World Library and Information Congress:
71th IFLA General Conference and Council

"Libraries - A voyage of discovery"

August 14th - 18th 2005, Oslo, Norway

Conference Programme:
http://www.ifla.org/IV/ifla71/Programme.htm

June 15, 2005

Code Number: 133-E
Meeting: 153 Audiovisual and Multimedia

Improving access to audiovisual and multimedia materials: the Moving Image Case Study of InterPARES2

James M Turner
École de bibliothéconomie et des sciences de l'information
Université de Montréal, Montréal, Québec, Canada

Mary Ide
Media Archives and Preservation Center, WGBH Educational Foundation
Boston, Massachusetts, USA

Randal Luckow
Turner Broadcasting System
Atlanta, Georgia, USA

Isabella Orefice
Associazione Nazionale Archivistica Italiana
Roma, Italia

Abstract

As the IFLA Guidelines for Audiovisual and Multimedia Materials in Libraries and other Institutions *remind us, access to audiovisual materials should be as open and free as access to print-based materials. However, such a reality is little more than a dream for the moment. The nature of audiovisual and multimedia materials is such that methods useful for organising print-based materials do not work well when applied to audiovisual and multimedia ma-*

1

terials, and much work remains to be done in developing adequate ways to provide good access to users. InterPARES2 is an international research project with 74 researchers in 15 countries on 4 continents. It focuses on records produced in complex digital environments in the course of artistic, scientific and e-government activities. As part of the work of Inter-PARES2, the Moving Image Case Study attempts to model the production processes of making moving images, to identify the documentation created at each stage of the process, and to reflect upon useful metadata for managing this documentation. The Moving Image Case Study looks at four types of production of moving image documents in three countries: a commercial Hollywood studio, a public national filmmaking institution, a public television producer and broadcaster, and an independent multimedia production company. Data was collected in all four environments, in order to obtain answers to the twenty-three research questions all InterPARES2 researchers must ask the participants in their studies. This paper gives an overview of the Moving Image Case Study within the context of InterPARES2, reports results of three of the four components of the study, and shows how this work contributes to improving access to moving images by providing a basis for better organisation of such collections.

Introduction

As the *IFLA Guidelines for audiovisual and multimedia materials in libraries and other institutions* remind us, access to audiovisual materials should be as open and free as access to print-based materials. However, such a reality is little more than a dream for the moment. The nature of audiovisual and multimedia materials is such that methods useful for organising print-based materials do not work well when applied to audiovisual and multimedia materials, and much work remains to be done in developing adequate ways to provide good access to users. This paper deals with one aspect of the vast world of audiovisual and multimedia materials, that of moving images. These in themselves constitute a rather large body of documentation, including film, television, and video, as well as web-based moving images and animation created with software such as animated GIFs, Flash objects, and SVG (Scalable Vector Graphics) files.

Our paper reports on the work carried out in the context of a larger research project. Inter-PARES2 (http://www.interpares.org) is an international research project with 74 researchers in 15 countries on 4 continents. It focuses on records produced in complex digital environments in the course of artistic, scientific and e-government activities. All the studies undertaken within InterPARES2 have to do with digital materials, so that in our study we were concerned only with digital moving images. The larger project is concerned with the broad goals of developing and articulating concepts, principles, criteria, and methods for the creation and maintenance of accurate, reliable records, and with the long-term preservation of authentic records.

The work of InterPARES2 is divided into 3 domains and 3 focuses. The domains are: 1. records creation and maintenance; 2. authenticity, accuracy and reliability; and 3. methods of appraisal and preservation. The focuses are: 1. artistic activities; 2. scientific activities; and 3. governmental activities. In addition, there are four cross-domains that cover all domains and focuses: 1. Terminology; 2. Policy; 3. Description; and 4. Modelling. Our case study fell into the area of work group 3.1, entitled Methods of appraisal and preservation of artistic works. The case study is entitled "Digital moving images: inputs, processes and outputs", but it became known as simply the Moving Image Case Study.

The study attempts to model the production processes of making moving images, to identify the documentation created at each stage of the process, and to reflect upon useful metadata

for managing this documentation. It looks at four types of production of moving image documents in three countries: a commercial Hollywood studio (USA), a public national filmmaking institution (Canada), a public television producer and broadcaster (USA), and an independent multimedia production company (Italy). Data was collected in all four environments, in order to obtain answers to the twenty-three research questions all InterPARES2 researchers must ask the participants in their studies (<http://www.interpares.org/ip2/ip2_23_questions.cfm>). Thus in practice it is actually four individual case studies. The objectives of the Moving Image Case Study are:

• to identify the digital entities created at each stage of production
• to model the structure of each entity
• to make explicit the contribution of each entity to the overall process
• to show relationships among the various digital entities

The data from the National Film Board of Canada has been collected, but it has not yet been compiled. Thus we cannot report it in this paper. However, in the next sections we will provide an overview of the three other types of production we studied. For each, we will discuss some of the aspects related to responses to the twenty-three research questions, insofar as they relate to the topic of this paper.

The commercial film studio
Our partner for this part of the study, a commercial Hollywood studio, required anonymity. The data was collected by the researchers in the project interviewing the studio's archivist and asking the twenty-three questions to which all InterPARES studies are required to respond. We investigated the sequence of steps used in creating the visuals for an animated film made using computer graphics. We studied the visual aspects of the process, from their conception to producing the final product, excluding administrative aspects such as memos, contracts, and spreadsheets. Although InterPARES2 is concerned with digital files only, we were obliged to include some analogue elements in order to carry out our study, since they are part of the process and cannot be isolated from the digital aspects. To get an understanding of how this works, let us consider the initial process of planning a film. Ideas are visualised in the form of artwork on paper, and this artwork is laid out in the form of a storyboard, many individual drawings pinned to a board in sequence, then shown to those who have the authority to decide whether the idea is interesting enough to make the film.

It is easy to imagine how the use of analogue drawings is advantageous. Artists prefer to make quick sketches on paper, because pencil-and-paper technology is very fast and easy to use. Laying the sheets of paper out on a board permits sequencing them, and they can then be shown easily in a presentation to pitch the film to the decision-makers. This method of consultation is much more efficient than manipulating many windows on a computer screen to show the same information. One can get an overview, but also zoom in on any particular drawing simply by moving the eyes or the body, as one does at a museum exhibition. Later, if the project is approved, the images are scanned, and these become the digital entities on which much further work is carried out to make the film.

In studying this process, we discovered that a relatively small number of digital entities were created. However, these entities evolve over the time of production by being constantly altered as new work is added to them before they are passed on to the next worker in the chain of production. This process involves adding new artwork in a new, transparent layer which leaves the preceding version intact but permits seeing through to it in order to new place new

elements where they need to go, for example taking the basic image of a character that has been approved for use in the film, and adding a costume to it. Once the newly-produced work is satisfactory, the artwork is composited and flattened at each stage in order to reduce the file size. This involves integrating the new artwork into the same layer as the previous version. Since the computer file is thus reduced to a single layer, it requires less storage space than a file with multiple layers. In this way, the number of files is kept down, but each file is constantly being modified as work progresses. All activity is focussed on producing the final artwork in TIFF format. These images are the ones that are seen in the film.

As part of our work in answering the twenty-three questions, we wondered how the digital entities are identified. We learned that at the beginning of production, everyone who will work on the film is required to learn the strict naming conventions for the files that are imposed. These include information such as the sequence and scene of the film, an identification of the particular digital object, the particular version of the object, and the file extension indicating the software needed to read the file. All this information is encoded in the file name, and as these files are passed from one person to the next, the version number changes. All storage and retrieval of digital entities is based on these naming conventions, and the sequence and scene numbers included in the file names permit assembling the images in the proper order to make the film.

These file naming conventions provide some stability to the procedures of filmmaking in this studio, since procedures are not otherwise documented. Since they can change radically from one film to the next, it is felt that documenting the procedures is not useful. Instead, training and other information about procedures is passed on from one production team to another by discussions in the early stages of pre-production.

We also wondered how reliability and authenticity of the digital files are assured. The version number of the files confers authenticity to them, and assigning this number correctly is assured through training in the use of the naming conventions adopted for the production. Thus if the document responds to the software by opening, and then by executing the commands given in the next stage of work, it is known to be reliable. At the broader system level, the backup copies of the files that are made each day constitute the reliability of the files. If need be, it is possible to revert to previous versions for which a backup copy is available.

In response to the research question concerning how the creator uses digital entities, we learned that apart from the initial development stage described above and which uses analogue materials, all production activities are carried out using digital entities. Once the film is released, these are archived for their legal value in proving ownership if need be. For example, publicity and merchandising materials surrounding the film are often contracted out to other companies, so the studio needs to be able to prove legal ownership of the material in case there is abuse of the contract permitting use of material from the film.

However, no rich metadata is created for longterm storage and retrieval, since the information only has value as current assets that will not be needed later. The context of commercial filmmaking is clearly not the same as that of patrimonial collections Interestingly, some of the analogue artwork is archived, because it can be sold to collectors of such work.

The independent multimedia production company
This part of the case study was carried out with the collaboration of the Italian National Archives Association (Associazione Nazionale Archivistica Italiana), which has organised a national working group of experts in the area of moving images that includes a number of na-

tional stakeholders in Italy. Among other things, the group is studying the question of digital moving images.

There were a number of reasons for undertaking this part of the case study. We wanted to include international partners. We wished to study a few types of production settings. We wanted to look at various types of moving images. We hoped to be able to compare various techniques in various settings. One member of our team had access to an interesting group of professionals who have a small multimedia production company, Altair4 Multimedia (<http://www.altair4.it/>), based in Rome. The company uses industrial design methodologies to create websites and information systems for museums and other institutions. In recent years, the company has focussed increasingly on promoting and popularising cultural and artistic heritage information, and has produced a large number of three-dimensional archaeologicial reconstructions for museums, for television broadcasts, and for interactive CD-ROMs.

The project of Altair4 Multimedia that we studied in this context was the virtual reconstruction using multimedia technology of an archaeological site at Pompeii, the house of Polybius, destroyed in 79 A.D. by the eruption of Vesuvius. The project is entitled The House of Julius Polybius. From Palladio via the French Academy's *envois des pensionnaires,* archaeological reconstruction using modern information technology is part of a long tradition, in which watercolours have now given way to pixels.

An interdisciplinary approach was essential in developing the virtual reconstruction of the House of Julius Polybius. The Altair4 team worked with archaeologists, botanists, mathematicians and vulcanologists to create a reconstruction of the greatest possible authenticity and accuracy. In modelling the entities (all the components of the villa), Altair4 began from the front and proceeded to the interior. This was necessary in order to follow the course of the archaeological excavation that was taking place. A digital camera was used to take photographs of all the pictorial walls, and the photos were then processed and restyled. The three-dimensional model of the architectural structures was made using a toolkit which manages the hierarchical organisation of spatial elements such as points, lines, and polygons. These entities are associated with sets of spatial coordinates and relational properties, including physical characteristics, modes of reflection, and light absorption. We investigated the steps involved in the virtual archaeological reconstruction of the house of Julius Polybius. The activities that generated digital entities that we studied were modelling the house, furniture, and objects, and producing textures.

As was the case with our study of the commercial film studio, file-naming conventions were established and all those involved in manipulating the digital files were required to adhere to these conventions. These included the project name, the name of the object of the file, and the version number. Files were organised into directories and sub-directories for storage while work progressed. As with the first part of our study, the workflow determined the way the digital entities are organised, as digital entities were passed from one work group to the next for manipulation and treatment.

The company's hardware and software were used to manage all the digital entities. The production is considered entirely digital, but again some analogue artefacts were created in the process. These took the form of images created on paper and then scanned into the computer. And again, the processes and procedures used were not documented. Since many moving image or multimedia productions are so unique, their creators do not feel it is useful to work out

procedures, because the next project will have different requirements and the procedures will not necessarily be the same.

Authenticity is again conferred on the digital entities by the version number, included in the file name. If the digital entity can be opened by the software for the next stage of the work, it is considered authentic and reliable. At the system level, the daily backups constitute the guarantee of reliability.

The digital entities are used by the production company in a variety of ways. During production, activities are centred on creating the end product. Once the product is completed, use of the entities depends on the specific contract with the client. The legal ownership is also an issue, and promotional and marketing activities constitute other uses of the digital entities. The question of long-term preservation is not considered. Files are stored on CD-ROM and DVD discs. In addition, there is a global backup, but no preservation strategies are worked out nor implemented. The company saves about 90% of the digital entities created. Selection is based on the importance of the entities. Because there is no system for archiving the digital entities, no descriptive metadata schemes or standards are used, and there is no cataloguing or indexing activity. Finding mechanisms available in the computer software are used if anything needs to be retrieved.

The public television producer and broadcaster
This partner in our study, WGBH public television in Boston, is in the process of converting to a digital asset management system, so that the current operating environment is both analogue and digital. The digital asset system is in operation, but the collection of analogue materials, including film, tapes, and audio content, dating back to the 1950s, is also being maintained. In this part of our case study, the activity studied is the production of a documentary television programme.

Two specific objects generated during the production process are examined, original footage, and the accompanying logs. Original footage is shot for a specific production either inside or outside the studio. Although some of the original footage includes sound, the sound elements are not included in this part of the case study, since they are often recorded separately. The material under study is usually recorded on a video cassette, and many formats of tape are used. The material can also be on film. The original footage log is a listing of every shot, and it records information such as the subject of discussion in the visual material, people who appear in the image, the location at which the material is shot, the environment, the context, and the recording time. The log references each shot, by listing the time codes during which it takes place. This material arrives at the television archive either as hard copy or in the form of a database.

The digital entities are identified using a unique identifier which links the catalogue record in the log with the original footage. The catalogue records are structured using a hierarchy, including the series, the season, and the programme. We can see that this is somewhat parallel to the systems used in the parts of our case study we discussed earlier, but in this case the information is recorded in a database and not just in the name of the file. In contrast with the commercial production environments, there is a concern with longterm management of the information, and much more effort is invested in this aspect. This has to do with the desire to find materials at a later date, while in the commercial production environments, little further use is made of the material once the production is completed. In the present environment, there is an investment in developing in-house standards for organising these materials.

The new digital asset management system can receive the original footage and logs in digital form, and the material can be used for reference and editing purposes. Users of the system must log in before they can use the system, so that the material remains secure. In addition, this permits the archives personnel to track use of the material. Once the production is completed, the digital files in final form will be sent to the archives for quality control review and final transfer to the institutional digital repository. Again, we can see here a much broader concern than in the commercial production environments for the longterm storage and retrieval of materials used in production.

The processes and procedures used to organise the information is documented. The archive service provides the production teams with training and with workbooks documenting procedures for completing the original footage logs. This stands in contrast to practices in the commercial environments we studied, and again reflects the desire for ongoing and longterm access to the material. Measures are in place to ensure the quality, reliability, and authenticity of the digital materials. They are compared to the data in the logs in order to check the accuracy of the information and to add information helpful in tracking the location of the physical asset. Since the logs can be read but not modified by most users, authenticity is assured. Since the original footage tapes circulate and are used for many different purposes, authenticity of the tapes cannot always be guaranteed in the analogue system. However, the new digital asset management system provides control over this aspect by assuring that the image files are such that they can be read but not altered. Users can make a copy and then work on the copy, but they cannot change the original material.

Again in contrast to our commercial partners, this public television producer and broadcaster makes use beyond the original production of the material it creates. As we mentioned, the system allows individual shots to be located, and these can be accessed by other production teams and used in creating new works. In addition, the new digital asset management system offers the capacity to create storyboards using clips from the system. Furthermore, material from this database can be repurposed for making compilations or other new products in a variety of media.

The metadata information used to manage the digital images can only be modified by the archives personnel, so that a good system of control is in place. The new system monitors changes to the digital files, and can also determine whether changes have been made to a file. Once again, we observe the contrast with commercial producers, in that creation and management of metadata and control of the materials are activities that are built into the system, again reflecting the different mission of the producer.

In the analogue system of this producer, the information created is turned over to the archives once the production is completed. The new digital asset management system can manage both the production activities and the archiving activities. Those responsible for the system are still working out procedures for exactly how current assets will become archived in the long term. However, we can see from the organisation of this system that because links will be made at the shot level between the original material and the archived material, materials will be easy to find. The new system allows greatly improved access to the material. The older system also permits access, but it is cumbersome, as is the working environment which presently requires using both systems. The new system includes tools for controlling vocabulary and makes a thesaurus available, so that additional benefits in retrieval can be expected.

Preservation of the material is also a conscious activity in this institution. In its older system, preservation is focussed on copying obsolete or deteriorating material to fresh supports, which are digital. Although procedures for the new digital system have not been finalised, they will take into consideration the necessity of migrating digital formats as new ones come into existence. Preservation activities in this institution are rigourous, in the form of check-lists of materials to be archived which production teams must respect, and of institution-wide records retention schedules which govern the length of time material is kept.

Metadata standards have long been in use, in the form of in-house descriptive standards (these are the norm in still and moving image collections worldwide, in contrast to practice in libraries, which have well-established procedures and standards) and use of the Library of Congress Subject Headings, adapted for use in the system. The new system will continue to make use of personnel with professional archives and library training, and will include Dublin Core metadata and PBCore reference resources.

Discussion

As we have noted in the preceding sections, the preoccupations with organisation and preservation of these digital materials vary greatly from one environment to another. However, the strongest pattern that emerges is that the commercial producers have little concern for organising materials in any systematic way, except insofar as is necessary to complete their productions. Once a production is completed, there is little concern for preservation of the materials. Since their purpose is to make a commerical product to be marketed by their client, all activity is focussed on getting the product to market. Once the product is completed, there is little motivation for preserving the production materials, except to be able to prove ownership for legal purposes. Longterm cultural heritage considerations are not necessarily part of their mission, and in this context it is easy to see that their viewpoint is reasonable. If anything should be archived, they can reason, it should be the final product they created and released. Public institutions that acquire the product, by purchase or by legal deposit, can then take over responsibility for preserving it.

Since they do not have a mandate to preserve the production materials for the long term, they see no reason to invest in this activity. Furthermore, even if they did, there is probably no payoff in more efficient production methods for future activities, since the technology for producing digital moving images and other multimedia elements is constantly changing. Thus even if the preserved materials were available, they would not necessarily be useful since any new production might well require different hardware and software than those used to create the archived product. However, there are some stirrings of concern over archiving material in the commercial environment because it is clear that it can be used for new re-purposing products that were previously not a part of the marketing strategy. The profit motive guides all activity in the commercial environment, and it may end up being the motor for undertaking some preservation activity.

Public institutions have long invested much more activity into developing good record keeping practices and better management of the materials under their care. They are not constrained by the profit motive, and some have mandates for the longterm preservation of cultural heritage materials. As part of that mandate, they are required to use standards and methods for organising the material for which they are responsible. Since these are only recently being developed in the world of moving images, ad hoc systems prevail for the moment and will for some time. However, there is increasing pressure, imposed by the relatively recent

networked environment that now affects us all, to do our work using common standards and methods. Therein lies quite a bit of hope for more rigorous organisation of collections in the future.

Although we do not yet have the data compiled from our fourth partner, the National Film Board of Canada, knowledge we have gained from our previous work with this institution, combined with its public nature, lead us to speculate that the patterns we will see in the responses to the twenty-three questions will more closely resemble those of our public television partner than those of our commercial partners. Whether we are eventually able to confirm this or not, we hope to report our findings at a later date.

Conclusions

The work of InterPARES2 will shed some light on present and future practices in the organisation of digital information of all kinds, including the moving images we concentrated on in our case study. We have seen some patterns emerging rather clearly, and our final research report will include comparison of these patterns not only within the components of our own case study but across the various case studies undertaken within the framework of this large international research project.

Our results will help draw an accurate portrait of the state of digital moving image production in the context of producing authentic and reliable documents, and of preserving them over the long term. Those who wish to draw on these findings with a view to improving present systems and practices will be able to do so. Over time, practices will become more standardised and more rigorous methods will be applied, so that information can be consulted using computer networks that will be both efficient and accurate.