Toward an Ontological Model for Audiovisual Archive Datafication

Yuchen YANG, EPFL, Lausanne, Switzerland

Abstract

With the initiatives like Collections as Data and Computational Archival Science, archives are no longer seen as a static documentation of objects, but evolving sources of cultural and historical data. This work emphasizes the potential updates in preserving and documenting digital audiovisual (AV) content from a data perspective, considering the recent developments in natural language processing and computer vision tasks, as well as the emergence of interactive and embodied experiences and interfaces for innovatively accessing archival content. As part of Swiss national scientific fund Sinergia project, this work was able to work end-to-end with real-world AV archives like Télévision Suisse Romande (RTS). Resorting to an updated narrative model for mapping data that can be obtained from the content as well as the consumer, this work proposed an experimental attempt to build an ontology to formally sum up the potential new paradigm for preservation and accessibility from a data perspective for modern archives, in the hope for nurturing a digital and data driven mind-set for archive practices.

Motivation

AV archives comprise a large and vital part of our cultural legacy. However, as recently pointed out in the UNESCO's Memory of the World programme report on AV archives, the increasing volume and the gap between technologies used for creation and new technologies available for preservation pose immense challenges for AV archives [1].

This challenge materialised during the wave of digitisation. Archivists have discussed in detail the interrelationships between the analogue and the digitised versions of AV materials, as well as how to carefully form digitisation strategies [2] to enhance the preservation in the digital age.

However, archives nowadays are not just facing a digital turn, but a data turn and the focus has shifted from digitisation to datafication [3]. As a growing number of new technologies (in computer vision and natural language processing) becomes available to archives, it is argued that there is a pressing need to understand what to preserve and how to manage these properties (on top of the traditional metadata) from a data management perspective [4].

On the other hand, the requirements for consuming such content has also evolved. Museums and archives, are emphasising the importance of a situated solution with a more intuitive interface using techniques such as visualisation, immersive and embodied interaction [5] for free exploration of their archives. Moving away from a keyword-based search web interface or single-themed exhibition, the objectives of using interactive and immersive experiences in physical space for accessing the archive are to "open up" the full potential of memory materials to the general public and to transform the access to archives to be more democratic, personalised, and explorative [6]. As one of the attempts echoing such notion, the SEMIA project [7] works on experimenting with alternative AV archive interfaces for exploring through features such as colour, shape, and visual clutter and encourages the serendipitous discovery of an archive rather than just search.

The scoping of new public-facing experiences, as well as the developing technology in breaking down content into more granular pieces of feature and data, raise the pressing needs to re-examine the current strategy to preserve and document cultural AV content [1].

Problem

The nature of audiovisual materials is rich in temporality and multimodality, making them a bit foreign to existing archival practices focused on text and images. The vast majority of their meanings are well hidden, creating barriers to managing and using such archives [8]. Although recent advances in natural language processing and computer vision have enabled some content level extraction and analysis tools, the effort of decoding AV archives into various data descriptors and features remains ad-hoc and is very much tool and method-driven. Moreover, due to the lack of systematic thinking, such newly available information (emotions, colours, gestures, semantic entities and etc.) is only used as tags to facilitate the search of materials in large archives and provide nothing much beyond what is available through standard full-text search [9]. The importance for re-examine the preservation and documentation for AV content emerge naturally under this circumstance. The critical question this work is trying to answer here is, "What new data should we preserve on top of the original AV materials and Why?"

Approach

Narrative as a tool

It is worth noting that narrative in the scope of this research means broadly a cognitive model that helps curators and designers to understand the audiences, the environment, and the cultural materials and their relationships.

Narrative cognition presents a two-step strategy:

- 1) *Break down*. It claims we interpret things by first breaking the original content into granular elements from different functioning perspectives [10, 11].
- 2) *Recreate.* It claims interpretations happen when elements are recombined and interact with the audience and the environment, and the cognition results depend on what is in the three aspects [12]

For memory institutions, such as archives and museums, narrative as a cognitive tool is adapted as core guidance for their curation practices to create [13, 14]. A quick example would be, given the same group of people (audience) and display (environment), a library presents bibles printed nowadays from all seven continents (content), showing the differences in wording and translations (content elements) will create a completely different narrative as opposed to the showing the differences in font and illustration styles (other content elements).

However, to keep up with the growing trend of datafication and adopting interactive and immersive techniques in memory institutions, this narrative-driven tool for curation needs an update to better understand and manage the abundance of data available in the original materials and to facilitate more detailed mapping for factors to consider in the audience and environment perspective when it comes to creating new public-facing experiences. Under this premise, this research tends to extend the philosophy of narrative to guide the data management and the creation of new experiences.

An ontological model for AV archive datafication

Although the original content (original videos, in the case of this research) has never changed, the evolving metadata, which is as indispensable as the original object, is becoming more diverse in archives. From the classic ones like production time to the more recent ones like 3D reconstruction of a specific scene [15], the increasing volume and complexity of data in AV archives make managing such assets a rough task. Based on existing works [16, 17, 18], a mapping of the complexity of AV archive metadata in various dimensions is made and summarised in Fig. 1.



Figure 1. Dimensions of the metadata of a multimedia object visualised in the level, the automation, the extraction time, the form, the retrieval, and the modality dimensions

For dealing with such complexity in multimedia data and metadata management, MPEG-7 [17] often used as a standard. However, they

are not entirely suitable for managing the content level metadata (descriptors) of multimedia content nowadays. The fundamental philosophy for such standards is that Low-level metadata are features that can be extracted automatically (like colour, texture, shape, timbre, pitch and rhythm), and High-level metadata are human annotations on a conceptual level (like emotions, content summaries) [19]. With the development of neural networks and deep learning, the semantic gap of the inability to create high-level conceptual descriptors in an automated way is closing down [20]. However, current standards have not reflected such changes. For example, descriptors such as body key points or text-video embedding vectors are not considered at all. Such standards also ignored the metadata (training data, model, etc.) related to the automated generated high-level descriptors, which are essential for understanding, improving, and managing the results [21], and a successful data management system integrating machine learning pipelines or results [22].

On the other hand, when considering narrative in the context of archive-based interactive and immersive experiences, it is argued that only considering the narrative formed from the content end is not enough and that the narrative is formed through the combination of the Content, Participants, and interaction. The ultimate goal is to allow participants to build their knowledge based on the interaction between them and the content [23, 24, 25]. According to such an updated narrative model for interactive and immersive experiences, these three parts are organically connected and cannot be separated. However, the connection and relationship between these three parts have never been formalised. Inspired by previous works to formalise textual narratives using ontology [26] and formalise the composition and relationship of multimedia big data [27], this part of the research aims to provide an ontological model to describe the interconnection between the three aspects.

The development of the ontology has been planned as follows. The first stage would be to pinpoint the core ontology to use for each of the three aspects and merge them to be a general ontology. Another focus at this stage is to extend the existing model with components considering the automated generated high-level metadata (such as text-video embeddings) and the coverage of new low-level features (such as skeleton). The second stage will then focus on the validation and improvements of the ontology by generating speculation-specific knowledge graphs from a data perspective and reflecting with domain experts. Another focus at this stage is to improve the data lineage perspective of the ontology both in terms of mapping data to experiences (e.g. colour features are used for an emotion-themed exploration) and low to high level descriptors (e.g. speech, face, and visual modalities are used for creating the text-video embeddings using what modal).

Results

Following the steps proposed in the previous section, this ontology is modulated along three corresponding dimensions - content, participant, and interaction. Fig. 2 provides a snapshot of the three dimensions and key classes with nodes and connections with edges. The implementation of the ontology is done with OWL 2 [28] to be adaptable and flexible enough for integrating with existing ontology models as well as for developing reasoning systems upon it if needed.



Figure 2. An exemplary look at some of its key classes and properties of it to illustrate the overall ontology concept

The proposed ontology uses HCIO [29] as a core and top-level reference for inheriting fundamental notions for describing the core aspects of the human-computer interaction phenomenon. Although particularities of types of interaction (e.g. haptic interaction), input (from participants), and content are not specified, it is advocated that downstream ontologies with a specific focus should reuse and specialise the core HCIO concepts.

Overall, this ontology merged the existing Ontology for Media Resources¹, COMM [30] and an ontology for harvesting user input in immersive environment [31] are used as starting point to populate the content, participant, and interaction dimensions. These three dimensions were then respectively mapped into the top concepts of Interactive Computer System, User Participation, and Human-Computer Interaction in HCIO.

Conclusions

Under the wave of undeniable trend of digitization and datafication, as well as the more interactive and divers innovations in archival experiences, this work answers the key question "What new data should we preserve on top of the original AV materials and Why?" by proposing an ontological approach. This attempt to build such and ontology formally sum up the potential new paradigm for preservation and accessibility from a data perspective for modern archives, in the hope for nurturing a digital and data driven mind-set for archive practices. However, it is worth noting that the ontology proposed here is far from mature and will be continuously updated and validated with domain experts working together in the Swiss

¹ https://www.w3.org/TR/mediaont-10

national scientific fund Sinergia project. However, this current ontology serves as a necessary material and technological foundation for starting such updates and validations, informing the following research with a holistic view of the sensemaking of narrative encoded AV archives. This ontology will also serve as the anchor for future connection and alignment of open datasets like IMDB or Wikidata, as well as any other AV databases, which, in turn, expand intra-archival access with sounding knowledge and cultivate inter-archival collaborations.

References

- [1] Edmondson, R. and Edmonson, R., 2004. Audiovisual archiving: philosophy and principles (p. 47). Paris: Unesco.
- [2] Cavallotti, D., 2018. From grain to pixel? Notes on the technical dialectics in the small gauge film archive. NECSUS. European Journal of Media Studies, 7(1), pp.145-164.
- [3] Manovich, Lev. 2020. Cultural Analytics. Mit Press.
- [4] Padilla, Thomas G. 2018. 'Collections as Data: Implications for Enclosure'. College and Research Libraries News 79 (6): 296.
- [5] Kenderdine, Sarah, Ingrid Mason, and Lily Hibberd. 2021.
 'Computational Archives for Experimental Museology'. In , 3–18. Springer.
- [6] Kenderdine, Sarah, 2021. 'Experimental Museology: Immersive Visualisation and Cultural (Big) Data'. Experimental Museology, 15.
- [7] Masson, Eef, Christian Gosvig Olesen, Nanne van Noord, and Giovanna Fossati. 2020. 'Exploring Digitised Moving Image Collections: The SEMIA Project, Visual Analysis and the Turn to Abstraction.' DHQ: Digital Humanities Quarterly, no. 4.
- [8] Tesic, J., 2005. Metadata practices for consumer photos. IEEE MultiMedia, 12(3), pp.86-92.
- [9] Brock, D. 2022. A MUSEUM'S EXPERIENCE WITH AI. Available at: https://computerhistory.org/blog/a-museums-experience-with-ai/ (Accessed: 20 Jan 2022).
- [10] Goodson, Ivor F, Gert Biesta, Michael Tedder, and Norma Adair. 2010. Narrative Learning. Routledge.
- [11] Bordwell, David. 2012. 'Three Dimensions of Film Narrative'. In Poetics of Cinema, 97–146. Routledge.
- [12] Brooks, Peter. 1992. Reading for the Plot: Design and Intention in Narrative. Harvard University Press.
- [13] Wolff, Annika, and Paul Mulholland. 2013. 'Curation, Curation'. In , 1–5.
- [14] Krukar, Jakub. 2014. 'Walk, Look, Remember: The Influence of the Gallery's Spatial Layout on Human Memory for an Art Exhibition'. Behavioral Sciences 4 (3): 181–201.
- [15] Božič, Aljaž, Pablo Palafox, Justus Thies, Angela Dai, and Matthias Nießner. 2021. 'TransformerFusion: Monocular RGB Scene Reconstruction Using Transformers'.
- [16] Vries, Arjen P. de, and H. M. Blanken. 1998. 'Database Technology and the Management of Multimedia Data in the Mirror Project'. In , edited by C.-C. Jay Kuo, Shih-Fu Chang, and Sethuraman Panchanathan, 443–53. Boston, MA.
- [17] Manjunath, B. S., P. Salembier, and Thomas Sikora, eds. 2002. Introduction to MPEG-7: Multimedia Content Description Interface. Berlin; New York: Wiley.
- [18] Rossetto, Luca, Matthias Baumgartner, Narges Ashena, Florian Ruosch, Romana Pernisch, Lucien Heitz, and Abraham Bernstein. 2021. 'VideoGraph – Towards Using Knowledge Graphs for Interactive Video Retrieval'. In MultiMedia Modeling, edited by Jakub Lokoč, Tomáš Skopal, Klaus Schoeffmann, Vasileios Mezaris, Xirong Li, Stefanos Vrochidis, and Ioannis Patras, 12573:417–22. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-67835-7_38.
- [19] Schallauer, Peter, Werner Bailer, Raphaël Troncy, and Florian Kaiser. 2011. 'Multimedia Metadata Standards'. In Multimedia Semantics, edited by Raphaël Troncy, Benoit Huet, and Simon Schenk, 129–44. Chichester, UK: John Wiley & Sons, Ltd.

- [20] Wan, Ji, Dayong Wang, Steven Chu Hong Hoi, Pengcheng Wu, Jianke Zhu, Yongdong Zhang, and Jintao Li. 2014. 'Deep Learning for Content-Based Image Retrieval: A Comprehensive Study'. In Proceedings of the 22nd ACM International Conference on Multimedia, 157–66. Orlando Florida USA: ACM.
- [21] Fourati, Manel, Anis Jedidi, and Faiez Gargouri. 2020. 'A Survey on Description and Modeling of Audiovisual Documents'. Multimedia Tools and Applications 79 (45–46): 33519–46.
- [22] Kumar, Arun, Matthias Boehm, and Jun Yang. 2017. 'Data Management in Machine Learning: Challenges, Techniques, and Systems'. In Proceedings of the 2017 ACM International Conference on Management of Data, 1717–22. Chicago Illinois USA: ACM.
- [23] Carter, Danielle. 2018. 'Narrative Learning as Theory and Method in Arts and Museum Education'. Studies in Art Education 59 (2): 126– 44.
- [24] Austin, Tricia. 2020. Narrative Environments and Experience Design: Space as a Medium of Communication. 1st ed. New York, NY: Routledge, 2020.: Routledge.
- [25] Dogan, Evinc, and M. Hamdi Kan. 2020. 'Bringing Heritage Sites to Life for Visitors: Towards A Conceptual Framework for Immersive Experience'. Advances in Hospitality and Tourism.
- [26] Bartalesi, Valentina, Carlo Meghini, and Daniele Metilli. 2016. 'Steps Towards a Formal Ontology of Narratives Based on Narratology'. Application/pdf, 10 pages.
- [27] Rinaldi, Antonio M., and Cristiano Russo. 2018. 'A Semantic-Based Model to Represent Multimedia Big Data'. In Proceedings of the 10th International Conference on Management of Digital EcoSystems, 31– 38. Tokyo Japan: ACM.
- [28] Pérez-Urbina, Héctor, Ian Horrocks, and Boris Motik. 2009. 'Efficient Query Answering for OWL 2'. In The Semantic Web - ISWC 2009, edited by Abraham Bernstein, David R. Karger, Tom Heath, Lee Feigenbaum, Diana Maynard, Enrico Motta, and Krishnaprasad Thirunarayan, 5823:489–504. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg.
- [29] Costa, Simone Dornelas, Monalessa Perini Barcellos, Ricardo de Almeida Falbo, Tayana Conte, and Káthia M. de Oliveira. 2022. 'A Core Ontology on the Human–Computer Interaction Phenomenon'. Data & Knowledge Engineering 138 (March): 101977.
- [30] Arndt, Richard, Raphaël Troncy, Steffen Staab, and Lynda Hardman. 2009. 'COMM: A Core Ontology for MultimediaAnnotation'. In Handbook on Ontologies, edited by Steffen Staab and Rudi Studer, 403–21. Berlin, Heidelberg: Springer Berlin Heidelberg.
- [31] Chera, Catalin-Marian, Wei-Tek Tsai, and Radu-Daniel Vatavu. 2012. 'Gesture Ontology for Informing Service-Oriented Architecture'. In 2012 IEEE International Symposium on Intelligent Control, 1184–89. Dubrovnik, Croatia: IEEE.

Author Biography

Yuchen Yang is a PhD candidate at EPFL, His research interests include Machine Learning on Heterogeneous Data, Knowledge Representation and Semantics, Audiovisual Narrative, Digital Curation, and HCI. He is especially interested in their application in the cultural and heritage sector in pursuit of the future for memory institutions