

# 1 Deelproject Spraakherkenning: SHoUT Audio Indexing Service

De in het CHoral project ontwikkelde audio-indexeringstechnologie op basis van automatische spraakherkenning (SHoUT) wordt beschikbaar gemaakt als service waar erfgoedinstellingen gebruik van kunnen maken. Naast het ontwikkelen van een centrale interface, worden de verschillende software componenten die bij audio-indexering een rol spelen schaalbaar gemaakt zodat makkelijker aan de specifieke eisen van collecties/erfgoedpartijen kan worden voldaan. Daarnaast worden tools ontwikkeld om niet-experts de mogelijkheid te bieden om de service aan te passen aan verschillende eisen van collecties.

De SHoUT Audio Indexing (SAI) service wordt gebouwd door het bedrijf Cross-Media Interaction, spinoff van de Universiteit Twente die opgezet is om de in onderzoeksprojecten ontwikkelde audio indexeringstechnologie voor de markt beschikbaar te maken. In Catch+ wordt audio indexering voor erfgoedpartijen beschikbaar gemaakt als een service. Het Gemeente Archief Rotterdam is 'first user' waarmee de SAI service wordt vormgegeven en getest. Raadsvergaderingen van de gemeente Rotterdam worden gebruikt als 'use case' in Catch+.

De SAI service wordt ingezet binnen/voor/i.s.m.:

- Gemeente Archief Rotterdam
- Beeld en Geluid
- NIOD
- IIAV

Daarnaast hebben een aantal erfgoedpartijen aangegeven geïnteresseerd te zijn in het gebruik van de tool, waaronder:

- Veteranen Instituut
- Dans

## 2 Functionele omschrijving/use cases

Op het allerhoogste niveau is het proces als volgt: een gebruiker/instelling met een account voor de service maakt vanuit de eigen instelling verbinding met de service en wordt in staat gesteld 1 of meerdere AV files in batch te uploaden, inclusief voor deze data relevante metadata en/of gerelateerde tekstdata. De SAI service gaat met deze files aan de slag en zal binnen bepaalde tijd een annotatie op basis van audio/spraakanalyse opleveren. Deze annotatie kan vervolgens ofwel worden geïntegreerd in een lokale metadata repository, ofwel worden geïndexeerd in de SAI service. In het eerste geval wordt ervan uitgegaan dat integratie in het eigen metadata model (indexering en search) door de erfgoedpartij zelf wordt geregeld. De tweede optie biedt erfgoedpartijen die geen eigen index & search op AV content willen of kunnen inrichten de mogelijkheid om rechtstreeks via een API (bijvoorbeeld vanuit een webpagina) te zoeken in eigen AV content via de SAI service. Ten behoeve van dit laatste scenario biedt de SAI service een aantal standaard oplossingen voor indexering en search. Desgewenst moet deze functionaliteit flexibel kunnen worden beheerd door of overgedragen aan een derde partij. Hieronder worden een aantal verschillende scenario's aan de hand van een voorbeeld uitgewerkt.

## 2.1 Scenario A

In figuur 1 wordt het eerste scenario visueel weergegeven. Erfgoedpartij A upload AV content naar de SAI service. Deze content wordt verwerkt door de audio analyse component van de service. Na verwerking wordt de annotatie aangeboden via een 'download interface' en vervolgens geïntegreerd in het eigen metadata model van Erfgoedpartij A, die tenslotte ook zorgt voor het aanbieden van zoek- en afspeelfunctionaliteit voor gebruikers via een lokale 'erfgoed portal'. In dit scenario is de erfgoedpartij volledig self-supporting --inclusief indexering, search, en content streaming--, op de audio analyse na.

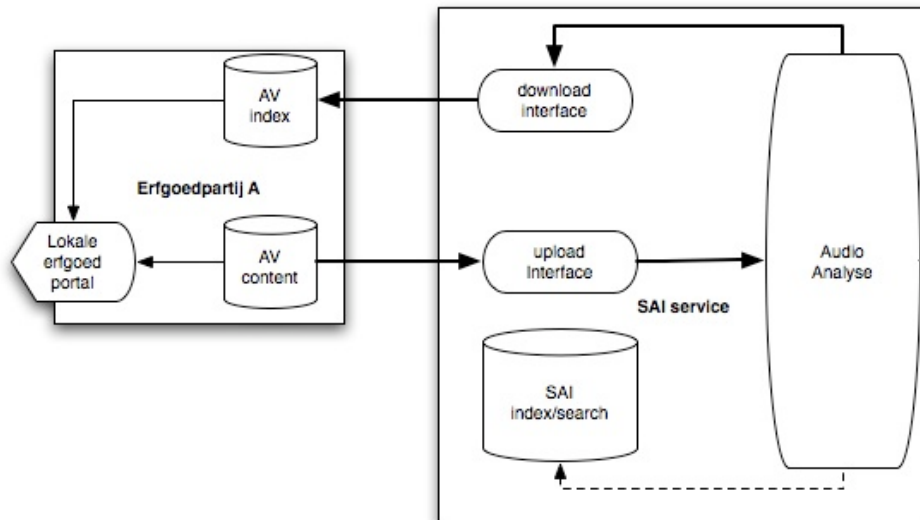


Figure 1 Scenario Erfgoedpartij A

## 2.2 Scenario B

In scenario B (zie Figuur 2) maakt de Erfgoedpartij ook gebruik van de indexering en search functionaliteit van de service. De AV content wordt geupload naar de service. Na audio analyse wordt de annotatie geïndexeerd binnen de SAI service. De Erfgoedpartij beheert zelf de portal waar gebruikers toegang kunnen krijgen tot het materiaal, maar maakt daarbij gebruik van de API op de index van de SAI service. In dit scenario out-sourced de erfgoedpartij de audio analyse, indexering en zoekfunctionaliteit. Gebruikersinterface en content streaming wordt door de partij zelf geregeld.

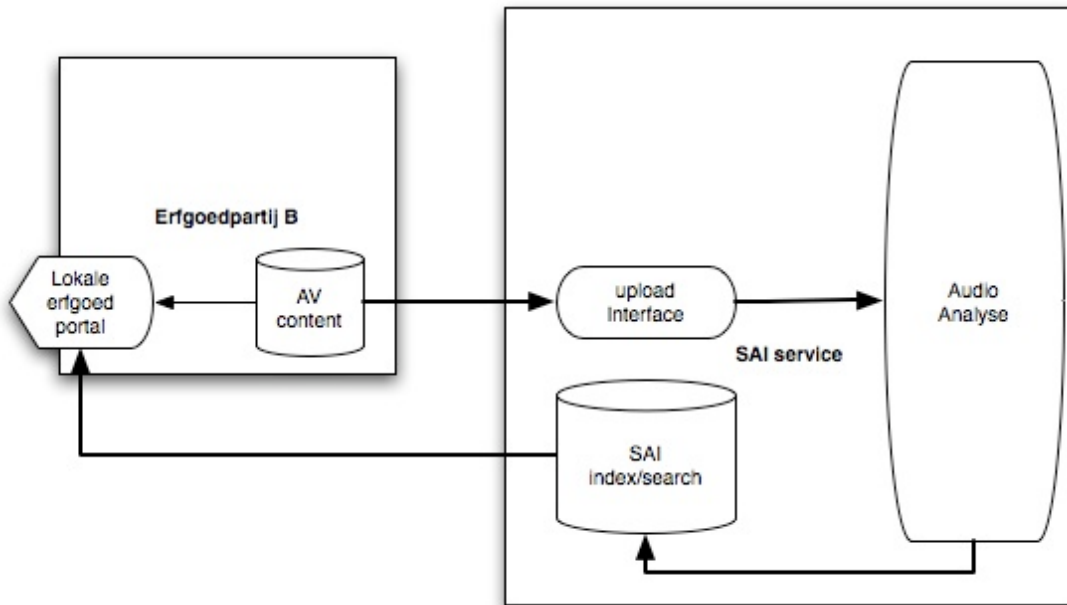


Figure 2 Scenario Erfgoedpartij B

### 2.3 Scenario C

In het laatste scenario wordt ervan uitgegaan dat de erfgoedpartij data opslag, indexering en search geheel overlaat aan 1 of meerdere derde partijen, visueel weergegeven in figuur 3. De audio analyse in de SAI service wordt in dit scenario aangesproken door de derde partij.

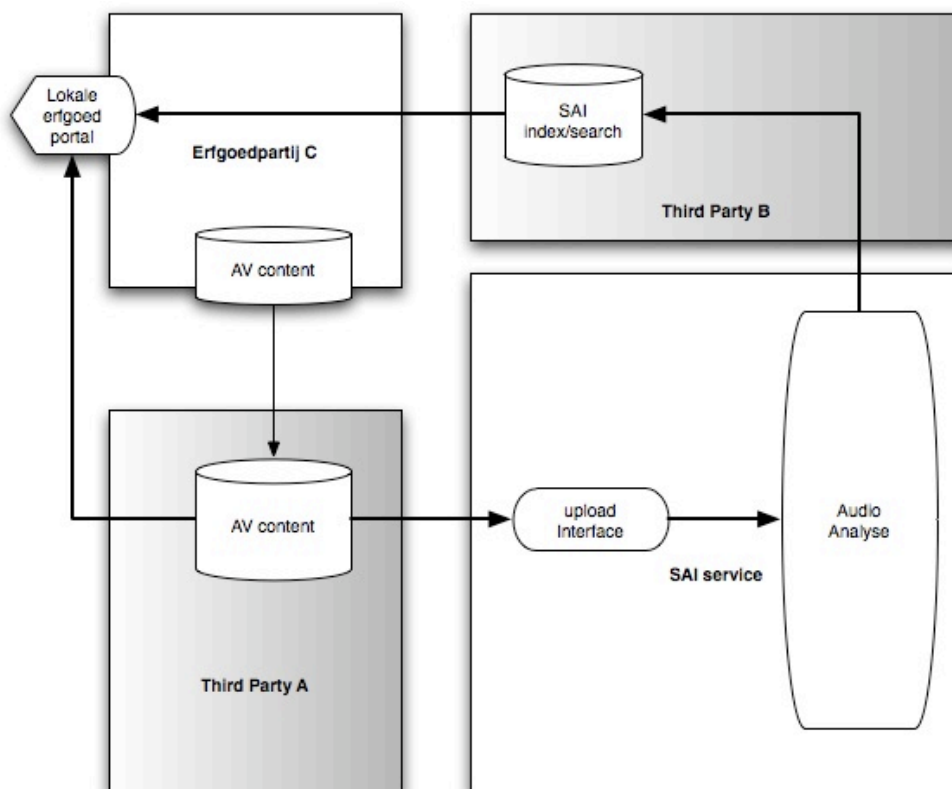


Figure 3 Scenario Erfgoedpartij C

## 2.4 Software componenten

De volgende software componenten maken deel uit van de SAI service:

### 2.4.1 Front-end

- upload API: moet voorzien in handmatige/automatische upload van AV content en gerelateerde data (metadata, relevante tekstbronnen)
- metadata 'download' API: moet voorzien in handmatig/automatisch transport van metadata uit audio analyse naar erfgoedpartijen
- search API: moet voorzien in het bevragen van geïndexeerde data op de SAI service
- module voor aanpassing audio analyse: voorziet in het aanpassen van audio analyse aan eigenschappen collectie door erfgoedinstellingen.

### 2.4.2 Back-end

- AV conversie module: audio extractie en conversie voor analyse
- Tekstuele (meta)data verwerkingsmodule: verwerkt binnenkomende tekstuele bronnen voor tunen spraakherkenning
- Audio analyse voorbereiding:
  - Taalmodel module
  - Uitspraakwoordenboek
- Verschillende modules voor audio analyse, oa:
  - spraak/niet-spraak detectie
  - sprekersegmentatie
  - spraakherkenning
- Audio analyse distributie module: regelt verwerkingsproces op basis van beschikbare hardware en taaklijst
- Indexering: regelt instroom analyse resultaat in database en indexering

## 3 Data sets

### 3.1 Brondata

In de GAR use case worden raadsvergaderingen van Gemeente Rotterdam verwerkt door de SAI service (volgens scenario A hierboven). In eerste instantie zullen alleen nieuwe vergaderingen worden geïndexeerd. In een later stadium zal retrospectieve indexering worden overwogen. Het zal gaan om circa X uur aan materiaal per maand.

### 3.2 Metadata

Indien beschikbaar ...

### 3.3 Annotaties

Voor het testen van spraakherkenningsperformance ...

### 3.4 Spraakherkenning specifiek

- Achtergrond lexicon uitspraken: beschikbaar als onderdeel van de service, niet los benaderbaar/beschikbaar
- Collectie specifieke uitspraken, gegenereerd binnen Catch+: in principe vrij beschikbaar