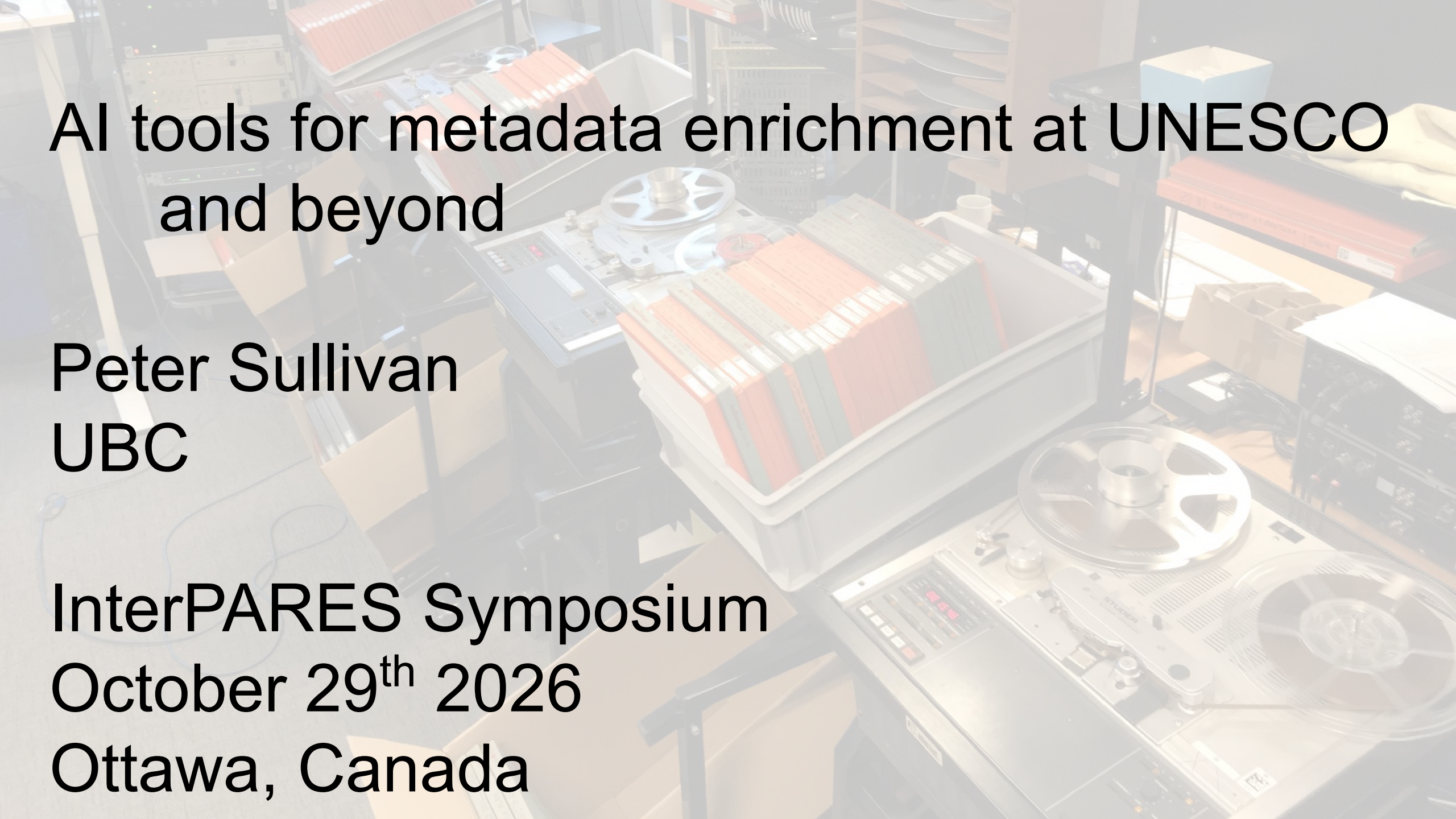


AI tools for metadata enrichment at UNESCO and beyond

Peter Sullivan
UBC

InterPARES Symposium
October 29th 2026
Ottawa, Canada

Agenda

Motivation

UNESCO Radio Archives

- **Automatic Speech Recognition**
- **Speaker Recognition and Language Identification**
- **Extractive Summarization**

Reflections

Motivation

- **Cultural value of audio recordings**
- **Lack of access to non-textual records**
- **Difficulty in describing audio recordings**
- **Archival & Epistemological Context**

How can AI enable better description of archival audio?



UNESCO Radio Archives

Archives At A Glance

Size:	16,000 Recordings (1000 described)
Time:	1950s – 1980s
Languages:	70+
Genre:	Diverse, but mainly one of five main genres
Topics:	Education Culture Natural Sciences Communication and Information Social and Human Sciences

Archives At A Glance

Size:	16,000 Recordings (1000 described)
Time:	1950s – 1980s
Languages:	70+
Genre:	Diverse, but mainly one of five main genres
Topics:	Education Culture Natural Sciences Communication and Information Social and Human Sciences

Archives At A Glance

Size:	16,000 Recordings (1000 described)
Time:	1950s – 1980s
Languages:	70+
Genre:	Diverse, but mainly one of five main genres
Topics:	Education Culture Natural Sciences Communication and Information Social and Human Sciences

Archives At A Glance

Size:	16,000 Recordings (1000 described)
Time:	1950s – 1980s
Languages:	70+
Genre:	Diverse, but mainly one of five main genres
Topics:	Education Culture Natural Sciences Communication and Information Social and Human Sciences

Archives At A Glance

Size:	16,000 Recordings (1000 described)
Time:	1950s – 1980s
Languages:	70+
Genre:	Diverse, but mainly one of five main genres
Topics:	Education Culture Natural Sciences Communication and Information Social and Human Sciences

Timeline

	Methods	Key Activities
1	Automatic Speech Recognition	Focus group analysis of transcripts
2	Speaker Recognition, Language ID	Benchmarking with proxy data & existing metadata
3	Audio Description, Sequence Labelling	Data annotation, and direct analysis



Phase 1: Focus Group

Automatic Speech Translation

Used `whisper-large` to transcribe a random sample

Focus group takeaways:

- Transcripts are high quality, as long as LangID is right
- Be wary of easily confused languages or missing context
  - Croatian vs Serbian, Galician vs Portuguese
- Anachronistic language
- Music and noise can cause unexpected behavior

Transcription Failures

<transcript>
00:00:01.004 --> 00:00:06.032
Music Round the World, a UNESCO presentation of song and dance from the five continents.

00:00:27.064 --> 00:00:40.084
Wann sie wann bombanier romantisch wandert bombanier, romantisch.

00:00:46.032 --> 00:00:52.036
Wem da Bomben? Levi's Tomb Raider haben in Number oder dreimal hab ich in der Welt.

00:00:55.052 --> 00:01:00.064
Ok und da vorbei mich umdrehen okay da wo man nicht unbezieht nur 3 Wochen.

00:01:23.012 --> 00:01:26.084
Nonne Renner Gottes Gott war Nonne Winkler.

00:01:59.088 --> 00:02:05.072
That song was recorded on La Margarita Island and is typical of the popular songs of the eastern states of Venezuela.



Phase 2: Benchmarking

Speaker Recognition and Language ID

Benchmarking using existing data, and proxy data

Language ID on accented English ✔
sampled UNESCO recordings ✔

Speaker recognition stability over time ✔ (up to 10 years)
and across language ✘

Sullivan, P. R., & Abdul-Mageed, M. (2024). On Barriers to Archival Audio Processing. In M. Abbas & A. A. Freihat (Eds.), *Proceedings of the 7th International Conference on Natural Language and Speech Processing (ICNLSP 2024)* (pp. 377–383). Association for Computational Linguistics. https://aclanthology.org/2024.icnlsp-1.39/

How good does Language ID need to be?

Assuming reliability of existing metadata
9% error rate on near optimal conditions

↳ 500 hours of gibberish transcripts

Consider:

- Narrowing down languages
- Triage (identify high resource languages)



Phase 3: Summarization

Trustworthiness of Summarization

LLM Summarization Pilot

Response:
Description: This recording is an episode of the UNESCO radio program "Music Round the World," which showcases traditional music and dance from various regions. The episode features a collection of songs, dances, and ceremonies from Venezuela, highlighting the country's rich cultural heritage and influences from Spain, Africa, and indigenous Indian traditions.

Speakers: The speakers mentioned in the transcript are Juan Lescano, associates, Pierre Gessos, and Keith Wood.

LangID: The languages spoken in the transcript are likely Spanish, French, and indigenous languages from Venezuela, although it is not explicitly stated. The transcript also includes musical lyrics and commentary in English.

Extractive Summarization

Observation: Existing metadata descriptions were often pulled directly from the transcript.

Instead of ***abstractive*** summarization using LLMs we can guarantee trustworthiness of summary by ***extracting*** salient parts of transcript.

Archival Diplomatics

To identify salient parts of transcripts, we apply archival diplomatics-based approach.

protocol

Contains: who is speaking, subject of recording

eschatocol

May contain: who was speaking, where it was recorded

LLM Sequence Labeling

Genre	(n)	Accuracy	Precision	Recall
Programme	8	97.61%	43.24%	75.92%
Report	6	94.55%	65.27%	88.60%
Interview	17	97.46%	40.53%	86.61%
Press Conferences	6	96.41%	35.17%	42.22%
Speech	3	93.18%	28.69%	47.22%
Total	40	96.11%	42.04%	73.33%

Word-level performance of Qwen3-30B-A3B-Instruct-2507



Reflections

Practical Concerns

How are AI tools being assessed?

How are archivists supported?

Is infrastructure available?



Acknowledgments

Eng Sengsavang
Trevor Wilkes


Thank You!