

## Příloha č. 5: Technická specifikace

### Slovníček pojmů

**Analytika mluveného slova (AMS)** je projekt Českého rozhlasu, který do sebe integruje nástroje speech2text a diarizaci a přes svoje API rozhraní tyto funkcionality nabízí systémům pro případnou integraci funkcionalit směrem k uživatelům nebo jiným systémovým výstupům.

**Speech2text, neboli rozpoznávání řeči (S2T)**, je technologie, která převádí mluvenou řeč do psané podoby. Tento proces využívá algoritmy umělé inteligence a strojového učení k analýze zvuku a rozpoznání jednotlivých slov a frází. Technologie se používá v různých aplikacích, jako jsou hlasoví asistenti, automatické přepisy záznamů nebo ovládání zařízení hlasem. Cílem je co nejpřesněji zachytit mluvený jazyk a převést ho na text s minimem chyb.

**Diarizace** mluvího je proces rozdělení a identifikace jednotlivých mluvčích v audiozáznamu. Jejím cílem je určit, kdo mluví a kdy, což je užitečné při analýze rozhovorů, schůzek nebo debat. Tento proces využívá algoritmy strojového učení, které rozpoznávají rozdíly ve způsobu řeči jednotlivých osob a rozdělují záznam na segmenty podle mluvčích. Diarizace je klíčová při zpracování přepisů, protože pomáhá správně přiřadit slova konkrétním mluvčím.

**D7:** současný redakční systém Drupal 7, který se používá pro iROZHLAS a mujRozhlas

**CMS Diego:** nový redakční systém ČRo pro budoucí správu iROZHLAS, mujRozhlas

**MDS:** divize Marketing a digitální služby

### Základní potřeby projektu AMS

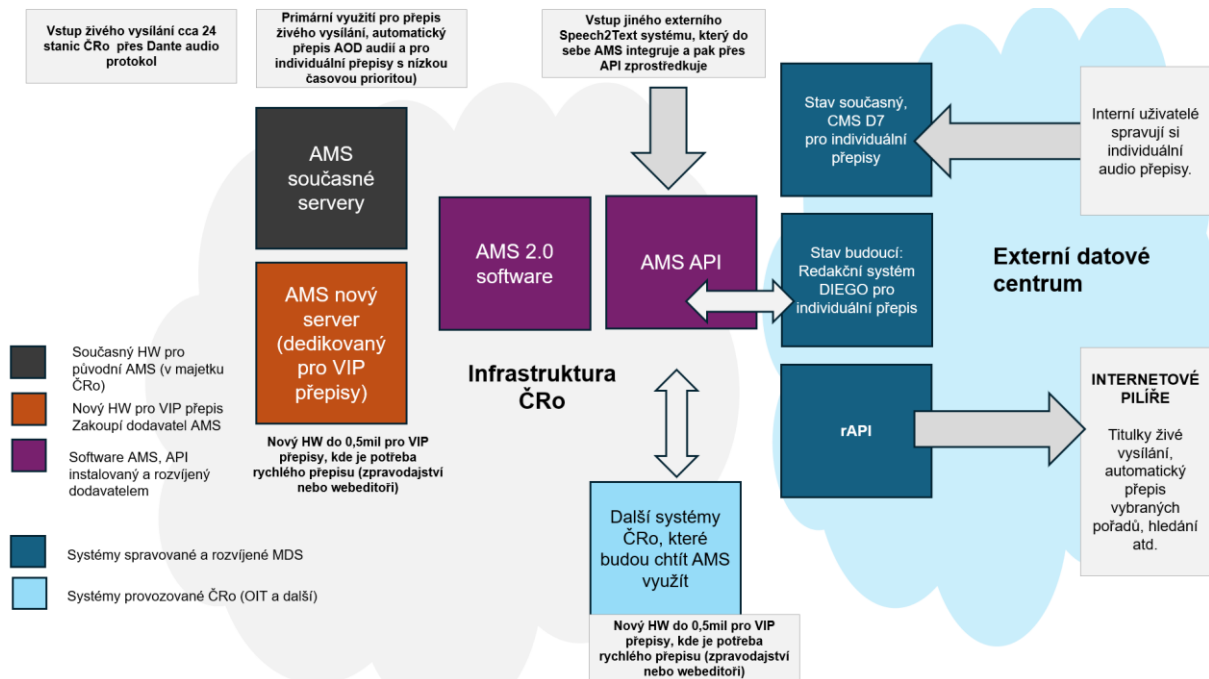
#### Potřeby ČRo:

- Automatický přepis živého vysílání pro titulkování
- Automatický přepis vybraných pořadů
- Individuální přepis vybraných audio souborů pro redaktory a podpůrné složky ČRo
- Využití pro další systémy ČRo

#### Funkce AMS

- Přepis audia do textu (on demand, live)
- Diarizace – označování mluvčího
- API rozhraní pro integraci do systémů ČRo
- API rozhraní pro integraci externích S2T systémů např. Whisper, Elevenlabs atd.

## Orientační schéma architektury AMS



## Technická specifikace přepis živého vysílání

- Přepis audia do textu pro živé vysílání. Ukázka současného stavu je dostupná na této adrese: <https://www.mujrozhlas.cz/zive/plus#panel-titulky>
- Audio signál živého vysílání se přebírá přes audio network platformu Dante
- Zpracování 24 stanic Českého rozhlasu
- Audioportál mujRozhlas přes API AMS přebírá přepis a zobrazuje ve svém webovém prostředí nebo v dalších výstupech

## Automatický on demand přepis

- AMS přebírá požadavky od redakčního systému D7, CMS Diego nebo případně dalších systémů požadavky na přepis konkrétního audio souboru
- Systémy ČRo komunikují s AMS přes jeho API rozhraní
- Přepis obsahuje informace o časové ose, aby přes API mohl předat informace např. v jakém čase proběhla změna mluvčího atd.
- Ukázka implementovaného přepisu pro návštěvníky mujRozhlas je zde: <https://www.mujrozhlas.cz/dvacet-minut-radiozurnalu/na-digitalizaci-stavebniho-rizeni-se-snazim-nekoukat-politicky-kazda>

## Požadavky systémové

- Dodávaný AMS systém není výpočetně závislý pouze na GPU, ale může využít také pouze výkonu CPU pro práci v alternativním „pomalejším“ režimu.

## Požadavky kvalitativní

- dodaný software musí mít chybovost maximálně 20 hrubých a maximálně 10 malých chyb pro průměrné audio z vysílání Českého rozhlasu ze stanice Radiožurnál či Plus o délce 20 minut.
  - Hrubou chybou se míní chyba, která mění význam nebo významně zhoršuje porozumění textu.
  - Malá chyba nemění význam textu (věty) a čtenář větě porozumí správně i bez znalosti kontextu (předchozích a následujících vět).

Podle experimentu Českého rozhlasu je požadavek na chybovost cca 3x nižší, než byla zjištěna testováním aktuálního open-source nástroje Whisper Large.

## Individuální VIP on demand přepis na

- Základní princip je stejný jako pro „Automatický on demand přepis“
- Pro tyto přepisy je dedikovaný speciální HW server, který je součástí dodávky
- Redakční systém volá API AMS a dle atributu určuje, že přepis bude zpracován na dedikovaném HW serveru

## Individuální on demand přepis

- Základní princip je stejný jako VIP přepisu, ale běží na běžném HW a není dedikován výkon

## Diarizace

- AMS zajišťuje identifikace mluvčího. Rozděluje mluvčí na A, B, C atd.
- Tato funkcionality je dostupná pro přepis živého vysílání i pro on demand přepisy

## Integrace jiných systémů pro přepis audia do textu

- AMS a jeho API počítá s případnou integrací jiných systémů jako je např. Whisper atd.
- API AMS musí být připraveno pro volání API jiných systémů a výsledný přepis přes své API nabízet systémům Českého rozhlasu
- Takové rozšíření by bylo případně řešeno v období trvání smlouvy

## Hardware specifikace serveru pro VIP on demand přepis

- server Supermicro AS – 4124GS-TNR+
- 2x CPU AMD EPYC Rome 7F72
- 8x GPU NVIDIA PNY Quadro RTX A4000 16GB GDDR6 PCIe 4.0
- 32x RAM 8 GB DDR4 288-PIN-3200MHz ECC
- 1x HDD 960GB NVMe PCIe4x4
- Případně obdobný HW v provedení 4U Rack mountable

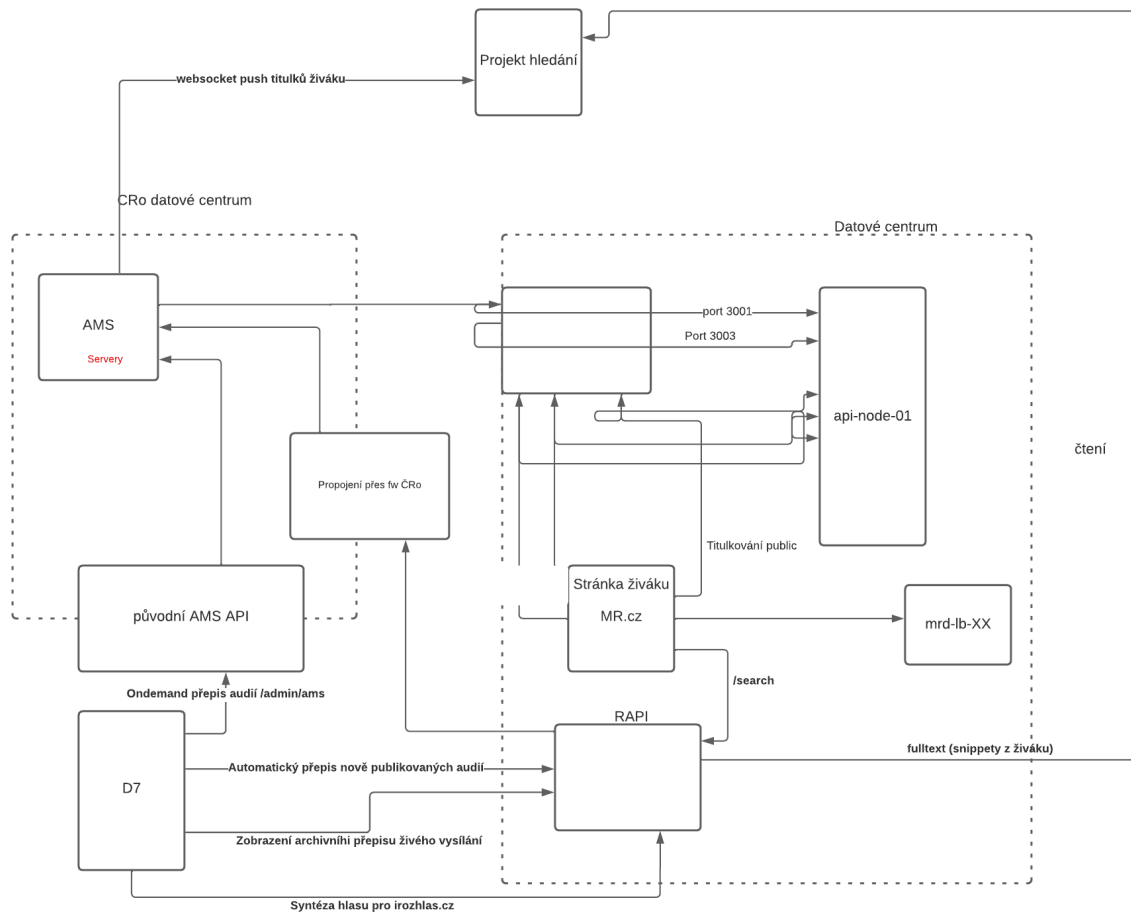
## Hardware nasazený pro přepis živého vysílání a běžných on demand audií

- HPE DL360 7x – využívají se na zpracování živého vysílání
- HPE DL20 (6x) – využívají se pro přepis On demand (zároveň část výkonu jako záložní řešení)

## Základní popis API

- V příloze č.6 - ČRo AMS REST API.pdf
- Strukturu API je nezbytné přesně dodržet, protože s ní počítají další interní systémy ČRo.

## Diagram AMS – anonymizovaný



## Harmonogram

31. 12. 2024

- Dodání licence pro software AMS včetně instalace na příslušný HW
- Dodání HW serveru pro VIP on demand přepisy a součinnost při instalaci v datovém centru Českého rozhlasu
- Instalace základního API rozhraní pro přebírání
  - Přepisu pro živé vysílání
  - Přepisu pro on „Automatický on demand přepis“
  - Příprava API pro přebírání „VIP on demand přepisu“

1. 1. 2025 – 31. 12. 2027

- Průběžný rozvoj API rozhraní a softwaru dle aktuálních požadavků Českého rozhlasu v odhadovaném počtu 400 hodin po dobu trvání smlouvy.