# Interactive multimedia ethnography: archiving workflow, interface aesthetics and metadata

P. MATTHEWS
University of the West of England, Bristol UK
AND
J. ASTON
University of the West of England, Bristol UK

Digital heritage archives often lack engaging user interfaces that strike a balance between providing narrative context and affording user interaction and exploration. It seems nevertheless feasible for metadata tagging and a "joined up" workflow to provide a basis for this rich interaction. After outlining relevant research from within and outside the heritage domain, we present our project FINE (Fluid Interfaces for Narrative Exploration), an effort to develop such a system. Based on content from Wendy James' archive of anthropological research material from the Sudan/Ethiopian borderlands, the FINE project attempts to use structural and thematic metadata to drive exploratory interfaces which link video, images, audio and text to relevant narrative units. The interfaces also benefit from the temporal and spatial variety of the collection to provide opportunities to discover contrasts and juxtaposition in the material across place and time.

Categories and Subject Descriptors: H.3.7 [INFORMATION STORAGE AND RETRIEVAL] Digital Libraries; H5.1 [INFORMATION INTERFACES AND PRESENTATION] Multimedia Information Systems --- Evaluation/methodology, Video; H.5.2 [INFORMATION INTERFACES AND PRESENTATION] User Interfaces --- Evaluation/methodology, Interaction styles, Prototyping, Theory and methods; J.5 ARTS AND HUMANITIES

General Terms: Human Factors, Design, Management

Additional Key Words and Phrases: ethnography, anthropology, memory, narrative, metadata, HTML5

**ACM File Format:**
MATTHEWS,P., AND ASTON, J. 2012. Interactive video ethnography: archiving workflow, interface aesthetics and metadata considerations. *ACM Computing and Cultural Heritage*.

## 1. INTRODUCTION

Multimedia has long been a popular method of recording and documenting the work of anthropologists, and is no longer considered a simple research tool. Instead, image, video and audio have become primary research outputs and when combined with more open access and standardized description, provide enormous potential for new uses, interpretations, and combinations within and between collections [Frisch 2006; Isaac et al. 2008].

Authors' addresses: P.Matthews, Department of Computer Science and Creative Technologies, Faculty of Environment and Technology, University of the West of England, Coldharbour Lane, Bristol, BS16 1QY UK. Paul2.matthews@uwe.ac.uk. J. Aston, Faculty of Arts, Creative Industries and Education, University of the West of England, Bristol, UK Judith.aston@uwe.ac.uk . 

The need for new interface tools and for the development of integrated workflows is clear if this potential is to be realized [Pea et al. 2007]. Not only is a persistent, reliable and discoverable archiving solution required, but also the methods for laying down content need to be easy, economical and timely. Coupled with this, we need user interfaces that allow for the exploration and flexible juxtaposition of media. Metadata, open content and description standards remain key enablers to these processes [Ntousias et al. 2008; Pitzalis et al. 2006; Zeitlyn 2005].

Archiving and storage needs have driven "back-end" developments to the stage where we now have open access repository solutions where ethnographic data can be deposited and accompanying metadata can be stored. Simultaneous with this work has been innovation in "front-end" displays and the development of tools which allow users working with multimedia collections the type of exploration, comparison and analysis that we have become used to with textual media. There is, however, still a gap between these two levels of operation [Beer et al. 2009]. More work is required in linking the two ends together to provide the 'middleware' that can draw content from the repository into the display in ways that can facilitate exploratory search and meaningful analysis [Shiri 2008].

This paper describes the practical development work that we have undertaken using Wendy James' collection of ethnographic recordings from the Sudan/Ethiopian borderlands, across which the communities she knew best were displaced several times by recent wars. Our work has been driven by prior work by Aston and James to create demonstrator displays that illustrate the enormous potential in juxtaposing video segments, audio and music to create aesthetic resonances across time, place and multiple points of view. In order for this to be an experience for the user which is both intellectually rigorous and pleasurable, we believe that a storage and retrieval system needs to support the discovery of such resonances through a structured metadata base and an interactive and adaptive/responsive interface.

Achieving an appropriate balance between structured content and free exploration within this system is one of the key challenges presented by this work. Whilst the aim is to create a fluid and exploratory system, the multimedia resources need to be contextualized within the wider framework of James' fieldwork. In this sense, our aim is to move beyond presenting a heterogeneous collection to creating a multi-layered narrative about an anthropologist's intermittent engagement with the same group of people over more than forty years. Whilst this narrative will initially place James at its centre as the main narrator, our longer term aim is to enable the people themselves to respond to the materials by adding their own recordings, creating their own juxtapositions, and adding narrated pathways to the system.

Before outlining our approach to addressing these aims, this paper will present relevant work looking at archiving workflow, interactive displays, and multimedia metadata standards both within and outside the heritage domain.

**Media exploration, user pathways and aesthetics**

The enormous potential of heritage multimedia collections to provide vivid, engaging experience of issues, contexts and personalities remains untapped, and promises to provide an authentic, less textually-biased access to materials [Frisch 2006]. By

"authentic" we mean the potential to evoke the subjects' culture, time and place and to provide necessary context while allowing for direct and meaningful interaction with the recordings themselves [Hickey 2005]. Video in particular can convey far more than audio and text alone, providing grounds for the truly collaborative creation of historical narrative [Sipe 2006]. From an anthropology perspective, advances in digital video technology and innovation in abstraction hold much promise for creating accessible archives [Zeitlyn 2005].

Pea and Hoffert for their work in the learning sciences, report on work with users to elicit requirements for analytical tools, and prioritized the features of acquisition, chunking, transcription, way-finding, organization / asset management, commentary, coding / annotation, reflection, sharing and presentation [Pea et al. 2007]. These requirements are expected to be similar in the anthropology domain, both for expert and novice user groups.

The former study's emphasis on collaboration drew out the features of annotation and sharing that have been further developed in other work which has emphasized the need for annotation to be easily to achieve and to give the user the impression of immediacy [Fraser et al. 2006].

In terms of way finding, in view of the enormous quantities and variety of evidence data that is becoming available in online research repositories, new, more visual ways of viewing the user query and relevant search results are needed [Shiri 2008]. These may more easily enable the user to view and understand the quantity of relevant information, including the visual signposting of related information that can serve to enhance the user's experience through providing opportunities for resource discovery.

Annotation of multimedia content to enable exploration and reassembly has been approached in a number of ways. Ontology and metadata schemes that provide a granular analysis of the narrative structure of video have been developed, with the aim of providing possibilities for user recombination and remixing [Lombardo et al. 2010; Kaiser et al. 2009], though this seems to have worked best for quite tightly structured, predictable stories, and where recombination has always been seen as a linear process.

Association of multimedia with a textual annotation track has been demonstrated to aid both retrieval and user understanding of the content. Experimental attempts to derive these using automated speech recognition (ASR) [Carmichael et al. 2008] or by extraction of descriptions from other sources [Kaiser et al. 2009] have had some success, and ASR technology is now mature enough to have been incorporated into enterprise digital asset management systems. Its utility to ethnographic data with poor (or no) sound and less widely used dialects remains questionable, however. Similarly the aggregation of textual data around multimedia content will have some good mileage in the anthropology field, but sources will generally be more limited and the need for hand annotation will remain.

While video annotation software such as the available qualitative data analysis tools permit the annotation of videos, the sharing and reuse of annotation has not been addressed until more recently [Pea et al. 2007]. Large scale projects to enhance metadata are increasingly looking to web 2.0 approaches such as user tagging, which can serve to both enrich the metadata and provide alternative views and vocabularies [Hauttekeete et al. 2009;Addis et al. 2010].

The "semantic gap" problem in automated video analysis and retrieval, where only low level features can be reliably identified, led to calls to shift the focus onto aesthetic features that might be gleaned computationally by analyzing the media in use

during the production process [Dorai et al. 2003]. That said, content-based techniques have remained relatively disappointing and failed to overcome difficult problems such as idiosyncrasy and context-specificity [Enser 2008].

In terms of user experience, creating space for reflexive interaction with media can greatly enhance enjoyment and prolong engagement with materials [Walsh et al. 2010]. In particular, co-presenting related or older materials alongside more contemporary recordings may add context and extend the playfulness and ability to create a narrative around a collection.

**Metadata usage and standards**

Conceptions as to the usefulness and application of metadata has varied between user and creator communities and might be classified according to "hunter" and "gatherer" purposes [Owens 2009]. The former may describe the classic information retrieval paradigm where an information manager designs a classification by which the expert user can find a known item. The "gatherer" paradigm, however, is more commonly associated with an end user browsing a collection to get an idea of the content or sample the topic.

Standardized approaches to metadata for digitized heritage collections are seen as a key to unlocking their potential, enabling the cross reference of materials between physical locations [Walsh et al. 2010]. As far as heritage and multimedia are concerned, the number and variety of standards to consider is considerable. Even within a single standard, the use and interpretation of the individual elements within it may vary such that interoperability is lost.

For video annotation and time-based segmentation, as well as the incorporation of low level and frame-based features, the MPEG-7 standard provides much of use. One shortcoming has been seen as its weakness at encoding semantic properties of the annotation, though for this some extensions have been proposed [Dasiopoulou et al. 2009]. It is also quite complex, meaning that for interoperability usage profiles are needed [Fraser et al. 2006]. Another limitation perhaps is the lack of widespread support in annotation tools, though tools such as Frameline and Advene do support it as an output. MPEG-7 has been recommended as a core enabler in large scale repository projects [Fraser et al. 2006].

Other standards which provide means for the standardized, potentially interoperable description of heritage digital data are the Metadata Encoding and Transmission Standard (METS), Encoded Archival Description (EAD) and the International Council of Museums' Conceptual Reference Model (CIDOC-CRM). These will become important as we look at the future connectivity and semantics of the James collection and are discussed further in Metadata Requirements below.

In terms of the presentation layer, standards are similarly heterogeneous and patchily applied. SMIL has been proposed as a simple format for encoding the outputs of user-based editing efforts [Kaiser et al. 2009]. While SMIL might provide the requisite structures for encoding media presentations, it unfortunately lacks modern browser support. Other interface project utilize homegrown XML formats which provide instruction to the interface code on how to configure the interface for a given media segment. As mentioned later, the HTML5 standard provides new media elements for presenting audio and video, and these may be combined with JavaScript on the browser to enable enhanced interaction.

While encoding of metadata/annotation, presentation parameters and retrieval of individual video segments remain outside the HTML5 provision, standards to enable these capabilities are being developed through the W3C Video in the Web Working Groups [W3C 2010]. The Media Fragments initiative promises to make retrieval of segments direct from the repository feasible, and the Media Annotations initiative promises a way to unify diverse metadata description formats.

### Workflow and "end to end" architectures

A key concern of this work is to establish working methods that tackle the combined requirements of user-centered curation, archiving and the creation of exploratory interfaces. We therefore need to consider the users and practices in addition to technical architectures that have been developed elsewhere.

With many digitization projects still pilot and exploratory in nature [Manžuch 2009], there remains much work to be done in developing reliable and stable workflows in taking the original work through the digitization approach, including not only the choice of metadata standards and subsets, but also the development of protocols and conventions for how these will be used. The latter is an underplayed, yet crucial area in the light of evidence that even on collaborative projects, individual cataloguers will select and apply metadata differently [Scifleet et al. 2009].

While archiving and information professionals are best placed and professionally oriented to metadata application, they are often interested in classification and access at the global level only, and see it as the role of the researcher to add meaning to the digitized media [Frisch 2006]. This perhaps heightens the need for cataloguing tools to enable the easy, productive preparation of metadata at digitization / ingest time. Some enterprise and desktop digital asset management packages are approaching this stage, but they often have significant shortcomings along dimensions of speed, usability and standards support.

Rather than a "write once" architecture, as mentioned above a key current concern for the development of more interactive, evolvable archives is the ability for user data to be associated with heritage objects [Hauttekeete et al. 2009][Addis et al. 2010]. This calls for the underlying system to afford the saving back of metadata to the collection following exploration, discovery and assembly sessions.

## 2.0 THE FINE PROJECT: "FLUID INTERFACES FOR NARRATIVE EXPLORATION"

The current project[1] is a collaboration with historical anthropologist Wendy James, who is in the process of digitizing and cataloguing her fieldwork materials with a view to creating a permanent archive. Unusually for an anthropologist of her generation, James has consistently used audiovisual media in her fieldwork. Whilst initially recording for research purposes, to act as *aides memoires* to her written analytical work, over the years she has come to recognize the wider relevance and value of these recordings.

In documenting aspects of James' fieldwork from the mid-1960s through to the present day, the recordings shed light on major world events in the latter half of the

[1] See http://www.spatialmontage.org/projects/fine

twentieth and early part of the twenty-first century, from the perspective of an anthropologist who has built a strong relationship over the years with her informants. These informants come primarily from the Uduk-speaking people of the Sudan./Ethiopian borderlands.

Achieving an appropriate balance between structured content and free exploration is one of the key challenges presented by this work. Our intention is to create a catalogued digital repository of James' materials in their entirety and to then give users access to an organized selection of these materials, within which further context will be embedded to reflect a more narrative-based approach to James' fieldwork.

In the first instance, we aim to create a system which places James at its centre as contextual narrator, but which also builds on Bakhtin's ideas about polyphony. Bakhtin proposed that Dostoevsky's novels did not combine a 'multitude of characters and fates into a single objective world, illuminated by a single authorial consciousness' but into 'a plurality of consciousnesses with equal rights and each with its own world'[Bakhtin 1984]. Such an approach can allow for different modes of speech and non-verbal communication to be presented through the characters, as opposed to there being an all-pervading style dictated by the author. At the same time it also recognizes the personality of the author, along with social and historical context, as being an important source of interpretation and meaning. Our aim is to present the story of the events and changes that have occurred in the region as far as possible 'through the words and experiences of the people themselves' [James 2007]. In this sense, the point of view of the anthropologist as narrator does not take final authority, with contradictory ideas and different styles of speech being able to co-exist in a more dialogic form.

As stated above, in one sense the collection used in this project may be considered as "pre-archive" stage. That is, it is material highly likely to have ongoing value but not yet lodged with an archival institution [Williams 2006]. We consider an important role of the project to be an intermediary between the collection owner and the future archiving institution. It is particularly important to develop and recommend a metadata model that goes beyond preservation to encompass important "secondary" informational uses [Williams 2006]. We have been in the fortunate position of being able to work with the collection owner and through experimentation to identify some of the metadata and narrative requirements for such uses.

## Project strategy

In seeking a sustainable, flexible and holistic solution, we see a need for a layered approach. Archival objects and associated descriptive, structural and administrative metadata are needed for long term objectives and need to be "loosely coupled" to interface technologies, as the latter are likely to have more limited lifespan. We also need to push as much descriptive and architectural detail as possible down to lower layers – to explicit metadata - so that this may also be available to future interface development.

Our aspirations for the interface are very similar to the design principles articulated by Goldman & Dong [2009]. That is, the system should allow layering of data, should match the subject domain and should allow movement between a high level view and specific media subsets.

Through the use of multiple windows, we aim to present different combinations of materials to serve different purposes. For example, users might want to compare recording across time according to a series of interlinked themes, or they might want to look at a specific set of recordings in one window and call up James' commentary on these materials in a separate window. Another example where juxtaposition works well is where we have recordings of actual events alongside interview footage of peoples' memories of these events.

Key project requirements, then, are as follows:-

**Interface**

- High level collection views in addition to curated pathways – initially provided by the collection owner but with scope to widen this to additional user groups;
- A thematic / sub-selection view which maximized the potential for discovery of interesting juxtaposition;
- Clear but non-invasive provision of contextual information through narrative video clips, audio commentary and text;
- With longer video clips containing several subjects, the ability to navigate to a subclip at a specific timecode;

**Storage/formats**

- Video, photos and audio will be digitized at high resolution, and linked to compressed versions of the same media at web resolution for interface display;

**Process/Workflow**

- An easy to use cataloguing process that is standardized across media formats (i.e. uses the same vocabularies and metadata fields across audio, video, image and text);
- Semi-automatic updating and export of the metadata to the interface;
- Addition of narrative video, audio commentary and text annotations using the same process;
- Media and metadata to be prepared in such a way to facilitate ingest into a digital repository at a future date;
- Media resources to use the web-friendly version for presentation but be linked to the associated high-resolution, repository-ready version.

## Media formats and structural content model

The project's requirement for a reasonably uniform treatment of content types called for a unified approach, which helped to simplify treatment of the media at the presentation layer – content could be simply rendered differently according to its type, with other description in common. In all cases the original digital format of the media was accompanied by a compressed and converted version for the interface (Table 1) along with a thumbnail image or poster frame for video.

**Table 1: Media formats in the James collection**

| Media | Formats |
|---|---|
| Video | Digitized Super-8 Cine and Hi8 video as DV-CAM files, subclipped into scenes of varying length. Converted to MP4/H264 and OGV for web presentation; |
| Images | Scans of slides and photographs as TIFF, compressed to JPG in two sizes for web; |
| Audio | Audio as AIFF, converted to MP3 for web; |
| Text | Conference papers and James' publications as PDF / MS Office format,converted to descriptive metadata elements. |

**Metadata requirements**

We found that, administrative aspects aside, the "whole object" aspects of the James collection could be quite thoroughly described using an organization that encompassed the location and time of recording (corresponding to research visits to Sudan and Ethiopia).In addition, a subject hierarchy needed to be developed around the topics of the sample of material originally used, which covered ritual, music and dance.. Table 2, below, shows the key descriptive elements, matched with elements from the international standards followed, those of MPEG-7 and the CIDOC-CRM.

MPEG-7 was found to meet the descriptive and structural requirements needed to form a foundation for the interfaces. An additional advantage of the scheme was that video sequences could be decomposed into timed segments with separate metadata applied through TextAnnotation elements. While the primary tool selected for media cataloguing did not support MPEG-7 directly, it did provide an XML-based export that could be converted and combined with temporal metadata from the video annotation tool. One key metadata requirement was to accommodate two types of semantics: those relating to the primary content of the media, and those relating to the subject of the conversation or focus of the activity portrayed. This was in order to enable reflexive juxtaposition of items such as an event paired with people talking about the same event. For instance, we have footage of dance events recorded in the 1960s which can be juxtaposed alongside footage shot thirty years later of people reminiscing on these events. A further requirement was to distinguish narration (audio and video) objects from primary content, in order that this might be dealt with differently in the interface. Narration would be another way to provide providing context, alongside descriptive text and metadata.

These two problems were approached by using two special-case subject descriptors – "reflection" and "narration". The co-presence of other descriptors then provided the connection between events or activities that they referred to. It was also useful to have more specific relations expressed between content items and this was achieved through the use of the "CreationInformation/RelatedMaterial/MediaLocator/MediaUri" MPEG-7 element.

**Table 2: Outline of descriptive metadata elements for the collection**

| Element Description | Notes / Examples | MPEG-7 Mapping | CIDOC-CRM class | Relevant Controlled Vocabularies |
|---|---|---|---|---|
| Title | | CreationInformation/Creation/Title | E73.InformationObject<br>E41.Appellation | |
| Description | General free text, including diary entries | CreationInformation/Creation/Abstract<br>TextAnnotation/FreeTextAnnotation | | |
| Theme | Music, Dance, Mission Church & Christianity, Dress, Diaspora | CreationInformation/Classification/Subject<br>TextAnnotation/KeywordAnnotation/Keyword<br>StructuralUnit (Pointer to ClassificationScheme) | E55.Type | AFS Ethnographic Thesaurus[2] |
| Pathway | A place / time period combination representing a phase of James' fieldwork | | E52.Timespan<br>E53.Place | |
| Transcript | | TextAnnotation/FreeTextAnnotation | | |
| Location | Country / Place Name / Lat & Long | CreationInformation/Creation/CreationCoordinates/Location<br>TextAnnotation/StructuredAnnotation/Where | E53.Place | Getty Thesaurus of Geographic Names (TGN) |
| Person(s) | Names / Ages / Titles | TextAnnotation/StructuredAnnotation/Who | E21.Person | |
| Date | | CreationInformation/Creation/CreationCoordinates/Date<br>TextAnnotation/StructuredAnnotation/When | E50.Date | |
| Event | | TextAnnotation/StructuredAnnotation/What Action | E7.Activity | |

With regard to future interoperability, the CIDOC-CRM was felt to provide the necessary abstraction from the detail of this collection to allow the possibility of integration of James' recordings with other material, particularly through its potential to map to MPEG-7 elements and to be expressed using semantic web structures [Angelopoulou et al. 2011]. While MPEG-7 was the primary format used for the interface, this was developed in such a way as to be export-compatible with the CRM[ICOM/CIDOC 2011].

METS and EAD are also considered important to the project as additional ways to expose the content and structure of the collection METS provides a common way to define OAIS compliant content packages [Addis et al. 2010] and provides placeholders for additional extant descriptive metadata. While there is some redundancy in terms of structural relations between MPEG-7 and METS, the latter is likely to provide a route by which the collection may be imported into long-term repository storage. EAD provides a means to connect the collection descriptions with biographic data and significant dates in

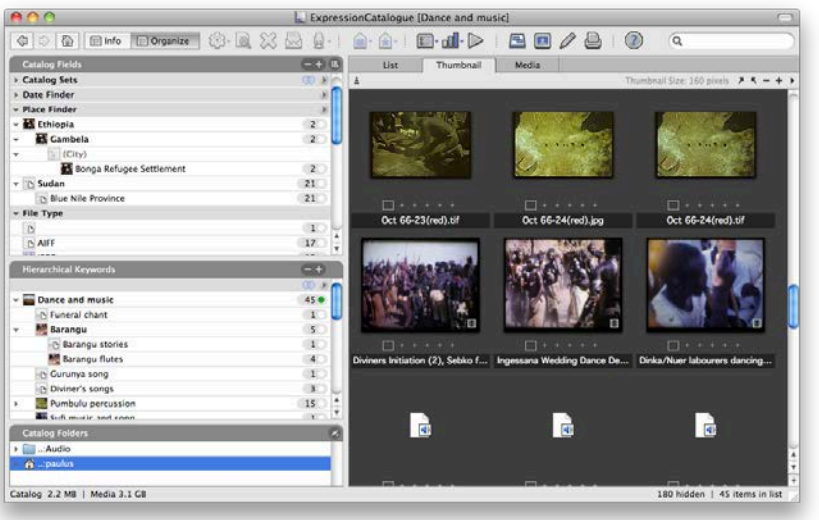
[2] http://www.afsnet.org/?page=AFSET

the life of the collection owner. As such procedures to generate METS and EAD compliant collection descriptions from the MPEG-7 files were needed.

## Standards and technologies employed

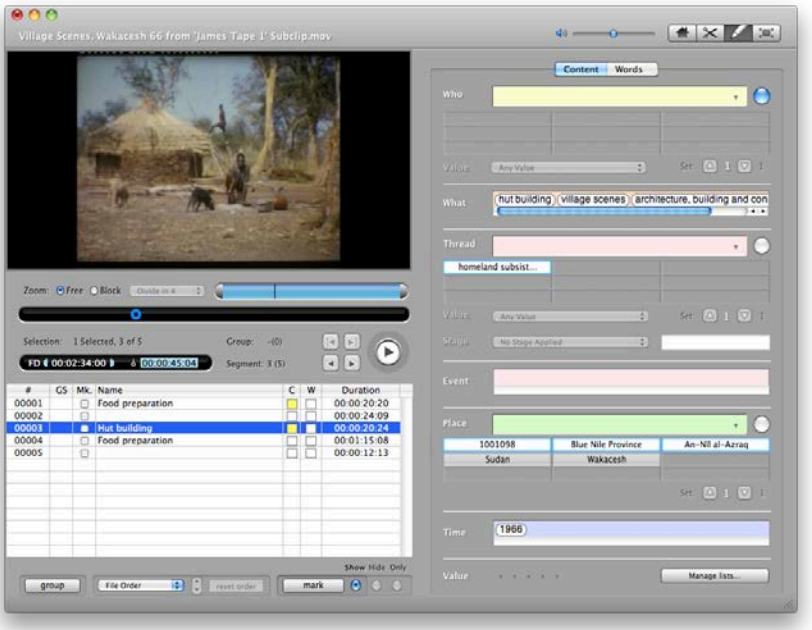
We began the cataloguing of high-quality image, audio and video using the desktop media asset management system "Expression Media"[3] ( Figure 1). This was chosen as it provided a relatively user-friendly way for batch annotation of all the project's media types in one environment. It also enabled metadata to be exported as a proprietory XML catalog file[4] which could then be transformed using XML stylesheets into MPEG-7. Additional segmentation for video clips was achieved using Frameline47[5] (Figure 2), and exported to MPEG-7.

**Figure 1: Capture of descriptive metadata using Expression Media (Images © Wendy James [17])**




[3] Now Media Pro: http://www.phaseone.com/en/Software/Media-Pro
[4] Though the program employs standard IPTC image description fields
[5] http://breadcrumb.tv/2010/11/15/frameline

**Figure 2: Annotation of video segments using Frameline 47 (Images © Wendy James [17])**




As regards the interface, we took a dual-pronged approach. We originally adopted Adobe Flash as a very useful front-end technology for manipulation and combination of media, and one which is widely supported. For this reason early prototypes were built using Flash with dynamic loading of media and metadata files from the MPEG-7 derived XML.

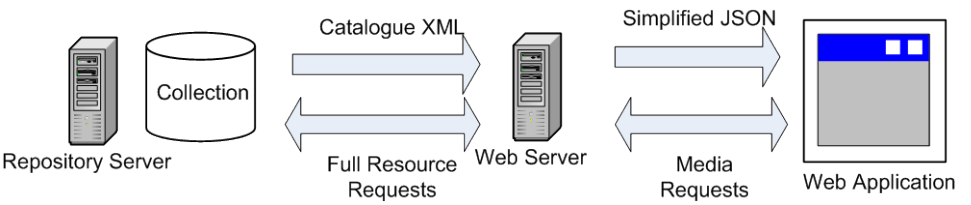
At the same time we explored the potential of HTML5 native video to support the kinds of interaction we have envisaged and have been encouraged by the Pad.ma [6] project which is based on open standards formats and enables the upload, annotation and retrieval of video, leveraging a range of media manipulation functions to generate abstractions from the loaded clips.

***Architecture***

An outline architecture of the prototype system is shown in Figure 3. The Expression catalogue was exported as XML which is transformed using XSLT (XML stylesheets) to intermediate storage as MPEG-7 and combined with video segment annotations, then transformed to a simplified subset as JSON (JavaScript Object Notation), which is brought into the web application where filtering and clustering can be achieved using JavaScript.The web application initially uses thumbnails generated by the collection, though once sub-selections are made, a web quality copy of the media is obtained from the collection.

[6] http://pad.ma/

**Figure 3: Architecture of FINE prototype**




In our HTML5 prototype, media can be selected according to theme and pathway, then brought into a media browser, which presents two main videos or images (Figure 4, Figure 5). A simple 2-means cluster of resources[7] according to metadata (a binary coding of presence or absence of a descriptor) results in the ability to juxtapose resources on similar themes by time and place. The user can select from the two clusters using a "Cover flow" selector to choose which resources to display. The resources are accompanied by sound and video narration to provide further context and an immersive experience.

The examples shown as screenshots illustrate juxtaposition across time (Figure 4, showing the single-stringed instrument the Dumbale recorded on two occasions 34 years apart) and across place (Figure 5, the Pumbulu, a percussion instrument made out of a pattern of holes in the mud).

***Evaluation and Further work***

We have begun by incorporating narrative video fragments applied at a low level, with commentaries on the materials at the level of themes down to particular recordings. The next phase of work will seek to provide longer introductions to the "pathways" or visits made by James to the Uduk communities at various periods since the 1960's. With metadata linking illustrative media to these longer narrations, the intention is to then provide a coherent guided trail through the material which the user may choose to sample or follow at length. We have successfully piloted the recording of narrative using a webcam while the collection is being viewed – an approach that can also be taken with other commentators. We also intend to develop the workflow and interfaces in a way that is content-independent, and future work will seek to use the system with a different collection, thereby testing and further developing the approach taken to date. With growing support and potential for cross-collection media retrieval, we also expect that such tools will enable visual comparison and discovery between previously disconnected materials.

An online evaluation of our prototypes will be conducted through the website referenced above, and will involve anthropology and multimedia researchers and other interested groups to experiment with the prototypes and provide feedback.

[7] Achieved using the figue.js javascript library http://code.google.com/p/figue/

**Figure 4: Dichotomous clustering of media using metadata to achieve temporal comparison (1966 on the left, 2000 on the right) Images © Wendy James**

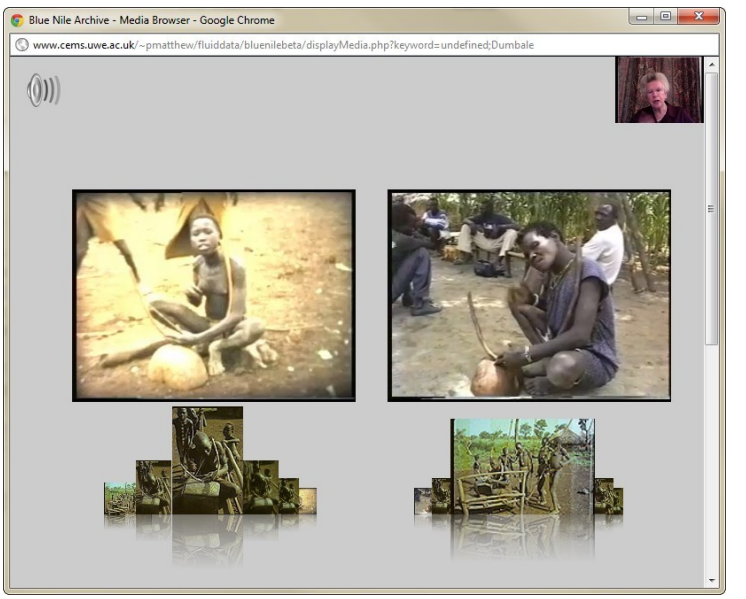
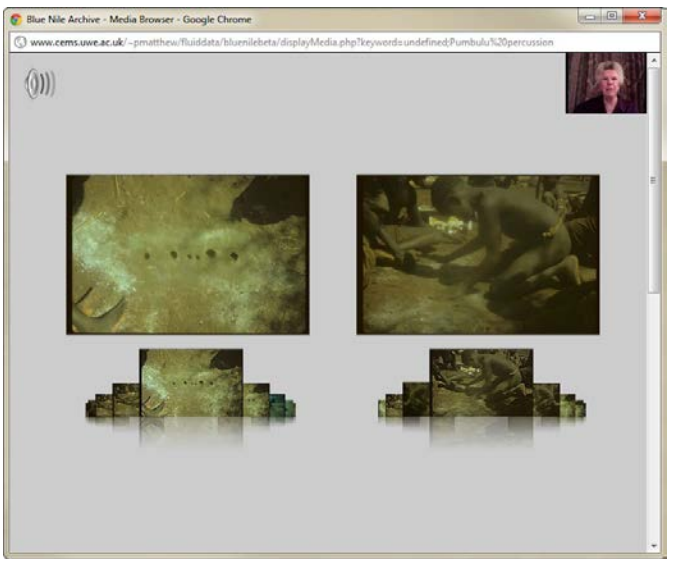



**Figure 3: Clustering by place: comparison of "Pumbulu" percussion in two locations (Images © Wendy James)**



## Conclusion

We have demonstrated an approach with potential for an end-to-end system for the archiving and future exploration of a relatively small yet important ethnographic collection. As such, we are addressing the need for collections to be "humanized" – for additional layers of meaning to be applied. We have also shown how metadata, carried through from the repository level, can be used to underpin media montage at the interface, with potential for the discovery of spatial and temporal contrasts and resonances. While the web technologies and software tools used are relatively low cost, they are also potentially scalable to larger collections with multiple curators – the direction in which we hope the work will proceed.

## Acknowledgements

This work was enabled by two inter-faculty collaboration grants from the Faculty of Environment and Technology, University of the West of England. The authors are grateful to Wendy James for her ongoing support and her feedback on the manuscript. Feedback from the reviewers was very valuable in helping the authors to improve and extend the article.

## References

ADDIS, M., BOCH, L., ALLASIA, W., GALLO, F., BAILER, W. AND WRIGHT, R. 2010. 100 Million Hours of Audiovisual Content: Digital Preservation and Access in the PrestoPRIME Project. In *First International Digital Preservation Interoperability Framework (DPIF) Symposium*, March, Anonymous

ANGELOPOULOU, A., TSINARAKI, C. AND CHRISTODOULAKIS, S. 2011. Mapping MPEG-7 to CIDOC/CRM. In *Research and Advanced Technology for Digital Libraries*, S. GRADMANN, F. BORRI, C. MEGHINI AND H. SCHULDT, Eds. Springer Berlin / Heidelberg, , 40-51

BAKHTIN, M.M. 1984. Problems of Dostoevsky's Poetics. University of Michigan Press, Minnieapolis

BEER, C.A., PINCH, P.D. AND CARIANI, K. 2009. Developing a flexible content model for media repositories: a case study. In *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*, Austin, TX, USA, Anonymous ACM, New York, NY, USA, 97-100

CARMICHAEL, J., LARSON, M., MARLOW, J., NEWMAN, E., CLOUGH, P., OOMEN, J. AND SAV, S. 2008. Multimodal indexing of digital audio-visual documents: A case study for cultural heritage data. In *Sixth International Workshop on Content-Based Multimedia Indexing*, London, UK, Anonymous IEEE,

DASIOPOULOU, S., TZOUVARAS, V., KOMPATSIARIS, I. AND STRINTZIS, M. 2009. Capturing MPEG-7 Semantics. In *Metadata and semantics*, M. SICILIA AND M.D. LYTRAS, Eds. Springer, New York, 113-122

DORAI, C. AND VENKATESH, S. 2003. Bridging the semantic gap with computational media aesthetics. *Multimedia, IEEE* 10, 15-17.

ENSER, P. 2008. The evolution of visual information retrieval. *Journal of Information Science* 34, 531-546.

FRASER, M., HINDMARSH, J., BEST, K., HEATH, C., BIEGEL, G., GREENHALGH, C. AND REEVES, S. 2006. Remote Collaboration Over Video Data: Towards Real-Time e-Social Science. *Computer Supported Cooperative Work (CSCW)* 15, 257-279.

FRISCH, M. 2006. Oral history and the digital revolution. In *The oral history reader,* R. PERKS AND A. THOMSON, Eds. Routledge, London, 102-114

GOLDMAN, R., DONG, C. AND LANSIQUOT, R. 2009. Software design principles for video research in the learning sciences and CSCL: two studies use the perspectivity framework \& Orion\&\#8482; In *Proceedings of the 9th international conference on Computer supported collaborative learning - Volume 2*, Rhodes, Greece, Anonymous International Society of the Learning Sciences, , 132-134

HAUTTEKEETE, L., EVENS, T., MECHANT, P. AND COURTOIS, C. 2009. Unlocking cultural heritage: a bottom-up approach. In *COST 298: The user and the future of information and communication technologies 13th - 15th May 2009,* Copenhagen, Denmark, Anonymous

HICKEY, N. 2005. Devising educational workshops and programmes that generate a more meaningful engagement with the public. In *Re-Thinking Technology in Museums: Towards a New Understanding of People's Experience in Museums* Interaction Design Centre, University of Limerick, Ireland, Anonymous

ICOM/CIDOC. 2011. Definition of the CIDOC Conceptual Reference Model. 2011, http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf

ISAAC, A., SCHLOBACH, S., MATTHEZING, H. AND ZINN, C. 2008. Integrated access to cultural heritage resources through representation and alignment of controlled vocabularies. *Library Review* 57, 187-199. http://dx.doi.org/10.1108/00242530810865475

JAMES, W. 2007. War and survival in Sudan's frontierlands: voices from the Blue Nile. Oxford University Press,

KAISER, R., HAUSENBLAS, M. AND UMGEHER, M. 2009. Metadata-driven interactive web video assembly. *Multimedia Tools and Applications* 41, 437-467; 467.

LOMBARDO, V. AND DAMIANO, R. 2010. An Intelligent Tool for Narrative-Based Video Annotation and Editing. In *Complex, Intelligent and Software Intensive Systems (CISIS), 2010 International Conference on,* Anonymous , 706-711

MANŽUCH, Z. 2009. Monitoring digitisation: lessons from previous experiences. *Journal of Documentation* 65, 768-796.

NTOUSIAS, A., GIOLDASIS, N., TSINARAKI, C. AND CHRISTODOULAKIS, S. 2008. Rich metadata and context capturing through CIDOC/CRM and MPEG-7 interoperability. In *Proceedings of the 2008 international conference on Content-based image and video retrieval,* Niagara Falls, Canada, Anonymous ACM, New York, NY, USA, 151-160

OWENS, C. 2009. When metadata worlds collide: The hunter/gatherer dichotomy. *Journal of Digital Asset Management* 5, 181-184.

PEA, R. AND HOFFERT, E. 2007. Video Workflow in the Learning Sciences: Prospects of Emerging Technologies for Augmenting Work Practices. In *Video Research in the Learning Sciences,* R. GOLDMAN, R. PEA, B. BARRON AND S.J. DERRY, Eds. Routledge, , 427-460

PITZALIS, D., LAHANIER, C., PILLAY, R., AITKEN, G., RUSSELL, A., SMITH, D.A., SINCLAIR, P.A.S., ADDIS, M.J., LOWE, R., HAFEEZ, S., LEWIS, P.H., MARTINEZ, K. AND SCHRAEFEL, M.C. 2006. Semantically Exposing Existing Knowledge Repositories: a Case Study in Cultural Heritage. In *The first international conference on Semantics And digital Media Technology (SAMT 2006),* Anonymous

SCIFLEET, P., WILLIAMS, S. AND COLE, C. 2009. The Human Art of Encoding: Markup as Documentary Practice. In *Metadata and semantics,* M. SICILIA AND M.D. LYTRAS, Eds. Springer, New York, 55-69

SHIRI, A. 2008. Metadata-enhanced visual interfaces to digital libraries. *Journal of Information Science* 34, 763-775.

SIPE, D. 2006. The future of oral history and moving images. In *The oral history reader*, R. PERKS AND A. THOMSON, Eds. Routledge, London, 406-415

W3C. 2010. Video in the Web Activity Statement. 2010, http://www.w3.org/2008/WebVideo/Activity.html

WALSH, R. AND HORNSBY, A. 2010. Improving the Value of Archived Personal Content with Aesthetic and Reflexive Qualities. In *Consumer Communications and Networking Conference (CCNC), 2010 7th IEEE*, Anonymous , 1-2

WILLIAMS, C. 2006. Managing archives : foundations, principles and practice. Chandos, Oxford

ZEITLYN, D. 2005. Multimedia Provocations and the future of ethnographic film. In Anonymous