

Analytika vysílání – technická specifikace

Obsah

Analytika vysílání – technická specifikace	1
1. Uvedení do situace	2
2. Stručný popis zakázky	2
2.1 Proč Český rozhlas potřebuje analýzu vysílání	2
2.2 Proč to Český rozhlas nedělá jinak?	3
2.3 Vývoj systému	3
2.4 Co je úkolem Dodavatele	4
3. Detailní popis zakázky	4
3.1 Předmět zpracování	4
3.2 Popis zakázky, cíle	5
3.3 Návrh na datový výstup analytiky živého vysílání	8
3.4 Kroky vývoje analytiky	10
3.5 Implementace a spuštění, zdrojový kód a dokumentace	10
3.6 API	11
3.7 Předání zpracovaných metadat z loggingu	11
4. Data k analýze vysílání	12
4.1 Stručný přehled zdrojů metadat dle předmětu zpracování	12
4.2 Vysílací schémata pro konkrétní den a stanici	12
4.3 Rundown pro konkrétní den a stanici:	13
4.4 Aktuální vysílání - XML	14
4.5 Práva	16
4.6 Zvuková grafika - znělky, jingly... ..	16
5. Logging - archiv kontinuálního vysílání	16
5.1 Archiv kontinuálního vysílání – logging ke zpracování	16
5.2 Přehled stanic s loggingem (jde o odhady)	17
5.3 Ukázka detailního popisu souborů loggingu	18
6. Živé vysílání – Dante protokol	20
6.1 Technologie:	20
6.2 Místo k připojení:	21
7. Serverové řešení	21
7.1 Základní parametry serverů:	21
8. Podklady pro návrh řešení zakázky	21
8.1 Logging – převzetí ukázek	21
8.2 Převzetí přístupů k datům	22
8.3 Převzetí podkladů	22

Analytika vysílání

1. Uvedení do situace

Český rozhlas (ČRo) disponuje archivem kontinuálního vysílání i jeho popisem metadaty, která jsou do určité míry použitelná, ale jsou nepřesná a v mnoha případech obsahují nesrovnalosti či často zcela chybí. Záměrem je tedy archiv kontinuálního vysílání analyzovat a přesně zpracovat tak, aby uživatelé internetu mohli přistupovat přímo ke konkrétním pořadům či jejich úsekům dle přidělených práv. Takto naučený systém by měl následně zpracovávat živé vysílání - tedy zpracovávat vysílání v reálném čase tak, aby tento archiv byl automaticky rozšiřován a zpřístupňován online.

Analýza vysílání je pouze první polovinou celkové analýzy, jejíž druhou částí je Analýza mluveného slova, tedy převod hlasu na text s rozlišením mluvčích či genderu a jejich případnou identifikací. Druhá část bude řešena samostatnou veřejnou zakázkou. Celkovým výsledkem, ve kterém dochází k propojení obou částí, je tedy podrobný popis zvukové stopy včetně převodu hlasu na text a to všech stanic ČRo. A tento výstup, tedy archiv, vytváření archivu přímo ze živého vysílání a streamy živého vysílání budou zpřístupněny dle práv uživatelům na internetu v rámci projektu můjROZHLAS.

Projekt můjROZHLAS je strategickým krokem ČRo a bude největším českým audio archivem na internetu.

První fáze projektu je definována milníkem, tedy spuštěním systému v květnu 2019 dle harmonogramu prací v příloze.

Analytika vysílání je založená na vývoji doposud neexistujícího softwaru, ČRo proto vyžaduje vývoj v postupných krocích (viz Use case bod 2.3). Vývoj takového software a jeho univerzální využití s sebou přináší i další obchodní příležitosti.

2. Stručný popis zakázky

Výroba a dodání systému pro zvukovou analýzu (dále systém), který rozpracovává vysílání / zvukový záznam na úseky, jež identifikuje, označuje, a případně zařazuje do vyšších celků, aby bylo možné k jednotlivým úsekům přistupovat napřímo. Přiřazení práv a dalších metadat provede ČRo.

Tento systém pro zpracování využívá dostupná metadata o vysílání (popis níže) případně obvyklé struktury pořadů i neočekávané situace a situace nezaznamenané v metadatech. Dodavatel zajistí podporu a vývoj systému po dobu trvání této zakázky.

2.1 Proč Český rozhlas potřebuje analýzu vysílání

Český rozhlas potřebuje získat přesný popis zvukové stopy odvysílaného proudu vysílání proto, aby s ním mohl dále nakládat dle práv a licencí, dále i pro zpětnou práci s archivem našich zaměstnanců či spolupracovníků.

Příklad 1. epizoda pořadu Meteor

Meteor se obvykle skládá z promluvy dvou moderátorů, reportáží z terénu, příspěvků spolupracovníků, hudebních vložek či skladeb, zvukové grafiky (znělka, jingly...), četbou na pokračování nebo rozhlasovou hrou. Podle práv a licencí může ČRo na internet umístit obvykle na 7 dnů celý pořad, po této lhůtě je možné zveřejnit vše kromě četby na pokračování, rozhlasové hry a hudební skladby. Tyto

*části je nutné „vystříhnout“ s tím, že v případě opětovného zakoupení práv je možné tyto části opět „vložit“. Analýza vysílání nám umožní s jednotlivými úseky vysílání nakládat automaticky, rychle a přehledně.
Bez tohoto vystřížení by byla celá epizoda pořadu nedostupná.*

Příklad 2. rekonstrukce událostí srpen 1968

Rekonstrukce probíhala 13 hodin a ČRo chce celý záznam zveřejnit, ale bez zpravodajských relací, které celé vysílání protínají v celou hodinu, ale i po půlhodinách. Systém nám umožní všechny relace jednoduše „vystříhnout“ a celý blok vysílání nabídnout posluchačům na internetu.

2.2 Proč to Český rozhlas nedělá jinak?

Český rozhlas disponuje metadaty (viz níže), pomocí kterých zvládne identifikovat většinu vysílání a tento systém si vyvine vlastními silami jako záložní (backup i fallback) systém.

V metadatech ale nejsou zahrnuty např. telefonáty do vysílání či jiné vstupy. Je-li vysíláno z předpřipravených pořadů (tzv. konzerv) tak známe začátek a případně konec, ale ne už obsah apod. Technika a technologie ČRo neumožňují vysílání jednoduše vícekanálově, stejně tak vkládání skrytých značek do zvukové grafiky. Takové postupy navíc neřeší zpracování archivu vysílání ani jeho otevřenost pro další vývoj.

Takto připravený systém připraví analytiku na budoucnost a umožní její uplatnění i pro jiné společnosti mimo Česko. Systém tedy v sobě zahrnuje byznysový potenciál nejen pro dodavatele řešení.

2.3 Vývoj systému

Níže popsaná specifikace v bodech 3. a více odkazuje k finálnímu produktu. Vývoj by měl probíhat na základě analýzy zvukové stopy, postupného a podrobnějšího rozlišování. Český rozhlas dodá veškerou dostupnou zvukovou grafiku s metadaty (znělky, jingly, podkresy...), které zjednoduší analytiku.

2.3.1 Use case vývoje analytiky na úseky (Dodavatel a ČRo):

1. Příjem vysílání (signálu) ano / ne
2. Ticho / zvuk
3. Promluva¹ / ostatní
4. Promluva / zvuková grafika / hudba / ostatní
5. Promluva / identifikovaná zvuková grafika / hudba / časové znamení / ostatní
6. Crossover²: promluva / identifikovaná zvuková grafika / hudba / časové znamení / promo a reklama / ostatní
7. Promluvy rozdílných úseků³ / identifikovaná zvuková grafika / hudba / časové znamení / promo a reklama / ostatní
8. Přiřazení práv úsekům⁴
9. Sloučení úseků do vyšších celků⁵
10. Přiřazení práv vyšších celků⁴

Následuje zpracování loggingu dle harmonogramu pro zpracování loggingu.

Poznámky:

¹ Promluva se řeší v rámci jiné veřejné zakázky, a to Analytiky mluveného slova (AMS). V rámci této analytiky budou rovněž řešeny identifikace úseků. Jde o identifikaci a označení zvuku jakožto lidského hlasu.

² Crossover je situace běžná ve vysílání, ve které přes sebe jde hlas, hudba či zvuková grafika.

Tzn., že promluva začíná během písničky, systém tak zaznamená konec písně až v průběhu skladby. Údaje o začátcích a koncích úseků se tedy budou překrývat.

³ Rozlišení nového úseku bez oddělovací znělky / jinglu. Např. Zprávy v celou hodinu se skládají z promluvy zprávaře a příspěvků redaktorů. Bude řešeno i s pomocí Analýzy mluveného slova.

⁴ Proveďte ČRo.

⁵ Proveďte dodavatel a ČRo, přesné vymezení vyplývá z vývoje systému.

2.3.2 Lidský faktor

Ideální systém by měl analyzovat vysílání bezpečně bez zásahu člověka. Ač je to cíl, nepředpokládáme takovou funkčnost. ČRo předpokládá, že bude nutné analytiku ladit ručně během provozu ať už na straně dodavatele nebo na straně ČRo. Nicméně součástí hodnocení bude nenáročnost na lidské úsilí při analytice vysílání a snižování podílu lidské práce na analýze.

2.3.3 Analýza mluveného slova

Pomocí analýzy mluveného slova (AMS), které je předmětem jiné veřejné zakázky, bude ČRo rovněž provádět analýzu vysílání, respektive ji upřesňovat zvláště tam, kde příspěvky nejsou jednoznačně odděleny zvukovou grafikou či v programu uvedeny. Analytika vysílání tedy není

2.3.4 Harmonogram a vývoj

Český rozhlas si je vědom složitosti vývoje a výzkumu systému pro AV. Dodavatel vnímá jako důvěryhodného partnera při vývoji zcela nového SW.

Use case a níže uvedená specifikace s načrtnutým harmonogramem představují ideální postup a scénář, který je vzhledem k neznámým překážkám spíše nemožné splnit. ČRo tedy vzhledem k povaze zadání předpokládá, že dodavatel vynaloží úsilí odpovídající finančním prostředkům a důvěře v něj vložené pro dosažení všech kroků uvedených ve 2.3.1.

2.4 Co je úkolem Dodavatele

- Dodavatel vyvíjí analytický software v úzké spolupráci s ČRo dle uvedených Use case (2.3.1).
- Dodavatel vyvíjí API dle níže uvedených požadavků v úzké spolupráci s ČRo.
- Dodavatel nadefinuje HW a OS, který ČRo zakoupí a bude na něm systém provozovat.
- Dodavatel průběžně implementuje software na HW v ČRo.
- Dodavatel poskytuje po dobu určenou support systému, v začátcích a při nasazování nových modulů 24/7 a posléze 12/7.

3. Detailní popis zakázky

3.1 Předmět zpracování

1. Živé vysílání všech stávajících i budoucích stanic ČRo (technický popis viz. bod 6, meta data bod 4)
2. Archiv kontinuálního vysílání ČRo od roku 2003, tzv. logging (technický popis viz bod 5, meta data bod 4)
3. On-demand požadavky zpracování zvuku se zavedením do online systému projektu můjROZHLAS. On-demand požadavky se budou vyžadovat při doplňování nebo úpravách online archivu. Obvykle se použije pro opravu již jednou analyzovaného zvuku.
 - a. Jednorázové zpracování s možností zavedení do systému projektu můjROZHLAS
 - b. Dávkové zpracování s možností zavedení do systému projektu můjROZHLAS

3.2 Popis zakázky, cíle

Cílem je vytvořit a zprovoznit inteligentní specializovaný systém pro analýzu vysílání. Systém by měl být vytvořený na základě strojového učení z dostupných audio souborů i metadat (*schéma stanice - program, rundown a XML s aktuálním vysíláním viz níže, přepisu mluveného slova i historie vysílání případně dalších zdrojů – popsány níže*), který bude dodávat níže popsané výstupy. Systém bude trénován na archivních datech tak, aby mohl reagovat i na nenadálé události, nebo v metadatech nezavedené změny ve vysílání s tím, že změnu dokáže sám identifikovat a označit. Ve výjimečných případech změnu označí odhadnutou identifikací či ničím a zařadí do řady požadavků ke schválení administrátorem. Ruční opravou nebo úpravou zavádí admin nové pravidlo do systému pro další analýzu zvuku. Součástí by měly být zvukové profily vysílání, které urychlí identifikaci zvuků, online doplňovaná databáze zvuků a případně hudby.

1. **Analytika živého vysílání** (*předmět č. 1*) je real-time vytváření popisu zvukové stopy a jeho předávání další službě ve strojově čitelné podobě. Tedy na časovou osu (timecode) s přesností na milisekundy budou vynášeny identifikační popisy obsahu s právy.
 - a. U živého vysílání bude v druhé části zakázky - AMS - probíhat tzv. hloubková kontrola. Tedy opětovná offline analýza s použitím větších slovníků tak, aby se původní analýza ze živého vysílání upřesnila, případně opravila.
V případě, že bude v Analýze vysílání docházet k nepřesnostem z důvodu nemožnosti přesnější analýzy živého vysílání, by systém měl po skončení pořadu provést hloubkovou a přesnější analýzu takovou, aby přesnost začátku a konců byla maximálně v řádech desítek ms.
 - b. Zpoždění výstupů je možné v řádu vteřin oproti živému vysílání.
2. **Zpracování archivu kontinuálního vysílání** (*předmět č. 2*) je popis obsahu zvukové stopy ve strojově čitelné podobě. Tedy na timecode s přesností na jednotky milisekund budou nanášeny popisy obsahu s právy a jejich kompletování do vyšších celků. Na zpracovaný archiv musí navazovat analytika živého vysílání. Tedy archivní záznamy se zpracovávají do té doby tak, aby navázaly na spuštěnou analytiku živého vysílání.
3. **On-demand** (*předmět č. 3*) zpracování zvuku je možnost zakomponovat opravy či výpadky analytiky záznamu vysílání a jejich automatizované zavedení do online vytvářeného archivu ať už doplněním nebo přepsáním již zavedeného. Případně zavést zcela nové části vysílání z archivu.
4. **Zvukové profily** jsou soubory informací o konkrétním pořadu, které usnadňují identifikaci zvuků. ČRo předpokládá, že takovou databázi musí dodavatel nutně připravit, pokud systém má být funkční.
5. **API** – viz bod 3.6

3.2.1 Návrh předpokládané workflow systému:

Jde pouze o návrh finálního produktu pro analýzu archivu a živého vysílání. Dodavatel může přijít s vlastním řešením, které bude lépe vyhovovat zadaným úkolům. Zde je popsána finální podoba systému propojující vývoj na straně dodavatele a na straně zadavatele.

Workflow systému / komponenty:

1. **Vytvoření pracovního rámce**
 - a. Ze schématu vysílání, systému Rundown a případně XML aktuálního vysílání si systém připraví pracovní rámec pro rychlejší analýzu. Rámec bude obsahovat meta data pořadu, předpokládaný zvukový profil pořadu pro rychlejší identifikaci zvukových úseků.

- b. Systém by měl primárně pro rychlou analytiku pracovat s daty v pracovním rámci. V případě, že se nepodaří úseky identifikovat, systém na pozadí zkusí dohledat potřebné zvuky. Nalezené zařadí jak do pracovního rámce a tak i zpětně online doplní do výstupů.
- c. U nenadálých změn ve vysílání – změny nejsou zaneseny metadatech ČRo, použije systém pracovní rámec vytvořený speciálně pro tyto události.
 - i. Pracovní rámec nenadálých událostí by měl obsahovat data získaná z analýzy archivu, obvykle se formát vysílání mění na all-talk, tedy bez hudby s promluvou expertů, zahraničních zpravodajů, svědků a reportérů z místa apod.
- d. Zvukový profil pořadu by měl systém aktualizovat automaticky nebo na základě zákroku admina.

2. Identifikace úseků

- a. Detailní technický popis výstupu viz bod 3.3 Identifikace úseků – Popis.
- b. Systém bude využívat k identifikaci dostupné online zdroje ČRo (viz bod 4)
 - i. Systém by měl být připraven i na případné využití i výstupů z Analýzy mluveného slova, tedy přepisů hlasu na text.
- c. Metoda identifikace dle návrhu dodavatele. Ideálně zavedená metoda, např. AudioPrint aj.
- d. Identifikace úseků a jeho nadřazených kategorií bude probíhat i zpětně, tedy tak, aby zpočátku neidentifikovaný zvuk byl doplněn o identifikaci a zařazen zpětně do online archivu systému můjROZHLAS s identifikací (zvláště u analytiky živého vysílání)
- e. Z pohledu základní kategorizace jde o:
 - i. Identifikátor promluvy
 - 1. Začátky a konce promluv, mohou se překrývat se zvuky
 - ii. Identifikátor zvuků a hudby
 - 1. Zvuková grafika pořadu
 - 2. Hudba (píseň ...)
 - 3. Zvuková grafika stanice
 - 4. Speciální zvuky (hromy, dusot kopyt...)
- f. Zvuky neidentifikované by měl systém umět odvozovat a řadit sám dle dostupných metadat, pracovního rámce schéma), případně předávat adminovi k odsouhlasení.

3. Zvukové profily

- a. Na základě analýzy archivu a podkladů dodaných ČRo může dodavatel vytvářet zvukové profily pořadů / bloků
- b. Zvukový profil se proměňuje v čase a systém tuto proměnu v profilu zachovává
- c. Zvukový profil může obsahovat další zvukový profil
- d. Obsah zvukového profilu:
 - i. Zvuková grafika:
 - 1. znělka úvodní,
 - 2. jingl,
 - 3. podkresová hudba XX,
 - 4. znělka závěrečná,
 - 5. časové znamení,
 - 6. ...
 - ii. Schéma pořadu bloku snižuje nebo rozšiřuje oblast vyhledávání, např. pořad Pro a proti má schéma: úvodní znělka, podkresová hudba, moderátorka, host 1, host 2, jingl, případně anonce (selfpromo - vzácně) či telefonní hovor (opět

vzácně), závěrečná znělka, reklamní blok, anonce... Pokud se v pořadu objeví neznámý zvuk, bude jeho identifikace na pozadí probíhat nejdříve v rámci schématu.

- iii. IDs
 - 1. Unikátní ID pořadu pro systém
 - 2. Jednotlivé zvuky
- iv. Verzování hlasového profilu
 - 1. Profil
 - 2. Zvuky
 - 3. Schémata

4. DB zvuků

- a. Ve spolupráci s ČRo vytvoří DB používaných zvuků (vše mimo písní / hudebních skladeb chráněných právy třetích stran – Instagram, Dilia, OSA...). ČRo dodá dostupné podklady, které má už v určité podobě částečně zpracované.
- b. Zvuky jsou přiřazovány do zvukových profilů, ke zvukům se přistupuje také napřímo mimo zvukové profily.
- c. Jeden zvuk může být v libovolném počtu zvukových profilů.

5. Hudba

- a. Většina pořadů / bloků obsahuje hudební části (vokální a vokálně instrumentální hudba, instrumentální hudba a případně scénická hudba).
- b. Hudební skladby použité v přednatočených pořadech nejsou vždy zaneseny do systémů ČRo před odvysíláním, často zapsány až zpětně, takže není možné je identifikovat online, často bývají zapisovány napřeskáčku (v rámci pořadu).
- c. Systém by měl rozpoznat začátky a konce jednotlivých hudebních (zvukových) částí.
 - i. Např. Záznam operního představení je jedno dílo a tedy jeden zvukový úsek, avšak ukázka z opery v rámci pořadu je rovněž jeden zvukový úsek.
- d. Cílem systému není, vytvořit databázi skladeb a identifikovat je dle např. AudioPrintu. Nicméně pokud existuje technické řešení, které takovou databázi umožní vytvořit a případně propojit s veřejně dostupnými zdroji pro identifikaci hudby, umožní tak systému pracovat efektivněji.
 - i. Systém tedy musí být připraven pro zavedení identifikátoru hudebních částí.

3.2.2 Přístupnost výsledků analýz

- API, které umožňuje získávat data ze systému a rovněž vkládat data do systému. Bližší technický popis níže bod 3.6.

3.2.3 Mimořádné a aktualizací situace

- 1. U všech tří předmětů zpracování systém kontroluje zvukovou stopu i po technické stránce tak, aby identifikoval technické chyby – především jde o výpadky zvuku (ticho, zkreslený zvuk), chybějící / nenavazující soubory atd.). Tyto chyby eviduje a upozorňuje administrátora.
- 2. Systém je připravený na změny ve vysílání ČRo, např.:
 - a. Zavedení nové vysílací stanice
 - b. Přejmenování či zavedení nové jmenné konvence úseků / pořadů / bloků
 - c. Zavedení nových úseků / pořadů / bloků
 - d. Změny ve vysílacích schématech
 - e. Zrušení vysílání úseků / pořadů / bloků
 - f. Zrušení / přerušení (dny / měsíce) vysílání stanice
 - g. Přejmenování vysílací stanice

3. Systému musí být připraven na případné výpadky a to tím způsobem, aby mohl využívat logging či minutové záznamy ke zpětnému doplnění dat.

3.2.4 Vývoj a podpora

Součástí dodávky je správa a vývoj systému pro Český rozhlas na dobu dokončení, předání kompletní zakázky. Systém bude provozován v Českém rozhlasu na serverech dle dohody se zadáním dodavatele a Oddělení informačních technologií Českého rozhlasu.

3.2.5 Systém – licence a vývoj

ČRo jakožto zadavatel a plátcé zakázky má zájem na dalším vývoji a komerčním uplatnění vyvinutého systému a jeho šíření. Prodej a implementace systému mimo ČRo budou možné pouze při splnění podmínky poplatku za celou uzavřenou zakázku splatnou do 3 měsíců od uzavření smlouvy se třetí stranou.

3.3 Návrh na datový výstup analytiky živého vysílání

3.3.1 Data

- Formát: JSON
- Metoda předání do systému: dle návrhu dodavatele
- Frekvence:
 - Data budou vystavovaná real-time, tedy ihned k okamžitému zpracování - při každé změně ve vysílání
 - Vystavování dat musí být ošetřeno tak, aby nedošlo ke ztrátě doposud nezpracovaného souboru
 - Lze řešit zavedením posledních pěti úseků do JSONu
- Struktura dat: dle návrhu dodavatele schválené zadavatelem
- Ochrana stability před změnami v čase:
 - zimní/letní čas
 - přestupný rok
 - „rozjetí času“ – technické chyby, při které by došlo k nanášení meta dat na nesprávný timecode

3.3.2 Identifikace úseků

Soubor s identifikačními údaji a jeho předpokládaná struktura – úpravy dle dohody. Soubor by obsahoval posledních pět úseků pro zajištění návaznosti v případě výpadku či rychlého sledu technických událostí v rámci vysílání.

1. Název souboru: jmenná konvence dle stanic <https://api.rozhlas.cz/mujROZHLAS/temp/>
2. ID stanice: jmenná konvence dle stanic <https://api.rozhlas.cz/mujROZHLAS/temp/>
3. ID úseku: inkrementální číslo identifikující úsek
 - a. Stav úseku – jmenná konvence dle dohody
 - i. Dokončenost úseku
 1. Dokončený
 2. Dokončený neúplný
 3. Rozpracovaný
 4. Dokončenost neznámá
 - ii. Identifikace úseku – jmenná konvence dle dohody
 1. Identifikován
 2. Odhad – nutný zásah admina
 3. Neidentifikován – nutný zásah admina

4. Identifikace neproběhla
- b. Datum a čas: publikace souboru
 - i. Formát YYYY-MM-DDTHH:MM:SS.MSS
- c. **Nadřazené kategorie** – do čeho úsek spadá
 - i. Nadřazená kategorie
 1. Pořad, reportáž v pořadu...
 2. Práva pořadu
 - a. Typ:
 - i. Stream, download, zakázané
 - b. Časové vymezení práva:
 - i. Bez omezení
 - ii. Od–do / do
 - iii. Geolokační omezení
 - c. Majitel práv
 - i. Agentura / ČRo / volné
 - ii. Nadřazená nadřazené kategorie
 1. Blok, pořad, reportáž v pořadu...
 2. Práva pořadu
 - a. Typ:
 - i. Stream, download, zakázané
 - b. Časové vymezení práva:
 - i. Bez omezení
 - ii. Od–do / do
 - iii. Geolokační omezení
 - c. Majitel práv
 - i. Agentura / ČRo / volné
 - iii. Původní kategorie
 1. ID pořadu / bloku, kterého byl úsek původně součástí
 - a. ID je dle jmenné konvence pro identifikaci konkrétního pořadu
- d. **Úsek**
 - i. Typ úseku
 1. Typ:
 - a. Promluva – případně rozlišení mluvčích
 - i. Mluvčí A
 1. ID mluvčího
 - ii. Mluvčí B
 1. ID mluvčího
 - b. Hudební skladba /píseň
 - i. Autor...
 - c. Znělková grafika
 - i. Úvodní
 - ii. Jingle
 - iii. Podkres...
 - d. Reklama
 - e.
 2. Označení typu: dle dohody na jmenné konvenci
 - ii. Práva úseku – úsek

1. Typ:
 - a. Stream, download, zakázané
2. Časové vymezení práva:
 - a. Bez omezení
 - b. Od–do / do
 - i. Formát: : YYYY-MM-DDTHH
3. Geolokační omezení
- iii. Délka úseku
 1. Start
 - a. Formát: YYYY-MM-DDTHH:MM:SS.MSS
 - b. Např.: 2018-03-08T17:26:10.656
 2. Konec
 - a. Formát: YYYY-MM-DDTHH:MM:SS.MSS
 - b. Např.: 2018-03-08T17:26:10.656

3.3.3 Zálohování dat

Data budou paralelně ukládána na jiné úložiště tak, aby byla v případě výpadku dohledatelná a použitelná pro opětovné zavedení do systému projektu mujROZHLAS. Data mohou být ukládána na úložiště ČRo, dle dohody.

3.4 Kroky vývoje analytiky

Vývoj analytiky zvuku ve vysílání proběhne na straně dodavatele na základě zpracování jednotlivých use case viz bod 2.3.1.

Na základě probíhajícího vývoje analytiky vysílání se bude

- rozšiřovat výstup API (3.6),
- měnit rozsah lidského zásahu (2.3.2).

3.5 Implementace a spuštění, zdrojový kód a dokumentace

3.5.1 Implementace analytického systému

- Dodavatel systém nainstaluje, spustí a aktualizuje dle dohodnutého harmonogramu, přičemž základní milníky jsou
 - únor 2019
 - Specifikace HW a (OS)
 - březen 2019:
 - Specifikace API
 - Use case 1-3 (beta)
 - únor 2019
 - Use case 1-3 (RC) a 4 (beta)
 - duben 2019:
 - První implementace systému na HW
 - Spuštění API
 - Use case 1, 2, 3, 4 (RC) a 5 (beta)
 - květen 2019
 - Use case 1-4 zavedení

- Spouštění projektu (ve verzi bez AV a částečné AMS)
- červen 2019
 - Use case 5 (RC) a 6 (beta)
- září 2019
 - Use case 5 zavedení, 6 (RC)
- prosinec 2019
 - Use case 6 zavedení

3.5.2 Zdrojový kód

- Zdrojový kód systému bude Českému rozhlasu předáván v poslední funkční verzi pravidelně každý první týden v kvartálu počínaje lednem 2019.
- Předání proběhne elektronickou formou stažením, zasláním nebo z přenosného disku.
- Zápis kódu bude veden pro snazší srozumitelnost sebepopisně.
- Převzetí zdrojového kódu bude stvrzeno Předávacím protokolem písemnou nebo elektronickou formou.

3.5.3 Dokumentace

Dokumentaci není nutné vést v plné podobě, ale v podobě rozepsání základní koncepce a struktury zdrojového kódu (do ca 4 stránek formátu A4), detaily funkcí a modulů by měly být zřejmé ze sebepopisného kódu.

3.5.4 Implementace analýzy loggingu

Dle termínů zavádění jednotlivých Use case.

3.6 API

Pro správu projektu připravuje zadavatel na vlastní straně administrační rozhraní provázané s přístupy do v ČRo používaného redakčního systému Drupal. Pro administraci je musí dodavatel připravit API tak, aby administrátoři mohli:

- zadávat on-demand požadavky a zpětně je zavádět do systému projektu můjROZHLAS,
- schvalovat, opravovat a zpětně upravovat identifikace úseků s možností hromadného schvalování,
- dávkově doplňovat a upravovat knihovny používané k identifikaci úseků, pořadů a příslušných práv.

Samotná Analytika vysílání bude rovněž přístupná přes API, je ale možné, po vzájemné dohodě a pro rychlejší zpracování dat zavést i jiný způsob. Návrh na formát dat viz 3.3

Z těchto důvodů musí dodavatel zprovoznit API dle harmonogramu prací, viz bod 3.4.

3.7 Předání zpracovaných metadat z loggingu

Předávání metadat získaných z analýzy loggingu bude prováděno v digitální podobě dle harmonogramu skrze rozhraní - API dle bodu 3.6 případně jiným způsobem dle vzájemné dohody.

4. Data k analýze vysílání

4.1 Stručný přehled zdrojů metadat dle předmětu zpracování

Český rozhlas přechází v roce 2019 na nový vysílací systém Anova, některé z níže uvedených zdrojů proto již nemusí být dostupné a k dispozici budou nové zdroje zmíněného systému Anova. Dodaný systém tedy musí s touto změnou zdrojů dat počítat, být připravený na změnu struktury a zdroje metadat.

1. Živé vysílání
 - a. Přístupy k živému vysílání
 - i. Dante protokol (AoIP) v místě zadavatele (6)
 1. Analytiku je možné provádět i na streamu ČRo, dostupné na www.rozhlas.cz nebo www.play.cz
 - b. Meta data k analytice zvuku živého vysílání
 - i. Vysílací schémata stanic (4.2), JSON, veřejně přístupné API
 - ii. Rundown (4.3) – podrobné schéma vysílání stanic, zpožděné o ca 1 min, JSON, veřejně přístupné API
 - iii. Real-time data o vysílání (4.4), real-time XML, připravíme dočasný veřejný přístup
 - iv. Annova – případné výstupy z plánovaného systému (ca od r. 2019) pro vysílání stanic ČRo (<https://www.annova.tv/en/products/index.php>)
 - c. Zvuková grafika stanic a pořadů
2. Archiv
 - a. Meta data k analytice zvuku živého vysílání
 - i. Vysílací schéma stanic, JSON, veřejné API
 - ii. Rundown – podrobné schéma vysílání stanic, zpožděné o ca 1 min, JSON, veřejné API
 - iii. Archiv real-time dat o vysílání, real-time, XML, od roku 2012 (4.4)
 - b. Zvuková grafika stanic a pořadů
3. On-demand
 - a. Meta data viz výše (4.1.2)

4.2 Vysílací schémata pro konkrétní den a stanici

Přístupné mimo síť ČRo. Aktualizuje se každých ca 12 minut.

Formát: JSON

Struktura URL pro získání výpisu:

1. <https://api.rozhlas.cz/mujROZHLAS/temp/schedule/day/>
2. + Rok – formát YYYY
3. + Měsíc – formát MM
4. + Den – formát DD
5. + Název stanice – formát xxxxxxxx.json
 1. Číselník názvů stanic: <https://api.rozhlas.cz/mujROZHLAS/temp/meta/stations.json>

Příklad: <https://api.rozhlas.cz/mujROZHLAS/temp/schedule/day/2018/03/01/radiozurnal.json>

Popis položek JSONu:

1. Timestamp: datum a čas vytvoření
2. Data:

1. Station: název stanice, bez diakritiky, viz <https://api.rozhlas.cz/mujROZHLAS/temp/meta/stations.json>
2. ID: identifikační číslo ze systému AIS (provázání mezi systémy)
3. Title: pracovní název, české znaky v UTF
4. Description: Obsah příspěvku / popis, české znaky v UTF
5. Since: spuštění příspěvku, formát: 2018-03-14T00:00:57+01:00
6. Till: ukončení příspěvku, formát: 2018-03-14T00:01:52+01:00
7. Type: data z AISu, nejspíš tam někde bude číselník
 - ID: ID v rámci číselníku
 - Code: třípísmenné označení
 - Name: pojmenování
8. Edition: pořad, české znaky v UTF, data z CMS (aktuálně RSCR)
 - ID: ID pořadu
 - Profile: URL stránky pořadu
 - Archive: URL archivu pořadu
 - Asset: URL obrázku pořadu
9. Persons: osoby v příspěvku, data z CMS (aktuálně RSCR)
 - ID: ID osoby v systému RSCR
 - Name: jméno osoby, české znaky v UTF
 - Profile: URL profilu osoby
 - Asset: URL obrázku osoby

4.3 Rundown pro konkrétní den a stanici:

Výpis odvysílaného programu, přístupné mimo síť ČRo. Aktualizuje se následovně:

Pro protokoly stanic, které vysílají z DaletPlus je aktualizace prováděna každou minutu, pro logy z CartMaster a Dalet 5.1 je aktualizace prováděna časně ráno a je importován minulý den.

Formát: JSON

Struktura URL pro získání výpisu:

1. <https://api.rozhlas.cz/mujROZHLAS/temp/broadcastlog/day/>
2. + Rok – formát YYYY
3. + Měsíc – formát MM
4. + Den – formát DD
5. + Název stanice – formát xxxxxxxx.json

Příklad: <https://api.rozhlas.cz/mujROZHLAS/temp/broadcastlog/day/2018/03/01/radiozurnal.json>

Popis položek JSONu:

1. Station: název stanice, bez diakritiky, <https://api.rozhlas.cz/mujROZHLAS/temp/meta/stations.json>
2. ID: identifikační číslo inkrementální ID z DB AISu
3. Title: pracovní název, české znaky v UTF
4. Author: autor příspěvku, bez diakritiky
5. Since: spuštění příspěvku, formát: 2018-03-14T00:00:57+01:00
6. Till: ukončení příspěvku, formát 2018-03-14T00:01:52+01:00
7. Code: kódové označení - tzv. itemkód, slouží pro jednoznačnou identifikaci zdrojového zvuku v příslušných systémech

4.4 Aktuální vysílání - XML

Změní se vždy při změně obsahu vysílání. Položka se přepíše a vystaví aktuální a následující. Zaznamenává pouze spuštění zvukových záznamů – nerozpozná telefonát.

Formát: XML, verze 1.0, UTF-8

URL pro získání výpisu

Data nejsou veřejně přístupná. Ukázka dat je v příloze zadávací dokumentace.

Popis položek XML:

- BroadcastMonitor – hlavní obalující tag
- Updated: čas posledního update souboru v ms
 - Formát: 2018-03-08T17:26:10.656
- StationName: název viz <https://api.rozhlas.cz/mujROZHLAS/temp/meta/stations.json>
- Current: aktuálně/naposledy vysílaný zvukový záznam
 - StartTime: čas spuštění záznamu v ms
 - Formát: 2018-03-08T17:26:10.656
 - ItemID: ID položky v playlistu, nepodstatné
 - Title ID: ID položky v DB, nepodstatné
 - ItemID: ID položky pro spárování s dalšími systémy
 - Řídí se jmenovou konvencí (*příloha: Příloha - Jmenná-konvence-ITEMCODE.doc a Příloha - Jmenná-konvence-typy-zvukovych-snimku .doc*)
 - Např.: DZB123456
 - TitleName: název příspěvku / sklady
 - ArtistName: název autora / interpreta
 - Author: jméno technika příspěvku, nikoliv autora, nepodstatné
 - AlbumName: název alba, nevyplňuje se
 - Label: nevyplňuje se
 - Year: rok vzniku
 - ItemDuration: celková použitelná délka příspěvku v sekundách
 - ItemDurationMS: celková použitelná délka příspěvku v milisekundách
 - CategoryID: ID kategorie v Daletu, nepodstatné
 - CategoryCode: kód kategorie, nepodstatné
 - CategoryName: název kategorie, číselník ca 7000 záznamů
 - Např.: Hudba – Selector
 - CategoryShortName: název kategorie s plnou cestou
 - CategoryLongName: název kategorie
 - CategoryFullName: cesta kategorie
 - SoundFileName: cesta ke zvukovému souboru
 - Comment1: případný komentář
 - Interface: interní data, nepodstatné
 - AMFiled type=Transfer_Reporting, nepodstatné
 - Name: název položky
 - Value: hodnota položky
 - AMFiled type=DateTime
 - Name: název položky
 - Value: datum založení souboru
 - AMField, nepodstatné

Příklad výpisu XML:

```
<?xml version="1.0" encoding="utf-8" ?>
<BroadcastMonitor>
  <updated>2018-03-08T17:26:10.656</updated>
  <stationName>RŽ 3R3 Proud</stationName>
  <Current>
    <startTime>2018-03-08T17:26:09.021</startTime>
    <itemId>2296970:1002787372:48602829</itemId>
    <titleId>59787</titleId>
    <itemCode>1006825</itemCode>
    <Reference/>
    <titleName>WORLD OF PROSE</titleName>
    <artistName>SUPPORT LESBIENS</artistName>
    <Author> </Author>
    <albumName> </albumName>
    <Label> </Label>
    <Year>2007</Year>
    <itemDuration>215</itemDuration>
    <itemDurationMS>215206</itemDurationMS>
    <CategoryId>258</CategoryId>
    <CategoryCode>CR1SL</CategoryCode>
    <CategoryName>Hudba - Selector</CategoryName>
    <CategoryShortName>Hudba - SelectorCATEGORIES/ČRo - Radiožurnál/Hudba/Hudba -
Selector</CategoryShortName>
    <CategoryLongName>Hudba - Selector</CategoryLongName>
    <CategoryFullName>CATEGORIES/ČRo - Radiožurnál/Hudba/Hudba - Selector</CategoryFullName>
    <SoundFileName>\\192.168.26.51\RZ_ZVUKY$\001\0502B29E.mp2</SoundFileName>
    <Comment1/>
    <Interface/>
    <AMField fieldType="Transfer_Reporting" type="string">
      <name>Transfer_Reporting</name>
      <value>2010-08-17T12:34:48.000 Export-dalet_admin Export \\DPS14UT-
TEST_01\RZ_Hudba_001$ 6d304688-78dd-46b4-8175-cadda00b4917 OK</value>
    </AMField>
    <AMField fieldType="DateTime" type="string">
      <name>Aired Start Time</name>
      <value>2015-07-20T03:05:53</value>
    </AMField>
  </Current>
  <Next>
    <startTime>2018-03-08T17:29:44.409</startTime>
    <itemId>2296970:1002800584:48562576</itemId>
    <titleId>3472424</titleId>
    <itemCode>RZG3271150</itemCode>
    <Reference/>
    <titleName>01B - REKLAMA</titleName>
    <artistName/>
    <Author> </Author>
    <albumName> </albumName>
    <Label> </Label>
    <Year>2015</Year>
    <itemDuration>1</itemDuration>
    <itemDurationMS>1944</itemDurationMS>
    <CategoryId>6845</CategoryId>
    <CategoryCode>R9D</CategoryCode>
    <CategoryName>Reklamy denní</CategoryName>
    <CategoryShortName>Reklamy denní</CategoryShortName>
    <CategoryLongName>Reklamy denní</CategoryLongName>
    <CategoryFullName>CATEGORIES/Reklamy/Reklamy denní</CategoryFullName>
    <SoundFileName>\\192.168.26.50\RK_ZVUKY$\001\053606b9.wav</SoundFileName>
    <Comment1/>
    <Interface>STANDARD</Interface>
    <AMField fieldType="Boolean" type="numeric">
      <name>Story Clip Timing</name>
      <value>0</value>
    </AMField>
    <AMField fieldType="Number" type="numeric">
      <name>Story Speaker Id</name>
      <value>236685720</value>
    </AMField>
    <AMField fieldType="Boolean" type="numeric">
      <name>Story Text Timing</name>
      <value>0</value>
    </AMField>
  </Next>
</BroadcastMonitor>
```

```

</AMField>
<AMField fieldType="Boolean" type="numeric">
  <name>Story TimeGap Timing</name>
  <value>0</value>
</AMField>
<AMField fieldType="Boolean" type="numeric">
  <name>Story Transcription Timing</name>
  <value>0</value>
</AMField>
<AMField fieldType="Boolean" type="numeric">
  <name>Story Use Timing</name>
  <value>0</value>
</AMField>
<AMField fieldType="Boolean" type="numeric">
  <name>StoryTemplate TP</name>
  <value>0</value>
</AMField>
<AMField fieldType="DateTime" type="string">
  <name>Aired Start Time</name>
  <value>2015-07-19T08:29:43</value>
</AMField>
</Next>
</BroadcastMonitor>

```

4.5 Práva

Výpis práv k jednotlivým pořadovým řadám a pořadům bude přístupný online. Identifikátor úseku / pořadu / bloku a typ práv, případné časové či lokační omezení. Formát bude upřesněn po dohodě s dodavatelem.

4.6 Zvuková grafika - znělky, jingly...

Soupis používaných jinglů a znělek v příloze.

Označení je provázáno s údaji ze systému Rundown (4.3).

Dostupná zvuková grafika bude dodána v elektronické podobě v běžných audioformátech.

5. Logging - archiv kontinuálního vysílání

5.1 Archiv kontinuálního vysílání – logging ke zpracování

Informace k souborům, které obsahují záznam kontinuálního vysílání stanic ČRo.

- Soubory jsou uloženy na diskovém poli ČRo, nepřístupné mimo interní síť, není k nim tedy možné přistupovat přes internet
- Předání dat dodavateli zakázky
 - Dodavatel si data přeuloží na vlastní zařízení
 - Dodavateli bude průběžně dodáván logging k datu spuštění projektu tak, aby zpracování živého vysílání a loggingu na sebe navazovalo. Logging tedy bude průběžně dodáván dle potřeby až do spuštění projektu mujROZHJLAS.
- Popis struktury souborů
 - Ke struktuře dat obdrží dodavatel seznam souborů v elektronické podobě, které identifikují jednotlivé soubory a umožňují tak soubory seřadit dle stanic a data tak, jak bylo zaznamenáváno kontinuální vysílání.
- Meta data – viz body 4.1-4.6
- Formát souborů pro zpracování:
 - Formát: mp3 (MPEG Audio), wav (MPEG Audio) ad.
 - Délka: Hodinové záznamy
 - Celková velikost: ca 100 TB

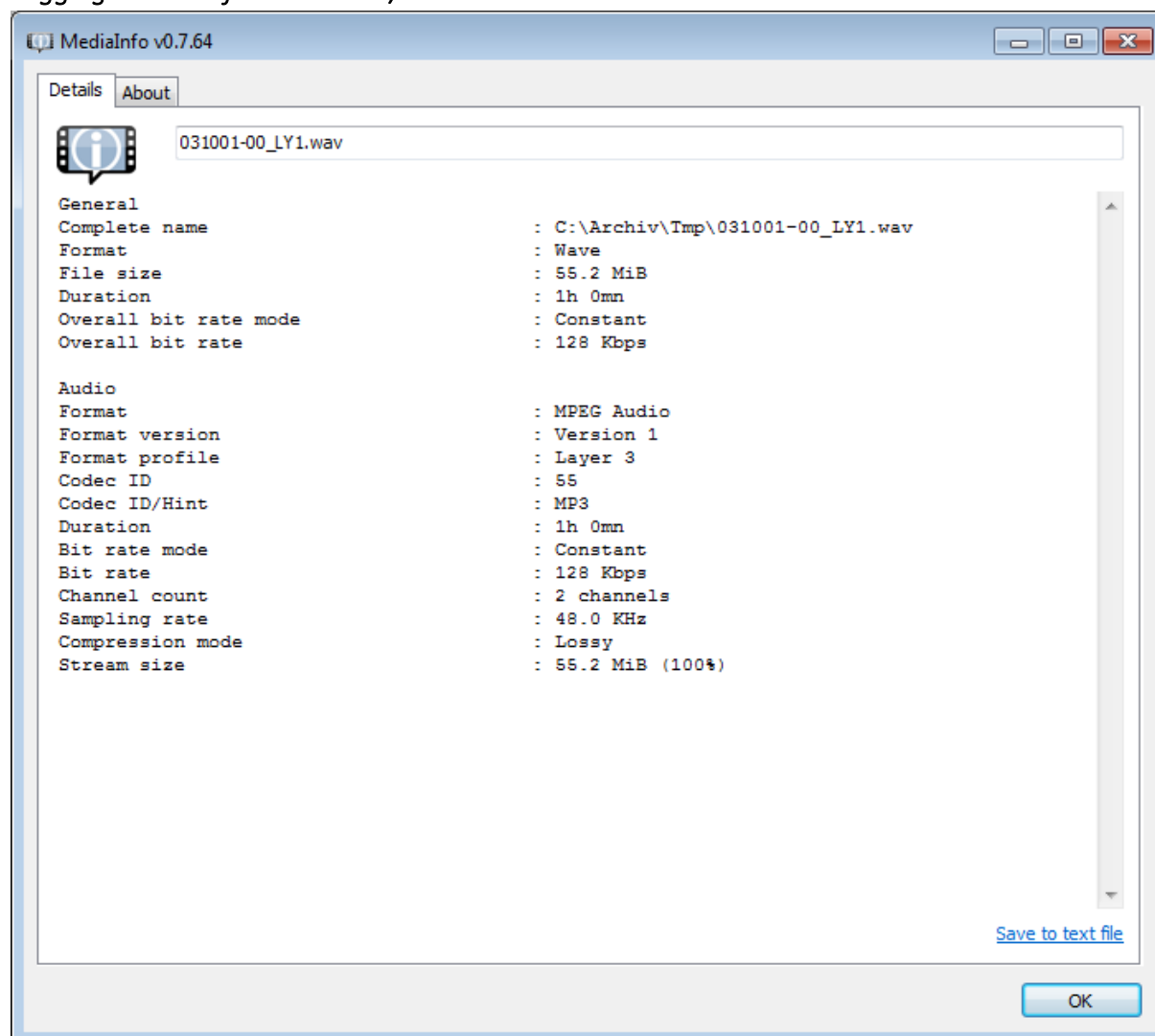
- Kvality:
 - Bitrate: 64 až 256 kbps
 - Vzorkovací frekvence: 48 kHz
 - Bit depth: obvykle 16 bit stereo, u starších mono

5.2 Přehled stanic s loggingem (jde o odhady)

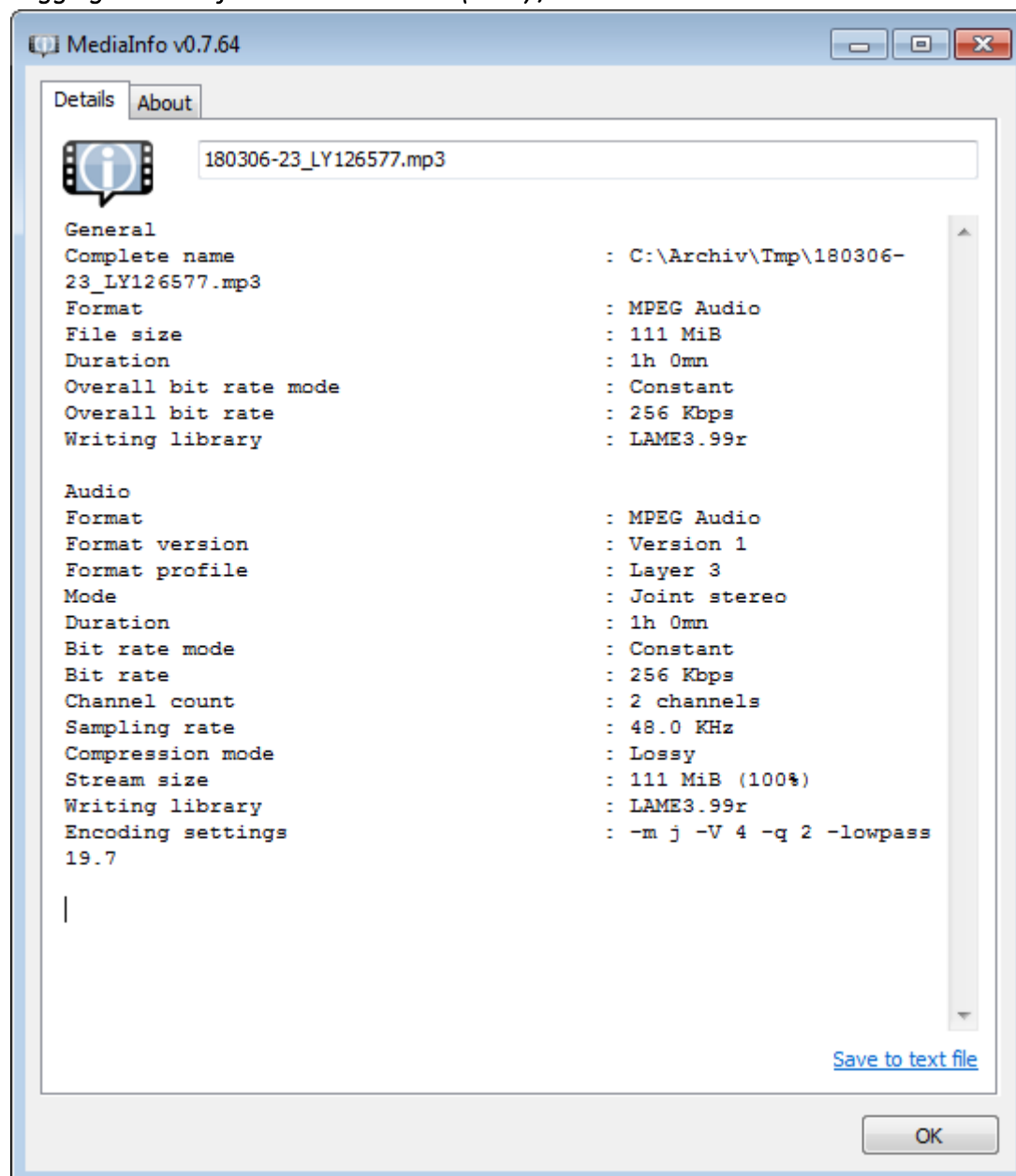
LOGGING dne:		21.2.2018		(počítáno z průměrné velikosti souborů)				
		ks	zahájeno	zahájeno	yní	změna bitrate	cca velikost	
Vysočina	LA	9075	1.7.2016	256kbps	256kbps		1 008,260	GB
Brno	LB	9065	1.7.2016	256kbps	256kbps		1 007,149	GB
Zlín	LC	1711	1.11.2017	256kbps	256kbps		190,097	GB
D-dur	LE	9074	1.7.2016	256kbps	256kbps		1 008,149	GB
Budějovice	LF	108066	1.11.2005	128kbps	256kbps	1.7.2016	6 796,841	GB
Pardubice	LG	9058	1.7.2016	256kbps	256kbps		1 006,371	GB
Hradec Králové	LH	9060	1.7.2016	256kbps	256kbps		1 006,594	GB
Liberec	LJ	34480	1.7.2016	256kbps	256kbps		3 830,833	GB
Karlovy Vary	LK	9040	1.7.2016	256kbps	256kbps		1 004,372	GB
Olomouc	LL	9085	1.7.2016	256kbps	256kbps		1 009,371	GB
RádioČesko	LM	61391	8.3.2006	192kbps	192kbps	do 13.3.2013	5 115,546	GB
Plzeň	LN	9068	1.7.2016	256kbps	256kbps		1 007,483	GB
Ostrava	LO	9041	1.7.2016	256kbps	256kbps		1 004,483	GB
Praha/Dvojka	LP	126280	1.10.2003	128kbps	256kbps	1.7.2016	7 812,767	GB
Regina	LR	107887	1.11.2005	128kbps	256kbps	1.7.2016	6 796,841	GB
Plus (Šestka)	LS	55226	1.10.2003	64kbps	256kbps	1.7.2016	4 707,659	GB
Wave	LT	106197	13.1.2006	192kbps	256kbps	1.7.2016	9 247,996	GB
Ústí nad Labem	LU	35218	1.7.2016	256kbps	256kbps		3 912,827	GB
Vltava	LV	126287	1.10.2003	128kbps	256kbps	1.7.2016	7 812,767	GB
Jazz	LW	87486	31.8.2005	128kbps	256kbps	1.7.2016	6 879,501	GB
Radiožurnál	LY	126257	1.10.2003	128kbps	256kbps	1.7.2016	7 812,767	GB
RadioPraha	LZ	126191	1.10.2003	128kbps	256kbps	1.7.2016	7 812,767	GB
Junior	L3	40012	1.3.2013	192kbps	256kbps	1.7.2016	4 040,374	GB

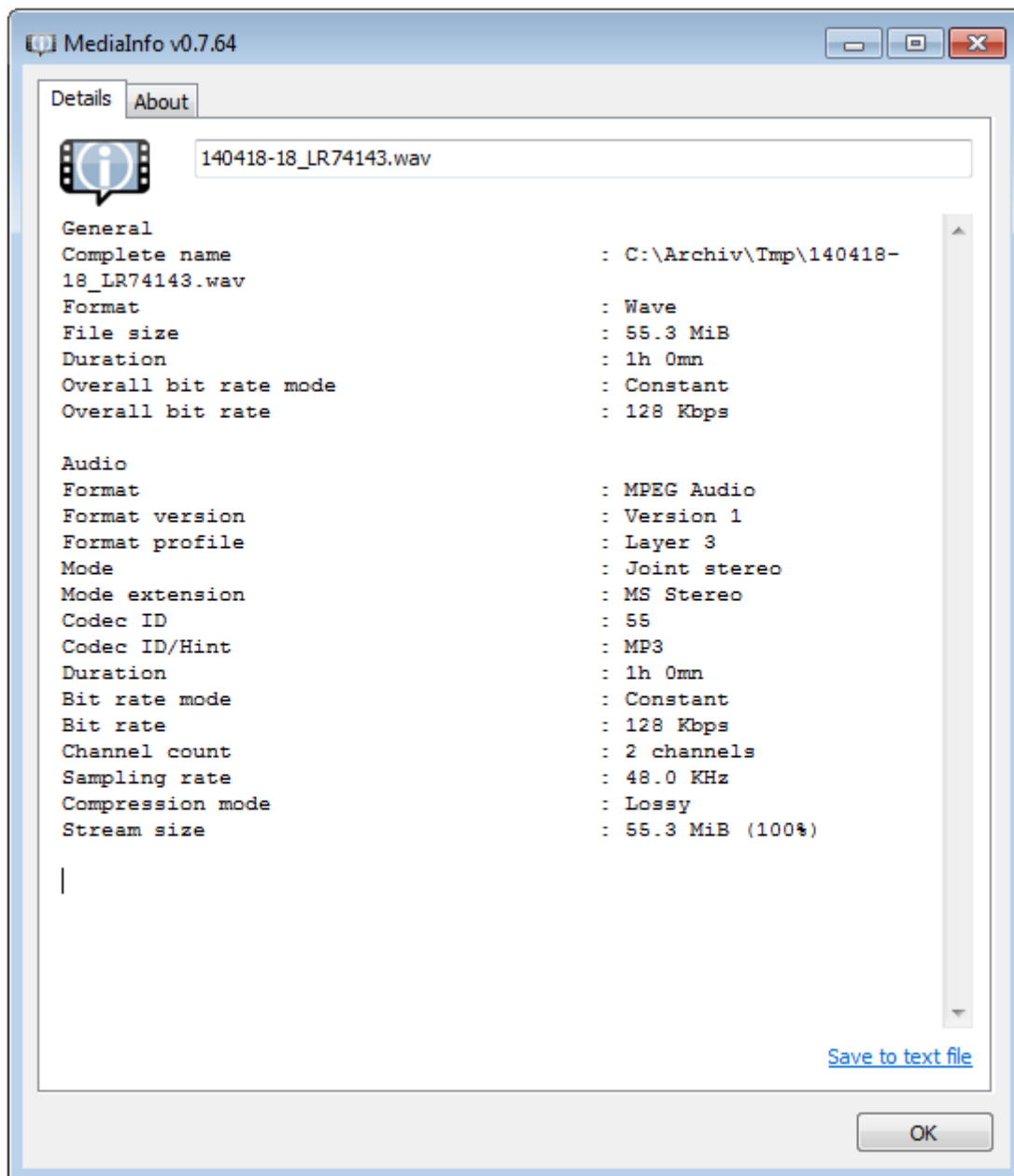
5.3 Ukázka detailního popisu souborů loggingu

Logging z r. 2003 formát Wave / MPEG Audio



Logging z r. 2018 formát MPEG Audio (MP3) / MPEG Audio





6. Živé vysílání – Dante protokol

Pro analytiku živého vysílání se musí dodavatel podřídit technickým podmínkám ČRo. Připojení na vysílání dodavatelem je možné pouze skrze rozhraní protoklu Dante v serverovně ČRo v a provádění analytiky živého vysílání je tek možné pouze sídle ČRo:

6.1 Technologie:

- Standard: AoIP – audio over IP
- Protokol: Dante (<https://www.audinate.com/products/software/dante-virtual-soundcard>)
- Vzorkovací frekvence: 16-bit audio ve frekvenci 48 kHz
- Vysílání: modulace pro jeden program je stereo
- V systému Dante je celkem 64 kanálů, tj. až 32 stereo programů

6.2 Místo k připojení:

- Serverovna v sídle Českého rozhlasu, Vinohradská 12, 120 99 Praha 2, ČR
- Serverové řešení, základní parametry viz bod 7.
- Pro testování jsou dostupné použít streamy ČRo

7. Serverové řešení

Systém bude provozován na serverech v budově Českého rozhlasu, Vinohradská 12, Praha a musí splňovat parametry níže uvedené. Ostatní nadefinuje dodavatel.

7. 1 Základní parametry serverů:

Zde uvádíme běžné provedení serverů, kterému by mělo odpovídat i řešení navržené dodavatelem. Usnadní to instalaci HW v serverovně ČRo.

- Izolované servery
- Rackové provedení, 2 zdroje
- Preferujeme Intel platformu standardních výrobců HP, Dell, Lenovo...
- 2x dedikované ethernetové porty DANTE HW kartu
- 2x ethernetové porty Dante pro síťové propojení
- Operační systémy: Windows, Linux
- Počet procesorů: 2/4
- Připojení SAN 8/16 Gb, zdvojené
- Síťovém připojení 1/10 Gb, zdvojené
- Napájení: 230 V
- Vzdálená správa
- Zdvojené řešení s minimální dobou výpadku?
- Řešení, které umožňuje zpracovávat vysílání z archivu – v případě výpadku serverů
 - Minimální doba výpadku.
 - Otevřené řešení pro možnost doplnění clusterovým řešením.

8. Podklady pro návrh řešení zakázky

ČRo nabízí zájemcům o veřejnou zakázku reálná data k přesnější kalkulaci prací a harmonogramu.

8.1 Logging – převzetí ukázek

Každý účastník má právo požádat o podklady v podobě ukázky loggingu čtyř stanic ČRo, které mu mohou umožnit lépe zpracovat nabídku. Tyto podklady vydává Český rozhlas na základě podepsaného protokolu, v němž se účastník zaváže, že poskytnutá data nebudou nijak komerčně využita, nebudou nijak šířena a budou k 31. červenci 2018 trvale a nenávratně smazána.

Logging za vybrané 3 měsíce stanic:

- Radiožurnál
- Plus
- Dvojka
- ČRo Brno

8.2 Převzetí přístupů k datům

Součástí podkladů jsou soubory s údaji o aktuálním vysílání:

- XML s aktuálním vysíláním v ZIP

8.3 Převzetí podkladů

Pro převzetí podkladů musí účastník delegovat pověřenou osobu, která se prokáže občanským průkazem a může za přihlášený subjekt vystupovat a převímat podklady. Zástupce účastníka se dostaví s přenosným 2TB diskem s USB konektorem, na který budou data nahrána. Delegování osoby lze písemnou formou nebo e-mailem s udáním **jména osoby a čísla občanského průkazu**. Osoby budou ověřeny zpětným kontaktováním dodavatele.

- **Kontaktní osoba pro předání podkladů na straně ČRo**
 - Jiří Špaček, +420 725 594 388, jiri.spacek@rozhlas.cz
 - Marietta Prajslerová, +420 602 311 282, marietta.prajslerova@rozhlas.cz
- **Místo předání podkladů**
 - Český rozhlas, Vinohradská 12, 120 99 Praha 2, ČR
- **Podklady lze převzít do:**
 - 3 kalendářních dnů před termínem uzavření veřejné zakázky