zmogljivosti je administratorjem dopuščeno veliko kreativne svobode, kar smo na RTV Slovenija s pridom izkoristili.

Ključne besede:

audio, digitalni arhiv, Mediateka, RTV Slovenija

Abstract:

Mediatheque - the Department of Digitization and Permanent Preservation of Audio Contents

RTV Slovenia's archives have large amounts of audio and video materials, various documents and music scores, and most of those are still in the analogue format. Widespread digitization has revolutionized processes and ways of creating content in the digital format, recorded on different media. At RTV Slovenia, we began digitizing the audio content first, because this content was the most critical due to the rapid technological development. The intensive work in the Mediatheque began in 2008, and after four years, RTV Slovenia can be proud of its modern system of permanent storage of audio content. Digital Archives meets all standards and regulations applicable to modern archival systems. Audio content is managed with the help of NOA-Mediarc software, which is used by many other broadcasting institutions worldwide. It is a high capacity and an extremely open system, which allows a good deal of creative freedom, and RTV Slovenia has been using it to its great advantage.

Key words:

audio, digital archives, Mediatheque, RTV Slovenia

V letu 2008 se je na RTV Slovenija pričelo sistematično digitaliziranje in arhiviranje zvokovnih vsebin, ki se nahajajo na analognih nosilcih v arhivih Radia Slovenija in regionalnih centrih v Kopru in Mariboru. Prav tako se zajemajo in arhivirajo tudi vsebine, ki so na digitalnih nosilcih zvoka (CD, CDR, DAT), in nove

* Bojan Kosi, Mediateka RTVSLO, RTV Slovenija, Kolodvorska 2, 1550 Ljubljana, Slovenija.

** Martin Žvelc, Mediateka RTVSLO, RTV Slovenija, Kolodvorska 2, 1550 Ljubljana, Slovenija.

vsebine, ki se vsakodnevno rojevajo v naših produkcijskih sklopih. Tako nastaja centralni digitalni arhiv zvokovnih vsebin, ki omogoča trajno hranjenje in enostavno uporabo v vseh programih, ki jih v eter pošilja nacionalni 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.

Zaradi zahtevnosti postopka digitalizacije zvokovnih arhivov hkrati s kataložko obdelavo podatkov je bila ustanovljena **Mediateka** - služba za digitalizacijo in zajem vsebin ter razvoj in upravljanje centralnega digitalnega arhiva zvokovnih vsebin RTV Slovenija.

Na osnovi projektne dokumentacije je bila že pred ustanovitvijo Mediateke izbrana, nabavljena ter 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). Poleg priporočil organizacije EBU so bili upoštevani tudi izsledki mednarodnih institucij, ki se ukvarjajo z zaščito pred propadanjem avdiovizualnih vsebin (IASA). 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 (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-odstotno varovanje digitalnih zapisov.

Trenutno je v Mediateki zaposlenih 20 sodelavcev, ki na različnih delovnih mestih opravljajo dela, potrebna za nemoten potek digitalizacije in kataložko obdelavo podatkov o digitaliziranih zvokovnih vsebinah.

Delo je organizirano tako, da so posamezna opravila znotraj različnih delovnih procesov med seboj čim bolj usklajena. Vsi glavni procesi dela so vodeni in nadzorovani preko računalniškega sistema, ki skrbi, da so posamezna opravila in naloge pravočasno dodeljeni posameznim delavcem ali avtomatiziranim modulom računalniškega sistema.

Računalniški sistem Mediateke je zgrajen na osnovi sistema, ki ga je za arhiviranje zvokovnih vsebin razvilo avstrijsko-nemško podjetje NOA-Audio Solutions, 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 nosilcih zvoka CD, CDR in DAT,
- zajem novih, vsakodnevno ustvarjenih digitalnih zvokovnih vsebin (povezava z vsemi produkcijskimi okolji na RTV Slovenija),
- kataložko obdelavo arhiviranih zvokovnih vsebin,
- iskanje, naročanje in distribucijo arhiviranih digitalnih vsebin za produkcijske potrebe vseh organizacijskih enot RTV Slovenija.

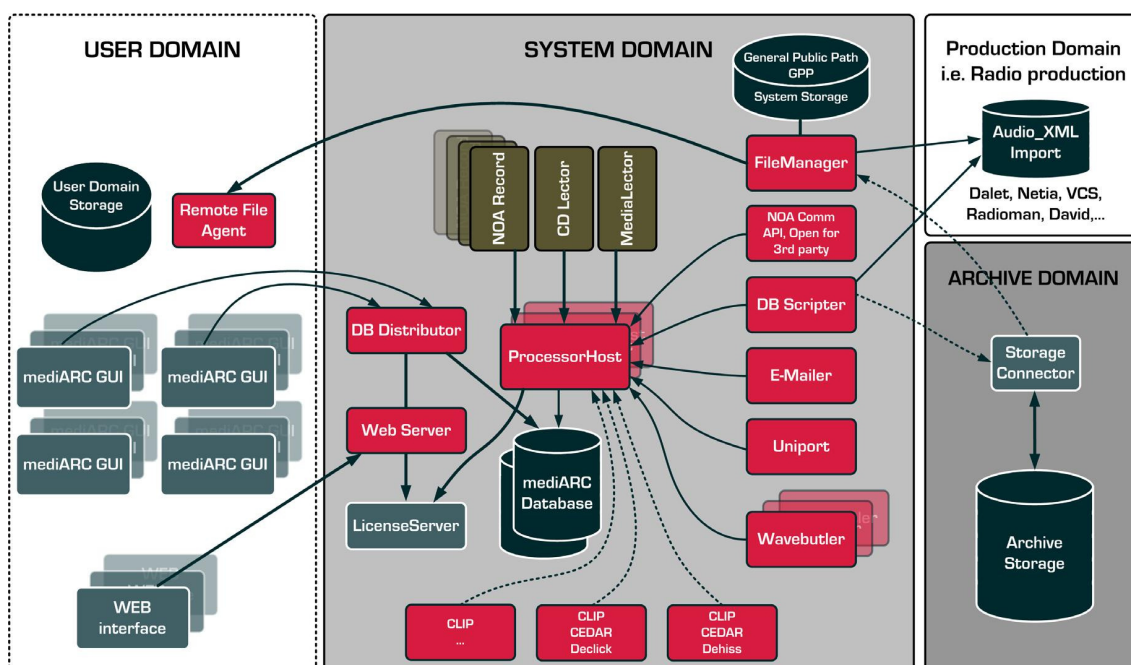
Jedro programske opreme predstavlja NOA mediARC - Sistem za upravljanje medijskih vsebin, ki ustreza modelu OAIS (Open Archival Information System).

Sistem lahko v osnovi nadzoruje različne vrste medijskih vsebin (avdio, video, slikovno gradivo, dokumente), še posebej pa je specializiran za upravljanje zvokovnih vsebin, v bližnji prihodnosti tudi video vsebin.

Pri upravljanju medijskih vsebin so seveda izjemno pomembni **metapodatki**, ki skupaj z avdio ali/in video datotekami 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 prilagajanja celotnega sistema NOA MediARC pa omogočajo njegovo učinkovito uporabo danes in v prihodnosti.

Za boljše razumevanje sistema mediARC si oglejmo še njegovo modularno zgradbo:

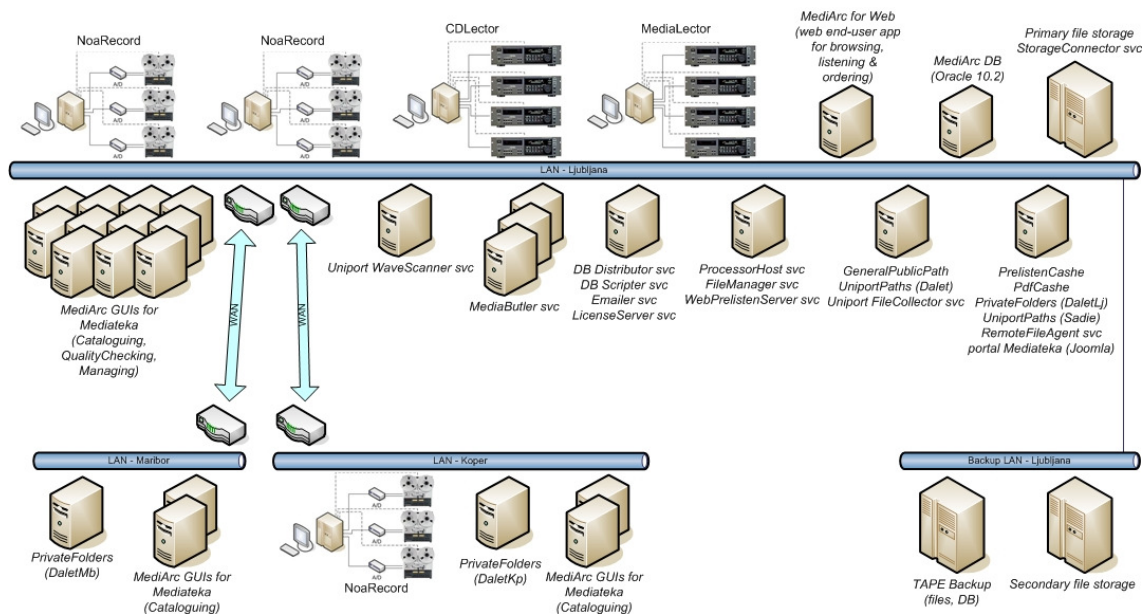


Slika 1: Shema sistema mediARC (primer radijskega arhiva)

- *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;
- *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 meta podatkov;
- *CLIP* je orodje za povezavo sistema z zunanjimi orodji, ki jih lahko vključimo v procese, nadzorovane preko 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*.

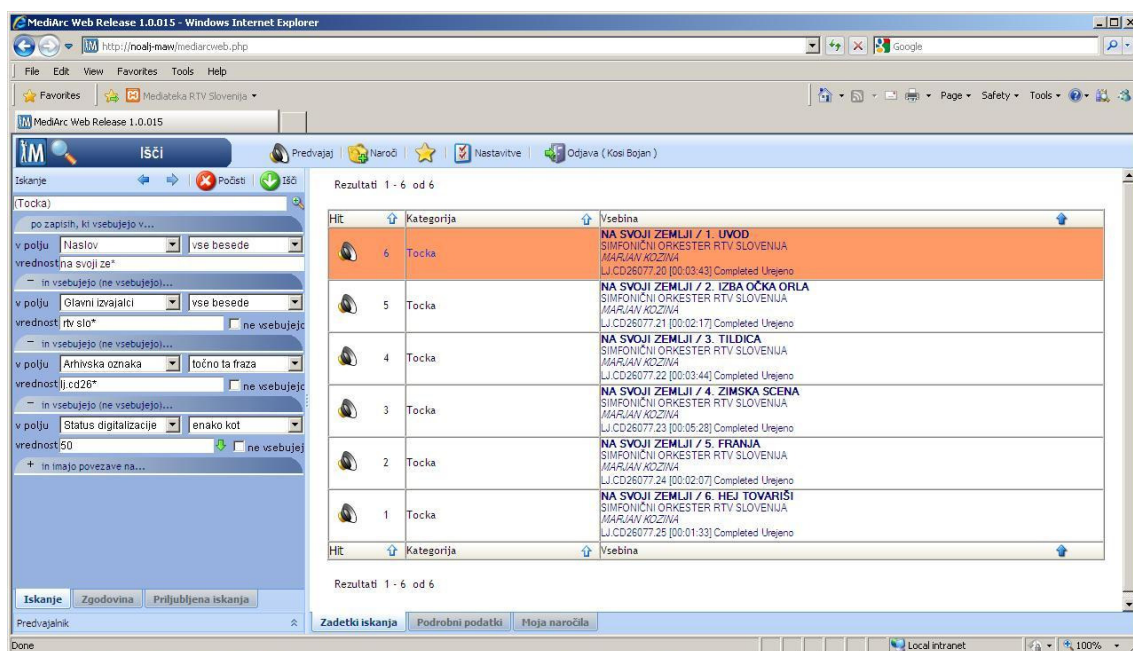
Spodnja shema prikazuje računalniške strežnike, delovne postaje in povezave, ki tvorijo računalniški sistem Mediateke RTV Slovenija:



Slika 2: Mrežna shema računalniškega sistema Mediateke RTV Slovenija

Večina vseh delovnih postaj za zajem, aplikacijskih strežnikov, podatkovna baza in primarni arhivski datotečni strežnik se nahajajo na primarni lokaciji v Ljubljani, kjer se na sekundarni lokaciji nahaja 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 LTO-kaset z arhivskimi datotekami in varnostno kopijo podatkovne baze.

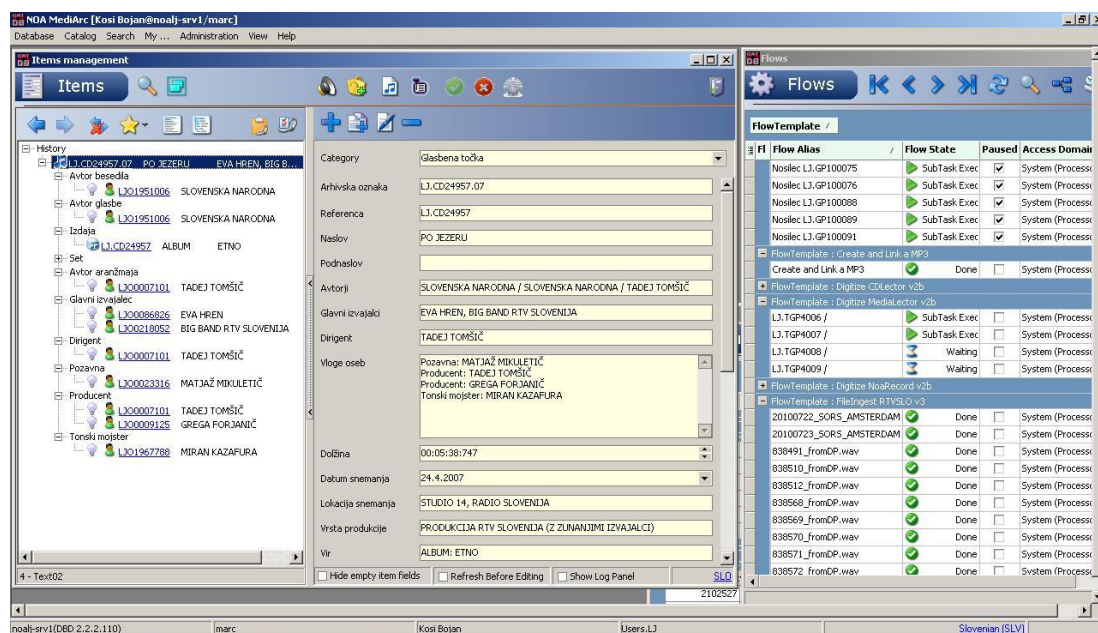
Končnim uporabnikom arhiva je namenjena spletna aplikacija "mediArc WEB", ki teče v spletnem strežniku Apache in omogoča iskanje, poslušanje in naročanje zelenih vsebin:



Slika 3: mediARC WEB

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 4: mediARC GUI

Uporabnik lahko poslušā arhivske posnetke v nizki ločljivosti (MP3, 128 kbps), pri naročilu posnetka pa se iz arhivske datoteke v visoki ločljivosti (BWF) ustvari datoteka v želenem 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 vsak zase točno določeno funkcijo.

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

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*.

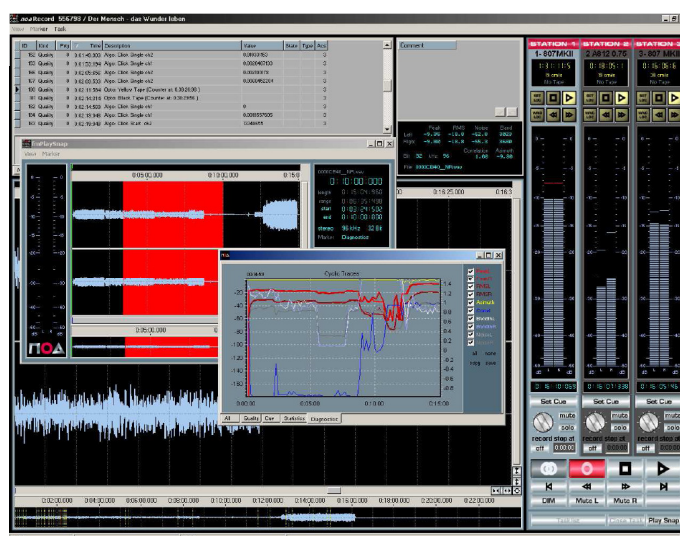
Tudi sam zajem in kataloško urejanje vsebin poteka preko vnaprej določenih *procesov* oz. njihovih *predlog*. Administrator sistema lahko obstoječe *predloge* popravlja, spreminja njihove *parametre* ali pa ustvari nove *predloge procesov*, po katerih bodo potekali posamezni procesi dela v Mediateki.

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 preko *procesov* skrbi za nemoteno in tekoče delo operaterjev:

- *CDLector* omogoča hkratni prepis štirih CD/CDR-jev,
- *MediaLector* omogoča hkratni prepis štirih DAT-ov,
- *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:



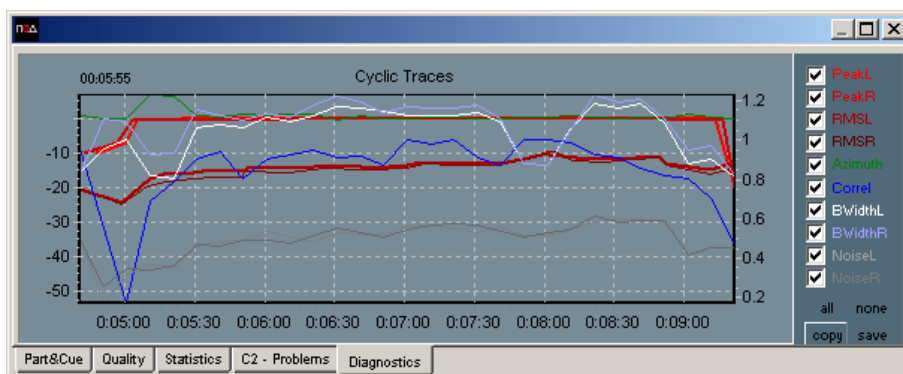
Slika 5: NoaRecord - aplikacija za zajem in digitalizacijo vsebin z magnetofonskih trakov

Pri zajemu vsebin je pomembna funkcija sistema tudi avtomatsko generiranje metapodatkov o kakovosti zajetega avdio signala.

Sistem med zajemanjem in digitaliziranjem analizira avdio signal in hkrati zapisuje podatke o različnih parametrih:

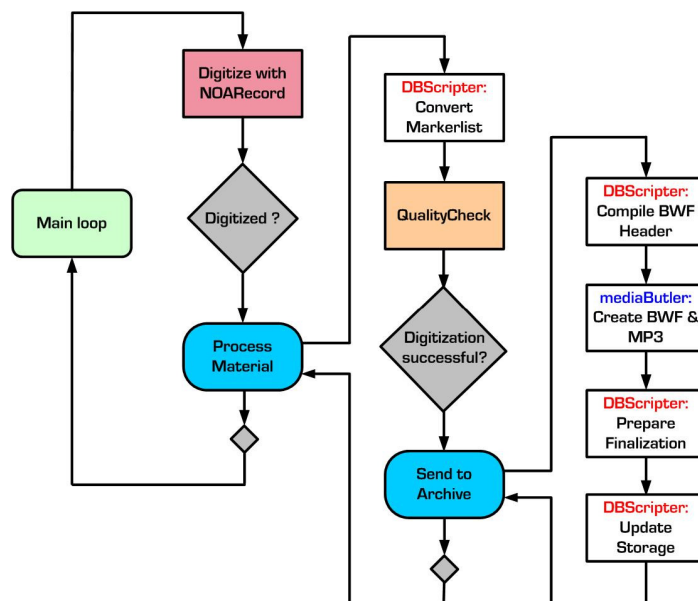
- kakovost (fizične poškodbe trakov, prekoračitev maksimalnega analognega in digitalnega nivoja signala),
- statistika (maks. jakost, povpr. jakost, razmerje signal/šum ...),
- diagnostika (azimut, amplituda, panorama, nivo šuma).

Te informacije se zapisujejo v posebno datoteko v obliki XML-zapisa, ki jo imenujemo *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 6: Diagnostika zajetih vsebin v aplikaciji NoaRecord

V nadaljevanju je prikazan in opisan proces zajema in arhiviranja zvokovnih vsebin z magnetofonskih trakov:



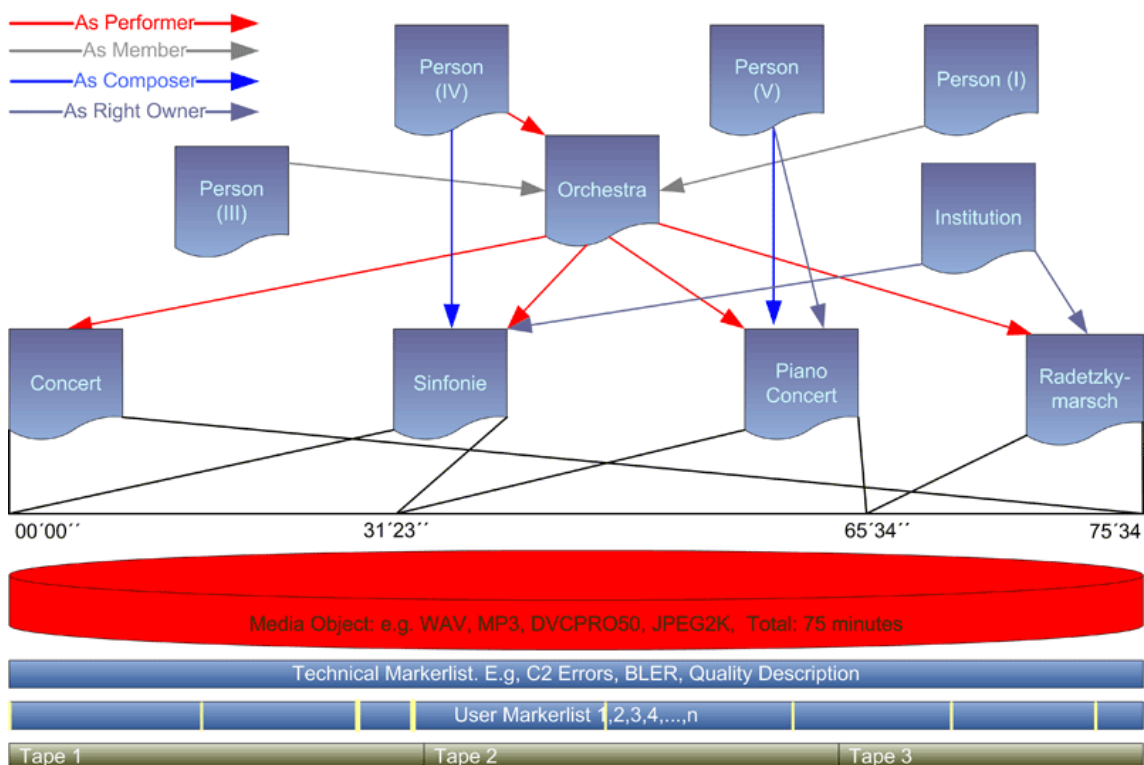
Slika 7: Proces zajema in arhiviranja vsebine z magnetofonskih trakov

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 arhivske datoteke BWF, sledi mu *MediaButler*, avtomatiziran servis za pretvorbo medijskih datotek, ki iz zajete surove avdio vsebine 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 datoteko BWF in MP3 v primarni arhivski datotečni strežnik (On-line NAS).

Za konec si pogledjmo še osnovni princip kataloške obdelave zajetih vsebin:



Slika 8: Princip relacij med kataloškiimi zapisi in arhiviranimi datotekami

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 ...).

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 pa je pri tem pomembno, da ostaja struktura kataloga dovolj pregledna in učinkovita za uporabo in nadzor arhiviranih vsebin.

ZAKLJUČEK

Po štirih 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 preko mikrovalovnih zvez. Uredniki, glasbeni uredniki, novinarji in drugi ustvarjalci lahko s svojega namiznega računalnika preko internih spletnih povezav iščejo po razkošni knjižnici, poslušajo in z enim samim klikom naročijo prepis želenega 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 prihodnjim rodovom.

LITERATURA

- IASA - TC04, *Second edition "Guidelines on the Production and Preservation of Digital Audio Objects"*, Contributing authors: Kevin Bradley - National Library of Australia, President IASA and Vice Chair IASA TC; Mike Casey - Indiana University; Stefano S. Cavagliari - Fonoteca Nazionale Svizzera; Chris Clark - British Library (BL); Matthew Davies - National Film and Sound Archive (NFSA); Albrecht Häfner - Südwestrundfunk, Germany; Franz Lechleitner, Phonogrammarchiv of the Austrian Academy of Sciences (OAW); Guy Maréchal, PROSIP; Michel Merten, Memnon; Greg Moss, NFSA; Will Prentice, BL; Dietrich Schüller, OAW; Lloyd Stickells and Nadia Wallaszkovits, OAW.
- *The Art of Digital Audio (third edition)*, Author: John Watkinson.

SPLETNI VIRI

- NOA Audio solutions - <http://www.noa-audio.com/>.
- IASA - <http://www.iasa-web.org/>.
- OAIS - <http://public.ccsds.org/publications/archive/650x0b1.pdf>.

SUMMARY

MEDIATHEQUE - THE DEPARTMENT OF DIGITIZATION AND PERMANENT PRESERVATION OF AUDIO CONTENTS

RTV Slovenia's archives have large amounts of audio and video materials, various documents and music scores, and most of those are still in the analogue format. Widespread digitization has revolutionized processes and ways of creating content in the digital format, recorded on different media. Such records also require new ways of permanent preservation. At RTV Slovenia, we began digitizing the audio content first, because this content was the most critical due to the rapid technological development.

In the middle of the 1980s, when a compact disk (CD) emerged as a new medium for recording sound in a digital form, RTV Slovenia began updating its technology, first in the segment of sound recording. A few years ago, the production of magnetic tapes became obsolete, and the same fate awaited tape recorders, which had been the technological backbone of all broadcasting institutions for decades. Those were years of intensive search and of intensive devising a new system that would provide permanent storage of digital recordings and enable the conversion of analogue media into digital format and thus their preservation. The project was compiled in 2004; and after the installation and testing of technical equipment a new department named Mediatheque was set up in 2007.

The intensive work in the Mediatheque began in 2008, and after four years, RTV Slovenia can be proud of its modern system of permanent storage of audio content. Digital Archives meets all standards and regulations applicable to modern archival systems. Given the specifics of planning and creating radio programs, it includes innovative, user friendly solutions that make searching and browsing the archives much easier, as well as listening to and selecting items for broadcasting on any of the three national radio channels in Ljubljana and on the channels of RTVSLO regional centres of Koper-Capodistria, Maribor and Lendava.

Mediatheque employees are organized in three groups. The first group is responsible for digitizing and capturing the content, the second for checking the technical quality of the content and META data, and the third for cataloguing. Work processes are largely automated and centrally controlled. High definition digital content of broadcasting quality is stored in the primary memory unit and for safety reasons also in the secondary unit, which is identical but in another location. In addition, copies of archived material are recorded also on tape and kept in the Library at the Regional Centre of Koper-Capodistria.

Audio content is managed with the help of NOA-Mediarc software, which is used by many other broadcasting institutions worldwide. It is a high capacity and extremely open system, which allows a good deal of creative freedom, and RTV Slovenia has been using it to its great advantage.