# Augmenting the Metadata of Audiovisual Archives with NLP Techniques: Challenges and Solutions

#### Alliata, Giacomo

giacomo.alliata@epfl.ch EPFL, Laboratory of Experimental Museology, Switzerland

#### Yang, Yuchen

yuchen.yang@epfl.ch EPFL, Laboratory of Experimental Museology, Switzerland

#### Kenderdine, Sarah

sarah.kenderdine@epfl.ch EPFL, Laboratory of Experimental Museology, Switzerland

### Research context and purpose

In the last decade, following the push of mass digitization, audiovisual archives have also been digitized, both by cultural institutions and broadcasting companies. The Radio Télévision Suisse (RTS) has for instance assembled more than 200,000 hours in its archive. However, as outlined by both archival scholars (Fossati et al., 2012) and the GLAM sector (Winesmith / Anderson, 2020), once digitized, these large collections require new frameworks to facilitate their access for the larger public and improve their civic value (Edmondson, 2004). Working with such large datasets proves to be challenging, as we observe in this research. In this paper, we tackle this challenge from the perspective of experimental museology (Kenderdine, 2021), with the goal of augmenting the metadata of the RTS Archives for supporting innovative modes of access through Natural Language Processing (NLP) techniques.

The main difficulty that arises is the potential lack of homogeneity in the archival process. This can result in overwhelming amounts of different tags and categories, unusable as is to cluster the items in the collection in a meaningful way. For instance, in the RTS Archives, we have 6528 topic tags, 3273 geographical tags and 8954 entity tags across 522103 videos. Figure 1 demonstrates that the distribution of the number of tags per video is heavily right skewed, with 75% of the videos having a maximum of 4 tags.



Figure 1: Distribution of the number of tags used per video (log scale).



Figure 2: Graph of the most common tags co-occurrences (red: locations, blue: entities).

Figure 2 displays a graph of tags co-occurrences, for the tags that have been used for at least 1000 videos. Unsurprisingly, Switzer-land ("suisse") and important Swiss cities are the most common tags.

The challenge is thus on how to infer a proper knowledge structure from this seemingly free and unconstrained text annotation, allowing for meaningful clustering of the videos and navigation of the archive. In parallel, the challenge of missing data can also highly influence the outcomes of visualizations attempts, as it can cause items to be isolated from the bulk of the collection, not because they are different but because they have not been properly tagged (Rodighiero et al., 2022). As the reader can see in figure 1, this is the case for many videos in the RTS Archives.

## Our approach

Our initial experiments used Bag-of-Words, TF-IDF, and word embeddings as video representations for clustering with the K-Means algorithm. Unfortunately, a qualitative inspection of the clusters revealed that well-defined groups could not be extracted. Therefore, to augment the metadata available and facilitate exploration of the entire archive, we decided to link it to the Wikidata database. We focused on geographical and entity tags, linking 10,618 Wikidata entities to tags in the archive using a simplified entity linking approach that took the first candidate from the Wikidata API. Since these tags lacked context, we also used video descriptions as an alternative source for tagging videos. To perform entity extraction, disambiguation, and linking using the video descriptions and Wikidata, we developed an end-to-end pipeline using spaCy and the extend project (Barba et al., 2022). A small-scale experiment on 15,305 sports-related videos (filtered by content category) resulted in an additional 4,679 unique tags linked to Wikidata entities, thereby expanding the scope of our analysis beyond the original set of tags.

Once this first phase of extraction is completed, the collection of Wikidata entities collected enables numerous approaches to further complement the available metadata. Since all the tags are processed and linked to a Wikidata entity, the RTS metadata automatically inherits the semantic and logical information from the Wiki knowledge graph. The audiovisual archive can thus be navigated through the Wikidata database. Properties of these entities can be queried in a simple manner, obtaining geographical coordinates for the location or citizenship of the persons for instance. The drawback of having a manual and arbitrary tagging system (often found missing tags, inconsistency in multi-lingual tags, etc.) is also solved by providing a holistic and connected higher structure to the tags.

## Conclusion

In conclusion, this research shows how the metadata of an audiovisual archive tagged in an unconstrained manner can be augmented by linking it to a well-structured and large collection: the Wikidata database. Future work will explore how tags can be automatically extracted from the videos, transcribing the audio for instance, and thus potentially resulting in a much more granular tagging, on the scale of short clips rather than entire videos.

## Acknowledgements

This research is part of the SNSF-funded research project *Narratives from the Long Tail: Transforming Access to Audiovisual Archives*, led by co-author Prof. Sarah Kenderdine ((grant number CRSII5\_198632, see https://www.futurecinema.live/project/ for a project description).

# Bibliography

**Barba, Edoardo; Procopio, Luigi and Navigli, Roberto** (2022). "ExtEnD: Extractive entity disambiguation". Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics.

Edmondson, Ray (2004). "Audiovisual Archiving: Philosophy and Principles". UNESCO Paris.

**Fossati, Giovanna** (2012). "Found Footage Filmmaking, Film Archiving and new Participatory Platforms". In *Found Footage. Cinema Exposed*. Amsterdam: Amsterdam University Press/EYE Film Institute Netherlands.

Kenderdine, Sarah (2021). "Experimental Museology: Immersive Visualisation and Cultural (Big) Data". In *Experimental Museology: Institutions, Representations, Users*. London & New York: Routledge, 15-34.

Rodighiero, Dario; Derry, Lins; Duhaime, Douglas; Kruguer, Jordan; Mueller, Maximilian C.; Pietsch, Christopher; Schnapp, Jeffrey T. and Steward, Jeff (2022). "Surprise Machines: Revealing Harvard Art Museums' Image Collection". Information Design Journal, 27 (1), 21–34.

**Winesmith, Keir and Anderson, Suse** (2020). "The Digital Future of Museums: Conversations and Provocations". Routledge.